# **Reference Manual on Scientific Evidence**

# **Reference Manual on Scientific Evidence**

Federal Judicial Center 1994

This Federal Judicial Center publication was undertaken in furtherance of the Center's statutory mission to develop and conduct education programs for judicial branch employees. The views expressed are those of the authors and not necessarily those of the Federal Judicial Center.

## Contents

Preface v Introduction 1 I Overview Management of Expert Evidence William W Schwarzer 7 **Evidentiary Framework** Margaret A. Berger 37 II. **Reference Guides** Comment on the Use of the Reference Guides 119 **Reference Guide on Epidemiology** Linda A. Bailey, Leon Gordis, and Michael Green 121 Reference Guide on Toxicology Bernard D. Goldstein and Mary Sue Henifin 181 **Reference Guide on Survey Research** Shari Seidman Diamond 221 Reference Guide on Forensic DNA Evidence Judith A. McKenna, Joe S. Cecil, and Pamela Coukos 273 **Reference Guide on Statistics** David H. Kaye and David A. Freedman 331 Reference Guide on Multiple Regression Daniel L. Rubinfeld 415 Reference Guide on Estimation of Economic Losses in Damages Awards Robert E. Hall and Victoria A. Lazear 471 III. **Extraordinary Procedures Court-Appointed Experts** Joe S. Cecil and Thomas E. Willging 525 **Special Masters** Margaret G. Farrell 575 List of Peer Reviewers 623 Index 629

## Preface

The Reference Manual on Scientific Evidence is the product of a cooperative effort by the Federal Judicial Center and the Carnegie Corporation of New York. The Center began its work on a manual to help federal judges deal with scientific evidence in 1990, shortly after the Federal Courts Study Committee recommended the preparation of such a manual. This work was done in furtherance of Center education programs on the subject; the purpose of the manual is to round out the Center's education effort in the area of scientific evidence.

The Center received substantial encouragement from the Task Force on Judicial and Regulatory Decision Making of the Carnegie Commission on Science, Technology, and Government. The interest of the Carnegie Corporation in furthering judicial education in the area of scientific evidence led to the establishment by the Center of a comprehensive program to develop the manual and produce related education programs. Funding by the Carnegie Corporation enabled the Center to support distinguished outside authors to prepare the papers and to have these papers reviewed by experts in science and its use as evidence in litigation.

We are grateful for the encouragement and support by David A. Hamburg, president of the Carnegie Corporation of New York, William T. Golden and Joshua Lederberg, co-chairs of the Carnegie Commission, and David Z. Robinson, executive director of the Carnegie Commission. We have benefited greatly from the advice of Helene Kaplan, chair of the Task Force on Judicial and Regulatory Decision Making. Steven Gallagher and David Beckler also provided valuable encouragement and assistance. We are especially grateful to the authors of the manual for their dedication, and to the many reviewers for their thoughtful suggestions. We would like to thank the staff of the Center's Information Services Office, in particular Rozzie Bell for helping us locate much source material. Finally, we have profited from the advice and assistance of the following members of the Center's Publications & Media Division: Susanna Carey, Geoff Erwin, Amy Hollander, Martha Kendall, and Kris Markarian.

Joe S. Cecil Carol E. Drew Marie Cordisco Dean P. Miletich

## Introduction

The purpose of this manual is to assist judges in managing expert evidence, primarily in cases involving issues of science or technology. Such issues may arise across the entire spectrum of litigation: from mass toxic tort and product liability cases to patent and trademark cases, from medical malpractice cases to contract cases, from environmental, security, and antitrust cases even to criminal cases. The context in which they arise varies widely, but generally they share one characteristic: They challenge the ability of judges and juries to comprehend the issues—and the evidence—and to deal with them in informed and effective ways. As a result, they tend to complicate the litigation, increase expense and delay, and jeopardize the quality of judicial and jury decision making.

Expert evidence<sup>1</sup> has, of course, long been a part of judicial proceedings. People qualified by skill, knowledge, education, or experience have been permitted to testify to help the trier of fact understand the evidence or determine a fact in issue. Increasingly, however, the issues coming before courts are more esoteric and complex. As a result, the resolution of such issues has become more dependent on the help of experts. No longer can judges and jurors rely on their common sense and experience in evaluating the testimony of many experts, as they could when evaluating the testimony of, say, a handwriting expert or an accident reconstructionist. Now they must assess expert testimony on such arcane subjects as the impact of altering genetic material, the toxic quality of littleknown substances, the similarity of computer operating systems, and the matching of DNA samples. The challenge the justice system faces is to adapt its process to enable the participants to deal with this kind of evidence fairly and efficiently and to render informed decisions.

The bedrock of that system is the adversary process, which depends on attorneys to present evidence on behalf of their clients, judges to make the necessary and appropriate rulings concerning admissibility, and juries to resolve disputed issues of fact. But when the adversary process yields conflicting testimony on complicated and unfamiliar issues and the participants cannot fully understand the nature of the dispute, courts may not be competent to make reasoned and principled decisions. Concern over this problem led the Carnegie Commission

<sup>1.</sup> The manual uses the inclusive term *expert evidence* to cover both testimony and nontestimonial evidence, such as demonstrative evidence presented by experts.

on Science, Technology, and Government to undertake a study of science and technology in judicial decision making. In the introduction to its final report, the Commission concluded:

The courts' ability to handle complex science-rich cases has recently been called into question, with widespread allegations that the judicial system is increasingly unable to manage and adjudicate science and technology (S & T) issues. Critics have objected that judges cannot make appropriate decisions because they lack technical training, that jurors do not comprehend the complexity of the evidence they are supposed to analyze, and that the expert witnesses on whom the system relies are mercenaries whose biased testimony frequently produces erroneous and inconsistent determinations. If these claims go unanswered, or are not dealt with, confidence in the judiciary will be undermined as the public becomes convinced that the courts as now constituted are incapable of correctly resolving some of the most pressing legal issues of our day.<sup>2</sup>

One need not fully share the opinions of critics to appreciate the existence of a problem that affects the administration of justice in the decision of particular cases and in the larger dimension of the public's perception of the courts. In 1990 the Federal Courts Study Committee, appointed by the Chief Justice to study the federal courts, noted the increasing importance of economic, statistical, technological, and scientific data and recommended that the judiciary enhance its ability to manage and adjudicate cases involving scientific and technological complexity. The committee specifically recommended that the Federal Judicial Center prepare a manual to assist judges in managing such cases. <sup>3</sup>

The recent decision by the Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc.<sup>4</sup> has heightened the need for judicial awareness of scientific reasoning and methods. In Daubert the Supreme Court held that Rule 702 of the Federal Rules of Evidence requires that to be admissible as "scientific knowledge," scientific testimony "must be derived by the scientific method."<sup>5</sup> "Evidentiary reliability," it explained, "will be based upon scientific validity."<sup>6</sup> The trial judge is assigned a "gatekeeping responsibility" to make "a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."<sup>7</sup>

Such a standard demands an understanding by judges of the principles and methods that underlie scientific studies and the reasoning on which expert evidence is based. This is a task for which few judges are adequately prepared when they arrive on the bench. Without a background in the sciences, many judges

4. 113 S. Ct. 2786 (1993).

<sup>2.</sup> Carnegie Comm'n on Science, Technology, & Gov't, Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges 11 (1993).

<sup>3.</sup> Federal Courts Study Comm., Report of the Federal Courts Study Committee 97 (1990).

<sup>5.</sup> Id. at 2795.

<sup>6.</sup> Id. at 2795 n.9 (emphasis omitted).

<sup>7.</sup> Id. at 2795 n.7, 2796.

find it difficult to master the many areas of expert evidence without neglecting the needs of the remainder of their caseload. This manual is intended to provide judges with quick access to information on specific areas of science in a form that will be useful in dealing with disputes among experts.

The manual is divided into three parts. The first part concerns management and admissibility of expert evidence. The paper on management of expert evidence addresses the need for early awareness of issues about which experts will testify and suggests several strategies under the recently amended Federal Rules of Civil Procedure for assessing the needs of the case, defining and narrowing issues addressed by expert evidence, controlling discovery of experts, and resolving before trial questions concerning admissibility of expert evidence.

The second paper in this part presents a framework for considering challenges to expert evidence by structuring the requirements of the Federal Rules of Evidence into four questions: (1) Is the expert qualified? (2) Is the expert's opinion supported by scientific reasoning or methodology? (3) Is the expert's opinion supported by reliable data? and (4) Is the expert's opinion so confusing or prejudicial that it should be excluded pursuant to Rule 403? This paper also notes emerging issues that courts may be called on to address as they seek to implement the standards of the *Daubert* decision.

The second and most novel part of the manual is composed of reference guides for seven areas of expert testimony-epidemiology, toxicology, survey research, forensic analysis of DNA, statistical inference, multiple regression analysis, and estimation of economic loss. The reference guides are intended to assist judges in identifying the issues most commonly in dispute in these selected areas and in reaching an informed and reasoned assessment concerning the basis of expert evidence. The reference guides do not instruct judges concerning the admissibility of specific types of expert evidence or conclusions of specific scientific studies, and they are not intended to establish minimum standards for acceptable scientific testimony. Instead, they present a primer on the methods and reasoning of selected areas of scientific evidence and suggest a series of questions that will enable judges to identify issues that are likely to be disputed among experts and to explore the underlying basis of proffered evidence. Citations in the guides identify cases in which specific issues were raised to give judges examples of other instances in which judges were faced with similar problems; each guide also contains a list of recommended references.

The authors of the reference guides were selected for their knowledge of substantive areas of science and an awareness of the use of the science as evidence in litigation. The reference guides will be most useful when used as the basis for defining disputes underlying expert evidence. They may be used to aid in the identification and narrowing of disputed issues before trial, to facilitate rulings on the admissibility of expert evidence during a pretrial proceeding, or to help in the drafting of jury instructions. For example, the Reference Guide on Forensic DNA Evidence identifies five pivotal issues and their material elements: the acceptance of the theory and technique of DNA analysis, the quantity and quality of the DNA sample, the performance of the specific sample analysis, the technique used to establish a match in DNA samples, and the statistical method used to estimate the probability of a random match. The judge will be able to use this outline to narrow the dispute, focus the lawyers' arguments, and come to a speedier and more informed ruling.

To inform the parties of the issues the judge is considering, the judge may want to distribute copies of relevant sections of the reference guide. This will also enable parties to direct the judge's attention to issues they believe should be considered, to supplement the material with more recent and specific information, to object to questions that are irrelevant or fail to account for recent developments, and to retain control over the presentation of critical evidence.

These reference guides should not be viewed as science textbooks. They serve the more limited purpose of outlining issues that may arise in litigation and improving the quality of the dialogue between the judge and the parties concerning the basis of scientific evidence. Nor should this manual diminish the role of the jury. The substantive law concerning the standards for the admission of expert evidence is still evolving as the courts interpret and apply *Daubert*. This manual is intended to aid the courts in this process.

The third part of the manual concerns the use of two extraordinary procedures to assist in problems of expert evidence—court-appointed experts and special masters. The Supreme Court in *Daubert* mentioned court-appointed experts as one technique that judges may use when faced with especially difficult expert testimony.<sup>8</sup> Court-appointed experts have traditionally been used to offer testimony at trial. Recently, court-appointed experts have also been used in a variety of pretrial procedures, such as educating judges concerning the fundamental concepts on which the experts differ and offering assessments of the methodology on which the parties' experts are basing their opinions. The paper on courtappointed experts considers the issues involved in using court-appointed experts and offers suggestions for their selection, instruction, and compensation.

Special masters may be appropriate in extraordinary cases in which the demanding nature of the scientific issues is combined with the need for special skills in fact finding. Special masters may also be appointed to conduct settlement negotiations in cases with difficult scientific testimony, or to manage the pretrial stages of cases in which problems of expert testimony may be common. The paper on special masters draws on the lessons learned in other forms of complex litigation to provide models for the use of special masters in cases involving complex scientific evidence. This manual represents an initial attempt to develop information that will aid judges in dealing with complex scientific and technical evidence. This is a difficult topic, and thoughtful observers may differ on the issues that should be addressed in such a manual. We need to learn more about the nature of problems that arise with such evidence and are eager to receive comments and suggestions for improvements in this manual. We also invite suggestions for additional topics that should be addressed. With such assistance we will be able to tailor future editions of the manual to fit the evolving needs of the judiciary.

This manual is intended to complement other manuals prepared by the Center: generic case management techniques are dealt with at length in the *Manual for Litigation Management and Cost and Delay Reduction*,<sup>9</sup> and suggestions for managing litigation that is procedurally complex are found in the *Manual for Complex Litigation*.<sup>10</sup> This manual focuses on the management of expert evidence. The management needs of cases differ; management is not an end in itself but should be designed to bring about the just resolution of cases. Although case management is a judicial responsibility, it is also the responsibility of attorneys, not only to serve their clients well but also to preserve the integrity and credibility of the justice system. This manual is intended to assist all parties to the litigation, attorneys as well as judges.

William W Schwarzer

Introduction

<sup>9.</sup> Manual for Litigation Management and Cost and Delay Reduction (Federal Judicial Center 1992). 10. Manual for Complex Litigation, Third (forthcoming 1995).

# Management of Expert Evidence

William W Schwarzer

**William W Schwarzer**, LL.B., is Director of the Federal Judicial Center and a senior U.S. district judge for the Northern District of California.

## Contents

- I. Introduction 11
- II. The Initial Conference 13
  - A. Assessing the Case 13
  - B. Defining the Issues 15
  - C. Narrowing the Issues—Use of Reference Guides 17
  - D. Limitations or Restrictions on Expert Evidence 19
- III. Use of Magistrate Judges, Special Masters, and Court-Appointed Experts 21
- IV. Discovery and Disclosure 23
  - A. Discovery Control and Management 23
  - B. Protective Orders and Confidentiality 26
  - C. Discovery of Nonretained Experts 27
  - D. Videotape Depositions 27
- V. Motion Practice 29
  - A. Motions in Limine 29
  - B. Summary Judgment 30
- VI. The Final Pretrial Conference 33
  - A. Trial Procedures 34
  - B. Presentation of Evidence 35

## I. Introduction

The purpose of this manual, and this paper in particular, is to assist judges in implementing effective management of expert evidence involving scientific issues. Depending on the nature, novelty, and complexity of such evidence, particular management measures and techniques may be necessary and appropriate. This paper deals with those kinds of measures and techniques. It does not deal with generic case management, or with case management of complex litigation generally, which will also often be necessary in such cases. For example, mass tort cases, which frequently involve scientific evidence, will also require the application of techniques to manage multiparty litigation. Those subjects are beyond the scope of this manual; they are covered in the Federal Judicial Center's Manual for Litigation Management and Cost and Delay Reduction, published in 1992, and the Manual for Complex Litigation, the third edition of which will appear in 1995.

### II. The Initial Conference

#### A. Assessing the Case

The court's first contact with a case will normally be at the initial Rule 16 conference. Note, however, that the attorneys should have previously met, as required by Federal Rule of Civil Procedure 26(f), "to discuss the nature and basis of their claims and defenses... and to develop a proposed discovery plan . . . and [to submit] to the court . . . a written report outlining the plan." Compliance with this "meet and confer" requirement is essential to effective case management. The report, prepared and submitted by the attorneys, together with the pleadings and other available materials, should give the judge useful insight into the case, including information about scientific issues and the likelihood of expert evidence, although this will not invariably be true. In addition, as a result of their conference, the attorneys should be reasonably well informed about the case and should be prepared for the initial conference. Expert testimony, and possible limitations or restrictions on its use, is specifically made a subject for the initial conference, as well as subsequent conferences by Federal Rule of Civil Procedure 16(c)(4).<sup>2</sup> Thus the judge should raise the subject of prospective expert evidence at the conference and begin to explore the issues bearing on it.

The range of subject matter addressed by expert evidence is virtually limitless. It covers the spectrum of the various sciences (both so-called hard and soft sciences), and it extends to other areas of technical or specialized knowledge in which people who have acquired special knowledge, skill, experience, training, or education may be able to give testimony that would assist in the resolution of disputed questions of fact.<sup>3</sup> Surveys indicate that expert testimony comes predominantly from physicians in various specialties, followed by economists, both of which are common in personal injury cases.<sup>4</sup> Engineers also frequently testify,

<sup>1.</sup> See Fed. R. Civ. P. App. of Forms, Form 35 (Report of Parties' Planning Meeting).

<sup>2.</sup> The Advisory Committee Notes state that the rule is intended to "clarify that in advance of trial the court may address the need for, and possible limitations on, the use of expert testimony." Fed. R. Civ. P. 16(c)(4) advisory committee's notes.

<sup>3.</sup> See Fed. R. Evid. 702.

<sup>4.</sup> See Molly Treadway Johnson & Joe S. Cecil, Problems of Expert Testimony in Federal Civil Trials (Federal Judicial Center forthcoming 1995). For a breakdown of experts appearing in state courts, see Anthony

mostly in patent and accident cases. Specialists in other areas of science, such as epidemiology, toxicology, microbiology, and statistics, testify less frequently though often in litigation involving numerous cases and parties. Persons in many other occupations may be offered as experts, such as law enforcement officers and other government agents, mechanics, and technicians.

The nature and degree of judicial management appropriate for the case will vary greatly with its particular circumstances. Much expert evidence can be entirely routine and require little judicial intervention or control. When experts disagree, however, the litigation may become more complicated, resulting in lack of comprehension and added cost and delay. For this reason, the judge should determine early on the nature of the conflict between experts, attempt to define and narrow the issues and initiate appropriate management procedures.

Although this manual is intended to be helpful in different kinds of situations involving expert evidence, its principal focus is on issues of science, where most of the difficulties with expert testimony are encountered. Cases involving issues of science do not necessarily create a unique need for judicial management; testimony from an economist about the extent of lost income due to a plaintiff's injuries is a routine occurrence in litigation. That the court has before it a seemingly ordinary single-plaintiff personal injury case, however, does not foreclose the presence of difficult questions of scientific proof. A medical malpractice case may, for example, present complicated and perhaps novel and controversial questions of the etiology of a cancer. Similarly, a two-party patent case may involve difficult questions concerning the state of the art. Whether a criminal case requires special attention may depend on whether experts use novel or only customary forensic techniques. And some cases may present difficulty if experts rely on nontraditional social science research.

Probably the greatest challenges are presented by multiparty litigation involving toxic torts or environmental harm, including product liability cases. Such cases often, although not necessarily, involve novel and controversial issues in which the science is still evolving and claims and defenses have not yet been shaken out in earlier litigation. Such cases also will impact numerous parties and potential litigants. Judges having cases of this kind need to take care to permit adequate development of emerging scientific issues and prevent the premature foreclosure of what may turn out to be meritorious theories while still performing their "gatekeeping function" with respect to expert evidence under Daubert v. Merrell Dow Pharmaceuticals, Inc.<sup>5</sup>

Champagne et al., Expert Witnesses in the Courts: An Empirical Examination, 76 Judicature 5 (1992), and Samuel R. Gross, Expert Evidence, 1991 Wis. L. Rev. 1113.

<sup>5. 113</sup> S. Ct. 2786, 2798 (1993). The Court stated that before admitting expert testimony the trial court must make a "preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid." *Id.* at 2796. The role of the District Court under the Supreme Court's interpretation of Rule 702 in *Daubert* in determining admissibility of expert testimony is addressed in detail in Margaret A. Berger, Evidentiary Framework §§ I, III, in this manual.

In contrast, large-scale litigation may involve scientific principles or theories that have become so well settled and widely accepted that relitigation may be minimized. These are the so-called mature torts, <sup>6</sup> of which harm caused by asbestos and DES (an anti-morning sickness drug) are examples. Management techniques, such as judicial notice, aggregation, compensation schedules for specific injuries, and other nontraditional means of processing claims, may need to be developed to avoid unnecessary litigation activity.

Cases involving scientific issues do not fall neatly into one or a set of preordained and categorical molds. The initial task for the judge is to determine the management needs of the case in light of all relevant factors, including the apparent characteristics of the prospective scientific issues. The initial assessment, though subject to reexamination and revision as more becomes known about the case and the issues, will guide the judge in defining and narrowing the issues, in discovery control, and in motion practice.

#### B. Defining the Issues

Meaningful case management must begin with defining the issues. Only when the issues are identified and understood can a fair and efficient case management plan be devised. Cases with scientific evidence present particular difficulty because often the parties will operate with inadequate information and the judge will be unfamiliar with the subject matter.

From the judge's perspective, the most effective way to start the process of identifying and defining issues is simply to ask questions. Counsel's responses should be followed by more questions in order to probe deeply into the nature of the claims, the theories of general and specific causation, the defenses, and in particular the bases for disagreement among experts. This process should be viewed as an occasion not for argument but for education, for the judge as well as for the attorneys, who will probably know little about their opponent's case. This approach is important, not only because it is most effective for laying bare the issues, but also because it helps set the right tone for the litigation. Expert witnesses have become intensely adversarial, thereby increasing the difficulty in arriving at fair and informed decisions and undermining civility. Although the litigation process is itself inherently adversarial, there is no reason why the judge should accept contentious advocacy by experts and their counsel at the cost of comprehension, efficiency, and fairness. By approaching the conference in a spirit of civil and enlightened inquiry, the judge can communicate to the participants how he or she expects the litigation to be conducted.

Cases with difficult issues of expert evidence will, of course, also involve traditional legal issues, the management of which will call for conventional case

<sup>6.</sup> See Francis E. McGovern, Toward a Functional Approach for Managing Complex Litigation, 53 U. Chi. L. Rev. 440 (1986).

management practices. Identifying issues in disputes over scientific evidence will be more difficult and complex. In approaching this task, the judge should keep in mind the following considerations:

- Because the attorneys may have difficulty communicating the necessary scientific information to the judge, it may be useful to retain experts (not necessarily prospective witnesses) who can explain the fundamentals necessary for a basic understanding of the subject matter without addressing the specific issues that divide the parties' experts.
- In cases in which the experts have not yet been retained or named as testifying experts and in cases in which expert testimony is an essential element, it may be helpful to defer further proceedings until the necessary expert evidence has been secured and exchanged by the parties. Frequently the parties may not retain experts, at least to testify at trial, until later in the litigation. (This can be for a number of reasons, such as the expectation that the case will settle, lack of sufficient familiarity with the facts, or difficulty in finding a suitable expert.) Sometimes parties retain experts as consultants and defer the decision to name them as testifying experts. The effect of such a delay depends on the role of the expert in the case. In some cases, the expert merely embellishes testimony of percipient witnesses; the expert's participation in the pretrial phase is therefore not critical to issue definition. In other kinds of cases, however, the expert is crucial to the case; this is true, for example, in medical malpractice litigation in which only an expert witness can supply the evidence of failure to conform to the applicable standard of practice, an essential element in a plaintiff's case.
- When experts have been retained and their positions are generally known, the critical task is to begin to identify the issues that divide opposing experts. In science-rich cases, it is likely that experts will have played a part in the preparation of the claims and defenses, and their theories can therefore be identified early in the litigation. If the process of issue definition is to be effective, it should not stop with a general statement of the experts' disagreement. The court should, with the assistance of the parties, probe deeper to identify the bases for their differences. Experts will often express diametrically opposed opinions on crucial issues in the case without explaining or disclosing the bases for their differences. Closer examination of the bases of their respective positions may well disclose that their differences are the products of different starting points. For example, experts may reason from different statistical or other databases or assumptions, leading them to different conclusions. If the controversy can be reduced to one about the appropriate selection of foundation data, it will be much more susceptible to a reasonable resolution. Experts may also operate from widely differing philosophical or

policy premises, such as the limits of acceptable risk or the nature of unacceptable harm. Finally, expert opinions may be the product of research or testing procedures, which, once disclosed, can be independently and objectively evaluated for adequacy.

Federal Rule of Civil Procedure 26(a)(2) establishes a procedure under which each party must, not less than ninety days before the trial date or at such other time as the judge may order, make detailed written disclosure with respect to each expert witness retained to testify at trial, including "a complete statement of all opinions to be expressed and the basis and reasons therefor [and] the data or other information considered by the witness in forming the opinions."7 Having those disclosures at hand should assist the parties and the judge in the process of identifying and narrowing issues. The time necessary for the parties to comply with the requirements of the rule, however-assembling all of the data and preparing complete written reports—is likely to delay the start of this process. The judge must consider how to make the most efficient use of this rule in each case. In most cases, however, the judge should be advised not to delay issue identification (particularly because a settlement may occur before the parties have incurred the expense of hiring experts and preparing their reports), but after disclosure has been completed, to consider further efforts to define and narrow the issues concerning expert evidence.

#### C. Narrowing the Issues—Use of Reference Guides

The process of defining issues should lead to the narrowing of issues. Some elements of the case may turn out not to be in dispute. For example, there may be no controversy about the plaintiff's exposure to the allegedly harmful substance, allowing that issue to be eliminated. Conversely, the plaintiff's ability to establish the requisite exposure may appear to be so questionable that it might usefully be singled out for early targeted discovery<sup>8</sup> and a possible motion for summary judgment.<sup>9</sup> Unless the judge takes the lead in probing for issues that may not be in dispute, or that may lend themselves to early resolution, the case is likely to involve much unnecessary work, cost, and delay.

The conclusions of a witness offering scientific testimony will generally be the product of a multistep reasoning process. By breaking down the process, the judge may be able to narrow the dispute to a particular step in the process, and

<sup>7.</sup> Fed. R. Civ. P. 26(a)(2)(B). Some courts have adopted alternative procedures. For a list of courts that have opted out of the provisions of Rule 26(a)(2), see Donna Stienstra, Implementation of Disclosure in Federal District Courts, with Specific Attention to Courts' Responses to Selected Amendments to Federal Rule of Civil Procedure 26 (Federal Judicial Center 1994).

<sup>8.</sup> Manual for Complex Litigation, Third, § 21.424 (forthcoming 1995) [hereinafter MCL 3d].

<sup>9.</sup> See, e.g., Celotex Corp. v. Catrett, 477 U.S. 317 (1986).

thereby facilitate its resolution. Those steps, while generally not intuitively obvious to the non-expert, may be identified in the process of issue identification. Once that is done, it can readily be determined which steps are in dispute. As noted, the initial Rule 16 conference may be too early for the parties to be adequately prepared for this process. Nevertheless, the stage should at least be set for the narrowing of issues, though the process may continue as the litigation progresses.

The reference guides in this manual are intended to assist in the process of narrowing issues in the areas they cover. <sup>10</sup> By way of illustration, the Reference Guide on Forensic DNA Evidence facilitates narrowing a dispute over whether proffered evidence may be received by dividing an issue into five distinct subsidiary issues:

- 1. the validity of RFLP (Restricted Fragment Length Polymorphism) analysis;
- 2. the quantity and quality of the specific forensic sample;
- 3. the proficiency and quality control of the laboratory;
- 4. the comparison of DNA profiles; and
- 5. the estimation of the probability that the DNA profiles match by coincidence.

For each subsidiary issue, there is a series of suggested questions that will enable the judge to explore the methodology and reasoning underlying the expert's opinion.

The remaining reference guides cover additional areas in which expert evidence is frequently offered and disputed:

- The Reference Guide on Epidemiology identifies issues concerning the appropriateness of the research design, the definition and selection of the research population, the measurement of exposure to the putative agent, the measurement of the association between exposure and the disease, and the assessment of the causal association between exposure and the disease.
- The Reference Guide on Toxicology identifies issues concerning the nature and strength of the research design, the expert's qualifications, the proof of association between exposure and the disease, the proof of causal relationships between exposure and the disease, the significance of the person's medical history, and the presence of other agents.
- The Reference Guide on Survey Research identifies issues concerning the purpose of the survey and the method of its design, selection of the population and sample and assessment of the responses, design of ques-

<sup>10.</sup> The reference guides are not intended to be primers on substantive issues of scientific proof or normative statements on the merits of scientific proof.

tions, selection of the control group, interviews, data entry, and disclosure and reporting.

- The Reference Guide on Statistics identifies three issues: the design of the data collection process, the extraction and presentation of relevant data, and the drawing of appropriate inferences.
- The Reference Guide on Multiple Regression identifies issues concerning the analysis of data bearing on the relationship of two or more variables, the presentation of such evidence, the research design, and the interpretation of the regression results.
- The Reference Guide on Estimation of Economic Losses in Damage Awards identifies issues concerning expert qualification, characterization of the harmful event, measurement of loss of earnings before trial and future loss, prejudgment interest, and related issues generally and as they arise in particular kinds of litigation.

The scope of these reference guides is necessarily limited, but their format is intended to suggest analytical approaches and opportunities that judges may use in identifying and narrowing issues presented by controversies over scientific evidence. A judge may, for example, ask counsel for both sides to exchange and provide to the court a step-by-step outline of the experts' reasoning processes (following generally the pattern of the reference guides) for use at the conference at which issue definition and narrowing is discussed. If the written statements of expert opinions required by Federal Rule of Civil Procedure 26(a)(2) have been exchanged, the judge could direct each side to identify specifically each part of the opposing expert's opinion that is disputed and to state the specific basis for the dispute. A further conference should then be held after receipt of these statements to attempt to narrow the issues.

#### D. Limitations or Restrictions on Expert Evidence

As noted, Federal Rule of Civil Procedure 16(c)(4) specifically makes "the avoidance of unnecessary proof and of cumulative evidence, and limitations or restrictions on the use of testimony under Rule 702 of the Federal Rules of Evidence" a subject for consideration and appropriate action by the court at any conference. The timing of such action will depend on the circumstances of each case. Not enough may be known at the initial conference for judicial action, although it may be clear that on certain issues on which expert testimony is proposed, the trier of fact should have no need for such assistance. As issues are defined and narrowed, the judge should consider whether expert evidence will aid the trier of fact on specific issues and should at least indicate tentative views based on the information provided, which are subject to revision if further information makes that appropriate. As issues are eliminated, the need for expert testimony on those issues is also eliminated. Experts increase the cost of litiga-

tion substantially, and permitting their proliferation in a case may place an unfair burden on the party with limited resources.

The judge should also consider the number of expert witnesses permitted to testify. Some local rules and orders limit a party to a single expert on a particular scientific discipline, that is, a single orthopedist, oncologist, or rehabilitation specialist. The judge may place the burden of showing necessity for additional experts on the party proposing to offer them. In cases in which multiple parties are litigating the same issue or in consolidated cases, duplication of expert testimony can be avoided, both by limiting the parties on one side to one expert per discipline and by avoiding repetition of the same testimony on multiple occasions.

In determining the need for expert testimony in the case, the judge should also consider whether the same issues have been previously tried and adjudicated. Scientific or technological facts may have become sufficiently well established to warrant taking judicial notice. Res judicata or collateral estoppel may be available to foreclose particular issues, or expert testimony from earlier cases may be directly on point and available for use in the case, at least on stipulation.<sup>11</sup>

11. MCL 3d, supra note 8, § 21.33.

## III. Use of Magistrate Judges, Special Masters, and Court-Appointed Experts

Federal Rule of Civil Procedure 16(c)(8) makes the referral of matters to a magistrate judge or a special master a subject for consideration at the conference. Although the rule does not specifically refer to court appointment of experts, subsection (c)(12) does call for consideration of "the need for adopting special procedures for managing potentially difficult . . . actions that may involve complex issues . . . or unusual proof problems." Cases involving scientific evidence may confront the court with the need to look for assistance.<sup>12</sup>

Many courts routinely refer the pretrial management of civil cases to magistrate judges. Some judges believe, however, that in complex cases, there are advantages in having pretrial management performed by the judge who will try the case; this promotes familiarity with the issues in the case and avoids the delay caused by appeals of magistrate judge rulings. <sup>13</sup> If pretrial management is nevertheless referred to a magistrate judge, he or she should keep the judge who will try the case apprised of developments affecting the complex issues in the case. A need for decisions by the trial judge may arise during the pretrial phase; for example, the decision to appoint an expert under Federal Rule of Evidence 706 or a special master under Federal Rule of Civil Procedure 53 is one the trial judge would have to make and therefore should not be deferred until the eve of trial.

The Supreme Court has taken a restrictive view of the trial judge's power to refer matters to a special master; reference to a special master under Rule 53(b) "shall be the exception and not the rule."<sup>14</sup> Nevertheless, masters have performed substantial services in complex litigation, including resolving privilege claims in massive document production, analyzing damage and other accounting data, and assisting in settlement negotiations. Appointment of a special master saddles the parties with additional and often substantial expense, however, and may therefore be expected to be viewed critically by appellate courts. <sup>15</sup>

<sup>12.</sup> For a discussion of issues surrounding the decision of a judge to invoke such assistance, see Jack B. Weinstein, Ethical Dilemmas in Mass Tort Litigation, 88 Nw. U. L. Rev. 469 (1994).

<sup>13.</sup> MCL 3d, supra note 8, § 21.53.

<sup>14.</sup> See La Buy v. Howes Leather Co., 352 U.S. 249, 256-58 (1957).

<sup>15.</sup> Prudential Ins. Co. of Am. v. United States Gypsum Co., 991 F.2d 1080, 1085 (3d Cir. 1993). For guidance with respect to the appointment and use of special masters in cases with scientific evidence, see Margaret G. Farrell, Special Masters, in this manual.

Under Rule 706, the court may on its own motion or the motion of a party appoint an expert witness. The court may appoint a person agreed on by the parties or make its own selection. Since the courts have no funds with which to compensate witnesses, the cost of a court-appointed expert is typically borne by the parties. The appointment of an expert may be for different purposes: it may be to testify, or it may be only to assist the judge in other ways in dealing with scientific issues. <sup>16</sup> Thus the functions of a court-appointed expert and those of a special master may well overlap. If the expert is to testify, it may be on an ultimate issue in the case or only on subsidiary scientific issues, such as the validity or reliability of methodology used by the parties' experts. <sup>17</sup> The timing of the decision whether to make an appointment can be critical. The appointment of an expert made too soon can result in needless expense; if an appointment is made too late, it may not be possible to locate, appoint, and instruct an expert without delaying the litigation.<sup>18</sup>

Reference Manual on Scientific Evidence

<sup>16.</sup> See In re Swine Flu Immunization Prods. Liab. Litig., 495 F. Supp. 1185 (W.D. Okla. 1980) (order appointing panel of medical experts to examine claimants and report to court).

<sup>17.</sup> See, e.g., Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1548 (D. Colo. 1990) (court-appointed expert testified to methodology used by plaintiffs to prove exposure to contaminated water), aff'd, 972 F.2d 304 (10th Cir. 1992).

<sup>18.</sup> For guidance with respect to the appointment and use of such experts, see Joe S. Cecil & Thomas E. Willging, Court-Appointed Experts, in this manual.

## **IV.** Discovery and Disclosure

#### A. Discovery Control and Management<sup>19</sup>

If the judge has the parties' report on their prediscovery conference and has their discovery plan in hand, as noted, he or she will be well situated to establish control over discovery. The basic control mechanism for testifying experts is provided by Federal Rule of Civil Procedure 26(b)(4)(A), which states that parties are entitled to depose experts identified as trial witnesses but may do so only after the expert's report under Federal Rule of Civil Procedure 26(a)(2)(B) has been provided if one is required.<sup>20</sup> That report may be dispensed with by order of the court or stipulation of the parties.<sup>21</sup> While the court probably cannot preclude the parties from entering into such a stipulation,<sup>22</sup> under its inherent power it may be able to override a stipulation and order the disclosures called for by Rule 26(a)(2)(B).<sup>23</sup> There are compelling reasons for requiring these disclosures with respect to expert witnesses:

• The process of complying with Rule 26(a)(2)(B) will compel attorneys to consider carefully whether to designate an expert as a witness at all, because of the need to fully prepare the witness before disclosure, the risk

19. With respect to discovery control and management, see generally MCL 3d, supra note 8, § 21.4.

21. Fed. R. Civ. P. 26(a)(2)(C). The report under Fed. R. Civ. P. 26(a)(2)(B) is presumptively required of any "witness who is retained or specially employed to provide expert testimony in the case or whose duties as an employee of the party regularly involve giving expert testimony." This would normally exclude a treating physician. The court may by order, or the parties may by stipulation, exempt a case from this requirement.

22. Fed. R. Civ. P. 29 gives the parties the right to modify, without court order, the procedures or limitations governing discovery except for stipulations that would interfere with any time set for completion of discovery, hearing of a motion, or trial.

23. In addition to disclosing the identity of any person who may be used as an expert witness, a party must also disclose

a written report prepared and signed by the witness. The report shall contain a complete statement of all opinions to be expressed and the basis and reasons therefor; the data or other information considered by the witness in forming the opinions; any exhibits to be used as a summary of or support for the opinions; the qualifications of the witness, including a list of all publications authored by the witness within the preceding ten years; the compensation to be paid for the study and testimony; and a listing of any other cases in which the witness has testified as an expert at trial or by deposition within the preceding four years.

Fed. R. Civ. P. 26(a)(2)(B).

<sup>20.</sup> In addition, Fed. R. Civ. P. 26(b)(2) gives the court broad authority to limit the frequency and extent of discovery, including the length of depositions.

of having to disclose the attorney's work product communicated to the witness, and the expense of preparing the requisite report and data;

- The information and materials required to be disclosed can facilitate the definition and narrowing of issues, both by enhancing the attorneys' preparation and by providing the judge with necessary information;
- Examination of the opposing expert witness's report may well lead to a decision that a deposition would serve no useful purpose; if a deposition is taken, however, having the report will expedite it;
- The disclosures will assist the court in making informed rulings limiting or restricting expert testimony;
- The disclosures will help counsel prepare for effective cross-examination and reduce the risk of surprise at trial, which often leads to delay and increased expense; and
- The disclosures may promote early settlement.

Thus, by following the scheme of the Federal Rules, the court will be able to reduce unnecessary discovery activity, control other activity directed at expert witnesses, and advance effective case management. In the scheduling order issued in connection with the initial conference, the court should prescribe the sequence and timing of these disclosures; generally the party with the burden on an issue should make its disclosure before other parties are required to make theirs on that issue.

Compliance with Rule 26(a)(2)(B) requires disclosure not only of data or information on which the expert relied in reaching the opinions but also of all data and material "considered by the witness in forming the opinions." As a result, "litigants should no longer be able to argue that materials furnished to their experts to be used in forming their opinions—whether or not ultimately relied upon by the expert—are privileged or otherwise protected from disclosure when such persons are testifying or being deposed." <sup>24</sup>

The obligation of disclosure under the rule highlights the importance of protecting and preserving records, documents, and other materials in the possession or under the control of the parties. Notes and records of tests and experiments that cannot be duplicated are an illustration of material of potentially crucial importance in cases with scientific evidence. The court may therefore want to consider the prompt issuance of an order providing for the preservation and nondestruction of documents and other materials potentially relevant to the litigation. Such an order should only be entered after consultation with counsel, and it should take into account the need to accommodate normal retention policies.<sup>25</sup>

Compliance with the rule also requires that the expert's report, as well as any information provided by the expert through a deposition, be supplemented if the

<sup>24.</sup> Fed. R. Civ. P. 26(a)(2)(B) advisory committee's notes.

<sup>25.</sup> MCL 3d, supra note 8, § 21.442.

party learns that the information so disclosed is in some material respect incomplete or incorrect (even if it was complete and correct when initially provided). Since it is not uncommon for an expert to modify an opinion in the course of litigation, the parties need to be reminded of their obligation to give timely notice to the other side. The court's scheduling order should make provision for periodic review and updates of discovery responses and disclosures.

Discovery by deposition or interrogatory may be directed at nontestifying experts, that is:

> an expert who has been retained or specially employed by another party in anticipation of litigation or preparation for trial and who is not expected to be called as a witness at trial [but] only as provided in Rule 35(b) [relating to physical or mental examinations] or upon a showing of exceptional circumstances under which it is impracticable . . . to obtain facts or opinions on the same subject by other means.<sup>26</sup>

The purpose of this restriction is to avoid penalizing a party that has sought expert assistance early in the litigation and to prevent the opponent from gaining the benefit of the other side's diligence. Exceptional circumstances may arise, however, where an expert, for example, has conducted destructive tests relevant to the issues but incapable of being repeated or where one side has retained all qualified experts.27

Use of court-appointed experts also raises difficult issues concerning discovery.<sup>28</sup> An expert appointed to testify as a witness under authority of Federal Rule of Evidence 706 is subject to deposition by any party under terms of the rule.<sup>29</sup> But when the expert is appointed as a technical advisor under the inherent authority of the court, there is no right to depose the expert.<sup>30</sup> The opportunity for discovery of an expert is less clear when the expert is appointed under Rule 706 and is not only offering testimony as a witness but also serving as a technical advisor. To the extent that the duties of the appointed expert depart from those of a testifying witness, courts have found that the appointment is similar to that of a technical advisor and have restricted the opportunity for discovery of the expert. 31

Rule 26(b)(4)(C) also requires payment of a reasonable fee to an expert for time spent responding to discovery and, in the case of a nontestifying expert, also

29. Fed. R. Evid. 706(a) ("[T]he [court-appointed] witness' deposition may be taken by any party; and the witness may be called to testify by the court or any party.")

30. Reilly v. United States, 863 F.2d 149, 154–56 (1st Cir. 1988). 31. Renaud v. Martin Marietta Corp., 972 F.2d 304, 308 n.8 (10th Cir. 1992); In re Joint E. & S. Dists. Asbestos Litig., 151 F.R.D. 540, 544 (E.D.N.Y. 1993), appeal dismissed, 14 F.3d 151 (2d Cir. 1994).

Management of Expert Evidence

<sup>26.</sup> Fed. R. Civ. P. 26(b)(4)(B).

<sup>27.</sup> For a discussion of discovery directed at experts appointed by the court under Fed. R. Evid. 706 or at special masters appointed under Fed. R. Civ. P. 53, see Joe S. Cecil & Thomas Willging, Court-Appointed Experts § V.C, and Margaret G. Farrell, Special Masters § II.B, in this manual.

<sup>28.</sup> Since special masters perform many of the duties of a judge, including oversight of discovery, the right of discovery concerning information considered by a special master is quite limited. Nevertheless, the order appointing the special master may specify the extent of access to information supporting the master's findings. See Margaret G. Farrell, Special Masters § IV.C, in this manual.

of a fair portion of the expenses incurred by the opposing party in obtaining facts and opinions from the expert. Expert discovery in science-rich cases may have other costly aspects, such as making computer runs or performing tests. The court has authority under Rule 26(c)(2) to condition such discovery upon payment of expenses by the party who should be appropriately charged.<sup>32</sup>

#### B. Protective Orders and Confidentiality

Protective orders may become an issue in expert discovery in two ways: a party may seek to bar public disclosure of matters disclosed in the course of an expert's deposition, or a party may seek access to discovery material from related litigation under protection of an order previously issued.<sup>33</sup>

Rule 26(c)(5) permits a court, on motion of a party or of the person from whom discovery is sought, and after the parties have conferred to attempt in good faith to resolve the dispute, to issue a protective order for good cause shown and as justice requires. A protective order may, among other things, bar disclosure of discovery (including limiting a person's presence at the deposition or other information. The rule specifically authorizes an order to protect trade secrets or other confidential research, development, or commercial information. When the information to be protected cannot be conveniently isolated from other information, the court may issue an umbrella order covering the entire deposition, subject to later order releasing information not entitled to protection. Umbrella orders expedite discovery and reduce disputes, but they can be controversial, as when requests are made for the release of information covered by the order. Since the order was entered without a particularized showing of need, little showing is required to obtain modification.<sup>34</sup>

Commonly, parties stipulate to such orders, in which case the question arises whether they can deny access by third parties to the information. Discovery materials that have not been used in trial or court proceedings are not subject to the public's First Amendment right of access.<sup>35</sup> However, the practice of sealing the record of a case as a part of a negotiated settlement is coming under increasing scrutiny.<sup>36</sup> While a guarantee of confidentiality facilitates settlement, it collides with other policy considerations, such as the interest in access to data affecting

34. In re "Agent Orange" Prod. Liab. Litig., 104 F.R.D. 559, 568–70 (E.D.N.Y. 1985), affd, 821 F.2d 139 (2d Cir.), cert. denied, 484 U.S. 953 (1987).

<sup>32.</sup> See MCL 3d, supra note 8, § 21.422.

<sup>33.</sup> MCL 3d, supra note 8, § 21.43.

<sup>35.</sup> Seattle Times Co. v. Rhinehart, 467 U.S. 20 (1984).

<sup>36.</sup> See Anne-Therese Bechampes, Note, Sealed Out-of-Court Settlements: When Does the Public Have a Right to Know?, 66 Notre Dame L. Rev. 117 (1990). See also Arthur R. Miller, Confidentiality, Protective Orders, and Public Access to the Courts, 105 Harv. L. Rev. 428 (1991); Richard L. Marcus, Myth and Reality in Protective Order Litigation, 69 Cornell L. Rev. 1 (1983).

public health and safety and assisting other litigation, government regulatory efforts, and public information.<sup>37</sup>

These considerations are relevant to the second prong of the issue: gaining access to discovery material in related litigation. Obtaining material such as the earlier deposition of an expert in the pending case may avoid duplicative discovery.<sup>38</sup> An analogous situation is presented in multidistrict litigation, in which transferee courts have vacated protective orders previously entered by a transferor court.<sup>39</sup>

#### C. Discovery of Nonretained Experts

A need for information in cases with scientific evidence may lead parties to seek discovery by subpoena from experts who have not been retained in the litigation. Federal Rule of Civil Procedure 45(c)(3)(B)(i) permits the court to quash a subpoena that "requires disclosure of an unretained expert's opinion or information not describing specific events or occurrences in dispute and resulting from the expert's study made not at the request of any party." However, if the party seeking the information shows a substantial need for it that cannot be otherwise met without undue hardship and assures that the person subpoenaed will be reasonably compensated, the court may order compliance under specified conditions. As the Advisory Committee Notes point out, this provision was intended to protect the intellectual property of nonretained experts: "The rule establishes the right of such persons to withhold their expertise, at least unless the party seeking it makes the kind of showing required for a conditional denial of a motion to quash . . . ; that requirement is the same as that necessary to secure work product under Rule 26(b)(3) and gives assurance of reasonable compensation."<sup>40</sup>

#### D. Videotape Depositions

Federal Rules of Civil Procedure 30(b)(2) and (3) permit a party, unless otherwise ordered, to record a deposition by audiotape, videotape, or stenographic means; any other party may designate on notice any other method to record the deposition in addition to the method specified by the person taking the deposition.<sup>41</sup> Videotape can be particularly useful for taking an expert's deposition in the following instances:

<sup>37.</sup> Legislation expanding public access has been adopted in some states and is under consideration in others and in Congress.

<sup>38.</sup> For orders granting access to previously discovered materials, see Wilk v. American Medical Ass'n, 635 F.2d 1295, 1301 (7th Cir. 1980); Cipollone v. Liggett Group, Inc., 785 F.2d 1108, 1121–23 (3d Cir. 1986). See Marcus, *supra* note 36, at 41–53.

<sup>39.</sup> In re Upjohn Co. Antibiotic Cleocin Prods. Liab. Litig., 664 F.2d 114 (6th Cir. 1981).

<sup>40.</sup> Fed. R. Civ. P. 45(c)(3)(B)(ii) advisory committee's note. See In re American Tobacco Co., 880 F.2d 1520, 1527 (2d Cir. 1989); see also Mark Labaton, Note, Discovery and Testimony of Unretained Experts, 1987 Duke L.J. 140, and Richard L. Marcus, Discovery Along the Litigation/Science Interface, 57 Brook. L. Rev. 381 (1991).

<sup>41.</sup> See MCL 3d, supra note 8, § 21.452.

- An expert may become unavailable for the trial because of other commitments, and a subpoena may be neither feasible nor desirable; videotape will provide a more interesting and meaningful presentation at trial than reading the transcript.
- The expert's testimony may be needed at separate trials in multiparty litigation or where the litigation has been bifurcated and the testimony is relevant to both phases.
- The expert's testimony may relate to matters that can be demonstrated on videotape but not in court, such as the operation of large equipment, the physical characteristics of a location, the conduct of a test, or the reconstruction of an accident; videotape permits the witness to point out relevant matter and illustrate the testimony.

When such depositions are contemplated, problems concerning their use at trial should be resolved before they are taken.

## V. Motion Practice

Scientific evidence raises two issues that may be addressed by motions:

- 1. admissibility under the rules of evidence; and
- 2. sufficiency as a matter of law to sustain a verdict for the proponent.

The two issues tend to become intertwined in the course of litigation but need to be considered separately. The exclusion of proffered evidence does not necessarily entitle the objector to judgment, although the result may be ultimately to leave the proponent unable to prove an essential element of its case. Even if admitted, however, the evidence may be legally insufficient, warranting entry of judgment as a matter of law before or at trial.<sup>42</sup>

Whether the ruling is on admissibility arising from a motion in limine or on summary judgment, the order should state the judge's findings (where appropriate) and reasons. Because such a ruling is likely to be reviewed on appeal, the court should provide a clear and complete statement of its legal and factual basis. The parties and the appellate court should not be left to guess which of several potentially applicable rules the court relied on and how it determined the factual issues.<sup>43</sup>

#### A. Motions in Limine

Objections to evidence raised before trial are best presented by a motion in limine under Federal Rule of Evidence 104(a). In its recent decision in *Daubert*, the Supreme Court stated:

Faced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset, pursuant to Rule 104(a), whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact in issue.<sup>44</sup>

<sup>42.</sup> See generally William W Schwarzer et al., The Analysis and Decision of Summary Judgment Motions: A Monograph on Rule 56 of the Federal Rules of Civil Procedure (Federal Judicial Center 1991). See also Margaret A. Berger, Evidentiary Framework § I.C.3, in this manual.

<sup>43.</sup> See In re Paoli R.R. Yard PCB Litig., 916 F.2d 829, 836 & n.3 (3d Cir. 1990), cert. denied, 499 U.S. 961 (1991).

<sup>44.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993) (footnotes omitted). Issues concerning the admissibility of such evidence are discussed at length in Margaret A. Berger, Evidentiary Framework § I.C.2, in this manual.

Rule 104(a) is the court's vehicle for determination of preliminary questions concerning the qualifications of a witness, the existence of a privilege, or the admissibility of evidence. The court may, if necessary, conduct a hearing (which must be outside the hearing of the jury), and it is not bound by the rules of evidence.<sup>45</sup> When the admissibility of expert evidence is pivotal to a motion for summary judgment, a Rule 104(a) hearing should precede consideration of the motion.<sup>46</sup> A ruling on admissibility may also be important in jurisdictions where the court may be precluded from granting judgment as a matter of law after trial on the ground that it had erroneously admitted expert testimony.<sup>47</sup>

By requiring the parties to follow the disclosure procedure under Federal Rule of Civil Procedure 26(a)(2), the court will have before it the complete statement of the opinions to which the expert will testify and their factual basis. This material, supplemented by memoranda addressed to the evidentiary issues, will provide a helpful record for rulings under Rule 104(a).<sup>48</sup>

#### B. Summary Judgment

The exclusion of critical expert evidence may leave the party bearing the burden of proof unable to prove an essential element of its case, thus laying the foundation for summary judgment;<sup>49</sup> or critical expert evidence may be so conclusory that it fails to raise a genuine issue of fact. As the Court stated in *Daubert*:

Additionally, in the event the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true, the court remains free to direct a judgment, Fed. Rule Civ. Proc. 50(a), and likewise to grant summary judgment, Fed. Rule Civ. Proc. 56.

At the initial and subsequent Rule 16 conferences, the court should consider whether a summary judgment motion is appropriate and, if so, when it should be made.<sup>51</sup> Discussion with counsel of the bases for a proposed summary judgment can forestall the filing of motions, which, because they implicate disputed facts, are a waste of resources.<sup>52</sup> Timing is important because if the motion is

<sup>45.</sup> Fed. R. Evid. 104(a), (c).

<sup>46.</sup> In re Paoli, 916 F.2d at 837, 854–55 (proponent of expert witness entitled to notice of grounds for exclusion and opportunity to remedy deficiency).

<sup>47.</sup> See Jackson v. Pleasant Grove Health Care Ctr., 980 F.2d 692, 695–96 (11th Cir. 1993).

<sup>48.</sup> For a discussion of the burden of demonstrating the need for a hearing under Rule 104(a) concerning deficiencies in expert testimony, see the discussion of judicial screening in Margaret A. Berger, Evidentiary Framework § I.C.2, in this manual.

<sup>49.</sup> See Celotex Corp. v. Catrett, 477 U.S. 317 (1986).

<sup>50.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2798 (1993).

<sup>51.</sup> Fed. R. Civ. P. 16(c)(5).

<sup>52.</sup> See Edward Brunet, The Use and Misuse of Expert Testimony in Summary Judgment, 22 U.C. Davis L. Rev. 93 (1988).
made too early, it may lack the necessary record for decision; if the motion is delayed, it loses the potential benefit of reducing cost and delay.<sup>53</sup>

When a summary judgment motion is properly supported, Federal Rule of Civil Procedure 56(e) requires the opposing party to present "specific facts [that would be admissible in evidence] showing that there is a genuine issue for trial." 54 Summary judgment motions turning on the sufficiency of scientific proof raise the question whether an expert's opinion may satisfy the requirement of Rule 56(e). Federal Rule of Evidence 705, as amended in 1993, permits an expert to testify "in terms of opinion or inference and give reasons therefor without first testifying to the underlying facts or data, unless the court requires otherwise." The purpose of the rule is to eliminate the much criticized practice of asking experts hypothetical questions, leaving it to cross-examination at trial to bring out relevant facts. <sup>55</sup> That purpose does not support importing the rule into summary judgment practice, and the rule's text, as revised in 1993, makes clear that the expert can be required to disclose the factual basis for an opinion. Conclusory expert affidavits therefore will not be sufficient to meet the burden on the party opposing the motion,<sup>56</sup> although an affidavit stating an adequately supported opinion may suffice to raise a triable issue.<sup>57</sup>

54. Under Fed. R. Evid. 703, an expert may base an opinion on hearsay evidence "[i]f of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject."

56. See Mendes-Silva v. United States, 980 F.2d 1482, 1488 (D.C. Cir. 1993).

57. Bulthuis v. Rexall Corp., 789 F.2d 1315, 1356–57 (9th Cir. 1985).

Management of Expert Evidence

<sup>53.</sup> See Celotex, 477 U.S. at 322 (the opponent of the motion is entitled to "adequate time for discovery" needed to oppose the motion); William W Schwarzer & Alan Hirsch, Summary Judgment After Eastman Kodak, 45 Hastings L.J. 1, 17 (1993). The disclosures required under Fed. R. Civ. P. 26(a)(2) should help in developing an adequate record.

<sup>55.</sup> Fed. R. Evid. 705 advisory committee's note.

## VI. The Final Pretrial Conference

The manner in which judges use pretrial conferences differs widely, and this manual offers no prescription for their effective use. A judge may conduct a series of conferences between the initial conference and the final pretrial conference or leave all unfinished business until the final conference. What is important is that the management issues affecting expert evidence be addressed and disposed of in the most effective manner appropriate for the case. The desired objective is that, if the case does not settle, the parties be fully prepared for trial and the trial be free of wasted effort.

Much of the subject matter discussed in connection with the initial conference may, as noted, carry over to subsequent conferences, including the final pretrial conference. Even if progress was made at the initial conference in the defining and narrowing of issues, developments during the discovery phase of the case will enlarge the parties' information and refine their positions. New issues may appear and others may disappear. It is therefore critical that the judge continue the effort to define and narrow issues and that the final pretrial conference result in a definitive statement of the issues to be tried.

The court may want to consider a number of possible techniques to identify and narrow the differences between opposing experts, including the following:

- Have each party mark for the opposition the parts of the opposing expert's report with which they agree and disagree, and indicate critical issues that the opposing expert has not addressed;
- Direct counsel to have their experts meet and prepare a joint statement summarizing the bases for their disagreement;
- Convene a conference attended by experts and counsel to identify and attempt to narrow the bases for their differences, leading to an appropriate preliminary instruction to the jury; and
- Explore the possibility of a joint report by the experts.

The final pretrial order should state clearly and specifically the issues of scientific evidence to be tried; it should include a preclusion order barring expert evidence not previously disclosed; <sup>58</sup> and it should make provision for trial proce-

<sup>58.</sup> Fed. R. Civ. P. 37(c)(1).

dures appropriate for the case that will enhance comprehension and expedite the trial (see the following section for discussion).

The final pretrial conference offers the last clear chance for settlement. Cases frequently settle at this stage, when the parties are fully informed about their case and their opponent's case. In cases with difficult scientific evidence, the court may want to consider appointing a mediator with relevant experience and expertise to conduct settlement negotiations. The court may also want to explore various alternative dispute resolution procedures.

In trials involving scientific evidence, the court and the parties are confronted with particular challenges, arising from the difficulties of presenting the case in a comprehensible and efficient manner. Techniques for enhancing comprehension and avoiding unnecessary cost and delay are generally known; while for the most part such techniques are not novel, they are not as widely used as they might be.<sup>59</sup> Judges as well as attorneys tend to resist change in their accustomed ways of doing things and often are disinclined to risk innovation even when the need for reform is demonstrable. What follows is a brief summary of the principal techniques judges have found useful in enhancing comprehension of the case and improving efficiency.

#### A. Trial Procedures

- Structure the trial. The trial may be bifurcated, separating the trial of issues, such as general causation, specific causation, and damages; or the trial may be structured to try one issue at a time, where the jury returns a verdict before the trial resumes; or the jury may be directed to return seriatim verdicts at the end of the trial, thereby deliberating on only one issue at a time.<sup>60</sup>
- Limit the scope of the trial. The judge can limit the scope of the trial by limiting the number of expert witnesses to avoid duplicate or unnecessary proof. Any reduction in the volume of proof presented to jurors will enhance their capacity to comprehend.
- Limit the length of the trial. Similarly, the judge can place limits on the amount of time allowed each side for direct examination and cross-examination. This, too, will reduce the volume of proof and enhance comprehension.
- Arrange a tutorial for the judge and jury before the trial begins, conducted by neutral experts or experts chosen by the parties, to explain noncontroversial fundamentals of complex scientific issues.

<sup>59.</sup> See generally MCL 3d, supra note 8, §§ 21.6, 22.2–22.4; William W Schwarzer, Reforming Jury Trials, 1990 U. Chi. Legal F. 119.

<sup>60.</sup> MCL 3d, supra note 8, § 21.68.

- Give the jury preliminary instructions at the start of the trial and explain the issues they will have to decide; this will make the evidence more intelligible to jurors.
- Permit jurors who want to take notes to do so.

#### B. Presentation of Evidence

- Eliminate legal and other jargon. Lawyers, judges, and experts use technical jargon, creating obstacles to jury comprehension. The judge should give instructions to participants before trial and repeat them from time to time as necessary. The judge may find it necessary to ask witnesses to translate their statements at trial into plain English.
- Have experts testify in succession; in lengthy trials, the jury's memory of earlier testimony may have faded when an opposing expert is called later, making it difficult for them to compare and evaluate the testimony.<sup>61</sup>
- Use summaries of voluminous data whenever possible.<sup>62</sup>
- Encourage stipulations by the parties on matters not reasonably disputable. A stipulated summary of a deposition, for example, can avoid the need for a lengthy reading of the transcript.
- Use visual and other teaching aids (models, pictures, films, or demonstrations) to explain complicated concepts.
- Provide jurors with notebooks containing glossaries of terms, fact stipulations, key exhibits, chronologies or time lines, a list of witnesses, and other reference material that will assist comprehension.
- Permit jurors to ask questions under controlled conditions; jurors may, for example, be permitted to ask for clarification when they do not understand some part of an expert's testimony.
- In bench trials, present the direct testimony of experts in written narrative form, subject to cross-examination.<sup>63</sup>

62. Fed. R. Evid. 1006.

Management of Expert Evidence

<sup>61.</sup> Fed. R. Evid. 611(a) permits the court to vary the order of calling witnesses.

<sup>63.</sup> The reports under Fed. R. Civ. P. 26(a)(2) can serve this purpose.

## **Evidentiary Framework**

Margaret A. Berger

**Margaret A. Berger** J.D., is Associate Dean and Professor of Law, Brooklyn Law School, Brooklyn, New York.

## Contents

- I Introduction 43
  - A. Impact of Daubert 45
  - B. A Note on Relevancy, or "Fit" 47
  - C. Related Procedural Issues 49
    - 1. Discovery issues 49
    - 2. Judicial screening 50
    - 3. Admissibility versus sufficiency 51
    - 4. Special problems in criminal cases 53
  - D. A Note on Appellate Review 53
- II. When Is a Person Qualified to Testify as an Expert? 55
  - A. General Approach: A Two-Pronged Test 55
  - B. Other Considerations Bearing on an Expert's Qualifications 56
  - C. Issues Bearing on an Expert's Minimal Qualifications 56
    - 1. Education or experience 56
    - 2. Expertise in particular field 57
    - 3. Meaning of minimal qualifications 58
    - 4. Discretion 58
  - D. Issues Bearing on Relationship of Expert's Qualifications to Subject Matter of Proposed Testimony 58
    - 1. How much of a specialist must the expert be? 58
      - a. Physicians 59
      - b. Engineers 61
    - 2. The "secondhand" expert 62
    - 3. The "professional" witness 62
  - E. Limiting Expert's Testimony 63
  - F. Lay Opinion Testimony on Scientific Issues 64
    - 1. Distinctions between Rule 701 and Rule 702 65
    - 2. Situations in which Rule 701 witnesses testify 65
      - a. The identifying witness 65

- b. Lay witnesses with special expertise 66
  - 1. Causation 66
  - 2. Economic issues 66
- III. Is the Expert's Opinion Supported by Scientific Reasoning or Methodology? 69
  - A. The Frye Test 70
  - B. The Daubert Test 71
  - C. Contexts in Which Questions Relating to Scientific Validity Arise 72
    - 1. Issues with regard to a particular discipline 73
      - a. Challenging a group of experts' methodology as lacking the characteristics of science 73
        - 1. Clinical ecology: Is a field scientific if its theories are not testable? 73
        - 2. Forensic techniques: How much inquiry into testing is required? 75
      - b. Challenging a methodology as lacking probative value 77
        - 1. Extrapolation problems; animal studies 78
        - 2. The need for probabilistic evidence: clinical medicine 80
    - 2. Issues with regard to the methodology and reasoning of a particular scientific theory 82
      - a. When does the expert's reasoning satisfy the Daubert test? 82
        - 1. Theories as to the cause of plaintiff's cancer 83
        - 2. Social science evidence 84
        - 3. Psychological syndrome evidence 87
      - b. Rejecting expert testimony because of skewed methodology 88
    - 3. Issues with regard to statistical estimates 92
      - a. Statistical significance: An issue for scientists or for the court? 93
      - b. Correlation of statistical results with the burden of proof 95
      - c. Confusing the probability of a sample identification with a probability of guilt 97
      - d. Reducing odds because of sampling uncertainties; DNA 98
      - e. Incorporating proficiency test performance results; DNA 101
- IV. Is the Expert's Opinion Supported by Reliable Data? 103
  - A. Rule 703 103
  - B. Rule 703; Scope of Rule 104
    - 1. The impact of Daubert 104

- a. Reclassifying issues under Rule 702 that some courts had classified under Rule 703 104
  - 1. Fit 104
  - 2. Methodology 104
- b. Rule 703 reference 104
  - 1. Standard of proof 104
  - 2. Function of Rule 703 as a rule of admissibility 105
- 2. Other theoretical issues about the function of Rule 703 105
  - a. Does the second sentence of Rule 703 apply only when an expert relies on inadmissible evidence? 105
  - b. Determining what is "reasonably relied upon" 106
  - c. Circumstances in which courts use a "reasonably rely" test to exclude 107
    - 1. Expert's failure to consider data that must be taken into account 107
    - 2. Expert's reliance on data that should not be taken into account 108
    - 3. Expert's reliance on data that are erroneous 110
    - 4. Expert's opinion does not rest on a foundation that experts would generally find reliable 111
- V. Is the Expert's Opinion Subject to Exclusion Under Rule 403? 113
  - A. The Interplay Between Rules 702, 703, and 403 113
  - B. Examples of Situations in Which Courts Apply Rule 403 114
    - 1. Prejudicial language 114
    - 2. "Aura of scientific infallibility" 115
    - 3. In-court demonstrations or evidence of experiments 115

### I. Introduction

The aim of this paper is to provide a framework for judges in considering disputes over the admissibility of various kinds of scientific evidence. The paper is not intended to be a review of the law of evidence; case citations are included for illustrative purposes primarily. The object is not to suggest that evidence is or ought to be admissible or excluded in any particular case. Instead, this paper is designed to assist judges in structuring inquiries necessary for making rulings on objections to expert evidence in pretrial proceedings, in connection with motions for summary judgment, or in connection with judgments as a matter of law at trial where the legal sufficiency of evidence is challenged.

Rules 702–705 of the Federal Rules of Evidence govern testimony by experts selected by the parties. These rules have a number of characteristics:

- 1. They were drafted as an integrated solution to the subject of expert testimony.
- 2. They abolished previous common-law constraints on expert testimony, such as the need for hypothetical questions, the bar on ultimate conclusions, and the *Frye* test.<sup>1</sup>
- 3. They were drafted in such general terms that the appellate courts have had to give content to the broad objectives mandated in the rules.
- 4. They accord a great deal of discretion to the trial courts to proceed on a case-by-case basis.

These characteristics have an impact when experts seek to testify about complex science and technology issues. The closely intertwined nature of the rules coupled with the lack of detailed content afford judges the possibility of approaching the same problem from different avenues. What one court has viewed as raising a Rule 702 issue is treated as a Rule 703 matter in a neighboring circuit. In addition, the meaning of particular phrases in the rules has been fleshed out by varying formulas in different courts. To complicate matters further, courts

<sup>1.</sup> In Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993), the Supreme Court applied to the expert testimony rules the plain-meaning approach it had previously applied to other Federal Rules of Evidence. Consequently other common-law doctrines that are not mentioned in Article VII of the Federal Rules of Evidence may also have been eradicated. For a discussion of other common-law clichés relating to expert testimony that are not referred to in the Federal Rules, see Margaret A. Berger, United States v. Scop: The Common-Law Approach to an Expert's Opinion About a Witness's Credibility Still Does Not Work, 55 Brook. L. Rev. 559 (1989).

have disagreed on how Rule 403 operates in conjunction with the rules on expert testimony. It is too soon to determine the extent to which these differences will be resolved in the aftermath of the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 

Coherence is at first glance difficult to discern when one surveys the case law on expert testimony. The disagreement among circuits, compounded by the great discretion afforded trial judges, results in a seeming lack of uniformity and consistency that surfaces whenever any two opinions on expert testimony are compared. Contributing to the want of cohesion is the fact that evidentiary rules are applied in a variety of procedural contexts, and courts differ as well in their procedural approaches when they implement evidentiary decisions.

If one looks at the body of recent cases dealing with expert testimony in cases with scientific evidence, however, a considerable amount of the variation turns out to be superficial. Although disparities in judicial methodology are common, there is much less divergence in result. While courts have approached the highly complex, intertwined legal and scientific issues presented by many recent cases from different starting points, the ultimate outcome with regard to expert testimony in groups of related cases has been remarkably consistent within the federal system and was so even before the *Daubert* decision.

As the first case in which the Supreme Court analyzed principles and rules of evidence and procedure governing expert testimony grounded in scientific knowledge, *Daubert* will be cited routinely whenever issues of scientific proof, or indeed any type of expert proof, arise. The majority's approach is, however, extremely general and does not address the many concrete interrelated scientific and legal issues that courts regularly must confront when a case revolves around scientific evidence. Furthermore, although the majority acknowledges that other rules bear on the admissibility of expert proof, its detailed analysis is concerned only with Rule 702 of the Federal Rules of Evidence; the appropriate scope of some of the other rules is not completely clear.

Rather than organizing the discussion in this paper about specific evidentiary rules or *Daubert*, therefore, it seems more fruitful to concentrate on specific problems that require a considerable investment of judicial time when experts seek to testify about scientific matters. Looking at how courts address frequently occurring fact patterns may identify the kinds of questions, scientific as well as legal, that must be considered, and evidentiary and procedural solutions, compatible with *Daubert's* objectives, that courts have used effectively. Although *Daubert* is concerned solely with scientific evidence, the scope of Rule 702 is considerably broader. In a number of sections, therefore, particularly in section II, which deals with an expert's qualifications, this paper considers experts who offer opinions on technological issues in addition to experts whose realm of expertise is classified as scientific knowledge.

After a number of background issues are surveyed, the body of this paper addresses four broad categories that seem to capture the central concerns that permeate judicial opinions:

- 1. Is the expert qualified?
- 2. Is the expert's opinion supported by scientific reasoning or methodology?
- 3. Is the expert's opinion based on reliable data?
- 4. Is the expert's opinion so confusing or prejudicial that it should be excluded pursuant to Rule 403?

The discussion in sections II–V examines particular issues that courts view as within the scope of these four questions and explores how courts analyze these issues from an evidentiary standpoint in the context of typical scientific fact patterns. Complicating the task of sorting out the various analyses is the fact that many opinions consider all four questions with regard to a particular expert. It may well be that failing to meet a combination of these requirements is what results in the exclusion of expert testimony. Consequently, although issues have been separated out for purposes of discussion, the reader should bear in mind that the distinctions made may at times be somewhat artificial and arbitrary. Cross references to further discussions of the same case have been added in the hope of obviating this problem somewhat.

#### A. Impact of Daubert

Before considering these four central problems, however, a few words are appropriate about the significance of Daubert in relationship to this organizational scheme and scientific expert proof in general. The first of the questions posed above—whether the expert is qualified—was not dealt with in Daubert; at each level of the litigation, the courts assumed that the proffered experts were adequately qualified pursuant to Rule 702. Clearly, however, Rule 702 mandates a qualified expert, and section II indicates that considerable case law exists dealing with a variety of problems in the context of qualifying scientific experts. The last of the categories to be discussed—when exclusion is warranted by Rule 403 also was not addressed by the Daubert court beyond an acknowledgment that the rule may operate to exclude expert testimony in some unspecified instances. Section V discusses the different approaches judges have used when relying on Rule 403 to exclude expert testimony. The ways in which the Daubert opinion may affect issues treated in sections III and IV, relating to the validity of the scientific methodology and reasoning and the reliability of the data on which the expert relies, are examined in connection with those sections.

The Daubert opinion is significant as well in a more general sense. In what is the first Supreme Court case to examine the governing legal principles that bear on expert scientific evidence, the justices made a number of statements that are

broadly applicable to the problems caused by disputed scientific proof. Of central significance is the Court's recognition both of the Federal Rules' "liberal thrust" with regard to the admissibility of expert testimony and the trial judge's "gatekeeping" role vis à vis expert proof on scientific issues.<sup>2</sup> Although stressing that in the usual case the evaluation of expert testimony must be left to the jury, the majority acknowledged the trial judge's responsibility pursuant to Rule 104(a) of the Federal Rules of Evidence to screen scientific evidence in order to keep unreliable evidence out of the courtroom.<sup>3</sup> The Court emphasized that a trial court must determine at the outset "whether the reasoning or methodology underlying the testimony is scientifically valid," and it discussed a number of nondefinitive factors that bear on the inquiry.<sup>4</sup> Rule 702 applies as well to forms of specialized knowledge other than scientific knowledge. Where courts will draw the line between scientific evidence and other types of evidence requiring expert proof is not yet clear.<sup>5</sup>

In *Daubert*, the majority's opinion concentrates primarily on the appropriate meaning of Rule 702, but advises trial judges to be mindful as well of Rules 703, 706, and 403 in handling scientific evidence. The Court also suggests that "conventional devices," like vigorous cross-examination, careful instruction on the burden of proof, grants of summary judgment, and directed verdicts, may be appropriate instead of the "wholesale exclusion" of scientific evidence under Rule 702.<sup>6</sup>

Finally, in a reprise to the "gatekeeping" role of the trial judge at the end of the opinion, the Court reminds the reader that the goals of science and the law differ. While acknowledging some similarities between the scientific and legal endeavors, the opinion recognizes that

<sup>2.</sup> Daubert, 113 S. Ct. at 2794, 2798-99.

<sup>3.</sup> *Id.* at 2796. Even Chief Justice Rehnquist and Justice Stevens, who dissented in part because they felt that "general observations" were not needed to dispose of the case, agreed that "Rule 702 confides to the judge some gatekeeping responsibility." *Id.* at 2800.

<sup>4.</sup> Id. at 2796-97. See discussion infra§ III.B.

<sup>5.</sup> See Richard D. Friedman, The Death and Transfiguration of Frye, 34 Jurimetrics J. 133, 140 (1994) (expressing hope that courts "will recognize that the dangers that led the court to impose such a requirement are very strong only in cases of great technical complexity and that, even in some fields of great difficulty, at least some issues are not readily susceptible to full exploration by the scientific method"). The American College of Trial Lawyers has suggested extending Daubert's approach to expert testimony in general. American College of Trial Lawyers, Standards and Procedures for Determining the Admissibility of Expert Evidence After Daubert, 157 F.R.D. (forthcoming Dec. 1994). See, e.g., Jacobelli Constr., Inc. v. County of Monroe, 32 F.3d 19 (2d Cir. 1994) (expert testimony in construction contract dispute does "not present the kind of 'junk science' problem that Daubert meant to address"); Tamarin v. Adam Caterers, Inc., 13 F.3d 51, 53 (2d Cir. 1993) (Daubert does not apply to testimony by accountant concerning the contexts of payroll records because "that case specifically dealt with the admissibility of scientific evidence"; "payroll records are straightforward lists of names and hours worked"); United States v. D'Ambrosio, No. 92-10526, 1993 U.S. App. LEXIS 27088, at \*6 (9th Cir. Oct. 14, 1993) (unpublished disposition) (expert testimony on clothing comparison was central factor in court's decision to sustain defendant's bank robbery conviction; court did not address whether there was a scientific basis for clothing comparison). See also discussion of social science evidence infra § III.C.2.a.2.

<sup>6.</sup> Daubert, 113 S. Ct. at 2798.

there are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly . . . . [The consequence is that] a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations. That, nevertheless, is the balance that is struck by Rules of Evidence designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes.<sup>7</sup>

The Daubert opinion's emphasis on the jury's role and recognition of the trial judge's responsibility to keep unreliable evidence out of the courtroom are fully consistent with this manual's approach of providing information about the ways in which the courts have dealt with representative and recurring scientific issues in pretrial and trial contexts. The objective is to ensure the fair and efficient resolution of legal controversies.

#### B. A Note on Relevancy, or "Fit"

Other than in this section, this paper does not treat relevancy issues. Although Rule 402 of the Federal Rules of Evidence provides that "all relevant evidence is admissible" and "[e]vidence which is not relevant is not admissible," courts often analyze relevancy problems with regard to expert proof pursuant to the expert testimony rules in Article VII of the Federal Rules of Evidence. The Supreme Court endorsed this approach in *Daubert* when it located within Rule 702 the obligation of the trial court to determine whether the proffered scientific evidence "properly can be applied to the facts in issue."<sup>8</sup> The Court, adopting terminology used by Judge Becker in *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985), characterized this consideration as one of "fit."<sup>9</sup> The Court placed the requirement of fit within Rule 702 because evidence or testimony that does not relate to any issue in the case cannot satisfy the rule's requirement of "assist[ing] the trier of fact to understand the evidence or to determine a fact in issue."<sup>10</sup>

Problems with fit occur independently of an expert's qualifications or deficiencies in the expert's scientific knowledge. The difficulty is that the proffered expert opinion may relate to facts or data that have not been adequately established in the case.<sup>11</sup> For instance, a plaintiff will not be able to succeed in a toxic

11. See, e.g., Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1113–14 (5th Cir. 1991) (plaintiff's expert premised his opinion on a twenty-year history of exposure, although the record indicated that Christophersen had worked in defendant's plant for only fourteen years; majority held that Rule 703 would permit re-

<sup>7.</sup> Id. at 2798-99.

<sup>8.</sup> Id. at 2796.

<sup>9.</sup> Id.

<sup>10.</sup> Fed. R. Evid. 702. The Court offers the example of the expert whose scientific training about the phases of the moon enables him or her to establish whether it was dark on a particular night. If that is the issue, the expert's testimony fits. Yet evidence that the moon was full on the night in question does not assist the trier on the issue of whether an individual is likely to be irrational when the moon is full. *Daubert*, 113 S. Ct. at 2796.

tort case unless he or she can prove adequate exposure to a toxic substance that was somehow connected to the defendant. Even if an expert testifies that Substance X can cause the plaintiff's injury, this testimony will not suffice if the plaintiff failed to produce evidence that he or she was exposed to Substance X, <sup>12</sup> or to a specific defendant's Substance X, <sup>13</sup> or at a significant level. <sup>14</sup>

In excluding an expert opinion as not based on the evidence, the court performs the same analysis in a science-rich case as in a routine motor vehicle accident case, although the complex nature of scientific evidence may make it more difficult in the former case to detect that the expert's testimony fails to provide "a valid scientific connection to the pertinent inquiry."<sup>15</sup> In an accident case, a court will exclude an expert's opinion that the defendant's speeding caused the accident when the record contains no evidence about this possibility—neither direct proof that the defendant was speeding, nor evidence, such as skid marks, from which an inference of speeding may be drawn.<sup>16</sup>

Prior to *Daubert*, a number of federal courts had analyzed the "opinion that does not fit the facts" problem pursuant to Rule 703. Because Rule 703 speaks of an expert's opinion being based upon the "facts or data" in the particular case, some courts had concluded that exclusion is warranted pursuant to Rule 703

12. The presence of signature diseases, such as mesothelioma, may, however, permit an inference of exposure.

13. See Celotex Corp. v. Catrett, 477 U.S. 317 (1986).

14. See Mateer v. United States Aluminum, No. 88-2147, 1989 U.S. Dist. LEXIS 6323 (E.D. Pa. June 2, 1989).

15. Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993).

16. Cf. Newman v. Hy-Way Heat Sys., Inc., 789 F.2d 269, 270 (4th Cir. 1986) ("nothing in the Rules appears to have been intended to permit experts to speculate in fashions unsupported by, and in this case indeed in contradiction of, the uncontroverted evidence in the case").

jection of an opinion "founded on critical facts that are plainly untrustworthy, principally because such an opinion cannot be helpful to the jury"), cert. denied, 112 S. Ct. 1280 (1992); Bernhardt v. Richardson-Merrell, Inc., 892 F.2d 440, 445 (5th Cir. 1990) (affirming grant of summary judgment for defendant when discovery established that drug exposure occurred after that point in pregnancy at which a particular birth defect could have occurred; the court refused to allow statements by the plaintiff's expert which "may be generally true" to create a genuine issue of material fact, relying on its "duty to scrutinize the probative value of the evidence"); Novak v. United States, 865 F.2d 718, 723–24 (6th Cir. 1989) (even if epidemiological evidence supported expert's contention of an increased incidence of persons contracting dermatomyositis/polymyositis (DM/PM) within fifteen days of a swine flu vaccination, evidence showed that decedent's symptoms commenced considerably after fifteen days); Peterson v. Sealed Air Corp., No. 86-C3498, 1991 U.S. Dist. LEXIS 5333, at \*21-22 (N.D. Ill. Apr. 22, 1991) (even if Cytoxan may cause transitional cell carcinoma, evidence in the case indicated that plaintiff's cancer was basaloid cell type and not transitional); Bailiff v. Manville Forest Prods. Corp., 772 F. Supp. 1578, 1583-84 (S.D. Miss. 1991) (summary judgment for defendants; plaintiff's expert concluded that plaintiff's respiratory problems were caused by exposure to defendants' products after reviewing Material Data Safety Sheets for chemicals manufactured at defendants' plant; court notes no listing in sheets for only Manville product to which plaintiff was allegedly exposed); Mateer v. United States Aluminum, No. 88-2147, 1989 U.S. Dist. LEXIS 6323, at \*21-22 (E.D. Pa. June 2, 1989) (summary judgment for defendant; experts' reports claimed that certain contaminants found in plaintiffs' well may be toxic to humans depending on dose and duration of exposure, but failed to state what level of exposure is hazardous to humans; plaintiffs failed to show that their levels of exposure created a risk); Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1089–90, 1093 (D. Md. 1986) (plaintiff's expert relied on study to conclude that there is a three- to five-fold increased rate of pelvic inflammatory disease in women wearing IUDs; only one participant in study, who did not develop pelvic disease, was wearing defendant's device; trial court ultimately directed verdict for defendant because of plaintiff's failure to prove causation after three-week trial and a jury that was unable to reach a verdict), aff'd sub nom. Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987).

when the expert's testimony is not tied to any facts or data in the case. These cases should now be resolved pursuant to Rule 702. Details about the expert's methodology may be needed to assess fit and at times, the line between lack of fit and a flawed methodology may be somewhat blurry.<sup>17</sup>

In terms of judicial efficiency, a problem in some cases is that the lack of correspondence between the expert's opinion and the facts of the case is not brought to the court's attention until trial. The increased opportunities for expert discovery under the 1993 amendments to the Federal Rules of Civil Procedure may result in objections based on lack of fit being raised prior to trial by a motion in limine or for summary judgment.

#### C. Related Procedural Issues

The *Daubert* opinion did not address many of the complex issues that will have to be elucidated in order to reconcile the Supreme Court's recognition of the Federal Rules' liberal admissibility policy for expert proof with its endorsement of the trial judge's gatekeeping function. Many of these issues raise procedural concerns that were not dealt with by the Court. In the future, courts will have to examine the interrelationship of discovery rules and *Daubert*, the nature of judicial screening pursuant to Rule 104(a), and the interplay between issues of admissibility and sufficiency when expert testimony is challenged. In addition, issues may arise as to whether the differing natures of criminal and civil litigation warrant procedural distinctions.

#### 1. Discovery issues

Less than six months after the Supreme Court's decision in *Daubert*, amendments to Rule 26(a)(2) and (b)(4) of the Federal Rules of Civil Procedure became effective that require a party, independently of any discovery request, to disclose the identity of all expert witnesses expected to testify at trial; to provide, among other things, the experts' written signed reports stating all opinions to be offered and support for opinions; and to make the expert available for deposition after the report is submitted.<sup>18</sup> In the absence of court order or stipulation, a party must disclose these items at least ninety days before the trial date or the date on which the case is to be ready for trial. Rule 16(a)(1)(E) of the Federal Rules of Criminal Procedure was simultaneously amended to provide that the

<sup>17.</sup> See, e.g., DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 955 (3d Cir. 1990), summ. *judgment granted on remand*, 791 F. Supp. 1042, 1050 (D.N.J. 1992) (in directing remand, appellate court had determined that fit was satisfied so that district court was not required to consider this factor; opinion on remand notes, however, that plaintiff's expert included in his chart studies that dealt with an ingredient that was not found in the two-ingredient formula of Bendectin ingested by Mrs. DeLuca; the inclusion of these data was treated as an aspect of the expert's suspect methodology), *aff'd without op.*, 6 F.3d 778 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 691 (1994). See further discussion of this case on remand *infra* §§ III.C.2.b, III.C.3.a. See also discussion *infra* §§ IV.B.2.c.2, IV.B.2.c.3.

<sup>18.</sup> As of this writing, a number of districts have opted out of these procedures.

government must disclose at the defendant's request "a written summary of testimony the government intends to use under Rules 702, 703, or 705 of the Federal Rules of Evidence during its case in chief at trial."

Neither rule specifically requires divulgence of the methodological details that according to *Daubert* bear on the admissibility of expert testimony.<sup>19</sup> It remains to be seen whether courts will require summaries and reports to disclose information bearing on *Daubert*'s nondefinitive checklist of factors and on additional factors that should be considered in particular kinds of cases.<sup>20</sup>

The timing of the disclosures, in the absence of order or stipulation, is geared to trial; yet *Daubert* suggests that in civil litigation, issues concerning the admissibility or sufficiency of expert testimony should be raised before trial. How timing requirements should be adjusted relates to other issues posed by judicial screening that *Daubert* does not address.

#### 2. Judicial screening

The *Daubert* opinion states that when expert scientific testimony is proffered, the district court must make a determination about admissibility "at the outset, pursuant to Rule 104(a)."<sup>21</sup> This Rule 104(a) inquiry requires the proponent of the expert to show by a preponderance of the evidence that the expert's opinion is admissible.<sup>22</sup>

Daubert does not, however, discuss the circumstances that will trigger in limine judicial screening pursuant to Rule 104(a), or the nature of an in limine hearing. While courts are unlikely to undertake the inquiry envisioned by Daubert whenever scientific evidence is proffered, <sup>23</sup> it is not yet clear when they must do so. The courts will have to determine whether judicial economy and the "liberal thrust" of the rules pertaining to experts justify placing a burden on

a complete statement of all opinions to be expressed and the basis and reasons therefor; the data or other information considered by the witness in forming the opinions; any exhibits to be used as a summary of or support for the opinions; the qualifications of the witness, including a list of all publications authored by the witness within the preceding ten years; the compensation to be paid for the study and testimony; and a listing of any other cases in which the witness has testified as an expert at trial or by deposition within the preceding four years.

See discussion infra§ III.B.

20. For example, courts might require divulgence of the background statistical information on which the probative value of an expert's opinion often depends. See discussion *infra* § III.C.3.c.

21. Daubert, 113 S. Ct. at 2796.

22. Id. at 2796 n.10 (citing Bourjaily v. United States, 483 U.S. 171 (1987)).

23. Indeed, much of the scientific evidence that is proffered in federal court undoubtedly falls into routine categories in which qualified experts disagree about the interpretation of data that were obtained through standard methodologies. A recent survey by the Federal Judicial Center concluded that orthopedists (17.9%) and neurologists (15.6%) are the two most prevalent types of experts testifying in federal civil cases. See Molly Treadway Johnson & Joe S. Cecil, Problems of Expert Testimony in Federal Civil Trials (Federal Judicial Center, forthcoming 1995). Daubert is unlikely to affect most of these cases.

<sup>19.</sup> Fed. R. Crim. P. 16(a)(1)(E) provides that the "summary must describe the witnesses' opinions, the bases and the reasons therefor, and the witnesses' qualifications." Fed. R. Civ. P. 26(a)(2)(B) requires the report to contain

the opponent of the expert proof to come forward with evidence showing deficiencies in the expert's testimony before the court has any obligation to engage in a Rule 104(a) analysis. If there is a burden, the courts will also have to consider the height of the burden, and the materials on which the opponent may rely in discharging its burden.<sup>24</sup>

Answering these questions will require consideration of the relationship between in limine screening and the discovery process. In light of the new discovery rules, for instance, must the opponent produce its experts' reports and make its experts available for deposition before a court will entertain an in limine motion?<sup>25</sup> May the opponent rely on affidavits either in seeking in limine consideration or on the motion itself, or should courts restrict their review to materials developed during discovery or at an evidentiary hearing? In a number of cases discussed elsewhere in this paper, judges have expressed concern that expert testimony will be excluded without the proponent of the expert testimony being provided an opportunity to develop an adequate record tested in an adversarial context.<sup>26</sup>

#### 3. Admissibility versus sufficiency

In *Daubert*, the majority acknowledges that scientific evidence that is admissible may not always suffice to discharge the plaintiff's burden of proof. The Court observed that even if evidence is ruled admissible, if "the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true, the court remains free to direct a judgment, Fed. Rule Civ. Proc. 50(a), and likewise to grant summary judgment, Fed. Rule Civ. Proc. 56."<sup>27</sup> Thus, the distinction between admissibility and sufficiency, though perhaps often blurred in

[W]e generally agree . . . that because under *Daubert* a judge at an *in limine* hearing must make findings of fact on the reliability of complicated scientific methodologies and this fact-finding can decide the case, it is important that each side have an opportunity to depose the other side's experts in order to develop strong critiques and defenses of their experts' methodologies. Given the 'liberal thrust' of the federal rules, it is particularly important that the side trying to defend the admissibility of evidence be given an adequate chance to do so.

<sup>24.</sup> See Margaret A. Berger, Procedural Paradigms for Applying the Daubert Test, 78 Minn. L. Rev. 1345 (1994).

<sup>25.</sup> See In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*31 (3d Cir. Aug. 31, 1994) (citations omitted):

<sup>26.</sup> See, e.g., In re Paoli R.R. Yard PCB Litig. (Paoli I), 916 F.2d 829, 855 (3d Cir. 1990) ("At least some process should have been devised to afford plaintiffs a surrogate for that trial scenario where the equivalent evidentiary exclusion and adverse judgment might occur."), *cert. denied*, 111 S. Ct. 1584 (1991) (see discussion *infra* § III.C.2.b); Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1122 (5th Cir. 1991) (en banc) (Reavley, J., dissenting) (objecting to exclusion of plaintiff's expert testimony where exclusion was based on affidavits of defendant's experts who were never deposed), *cert. denied*, 112 S. Ct. 1280 (1992) (see discussion *infra* § IV.B.2.a).

<sup>27. 113</sup> S. Ct. at 2798. See also Joseph Sanders, Scientific Validity, Admissibility, and Mass Torts After Daubert, 78 Minn. L. Rev. 1387, 1433 (1994) (urging courts to distinguish between decisions based on the in-admissibility of evidence and decisions based on the insufficiency of evidence).

the past by courts when handling issues relating to scientific evidence, is clearly reaffirmed in *Daubert*.<sup>28</sup> Of course, whether a particular issue should be resolved in terms of the admissibility of expert testimony or the insufficiency of the expert proof to discharge the plaintiff's burden will depend on the circumstances of each case. But it is important for courts to have in mind the differences in the applicable standards depending on which procedure is followed.

The standards that apply to resolution of a motion in limine, primarily Rules 702, 703, and 403 of the Federal Rules of Evidence, are governed by the principles discussed in this paper.<sup>29</sup> The standard that applies under Rule 56 (and its functional equivalent, Rule 50) is quite different. As stated in *Celotex Corp. v. Catrett*, the moving party must demonstrate the absence of a triable issue of fact.<sup>30</sup> Expert evidence may be admissible under the rules of evidence but fail to be sufficient to raise a triable issue.<sup>31</sup> Thus, while in passing on admissibility a judge under *Daubert* may have to rule on whether the methodology or reasoning relied on by an expert in arriving at an opinion was scientifically sound, on summary judgment the judge may have to determine whether the proponent to trial.<sup>32</sup>

Even though a defendant may in some instances be able to discharge its burden of production on a summary judgment motion by merely "pointing" to deficiencies in the plaintiff's case, <sup>33</sup> a higher burden may be more appropriate when the defendant is attacking the plaintiff's scientific evidence. Evaluating the validity and sufficiency of a scientific expert's methodology and reasoning may require a more complex determination than that required when the judge merely has to ascertain the availability of evidence on an issue. In making a summary judgment ruling that turns on expert scientific evidence, the court may need to be informed about the kinds of factors discussed in *Daubert*. Affidavits may not suffice to apprise the judge adequately. If a defendant must satisfy a higher burden than merely pointing to alleged deficiencies in the plaintiff's scientific proof, the defendant may have more of an incentive to depose the

<sup>28.</sup> See, e.g., Brock v. Merrell Dow Pharmaceuticals, Inc., 884 F.2d 167, 169 (5th Cir. 1989) (Higginbotham, J., dissenting from the majority's refusal to rehear en banc an appeal granting judgment n.o.v. to defendant in a Bendectin case because the panel had shied away from addressing the crucial issue—the admissibility of the evidence in the first place rather than its sufficiency: "Yet, while skepticism permeates its opinion, the panel does not seem to engage the question at this juncture. Rather, the panel chooses to accept the admissibility of the testimony and to quarrel with its effect.").

<sup>29.</sup> For a discussion of the relative burdens of the parties on a Rule 104(a) in limine motion, see Berger, supra note 24.

<sup>30. 477</sup> U.S. 317, 322–23 (1986). See also Justice Brennan's dissenting opinion elaborating on the procedure for demonstrating the absence of a triable issue. *Id.* at 328–37; William W Schwarzer et al., The Analysis and Decision of Summary Judgment Motions: A Monograph on Rule 56 of the Federal Rules of Civil Procedure 45–47, 53–57 (Federal Judicial Center 1991). For particular reference to the parties' relative burdens, see William W Schwarzer & Alan Hirsch, *Summary Judgment After Eastman Kodak*, 45 Hastings L.J. 1, 4 (1993).

<sup>31.</sup> See, e.g., Maffei v. Northern Ins. Co., 12 F.3d 892, 897–900 (9th Cir. 1993); Mid-State Fertilizer Co. v. Exchange Nat'l Bank, 877 F.2d 1333, 1339 (7th Cir. 1989).

<sup>32.</sup> Šee, e.g., Bulthuis v. Rexall Corp., 789 F.2d 1315 (9th Cir. 1985).

<sup>33.</sup> See Celotex Corp., 477 U.S. at 323.

plaintiff's experts in order to substantiate its claims about the defects in the plaintiff's expert proof. Consequently, the court will have the benefit of a record developed through the adversarial process in making its Rule 56 determination. It is not yet clear at this time, however, how courts will handle the procedural issues stemming from the *Daubert* case.

#### 4. Special problems in criminal cases

The *Daubert* opinion deals with the admissibility of scientific evidence in a civil case. With a few exceptions, this paper discusses issues that arise in civil litigation.

Judges may want to consider whether special procedures with regard to scientific evidence need to be devised for criminal cases. The accused may be more handicapped in challenging expert scientific proof proffered against him or her than the civil litigant because of less extensive discovery rights and fewer resources. In addition, the prosecution may have considerable control over the expertise if it participated in creating and applying the forensic technique in question. In light of these factors, burdens of production with regard to in limine hearings might be allocated differently in criminal cases than in the civil context discussed above.<sup>34</sup> When novel scientific evidence is offered, courts might consider the desirability of obtaining more information by appointing experts pursuant to Rule 706, or referring the motion to a magistrate judge for fact-finding and recommendation.<sup>35</sup>

#### D. A Note on Appellate Review

It must also be noted that the different levels of determinations trial judges make with regard to expert testimony—on the expert's qualifications, reasoning, and methodology, and on underlying data and the applicability of Rule 403—perhaps require different standards of review by the appellate courts. The *Daubert* opinion does not address this issue. The Ninth Circuit, in its opinion below, had applied a de novo standard in finding that the plaintiffs' expert opinion did not satisfy the *Frye* test.<sup>36</sup> Although *Daubert* rejects *Frye*, the opinion does not address the issue of the standard of review.

The Ninth Circuit treated determinations about scientific validity as akin to rulings on matters of law, to which de novo standards customarily apply, reason-

<sup>34.</sup> Cf. Ake v. Oklahoma, 470 U.S. 68 (1985) (expert must be provided for indigent defendant in capital case).

<sup>35.</sup> See, e.g., United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991) (adopting a magistrate judge's 120-page report and admitting DNA evidence), aff'd sub nom. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993).

<sup>36.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 951 F.2d 1128, 1130 (9th Cir. 1991), rev'd, 113 S. Ct. 2786 (1993). See David L. Faigman et al., Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence, 15 Cardozo L. Rev. 1799 (1994). Of course, subsequent developments may warrant a changed ruling. The Supreme Court acknowledged that "[s]cientific conclusions are subject to perpetual revision." Daubert, 113 S. Ct. at 2798.

ing that the appellate court is in as good a position as the trial court to make this determination.<sup>37</sup> Other circuits have applied an abuse-of-discretion standard when reviewing a trial court's exclusion of an expert's testimony.<sup>38</sup> Some issues that courts address with regard to the admissibility of expert testimony may present more of a mixed question of law and fact.<sup>39</sup> Even issues regarding an expert's qualifications may perhaps be classified as raising mixed questions, since the court is assessing the expert's qualifications in light of a scientific theory that the court considers relevant.<sup>40</sup> The courts have not yet clarified the appellate courts' role vis à vis expert testimony in instances when the court has to deal with mixed issues of fact and law.<sup>41</sup>

Finally, the Supreme Court acknowledges in *Daubert* that Rule 403 may play a role in the exclusion of expert testimony.<sup>42</sup> Decisions under this rule are clearly viewed as committed to the discretion of the trial court and therefore are reviewed under an abuse-of-discretion standard that examines whether the court below took into account the appropriate factors in arriving at its conclusion.

37. See also In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*69, 71 (3d Cir. Aug. 31, 1994) (emphasis added) (citations omitted):

[E]valuating the reliability of scientific methodologies and data does not generally involve assessing the *truthfulness* of the expert witnesses and thus is often not significantly more difficult on a cold record." The court concludes that "when the district court's exclusionary evidentiary rulings with respect to scientific opinion testimony will result in a summary or directed judgment, we will give them a 'hard look' (more stringent review) to determine if a district court has abused its discretion in excluding evidence as unreliable.

38. See, e.g., Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1109 (5th Cir. 1991) ("A trial court's ruling regarding admissibility of expert testimony is protected by an ambit of discretion and must be sustained unless manifestly erroneous."), *cert. denied*, 112 S. Ct. 1280 (1992); United States v. Bonds, 12 F.3d 540, 554 (6th Cir. 1993) ("We review the trial court's admission of testimony and other evidence under the abuse of discretion standard"; post-Daubert review of admissibility of DNA evidence admitted at trial pursuant to a *Frye* standard). *See also* DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 944 (3d Cir. 1990) ("Our review of a district court's decision to exclude the testimony of an expert is ordinarily limited to ensuring there has been no abuse of discretion, but to the extent the district court's ruling turns on an interpretation of a Federal Rule of Evidence our review is plenary.").

39. See discussion of Rule 703 infra § IV.

<sup>40.</sup> See also infra § II.

<sup>41.</sup> See Ursula Bentele & Eve Cary, Appellate Advocacy: Principles and Practice 93 (1990):

Courts have sent decidedly mixed signals about what is the appropriate standard of review for such hybrid questions, with some courts announcing that a de novo standard should apply, others deciding that mixed findings are essentially factual, and therefore entitled to great deference, and several courts swinging back and forth between the two positions.

<sup>42.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2798 (1993).

# II. When Is a Person Qualified to Testify As an Expert?

The courts generally agree that issues with regard to an expert's qualifications are governed by Rule 702 of the Federal Rules of Evidence. Rule 702 provides:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

#### A. General Approach: A Two-Pronged Test

To ascertain whether a proposed expert is qualified to act as a witness, a court must undertake a two-step inquiry:

- 1. The court should determine whether the proffered expert has minimal educational or experiential qualifications in a field that is relevant to a subject which will assist the trier of fact.
- 2. If the expert passes this threshold test, the court should further compare the expert's area of expertise with the particular opinion the expert seeks to offer. The expert should be permitted to testify only if the expert's particular expertise, however acquired, enables the expert to give an opinion that is capable of assisting the trier of fact.<sup>43</sup> The more difficult question—the extent to which a court may have to inquire into the methodological underpinnings of the theory on which the expert is relying in order to determine whether the expert's opinion is admissible—is discussed in section III. It should be noted, however, that the two categories may overlap. In determining whether the expert is relying on a methodologically sound theory pursuant to Rule 702, the court may take

<sup>43.</sup> See the helpful discussion in Carroll v. Otis Elevator Co., 896 F.2d 210, 214–15 (7th Cir. 1990), as to why a specialist in experimental psychology and visual perception would be able to assist the trier in determining whether children would be likely to push a particular button on an escalator. See also Kloepfer v. Honda Motor Co., 898 F.2d 1452, 1458–59 (10th Cir. 1990), which held that the lower court had properly excluded the testimony of a pediatrician who was experienced as a children's accident preventionist. The lawsuit involved the death of a child while a passenger on an all-terrain vehicle manufactured by the defendant. The excluded testimony, however, related to the conduct of the adult driver and had no bearing on the behavior of the child passenger.

into account the degree of specialized knowledge the expert possesses about the particular issues in dispute.<sup>44</sup>

#### B. Other Considerations Bearing on an Expert's Qualifications

A combination of the factors discussed in section C below may suffice to disqualify an expert even when a particular factor standing alone would not. Even if the court finds the expert qualified to offer some opinions, it may preclude the expert from offering others because of a lack of expertise with regard to certain issues.<sup>45</sup>

Although rarely explicitly discussed, another factor that may affect the court's determination is the degree to which experts are available to all the parties. When the experts in a field are all arrayed on one side of the case—typically the defendant's—a court may have to allow some leeway in the plaintiff's choice of an expert in order to provide the plaintiff with fair access to the courts. This is especially true if virtually all of those with the requisite expertise are persons currently or formerly associated with the defendant.

#### C. Issues Bearing on an Expert's Minimal Qualifications

1. Education or experience

The Federal Rules of Evidence state that an expert may be qualified by virtue of education or practical experience, or some combination of the attributes stated in Rule 702. An expert should not be excluded from testifying merely because he or she lacks an educational background if the requisite expertise has been acquired through training or experience. For example, in *Circle J. Dairy, Inc. v. A.O. Smith Harvestore Products, Inc., a* witness was found qualified to testify as to cattle's injuries, since he had "significant practical experience with feed-related health problems in dairy cattle" even though he was not a veterinarian and held no advanced degrees. <sup>46</sup> But the court may exclude an expert who does not

46. 790 F.2d 694, 700 (8th Cir. 1986). See also Davis v. United States, 865 F.2d 164, 168 (8th Cir. 1988) (witness with university degree in journalism qualified to testify about likelihood of female to male transmis-

<sup>44.</sup> See, e.g., O'Conner v. Commonwealth Edison Co., 13 F.3d 1090, 1107 & n.19 (7th Cir. 1994) (in affirming district court's exclusion of plaintiff's expert, who claimed that plaintiff's cataracts were radiation-induced, because he lacked a proper methodology (see infra § III), the court noted that the expert had treated only five cases of radiation-induced cataracts in twenty years: "We do not believe that this limited exposure . . . qualifies as a basis for a scientifically sound opinion."); Chikovsky v. Ortho Pharmaceutical Corp., 832 F. Supp. 341, 344–46 (S.D. Fla. 1993) (plaintiff alleged that defendant's product, Retin-A, caused birth anomalies; on defendant's motion for summary judgment, the court, citing *Daubert*, found that testimony of plain tiff's sole expert, an obstetrician-gynecologist, would not be admissible; the court noted that the expert had no specialized training in embryology or teratology, did not know if genetic explanations existed for the child's birth defects, and did not know how much Retin-A the mother might have absorbed through topical applications; the court also stressed that expert's theory that topical applications of Retin-A during pregnancy can cause birth defects had not been tested; the court pointed to total lack of data; the court, citing *Daubert*, stated: "This is precisely the kind of evidence that the trial judge must exclude in performing the gatekeeper function.").

<sup>45.</sup> See infra§II.E.

have the appropriate experience, education, or training to offer a helpful opinion with regard to the controverted issue.<sup>47</sup>

#### 2. Expertise in particular field

Courts recognize that experts in a variety of fields may be helpful with regard to a particular issue. For instance, a Ph.D. who is a toxicologist may be as qualified as an M.D. to express an opinion about causation in a toxic tort case.<sup>48</sup> Furthermore, different fields of expertise may be relevant to different aspects of an issue. For instance, in *Williams v. Pro-Tec, Inc.*, a products liability action in which the plaintiff claimed that an eye guard produced by the defendant was unreasonably dangerous, the appellate court agreed that a mechanical engineer was properly qualified.<sup>49</sup> The engineer testified with regard to "the factor by which the eye guard reduced the force that a racquetball exerted upon a simulated eyeball at different speeds." An ophthalmologist would have been able to testify about the force necessary to injure an eye.

Some issues, however, clearly require expertise in a particular field. For example, in *Edmonds v. Illinois Central Gulf Railroad*, the district court committed reversible error in permitting a clinical psychologist to testify that stress worsened the plaintiff's preexisting heart condition, since causation of a heart condition is a medical issue.<sup>50</sup> Similarly, in *Stull v. Fuqua Industries, Inc.*,<sup>51</sup> a mechanical engineer was found not qualified to state that the plaintiff's leg would have broken had the accident occurred in the manner claimed by the plaintiff, since the expert lacked expertise in human anatomy.<sup>52</sup>

sion of gonorrhea; witness had worked for more than eight years as a public health investigator and had received Centers for Disease Control training).

47. See, e.g., Thomas J. Kline, Inc. v. Lorillard, Inc., 878 F.2d 791, 800 (4th Cir. 1989) (abuse of discretion for trial court to have allowed testimony about credit discrimination by witness who was not an economist and whose general business education did not indicate "any training in the area of anti-trust or credit" and who admitted "that she lacked any other experience in such matters.") (emphasis in original), *cert. denied*, 493 U.S. 1073 (1990); Hughes v. Hemingway Transp., Inc., 539 F. Supp. 130, 133 (E.D. Pa. 1982) (exclusion of witness's opinion testimony was proper because deposition revealed that witness could not calculate the coefficient of friction on the roadway at the time of the accident and therefore could not determine whether the driver of a tractor-trailer was using the proper technique for coping with a skid during icy conditions).

48. Genty v. Resolution Trust Corp., 937 F.2d 899, 916–17 (3d Cir. 1991); Peterson v. Sealed Air Corp., Nos. 86-C3498, 88-C9859 consolidated, 1991 U.S. Dist. LEXIS 5333, at \*31 (N.D. Ill. Apr. 22, 1991).

49. 908 F.2d 345, 348 (8th Cir. 1990).

50. 910 F.2d 1284, 1287 (5th Cir. 1990).

51. 906 F.2d 1271 (8th Cir. 1990).

52. *Id.* at 1275. *See also* Livshits v. Natural Y Surgical Specialties, Inc., No. 87-C2403, 1991 U.S. Dist. LEXIS 17245, at \*23 (S.D.N.Y. 1991) (certified toxicologist with a doctorate in experimental pathology was qualified to testify about possible dangers posed by breast implant, but was not qualified to express a diagnostic opinion as to cause of acceleration of cancer in plaintiff's breast; he admitted that he was not qualified to render diagnoses in humans), *reaff'd*, No. 87-C2403, 1991 U.S. Dist. LEXIS 18445 (S.D.N.Y. Dec. 19, 1991); Owens v. Concrete Pipe & Prods. Co., 125 F.R.D. 113, 115 (E.D. Pa. 1989) (although nonphysicians who are doctors of pharmacology and chemistry are qualified to testify as to risks associated with exposure to certain chemicals, they "may not be qualified to diagnose [plaintiff's] medical condition"). *Cf.* Fox v. Dannenberg, 906 F.2d 1253, 1256–57 (8th Cir. 1990) (two engineers who had more than twenty years of experience in accident reconstruction could offer opinion on who was driving even though one factor entering into their opinion was the pattern of injuries; court concluded that as a consequence of their long practical training, they had undoubtedly acquired some knowledge of the medical aspects of traffic injuries).

#### 3. Meaning of minimal qualifications

The fact that an expert has a particular title or degree is not dispositive in either qualifying or disqualifying the expert. The lack of a title or degree does not require exclusion of the expert; knowledge or skill, however obtained, is what counts.<sup>53</sup> Nor is the expert automatically qualified merely because he or she possesses a particular degree or title. In *Genty v. Resolution Trust Corp.*, for instance, the court held that the district judge had not erred in excluding a proffered witness where nothing appeared in the record to substantiate his credentials other than the bare assertion that he was a scientifically trained toxicologist holding a Ph.D.<sup>54</sup> The appellate court noted the absence of a curriculum vitae and the failure to recite studies conducted or methods used, or to include articles published.<sup>55</sup>

#### 4. Discretion

District courts are accorded considerable deference with regard to their rulings on qualification. Consequently, the same appellate court may affirm a ruling excluding an expert who has received only academic training and lacks practical experience, and a ruling excluding an expert with extensive practical experience who lacks academic training.<sup>56</sup>

#### D. Issues Bearing on Relationship of Expert's Qualifications to Subject Matter of Proposed Testimony

The expert's credentials or experience, or both, may enable the expert to meet a threshold test. But before the expert is found qualified to offer an opinion about a particular issue, the court must also decide whether the actual qualifications of the expert enable him or her to assist the trier of fact with regard to each controverted issue about which the expert seeks to testify.<sup>57</sup>

#### 1. How much of a specialist must the expert be?

A recurring problem concerns the requisite level of specialization required of the expert. In 1954, Professor Charles McCormick wrote: "While the court may rule that a certain subject of inquiry requires that a member of a given profes-

<sup>53.</sup> See supra § II.C.1.

<sup>54. 937</sup> F.2d 899, 917 (3d Cir. 1991).

<sup>55.</sup> See discussion of specialization infra § II.D.1.

<sup>56.</sup> Compare Lavespere v. Niagara Mach. & Tool Works, Inc., 910 F.2d 167, 177 (5th Cir. 1990), cert. denied, 114 S. Ct. 171 (1993) with Sullivan v. Rowan Cos., 952 F.2d 141, 145–46 (5th Cir. 1992). In both cases, the circuit court acknowledged that a contrary decision by the district court would not necessarily have required a reversal.

<sup>57.</sup> See, e.g., United States v. Roldan-Zapata, 916 F.2d 795, 805 (2d Cir. 1990) ("A witness may be qualified as an expert on certain matters and not others."), *cert. denied*, 111 S. Ct. 1397 (1991). See also Livshits v. Natural Y Surgical Specialties, Inc., No. 87-C2403, 1991 U.S. Dist. LEXIS 17245 (S.D.N.Y. 1991), *reaff'd*, No. 87-C2403, 1991 U.S. Dist. LEXIS 18445 (S.D.N.Y. Dec. 19, 1991). See *infra* § II.E for examples of cases in which courts have limited the scope of the expert's testimony.

sion, such as a doctor, an engineer or a chemist, be called, *usually* a specialist in a particular branch within the profession will not be required."<sup>58</sup> Some courts quote the sentence without reflecting on whether the usual nonspecialization rule is applicable given the scientific issue posed in the particular case before the court. <sup>59</sup> The governing principle should be whether the expert can assist the trier of fact. How much of a specialist the proffered witness needs to be will depend on the relationship between the expert's particular expertise and the subject matter of the opinion that is being offered. For example, in *Wilkinson v. Rosenthal & Co.*, a professor of finance who taught a basic course at the Wharton School at the University of Pennsylvania was not sufficiently qualified to testify about what constitutes excessive trading in commodity futures, even though he was permitted to testify about basic principles of commodity investing.<sup>60</sup> From the reported cases, it appears that the issue of specialization arises primarily with regard to physicians and engineers.

#### a. Physicians

Language in some cases suggests that the holder of an M.D. degree is qualified to render an opinion about anything possibly characterized as a medical question. For example, in *Payton v. Abbott Labs*, the court stated, "The fact that the physician is not a specialist in the field in which he is giving his opinion affects not the admissibility of his opinion but the weight the jury may place on it."<sup>61</sup> The facts of such cases do not necessarily support such a broadly stated rule. In *Payton*, for example, the physicians in question testified that the drug diethyl-stilbestrol (DES) is a teratogen and that the plaintiff's injuries were caused by her mother's ingestion of DES during pregnancy. The experts were board-certified obstetrician–gynecologists who served as clinical instructors at Harvard Medical School. Although they were not research scientists, both had studied the literature on DES and embryology and had treated numerous DES daugh-

<sup>58.</sup> Charles McCormick, Evidence § 14, at 29 (1954) (emphasis added). This statement also appears in subsequent editions. See McCormick on Evidence § 13, at 34 (Edward W. Cleary ed., 3d ed. 1984), quoted in Peteet v. Dow Chemical Co., 868 F.2d 1428, 1432 (5th Cir. 1989), cert. denied, 493 U.S. 935 (1989).

<sup>59.</sup> See, e.g., Mateer v. United States Aluminum, No. 88-2147, 1989 U.S. Dist. LEXIS 6323, at 21 n.7 (E.D. Pa. June 2, 1989) (quoting McCormick in assuming that an osteopath would be qualified to express an opinion as to health effects of exposure to contaminants in well water; dictum).

<sup>60. 712</sup> F. Supp. 474, 477–78 (E.D. Pa. 1989). See also LeMaire v. United States, 826 F.2d 949, 951–52 (10th Cir. 1987) (in medical malpractice case in which plaintiff claimed that treatment led to fatal episode which may have been stroke, court found no error in permitting opinion testimony on the subject of neurology by the defense witness who "was endorsed at trial, without objection, as an expert on internal medicine and cardiology" because "plaintiff's counsel should have foreseen the general nature of . . . [the expert's] testimony in light of his endorsement . . . and the undisputed relationship between the patient's neurological and cardio-vascular condition").

<sup>61. 780</sup> F.2d 147, 155 (1st Cir. 1985) (citing Alvarado v. Weinberger, 511 F.2d 1046, 1049 (1st Cir. 1975)). See also Quinton v. Farmland Indus., Inc., 928 F.2d 335, 337 (10th Cir. 1991) (in rejecting contention that a doctor of veterinary medicine, as opposed to a toxicologist, is unqualified to proffer opinion regarding toxic effects of substances on dairy cows, the court stated: "This assumption about the insufficiency of general medical study, which reflects the implausible view that such training qualifies a doctor to diagnose and treat a wide range of physical disorders in the real world but not to render expert opinions about particular examples in the courtroom, has been expressly rejected in the case of physicians.").

ters.<sup>62</sup> They had far more specialized knowledge about DES than a physician whose knowledge about DES was acquired for the purpose of becoming an expert witness.<sup>63</sup>

Other opinions focus on the actual expertise of the physician in light of the issue on which expert assistance is sought. For instance, in *Christophersen v. Allied-Signal Corp.*, the court

caution[ed] . . . that although credentials can be significant, they alone are not necessarily determinative. The questions, for example, do not stop if the expert has an M.D. degree. That alone is not enough to qualify him to give an opinion on every conceivable medical question. This is because the inquiry must be into actual qualification—sufficient to assist the trier of fact. The trial judge here rightly scrutinized Dr. Miller's lack of specialized experience and knowledge.<sup>64</sup>

In a number of cases, courts have excluded the testimony of a physician on the ground that he or she lacked adequate knowledge about the issue before the court. For example, in *Will v. Richardson-Merrell, Inc.*, a Bendectin case, the court refused to admit testimony on causation by a plastic surgeon with "relatively little, if any, scientific knowledge regarding Bendectin, its components, or its effects." <sup>65</sup> Similarly, in *Chikovsky v. Ortho Pharmaceutical Corp.*, a post-*Daubert* case, the court found that the testimony of the plaintiff's sole expert that the defendant's product caused birth defects would not be admissible, noting that the expert, an obstetrician–gynecologist, had no specialized training in embryology or teratology, did not know if genetic explanations existed for the child's birth defects, and did not know how much Retin-A the mother might have absorbed through topical applications, and that the theory that topical applications of Retin-A during pregnancy can cause birth defects had not been tested. <sup>66</sup>

66. 832 F. Supp. 341, 344–46 (S.D. Fla. 1993). See also O'Conner v. Commonwealth Edison Co., 13 F.3d 1090, 1107 & n.19 (7th Cir. 1994), and discussion *supra* note 44. But see Rubinstein v. Marsh, No. CV-80-0177, 1987 WL 30608, at \*6–7 (E.D.N.Y. Dec. 10, 1987) (in action claiming that infants' birth defects were caused by the defendant's product, court found that plaintiffs' experts "were qualified by virtue of the fact that each was a doctor"; court ultimately granted judgment for defendants in this bench-tried case on the ground that plaintiffs had completely failed to prove causation; the court stated that one of plaintiffs' experts was a pediatrician who had never diagnosed a drug-related birth defect in his own practice, had no experience in obstetrics or gynecology, did not know when hands and fingers differentiate in embryo (one infant had suffered a severe hand malformation), and did not know the properties of defendant's drug; the second expert's testimony was characterized as even less compelling). *Cf.* Payton v. Abbott Labs, 780 F.2d 147, 157 (1st Cir. 1985) (see

<sup>62.</sup> Payton, 780 F.2d at 155-56.

<sup>63.</sup> See discussion of the professional witness or the physician whose expertise is derived solely from the work of other experts *infra* § II.D.3. See also discussion of the secondhand expert *infra* § II.D.2.

<sup>64. 939</sup> F.2d 1106, 1112-13 (5th Cir. 1991) (en banc) (citation omitted), cert. denied, 112 S. Ct. 1280 (1992).

<sup>65. 647</sup> F. Supp. 544, 548–49 (S.D. Ga. 1986). See also Maddy v. Vulcan Materials Co., 737 F. Supp. 1528, 1533 (D. Kan. 1990) (osteopath specializing in pulmonary medicine had "no special skill or expertise in determining the health effects of toxic chemical exposure"; court granted summary judgment because of plaintiff's inability to establish causation); Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at \*31–38 (W.D.N.C. Oct. 29, 1990) (doctor not permitted to opine about length and degree of plaintiff's exposure to substance; doctor lacked training and knowledge about substance and was not familiar with relevant scientific literature).

It is the actual knowledge of the physician and how it relates to the controverted issue that must be examined, rather than credentials bearing on specialization. For instance, a physician in general practice who is not a board-certified psychiatrist may express an opinion about the mental condition of a patient for whom the physician is prescribing medication to counter depression.<sup>67</sup> A treating physician may express an opinion about whether his or her patient's exposure to benzene resulted in leukemia if the physician is acquainted with the body of epidemiological literature relating benzene exposure to leukemia.<sup>68</sup> In *In re Joint Eastern & Southern District Asbestos Litigation*, the Second Circuit found that the district judge had been "overly harsh" in rejecting as an expert a specialist in internal medicine who had been retained to testify that the plaintiff's colon cancer was caused by asbestos exposure.<sup>69</sup>

#### b. Engineers

The opinions indicate that in some cases a court will find that the proffered expert's knowledge of general engineering principles does not entitle the expert to render a particular opinion about a specialized topic. For example, in *Perkins v. Volkswagen of America, Inc.*, a specialist in mechanical engineering with no experience in designing entire automobiles was properly permitted to express opinions on general mechanical engineering principles, but prohibited from testifying as an expert in automotive design.<sup>70</sup> In other cases, courts have found an engineer's knowledge adequate in light of the subject matter of the testimony and the engineer's education and training. For example, in *Martin v. Fleissner GmbH*, experts who had no direct experience with the particular crimper machine involved were permitted to testify because they were specialists in machine design and were familiar with the general principles of the machine's rollers as a result of experience with similar machines.<sup>71</sup>

71. 741 F.2d 61, 63–64 (4th Cir. 1984). See also Coleman v. Parkline Corp., 844 F.2d 863, 865–66 (D.C. Cir. 1988) (although expert had no practical experience with loading domes for elevator cab interiors, he was

discussion within this section); in *Payton*, court denied defendants' motion for a directed verdict, noting that the uncontradicted testimony provided by plaintiff's experts, that the DES-affected organs developed between the sixth and twenty-second weeks of pregnancy, and that mother of plaintiff took DES commencing in the fifteenth week, was sufficient evidence to allow the jury to have found causation.

<sup>67.</sup> Sprague v. Bowen, 812 F.2d 1226, 1231-32 (9th Cir. 1987) (patient was seeking disability payments).

<sup>68.</sup> Mason v. Texaco, Inc., 741 F. Supp. 1472, 1496–97 (D. Kan. 1990), aff'd in part and modified in part, 948 F.2d 1546 (10th Cir. 1991), cert. denied, 112 S. Ct. 1941 (1992).

<sup>69. 964</sup> F.2d 92, 97 (2d Cir. 1992).

<sup>70. 596</sup> F.2d 681, 682 (5th Cir. 1979). See also Hoban v. Grumman Corp., 717 F. Supp. 1129, 1133–34 (E.D. Va. 1989) (licensed professional engineer was not permitted to testify as an expert regarding aircraft engines or fuel systems where his only formal education in aerodynamics was as an undergraduate and he had never worked in the field), *aff'd without op.*, 907 F.2d 1138 (4th Cir. 1990); Tokio Marine & Fire Ins. Co. v. Grove Mfg. Co., 762 F. Supp. 1016, 1017–18 (D.P.R. 1991) (proposed witness's work as civil engineer in construction field did not qualify him as an expert concerning the design and manufacture of cranes), *aff'd*, 958 F.2d 1169, 1173–75 (1st Cir. 1992) (court agreed that trial judge's refusal to permit someone of expert's back-ground to offer opinion as to "defect" in crane was not clear error; court stated that it was a closer question whether the expert, who had investigated the cause of crane accidents, should have been permitted to render opinion about how accident occurred; but court affirmed, noting that the expert had never inspected crane or spoken to operator, and that he had a "hired gun" background). See further discussion *infra* § II.D.3.

#### 2. The "secondhand" expert

May an expert testify when his or her expertise is based solely on work done by others so that the expert is summarizing other people's work? In an extreme case, the court may conclude that the testimony amounts to nothing more than "a conduit for hearsay testimony." 72 At other times, it may be much more difficult to determine the extent to which the proffered witness is adding something of his or her own to information derived from others. For example, in Loudermill v. Dow Chemical Co., the plaintiff claimed that the decedent's cirrhosis of the liver was a direct result of the decedent's exposure to a halogenated hydrocarbon while working at the defendant's plant.73 The plaintiff's expert on causation had extensive academic and practical knowledge in the field of toxicology, but admitted on voir dire that he was not specifically familiar with the relationship between halogenated hydrocarbons and liver toxicity. The appellate court stated that "Dr. Lowry's credentials are not unassailable in the specific area of the relationship between halogenated hydrocarbons and liver injury," but found no abuse in discretion in permitting opinion on causation based on examination of microscopic specimen slides, pathology and autopsy reports, government records, and publications concerning liver injuries caused by halogenated hydrocarbons.74

#### 3. The "professional" witness

Closely related to secondhand witnesses are the "professional" witnesses who spend the bulk of their time testifying in court rather than working in their alleged field of expertise, particularly those who have testified as an expert "in an extraordinary array of dissimilar fields."<sup>75</sup> The fact that proffered experts spend substantially all of their time in connection with litigation is not in itself a disqualification.<sup>76</sup> The time spent in court does not, however, add to the witness's qualifications.<sup>77</sup>

77. See Thomas J. Kline, Inc. v. Lorillard, Inc., 878 F.2d 791, 800 (4th Cir. 1989) ("Although it would be incorrect to conclude that Gordon's occupation as a professional expert alone requires exclusion of her testi-

experienced in the investigation of accidents involving the loading of industrial materials and knowledgeable about OSHA regulations and the safety literature on loading); Exum v. General Elec. Co., 819 F.2d 1158, 1163–64 (D.C. Cir. 1987) (professional engineer with special expertise in the area of safe industrial design was qualified to testify on feasible and economical alternatives to french fryer although he had no experience with kitchen equipment; he had worked at OSHA and Institute of Safety Analysis); Knight v. Otis Elevator Co., 596 F.2d 84, 87–88 (3d Cir. 1979) (engineer specializing in materials engineering and safety could testify even thought he was not a specialist in elevators).

<sup>72.</sup> Hutchinson v. Groskin, 927 F.2d 722, 725 (2d Cir. 1991) (defendant's medical expert testified that he had reviewed three letters from eminent oncologists that had been sent to defense counsel).

<sup>73. 863</sup> F.2d 566, 568-570 (8th Cir. 1988).

<sup>74.</sup> Id.

<sup>75.</sup> See Tokio Marine & Fire Ins. Co. v. Grove Mfg. Co., 958 F.2d 1169, 1174–75 (1st Cir. 1992) (expert had testified on behalf of insurance companies in eighteen dissimilar fields).

<sup>76.</sup> See In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*84 (3d Cir. Aug. 31, 1994) ("The fact that most of [the expert's] work since 1976 has been for plaintiffs in litigation may undermine her credibility but does not eradicate her expertise. For litigants to have access to experts, it may be necessary for some experts to concentrate on litigation.").
77. See Thomas J. Kline, Inc. v. Lorillard, Inc., 878 F.2d 791, 800 (4th Cir. 1989) ("Although it would be

Some courts have viewed an expert's career as a professional witness as a reason for scrutinizing the expert's opinion carefully to see whether it should be excluded on grounds discussed above. In *Tokio Marine & Fire Insurance Co. v. Grove Manufacturing Co.*,<sup>78</sup> the court stated, "In a field like accident reconstruction that is more art than science, the trial judge has particular liberty to eschew 'professional witnesses.'" The court agreed with the district judge below that the expert's "'hired gun' background as an instant expert in an astonishing number of other areas suggested he 'would not possess the professional safe-guards ensuring objectivity."<sup>79</sup>

#### E. Limiting Expert's Testimony

Although the expert may be qualified, the court may impose restrictions on the opinions that the expert will be allowed to express. When the proffered witness's expertise with regard to the relevant issues is of a generalized nature, the court may decide that the expert is incapable of assisting the trier with regard to the ultimate issues in a case. Instead, the court may, for instance, limit the scope of the testimony to foundational or background matters.<sup>80</sup>

Courts may also restrict an expert's testimony to the field in which he or she has specialized knowledge, and refuse to allow the expert to testify to related matters in a field in which the expert has no special expertise. This issue arises with regard to probabilistic evidence that may require a statistical analysis in addition to testimony about the principles of some other scientific field. For the results of DNA testing to be admitted, for instance, testimony might be required from a population geneticist or statistician in addition to testimony from someone knowledgeable about DNA testing techniques.<sup>81</sup>

80. See, e.g., Perkins v. Volkswagen of Am., Inc., 596 F.2d 681, 682 (5th Cir. 1979) (mechanical engineer with no experience in designing automobiles permitted to testify about general mechanical engineering principles but not as an expert in automotive design); *In re* Related Asbestos Cases, 543 F. Supp. 1142, 1149–50 (N.D. Cal. 1982) (environmental consultant could testify as a foundational witness and identify articles that he located written on asbestos hazards but could not qualify the articles as evidence or render an opinion about controverted issues); Wilkinson v. Rosenthal & Co., 712 F. Supp. 474, 478 (E.D. Pa. 1989) (expert permitted to testify about basic principles of commodity investing but not about what constitutes excessive trading).

81. See Judith A. McKenna et al., Reference Guide on Forensic DNA Evidence § III.C, in this manual.

mony, it would be absurd to conclude that one can become an expert simply by accumulating experience in testifying."), cert. denied, 493 U.S. 1073 (1990).

<sup>78. 958</sup> F.2d 1169 (1st Cir. 1992).

<sup>79.</sup> *Id.* at 1174–75. Courts have hinted that they might reject an opinion if the expert reached his conclusion before having acquired the expertise needed to form the opinion. See Viterbo v. Dow Chem. Co., 826 F.2d 420, 423 n.2 (5th Cir. 1987) ("We agree that an expert who forms an opinion before he begins his research is biased and lacking in objectivity. See Perry v. United States, 755 F.2d 888 (11th Cir. 1985). Because we reject Dr. Johnson's opinion on other grounds, it is not necessary to resolve this question. We would note, however, that this could be an additional ground indicating lack of reliability of his opinion."). *Cf. In re* Air Crash Disaster at New Orleans, 795 F.2d 1230, 1234 (5th Cir. 1986) ("[E]xperts whose opinions are available to the highest bidder have no place testifying in a court of law, before a jury, and with the imprimatur of the trial judge's decision that he is an 'expert."). The expert's professional witness status, when combined with other problems, may contribute to a decision to exclude expert testimony.

It should also be noted that the judge's determination that an expert is qualified does not require the judge to make a finding in open court in the hearing of the jury. Some judges believe that such a finding by the court might unduly influence the jury "and the better procedure is to avoid an acknowledgment of the witnesses' expertise by the Court."<sup>82</sup>

#### F. Lay Opinion Testimony on Scientific Issues

Rule 701 of the Federal Rules of Evidence may permit lay witnesses to express opinions relating to scientific issues that could also be the subject of expert proof. It provides:

If the witness is not testifying as an expert, the witness' testimony in the form of opinions or inferences is limited to those opinions or inferences which are (a) rationally based on the perception of the witness and (b) helpful to a clear understanding of the witness' testimony or the determination of a fact in issue.<sup>83</sup>

The distinctions that once existed between lay and expert testimony have been blurred by the liberalization of Rule 701.

No longer is lay opinion testimony limited to areas within the common knowledge of ordinary persons. Rather, the individual experience and knowledge of a lay witness may establish his or her competence, without qualification as an expert, to express an opinion on a particular subject outside the realm of common knowledge.<sup>84</sup>

Consequently, as many of the opinions discussed below acknowledge, the witness in question could have been qualified pursuant to either Rule 701 or Rule 702.<sup>85</sup> At times, however, a proffered lay witness will not have the experience and knowledge required to render the desired opinion. For example, in *Willard v. Bic Corp.*, the court, in granting summary judgment to the defendant in a product liability action, stated that a water patrolman who was present at the accident scene and who conceded that he was not an expert in the reconstruction of boat fires would not be permitted to testify that he had concluded that plaintiff's lighter was the origin of the fire.<sup>86</sup>

<sup>82.</sup> United States v. Bartley, 855 F.2d 547, 552 (8th Cir. 1988). *See also* Charles R. Richey, Proposals to Eliminate the Prejudicial Effect of the Use of the Word "Expert" Under the Federal Rules of Evidence in Civil and Criminal Jury Trials, 154 F.R.D. 537 (1994).

<sup>83.</sup> Fed. R. Evid. 701.

<sup>84.</sup> United States v. Paiva, 892 F.2d 148, 157 (1st Cir. 1989).

<sup>85.</sup> See, e.g., Eckert v. Aliquippa & Southern R.R. Co., 828 F.2d 183, 185 n.5 (3d Cir. 1987); Ernst v. Ace Motor Sales, Inc., 550 F. Supp. 1220, 1224 (E.D. Pa. 1982), *aff'd without op.*, 720 F.2d 661 (3d Cir. 1983). See *also* Farner v. Paccar, Inc., 562 F.2d 518, 529 (8th Cir. 1977) (court emphasized that it was unnecessary to decide whether witness could be qualified as an expert, but hinted that he could have testified pursuant to Rule 702).

<sup>86. 788</sup> F. Supp. 1059, 1066-67 (W.D. Mo. 1991).

#### 1. Distinctions between Rule 701 and Rule 702

The choice of the rule pursuant to which the witness testifies may make a difference in some instances. A lay witness's opinion must be rationally based on the witness's personal perception; consequently, the nonexpert may not express an opinion until adequate personal knowledge is demonstrated.<sup>87</sup> Because the opinion must be based on facts or data personally perceived, the lay witness cannot be asked hypothetical questions.<sup>88</sup>

At times, a witness may be precluded from testifying as an expert because the party calling the witness failed to list him or her as required in a pretrial order. If the court did not also require listing the names of lay witnesses, the witness may be able to testify pursuant to Rule 701.<sup>89</sup>

#### 2. Situations in which Rule 701 witnesses testify

Testimony by lay witnesses concerning scientific and technical issues falls into two general categories:

- 1. when the witness's personal knowledge is used to authenticate or identify something that would otherwise be established by expert proof; and
- 2. when the witness's experience, combined with personal knowledge of facts being litigated, amounts to sufficient expertise to support an opin-ion that could also be provided by a Rule 702 witness.

#### a. The identifying witness

Lay witnesses routinely testify as to whether a handwriting sample<sup>90</sup> or voice sample<sup>91</sup> is that of a particular person. The Federal Rules of Evidence expressly contemplate authenticating testimony of this type as an alternative to expert testimony.<sup>92</sup>

87. See United States v. Rea, 958 F.2d 1206, 1216–18 (2d Cir. 1992) (it was error, though harmless, for trial court to admit co-worker's opinion that defendant must have known that he was participating in a tax eva sion scheme; judge did not permit inquiry into the basis for the opinion, so that there was no way to know if the opinion was based on the perception of the witness); United States v. Paiva, 892 F.2d 148, 157 (1st Cir. 1989) ("the individual experience and knowledge of a lay witness may establish his or her competence, without qualification as an expert, to express an opinion on a particular subject outside the realm of common knowledge"; district court did not abuse its discretion in admitting lay witness's opinion that substance she found was cocaine because her opinion was "rationally based on her own perceptions").

90. Experts frequently testify to the same issue, although the proficiency of handwriting experts has been questioned. See D. Michael Risinger et al., Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise," 137 U. Pa. L. Rev. 731 (1989). But see Moshe Kam et al., Proficiency of Professional Document Examiners in Writer Identification, 39 J. Forensic Sci. 5 (1994).

91. For a discussion of issues that arise with regard to voiceprint evidence, see infra § III.C.1.a.2.

92. See Rule 901(b)(2) (nonexpert opinion on handwriting) and Rule 901(b)(5) (voice identification). See, e.g., United States v. Tipton, 964 F.2d 650, 655 (7th Cir. 1992) (co-worker identified defendant's handwriting); United States v. Barker, 735 F.2d 1280, 1283 (11th Cir.) (two co-workers testified that defendant's handwriting matched that on checks), *cert. denied*, 469 U.S. 933 (1984); United States v. Vega, 860 F.2d 779, 789–90 (7th Cir. 1988) (police officer permitted to identify speaker on a recorded telephone conversation conducted primarily in Spanish on the basis of a two-hour conversation with defendant in English two years previously).

<sup>88.</sup> Teen-Ed, Inc. v. Kimball Int'l, Inc., 620 F.2d 399, 403–04 (3d Cir. 1980).

<sup>89.</sup> Id. at 404. See also MCI Telecommunications Corp. v. Wanzer, 897 F.2d 703, 706 (4th Cir. 1990).

Courts also allow a nonexpert to state an opinion as to whether the person depicted in a surveillance photograph is a particular person.<sup>93</sup> Courts have also permitted drug users associated with the defendant to identify a substance as a particular illegal drug.<sup>94</sup>

#### b. Lay witnesses with special expertise

1. Causation. Provided the witness has sufficient experience, courts have allowed a lay witness to express an opinion about the cause of the accident or damage which is the subject of the suit. For example, in *Hurst v. United States*, a pilot who flew over the scene of a river flooding was permitted to testify that flooding had not been caused by jetties built by one of the defendants.<sup>95</sup> In affirming the jury verdict for that defendant, the appellate court stressed the witness's unique background in having had thirty-nine years of experience in flying over that particular river to monitor ice jams and floods.

2. Economic issues. Although experts frequently furnish valuations which may require complex calculations, courts also allow lay valuation testimony. In *MCI Telecommunications Corp. v. Wanzer*, the court ordered a new trial on damages because the trial court excluded the testimony of a bookkeeper as to the profits made by a company with whom the defendant negotiated in breach of his fidu-

94. See, e.g., United States v. Paiva, 892 F.2d 148, 156–57 (1st Cir. 1989) (defendant's stepdaughter, a cocaine user, was permitted to testify that a few years previously she had discovered a bag of a white powder in his shoes, that the substance looked and tasted like cocaine, and that in her opinion it was cocaine; court admitted the testimony under Rule 701). See also United States v. Zielie, 734 F.2d 1447, 1456 (11th Cir. 1984) (chemical analysis not essential to conviction; two experienced marijuana dealers permitted to testify that the substance given to the defendant was marijuana), cert. denied, 469 U.S. 1189 (1985); United States v. Sweeney, 688 F.2d 1131, 1145 (7th Cir. 1982) (prior use, knowledge, and sampling of drug identified sufficient to qualify witness to testify as to identity of a drug under Rule 701).

95. 882 F.2d 306, 311–12 (8th Cir. 1989). See also Soden v. Freightliner Corp., 714 F.2d 498, 510–12 (5th Cir. 1983) (a lay witness was permitted to testify to dangerousness of truck design; witness had eighteen years of experience working on large trucks and worked as a service manager in charge of repairs and preventive maintenance on a fleet of 500 trucks, mainly defendant's); Eckert v. Aliquippa & Southern R.R. Co., 828 F.2d 183, 185 n.5 (3d Cir. 1987) (witness who was "employed by the railroad for thirty years and fully familiar with railroad procedures" was permitted to state whether proper coupling of railway car would have prevented injuries); Ernst v. Ace Motor Sales, Inc., 550 F. Supp. 1220, 1222–24 (E.D. Pa. 1982) (police officer who arrived on scene five to ten minutes after accident permitted to opine on cross-examination as to point of impact; court noted that testimony would have been admissible under either Rule 701 or 702), *aff'd without op.*, 720 F.2d 661 (3d Cir. 1983); Gravely v. Providence Partnership, 549 F.2d 958, 961 (4th Cir. 1977) (witness with twenty-six years of experience in stairway construction allowed to express an opinion regarding safety of conventional and spiral staircase construction); Farner v. Paccar, Inc., 562 F.2d 518, 528–29 (8th Cir. 1977) (witness with thirty years' experience in trucking industry could testify as to the proper design of a truck suspension system).

<sup>93.</sup> See United States v. Wright, 904 F.2d 403, 404–05 (8th Cir. 1990); United States v. Langford, 802 F.2d 1176, 1179 (9th Cir. 1986) ("such testimony is particularly valuable where . . . lay witnesses are able to make the challenged identifications based on their familiarity with characteristics of the defendant not immediately observable by the jury at trial"), *cert. denied*, 483 U.S. 1008 (1987). *Cf.* United States v. Stanley, 896 F.2d 450, 451–52 (10th Cir. 1990) (in trial for receiving child pornography, court allowed testimony of postal inspector that photographs in seized magazines were of children under eighteen years of age). Expert proof comparing the defendant with the person depicted in the photograph has also been allowed. *See, e.g.*, United States v. Alexander, 816 F.2d 164, 166–69 (5th Cir. 1987) (trial court erred in excluding proffered testimony by an orthodontist specializing in cephalometry, the scientific measurement of the dimensions of the head, and by an FBI agent with expertise in photographic comparisons), *cert. denied*, 493 U.S. 1069 (1990).

ciary duty to his employer.<sup>96</sup> The bookkeeper's testimony would have been based on records she kept, and her projection of profits would have been predicated on her personal knowledge and perception. If the jury credited her testimony, the amount of damages awarded might have been lower.97

<sup>96. 897</sup> F.2d 703, 706 (4th Cir. 1990). 97. See also Joy Mfg. Co. v. Sola Basic Indus., 697 F.2d 104, 111–12 (3d Cir. 1982) (court reversed and ordered new trial when trial judge excluded plaintiff's supervisor of production control from testifying as to the percentage of increased cost and downtime the company incurred that was attributable to the failure of the defendant's furnaces; no statistical expert was required).
# III. Is the Expert's Opinion Supported by Scientific Reasoning or Methodology?

Probably the thorniest problems surrounding expert proof center on a court's scrutiny of an expert opinion to determine if the expert's reasoning and methodology are scientifically valid. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the Supreme Court recognized the importance of this question; it termed the "scientific validity . . . of the principles that underlie a proposed submission" as the "overarching subject" of the inquiry the trial judge must undertake.<sup>98</sup> It confirmed the trial judge's responsibility to make a preliminary assessment pursuant to Rule 104(a) "of whether the reasoning or methodology underlying the testimony is scientifically valid" and recognized that "[m]any factors will bear on the inquiry." <sup>99</sup>

Disputes as to whether evidence is based on "scientifically valid princi ples" <sup>100</sup> arise primarily with regard to novel scientific evidence: Scientific principles gradually gain recognition until they are viewed as incontestable and become the subject of judicial notice. <sup>101</sup> When, however, experts seek to substantiate their conclusions by reference to as yet disputed scientific theories, a number of pervasive and related questions have to be considered by the court:

- Under what circumstances can judges with limited scientific expertise exclude an expert's opinion because of flaws in the scientific reasoning or methodology on which it rests? The expert, after all, is an expert precisely because he or she has specialized knowledge that a nonexpert in the relevant field lacks.
- 2. When is scientific validity a question of law for the court rather than a question of fact to be resolved by the trier of fact?
- 3. When will a lack of scientific validity result in the inadmissibility of expert testimony, and when will it lead to a finding of insufficiency?

98. 113 S. Ct. at 2797.
99. *Id.* at 2796.
100. *Id.* at 2799.
101. *Id.* at 2796 n.11.

## A. The Frye Test

Before the enactment of the Federal Rules of Evidence, federal courts typically approached questions relating to the validity of an expert's theory by applying the "general acceptance," or Frye, test.<sup>102</sup> In 1993, in Daubert v. Merrell Dow Pharmaceuticals, Inc.,<sup>103</sup> the Supreme Court unanimously concluded that the Frye test did not survive the enactment of the Federal Rules of Evidence.<sup>104</sup> The Court's determination wrote finis to an enormous judicial and scholarly output devoted to discussing the applicability of Frye, the meaning of Frye, and alternatives to Frye. When one looks at the actual results in comparable cases, however, it is considerably less clear how much it mattered whether a circuit purported to employ Frye or some other test.<sup>105</sup> What was significant and continues to be significant under Daubert is the extent to which a court is willing to look at the methodological underpinnings of the scientific principles being espoused and the circumstances in which courts find that a flaw in scientific reasoning leads to exclusion of the expert's opinion, or takes an issue from the jury as a matter of law. As will be seen in the discussion below, pre-Daubert courts scrutinized and screened scientific testimony in a variety of situations regardless of whether they subscribed to Frye or to other tests.

As Daubert acknowledges and the cases decided before Daubert illustrate, the scientific issues and the differing procedural postures in which these issues arise are too complex to be amenable to resolution by precise verbal formulas. Furthermore, judicial attitudes toward issues of scientific validity may change over time. In toxic tort litigation, this evolution appears attributable to two simultaneously occurring phenomena:

1. Courts become more conversant in general with the parameters of scientific and probabilistic reasoning as they are exposed to complex statis-

<sup>102.</sup> The "general acceptance" test had its genesis in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). In that case, in the course of discussing whether polygraph evidence should be admitted, the court made the following statement: "[W]hile courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained *general acceptance* in the particular field in which it belongs." *Id.* at 1014 (emphasis added).

<sup>103. 113</sup> S. Ct. 2786 (1993).

<sup>104.</sup> The Court employed the plain-meaning approach it had previously used when interpreting the Federal Rules of Evidence in construing Rule 702, which does not mention "general acceptance." *Id.* at 2793–94. See, e.g., Bourjaily v. United States, 483 U.S. 171 (1987). Consequently, *Frye*, upon which the Ninth Circuit, as well as other circuits, had relied, is dead as the talisman for determining when scientific evidence is admissible, although general acceptance remains a factor that may be considered in assessing the validity of reasoning and methodology. *Daubert*, 113 S. Ct. at 2797. See discussion *infra* § III.C.1.

<sup>105.</sup> See, e.g., United States v. Smith, 869 F.2d 348, 351–54 (7th Cir. 1989), in which the court relied on Frye to admit voiceprint evidence (despite a National Research Council study showing an absence of scientific consensus) by using factors previously used in United States v. Williams, 583 F.2d 1194 (2d Cir. 1978), cert. denied, 439 U.S. 1117 (1979), the Second Circuit's leading case on abandoning the Frye methodology.

tical issues and problems of causation in cases rife with scientific uncertainty.  $^{\rm 106}$ 

2. Courts become more knowledgeable about particular factual issues through the gradual accumulation of evidence as categories of related cases work their way through the litigation process and mature. <sup>107</sup>

Consequently, the judicial desire for efficiency must be balanced against the need to allow scientific issues an opportunity to develop.

### B. The Daubert Test

The opinion for the majority commenced its discussion of the trial judge's obligation to screen "purportedly scientific evidence" by construing the words "scientific" and "knowledge," which appear in Rule 702.<sup>108</sup> It explained that "scientific' implies a grounding in the methods and procedures of science," while "the word 'knowledge' connotes more than subjective belief or unsupported speculation." <sup>109</sup> When the Court put these two words, "scientific" and "knowledge," together, it concluded that Rule 702 limits expert testimony on scientific issues to opinions that are the product of a scientific thinking process. The Court wrote:

[I]n order to qualify as "scientific knowledge," an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—*i.e.*, "good grounds," based on what is known. In short, the requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability.<sup>110</sup>

The Court went on to explain that in order to determine whether the expert's proffered testimony pertains to "scientific knowledge," the trial judge must assess "whether the reasoning or methodology underlying the testimony is scientifically valid." <sup>111</sup> The Court stressed that "[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate." <sup>112</sup> It also provided a list of illustrative factors that bear on the trial judge's inquiry. This inventory, which the Court cautioned should not be considered definitive, corroborated the Court's conception of science as an empirical endeavor in which testing plays a crucial role. Mentioned by the Court were

109. *Id.* When an expert seeks to testify about scientific knowledge pursuant to Fed. R. Evid. 702, the inferences or assertions that the expert is making "must be derived by the scientific method." *Id.* 

110. 113 S. Ct. at 2795.

<sup>106.</sup> For a comprehensive discussion of why legal and scientific approaches to the issue of causation may differ, see Troyen A. Brennan, Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation, 73 Cornell L. Rev. 469 (1988).

<sup>107.</sup> See Joseph Sanders, The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts, 43 Hastings LJ. 301 (1992).

<sup>108.</sup> Daubert, 113 S. Ct. at 2795.

<sup>111.</sup> Id. at 2796.

<sup>112.</sup> Id. at 2797.

- 1. "falsifiability" (whether the theory or technique can be, and has been, tested);
- 2. peer review and publication (submission to peer review is not dispositive, but is viewed as a component of good science);
- 3. the known or potential rate of error and the existence and maintenance of standards controlling the technique's operation; and
- 4. general acceptance of the methodology in the scientific community (still a factor to be considered but not dispositive).<sup>113</sup>

The trial court must also decide whether the expert's testimony fits the facts of the case. This condition, as the Court recognized, is essentially one of relevance.<sup>114</sup> The "helpfulness" standard incorporated in Rule 702 means that the expert's opinion must relate to an issue that is actually in dispute and must provide a valid scientific connection to the pertinent inquiry.<sup>115</sup>

# C. Contexts in Which Questions Relating to Scientific Validity Arise

The discussion which follows is broken down into three broad areas which encompass the situations in which courts have confronted the issue of scientific validity:

- 1. issues with regard to a particular discipline;
- 2. issues with regard to the methodology and reasoning of a particular scientific theory; and
- 3. issues with regard to statistical estimates.

Although pre-Daubert cases did not necessarily frame the evidentiary issues in these terms, the cases are useful in illustrating the kinds of fact patterns that arise.

These questions differ from the qualification problems discussed in section II, which focus on whether the expert knows enough about the particular theory he or she is seeking to espouse; the emphasis here is on whether the alleged science has something to offer the judicial system. The line between qualification questions and validity questions is at times blurred, as the discussion of treating physicians and the *Christophersen* case indicates, <sup>116</sup> and may be even more indistinct after *Daubert*. <sup>117</sup> The boundary between discipline and theory, drawn

<sup>113.</sup> Id. at 2796–97. See Bert Black et al., Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, 72 Tex. L. Rev. 715, 782–86 (1994) (other factors a court might consider).

<sup>114.</sup> Daubert, 113 S. Ct. at 2795-96.

<sup>115.</sup> *Id.* The Court offered an example of the expert whose scientific training about the phases of the moon enables the expert to establish whether it was dark on a particular night. If that is the issue, the expert's testimony fits. But evidence that the moon was full on the night in question does not assist the trier on the issue of whether an individual is likely to be irrational when the moon is full. *Id.* at 2796. See discussion of this as pect of *Daubert supra* § I.B.

<sup>116.</sup> See discussion of Christophersen v. Allied-Signal Corp., 939 F.2d 1106 (5th Cir. 1991), cert. denied, 112 S. Ct. 1280 (1992) supra § II.D.1.a.

<sup>117.</sup> See supra § II.A.

in Questions 1 and 2 below, was generally ignored by the courts before *Daubert*. An attempt is made to separate the two types of questions, however, for reasons which are discussed below. The discussion proceeds as to each category by considering the different ways in which courts have analyzed the relevant issues and the consequences of their approach.

1. Issues with regard to a particular discipline

A basic question that courts may have to resolve is whether the expert's discipline or field can make any contribution to the resolution of the controverted issue to which the expert proof is directed. Prior to *Daubert*, the *Frye* test, with its "general acceptance" formulation, was not well suited to resolving this issue. Even though experts may have been relying on generally accepted theories in their field, as required by *Frye*, the field was perhaps not capable of providing assistance with regard to the controverted issue before the trier of fact. After considering some of the contexts in which these problems arose, the following discussion considers a post-*Daubert* approach.

a. Challenging a group of experts' methodology as lacking the characteristics of science

Judges would undoubtedly exclude certain evidence—such as predictions based on astrology—as incapable of proving a fact in issue. But on what basis does a court reach this conclusion? The expert who acknowledges reliance on a theory that has not been validated by methods accepted by his or her acknowledged peers is discussed in the next subsection. But suppose the proffered expert belongs to an organized discipline that holds regular meetings and publishes journals to put forth its theories. The proposed expert is clearly qualified in terms of the tenets of this group. Under these circumstances, what evidentiary test must the expert's testimony satisfy? Relatively few cases to date have confronted this issue directly.

1. Clinical ecology: Is a field scientific if its theories are not testable? Clinical ecologists claim that various kinds of environmental insults may depress a person's immune system so that the exposed person develops a "multiple chemical sensitivity," that is, becomes hypersensitive to other chemicals and naturally occurring substances. According to this theory, not all persons will necessarily develop the same symptoms as a consequence of this hypersensitivity; each person exposed may present a distinctive profile. Clinical ecologists have not been recognized by traditional professional organizations within the medical community,<sup>118</sup> al-

<sup>118.</sup> See discussion in Sheila Jasanoff, *Science on the Witness Stand*, 6 Issues in Sci. & Tech. 80, 86 (1989) (identifying criticisms of the theory and methods of clinical ecology by the American Academy of Allergy and Immunology and the California Medical Association).

though other authorities are somewhat more open about their contributions, <sup>119</sup> and recent studies may provide some support for some of their claims. <sup>120</sup> The leading professional societies in the fields of allergy and immunology have rejected clinical ecology "as an unproven methodology lacking any scientific basis in either fact or theory," but "numerous other professional organizations and societies . . . have not discredited completely the potential usefulness of clinical ecology." <sup>121</sup> According to the reported cases, federal courts have rejected the opinions of clinical ecologists, although clinical ecologists have fared better in some state courts.

In Sterling v. Velsicol Chemical Corp., a class action in which the plaintiffs claimed that hazardous chemicals from the defendant's landfill had contaminated the water supply, damages had been awarded to the plaintiffs for alleged impairment to their immune systems and to one plaintiff for additional learning disorders resulting from immune system impairment.<sup>122</sup> On appeal, the defendant argued that the district court had improperly admitted the testimony of clinical ecologists supporting these claims "because the principles upon which the experts based their conclusions were not in conformity to a generally accepted explanatory theory."<sup>123</sup>

The appellate court agreed with the defendant and reversed the entire award of damages related to immune system impairment. The court stated the following test for confirming the existence of a "generally accepted explanatory theory":

In order to prevent deception or mistake and to allow the possibility of effective response, there must be a demonstrable, objective procedure for reaching the opinion and qualified persons who can either duplicate the result or criticize the means by which it was reached, drawing their own conclusions from the underlying facts.<sup>124</sup>

In applying this test, the court pointed to the lack of replication ("plaintiffs' experts neither performed nor could identify any studies of the effects of carbon tetrachloride or chloroform on the immune system" <sup>125</sup>) and the lack of standard, objective procedures ("plaintiffs' experts neither personally examined or interviewed plaintiffs, nor performed the requisite medical tests" <sup>126</sup>).

The court's language in *Sterling* is consistent with the Supreme Court's reference to testability in *Daubert*. "General acceptance," which the Supreme Court listed as a relevant factor, also played a role in *Sterling*, as the court buttressed its

<sup>119.</sup> See discussion in Troyen A. Brennan, Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts, 51 U. Pitt. L. Rev. 1, 59–62 (1989).

<sup>120.</sup> See Jon R. Luoma, New Effect of Pollutants: Hormone Mayhem, N.Y. Times, Mar. 24, 1992, at C1.

<sup>121.</sup> Sterling v. Velsicol Chem. Corp., 855 F.2d 1188, 1208 (6th Cir. 1988).

<sup>122.</sup> Id. at 1188.

<sup>123.</sup> Id. at 1208.

<sup>124.</sup> Id.

<sup>125.</sup> Id. at 1208-09.

<sup>126.</sup> Id. at 1209.

conclusion by stating that plaintiffs' experts lacked "a widely accepted medical basis for reaching" their conclusions.<sup>127</sup> Although the court's discussion of methodology and general acceptance seems couched in terms of the admissibility of the expert testimony, the court does not ultimately state that the testimony should not have been admitted, only that it was "insufficient to sustain plaintiffs' burden of proof."<sup>128</sup>

A theory whose central feature is that persons react differently to various substances because of their individual peculiarities is of course difficult to test. Whether the courts in the post-*Daubert* era will treat inability-to-test cases differently from failure-to-test cases and whether they will rule on admissibility or sufficiency grounds remains to be seen.

2. Forensic techniques: How much inquiry into testing is required? Over the years a number of forensic techniques that initially found their way into the courtroom have subsequently fallen into disfavor.<sup>129</sup> The original judicial approval of these techniques was perhaps facilitated by the Frye test. Because Frye emphasized "general acceptance" in a particular field, a well-organized group of expert witnesses in some instances became "the field." "General acceptance" by these experts then verified the reliability of the evidence. For instance, voiceprint evidence was introduced into the courts through the efforts of a small number of experts who were former employees of Bell Labs and the Michigan state police even though their conclusions had never been proven by empirical evidence.<sup>130</sup> Voiceprint evidence gradually faded from the courtroom after a blue-ribbon Committee on Evaluation of Sound Spectrograms, appointed by the National Academy of Sciences, concluded that the scientific results reported to date did not provide quantitative information about improvements in accuracy of voice identifications associated with the use of voice spectrograms.<sup>131</sup> The "paraffin test" is another example of a technique that passed Frye and is now discredited.<sup>132</sup> Handwriting analysis is currently the subject of de bate.<sup>133</sup>

Whether the *Daubert* case, with its emphasis on testing, will cause courts to be more cautious before admitting evidence produced by a new forensic tech-

<sup>127.</sup> Id.

<sup>128.</sup> Id.

<sup>129.</sup> See, e.g., Michael J. Saks, Implications of the Daubert Test for Forensic Identification Science, 1 Shepard's Expert & Sci. Evidence Q. 427 (1994); Randolph N. Jonakait, Forensic Science: The Need for Regulation, 4 Harv. J.L. & Tech. 109 (1991).

<sup>130.</sup> Andre A. Moenssens, Admissibility of Scientific Evidence—An Alternative to the Frye Rule, 25 Wm. & Mary L. Rev. 545, 556–57 (1984).

<sup>131.</sup> Committee on Evaluation of Sound Spectrograms, National Research Council, On the Theory and Practice of Voice Identification 10 (1979) (technique "lacks a solid theoretical basis of answers to scientific questions concerning the foundations of voice identification. This disparity between practice and theory appears to be recognized by practitioners and scientists involved in the field of voice identification.").

<sup>132.</sup> Paul C. Giannelli, The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later, 80 Colum. L. Rev. 1197, 1224–25 (1980).

<sup>133.</sup> See the argument in Risinger et al., *supra* note 90. But see Kam et al., *supra* note 90, at 7, 13 (conceding "a lamentable lack of empirical evidence" but finding that FBI document examiners performed significantly better than college-educated nonexperts).

nique remains to be seen. <sup>134</sup> Certainly, however, *Daubert* requires the proponent to bear the burden of demonstrating the technique's capacity to produce a reliable result. Recent experience with DNA evidence indicates that prosecutors may have overstated their claims in the early cases, <sup>135</sup> and that the defense may lack the training and resources to make the inquiries that *Daubert* requires. It takes time, money, and effort to understand a new technique sufficiently to ask the right questions about how it has been tested and how it works. <sup>136</sup>

Commentators on forensic evidence have noted that many courts are reluctant to provide expert assistance to indigent defendants.<sup>137</sup> A failure to do so may be especially problematic when the prosecution is relying on a novel forensic technique that must pass the *Daubert* test. When the defense is unable to put forward an expert, questions about the methodological validity of the new technique may not be adequately explored.<sup>138</sup> Whether or not funds are made available to the defense to hire experts, the court might want to seek assistance for itself. It could either direct a magistrate judge to conduct an appropriate inquiry<sup>139</sup> or appoint a methodological expert or experts pursuant to Rule 706 of the Federal Rules of Evidence to assist the court in understanding the issues that are likely to arise.

135. Compare Andrews v. State, 533 So. 2d 841, 843 (Fla. Dist. Ct. App. 1988) (first case in which DNA evidence was admitted to prove guilt in a criminal trial; prosecution expert testified that the probability that the DNA in question came from someone else was 1:839,914,540), *affd*, 533 So. 2d 851 (Fla. Dist. Ct. App. 1988) with United States v. Yee, 134 F.R.D. 161, 164 (N.D. Ohio 1991) (FBI first stated the likelihood of a match at 1:270,000 and then recalculated the odds at 1:35,000), *affd sub nom*. United States v. Bonds, 12 F.3d 540, 552 (6th Cir. 1993) (on appeal, defense argued that under the ceiling principle advocated by the report of the National Research Committee of the National Academy of Sciences (issued after trial), probability of defendant's DNA being found in the relevant population was 1 in 17; government rejoined that even under the ceiling principle, the odds would be 1 in 6,200).

136. See, e.g., Peter J. Neufeld & Neville Colman, When Science Takes the Witness Stand, Sci. Am., May 1990, at 46, 53.

137. See Paul C. Giannelli, "Junk Science": The Criminal Cases, 84 J. Crim. L. & Criminology 105, 122–23 (1993) (concludes that "[w]ithout an effective right to defense experts, the accused often lacks the resources to combat junk science"; provides numerous examples).

138. For instance, the defense offered no expert in the first DNA case, Andrews v. State, 533 So. 2d 841 (Fla. Dist. Ct. App. 1988), aff'd, 533 So. 2d 851 (Fla. Dist. Ct. App. 1988), which is discussed supra note 135.

139. In United States v. Yee, the magistrate judge conducted a six-week *Frye* hearing to determine the admissibility of DNA evidence; at the hearing, the government called six expert witnesses, the defendants called five expert witnesses, and the court called its own witnesses. The magistrate judge issued a 120-page report and recommendation, which was adopted by the district court. 129 F.R.D. 629 (N.D. Ohio 1990), adopted, 134 F.R.D. 161 (N.D. Ohio 1991), aff'd sub nom. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993).

<sup>134.</sup> Randolph N. Jonakait, *Real Science and Forensic Science*, 1 Shepard's Exp. & Sci. Evidence Q. 435 (1994) (if *Daubert* is taken seriously, a "dramatic change" will occur with regard to scientific evidence in criminal cases). *Cf.* Sheila Jasanoff, *What Judges Should Know About the Sociology of Science*, 77 Judicature 77, 81 (1993) (pointing out that "[w]hether or not a theory or technique has been adequately tested is as much a social as a scientific question," and that a particular community of experts is unlikely to question testing if an issue is not contentious within a given community). This criticism suggests the possibility of having a court employ a methodological expert outside the particular community in question.

### b. Challenging a methodology as lacking probative value

The Court did not discuss in *Daubert* an issue that may prove critical in the post-*Daubert* era—may a court exclude an opinion that is the product of a standard, reliable methodology on the ground that the opinion is not sufficiently probative with regard to the issue for which it is being offered? The problem is not—as in the previous section—whether the expert's conclusions were adequately tested in accordance with the types of factors discussed in *Daubert*. In the cases now being discussed, the expert has reached a conclusion that was empirically verified according to the expert's discipline. That conclusion is offered to prove a consequential, material issue in controversy. The opponent claims, however, that the expert's opinion does not adequately tend to establish the controverted issue and should therefore be excluded pursuant to an evidentiary rule<sup>140</sup> or on sufficiency grounds.

The controversy centers on the Supreme Court's statement in *Daubert* that "[t]he focus, of course, must be solely on principles and methodology, not on the conclusions that they generate."<sup>141</sup> Some commentators interpret this comment to mean that an opinion must be admitted once an expert demonstrates reliance on a standard scientific methodology; otherwise, the court would be second-guessing the expert's conclusion contrary to *Daubert*.<sup>142</sup> According to this approach, if the expert uses a standard methodology, the court may not exclude the opinion as not adequately probative of an issue in controversy.

Others, however, view the Supreme Court's remark as directed to an entirely different concern—as not permitting a court to choose between competing conclusions when both are based on a reliable methodology and the probative value of the conclusion in question is established.<sup>143</sup> But reliability alone does not make evidence probative. In the arena of nonexpert proof, for instance, courts often reject evidence not because it is untrustworthy but because no valid evidential hypothesis connects the evidence to the proposition for which it is offered.<sup>144</sup> Similarly, "a scientist may reach the wrong conclusion because the prediction being tested is not really a logical consequence of the hypothesis or

<sup>140.</sup> Objections might be phrased in terms of Fed. R. Evid. 401 (the opinion is not relevant), Rule 702 (it does not assist the trier), Rule 703 (experts would not reasonably rely on such an opinion), or Rule 403 (the probative value of the opinion is substantially outweighed by "the danger of unfair prejudice, confusion of the issues, or misleading the jury . . . . ").

<sup>141.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2797 (1993).

<sup>142.</sup> See Kenneth J. Chesebro, Taking Daubert's "Focus" Seriously: The Methodology/Conclusion Distinction, 15 Cardozo L. Rev. 1745 (1994); Anthony Z. Roisman, Conflict Resolution in the Courts: The Role of Science, 15 Cardozo L. Rev. 1943 (1994).

<sup>143.</sup> If, for instance, two physicians reach differing prognoses with regard to the permanency of a plaintiff's injuries after using standard tests and employing standard medical practices, both opinions will be sufficiently reliable to gain admission. According to *Daubert*'s analysis, a court may not decide that it prefers one of the physician's conclusions—the issue must be left to the jury.

<sup>144.</sup> Impeccably reliable evidence that a defendant in an accident case was speeding before the accident at a point twenty miles from where the accident occurred may still be rejected to prove that the defendant was driving too quickly when the accident occurred.

because of erroneous assumptions in an experiment's design."<sup>145</sup> Experts, however, are permitted to testify about scientific matters because of their specialized scientific knowledge. Does this mean that the court must defer to the expert's assessment about what the opinion proves without scrutinizing the expert's assumptions? Should the expert's conclusion about what the evidence proves in the world of science be dispositive in determining what the evidence proves in a court of law? These are questions the courts will have to decide after *Daubert*.

1. Extrapolation problems; animal studies. <sup>146</sup> To what extent may a court reject an expert's opinion on the ground that it rests on unfounded extrapolation? The question arises with some frequency in toxic tort litigation, when plaintiffs seek to prove that exposure to a defendant's product caused the nonsignature disease or birth defect that is the subject of the suit. In support, the plaintiff offers studies that show a correlation between the product and the disease in a number of animal species. For the results of these studies to be probative, at least two assumptions must be made: (1) that if a substance is toxic in these species of animals it must also be toxic in humans; and (2) that one can extrapolate from the higher and more intense dosage level used in the study to the lower level to which the plaintiff was actually exposed. Although scientists are willing to make these assumptions, and animal studies are routinely used in risk assessment,<sup>147</sup> a number of courts have rejected this evidence to prove causation on the ground that the underlying premises cannot be confirmed. These courts view the discrepancy between humans and animals not as a weight-of-the-evidence question for the jury, but as a matter of law requiring judicial resolution.

The contours of the courts' conclusions are still somewhat vague. For although a number of recent federal court opinions deal with animals studies, many in the context of Bendectin litigation, a number of issues have not been clearly or uniformly resolved. For instance, under what circumstances, if any, is such testimony inadmissible, and when is it insufficient? Should this problem be handled pursuant to Rule 702 or 703, or the relevancy rules? To what extent does it matter that in all of these cases contrary epidemiological evidence was available? In the Bendectin litigation, for example, the defendants introduced two types of evidence pointing to no effect in humans: epidemiological studies

<sup>145.</sup> Bert Black et al., Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, 72 Tex. L. Rev. 715, 773 (1994). The authors use the term "implicit error" for this problem, which they illustrate with the example of a scientist testing whether a parasite causes a particular disease in humans by injecting the parasite into an animal species that is immune to the disease in question. The problem is with the assumption "that animal models, particularly mammals, are generally appropriate for testing harmful effects on humans." *Id.* at 774.

<sup>146.</sup> For a more detailed discussion of animal studies in the context of toxicological proof, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § III, in this manual. Extrapolation issues also arise in other contexts. For instance, it has been pointed out that the epidemiological data on asbestos have been derived almost entirely from studies of workers. In estimating risk to the general population, one must extrapolate from effects at the high exposures seen in the workplace to the much lower doses typically seen in the ambient environment. Devra Lee Davis et al., Assessing the Power and Quality of Epidemiologic Studies of Asbestos-Exposed Populations, 1 Toxicol. & Indus. Health 93, 94 (1985).

<sup>147.</sup> Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § I.E, in this manual.

showing no statistically significant increased risk of birth defects in children born to mothers who had taken Bendectin during pregnancy, and secular trend studies showing no decrease in birth defects after Bendectin was taken off the market. Is it significant that the expert in some of these cases acknowledged the limits of animal study methodology, or may a court take judicial notice of these limitations? Will studies showing toxicity in animals by themselves discharge a plaintiff's burden of proof sufficiently to make out a question for the jury?<sup>148</sup>

In Brock v. Merrell Dow Pharmaceuticals, Inc.,<sup>149</sup> a Bendectin case, the court discussed the methodological flaws in animal studies in general<sup>150</sup> and the Bendectin studies in particular, pointing out the various extrapolations that have to be made. The court's discussion is in the context of granting judgment n.o.v. to the defendant; the court found the plaintiff's evidence insufficient and did not consider whether it was admissible.<sup>151</sup>

The Sixth Circuit in *Turpin v. Merrell Dow Pharmaceuticals, Inc.* rested its affirmance of summary judgment for the defendant in large measure on the insufficiency of the particular animal studies relating to Bendectin.<sup>152</sup> The opinion suggests that the court does not view animal studies as inherently incapable of proving causation because of the extrapolation problems discussed above. It left open the possibility that "there may be other animal experiments which . . . because of the extreme toxicity of the substance tested, would permit a reasonable jury to find that it is more probable than not that the substance causes a similar harm to humans."<sup>153</sup> Rather than relying on methodological flaws, the court seems to be evaluating the probative value of the evidence.<sup>154</sup> The court also emphasized, however, that with one exception,<sup>155</sup> all of the experts claimed only that the animal studies showed that Bendectin could possibly cause birth

148. Cf. discussion of Ferebee v. Chevron Chem. Co., 736 F.2d 1529 (D.C. Cir.), cert. denied, 469 U.S. 1062 (1984), infra § III.C.1.b.2.

149. 874 F.2d 307 (5th Cir.), modified, 884 F.2d 166 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990).

150. The court began its discussion with a case in which it had rejected animal study results used in risk analysis and commented that the "circuit has previously realized the very limited usefulness of animal studies when confronted with questions of toxicity." 874 F.2d at 313.

151. *Id.* at 315. On petition for rehearing en banc, Judge Higginbotham concurred in the dissent from the refusal to rehear the case en banc. 884 F.2d 167, 168–69 (5th Cir. 1989). Judge Higginbotham dissented on the ground that the panel had shied away from addressing the crucial issue—the admissibility of the evidence in the first place rather than its sufficiency after it is admitted. *Id. Cf. In re "Agent Orange"* Prod. Liab. Litig., 611 F. Supp. 1267, 1273–83 (E.D.N.Y. 1985) (granting summary judgment to defendants; plaintiff's expert's "resort to inappropriate studies of animals . . . cannot redeem his unfounded opinion. The conclusions set forth in the Carnow affidavit would be excluded at trial under Rule 703 of the Federal Rules of Evidence."), *aff'd*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988). *See also* Villari v. Terminix Int'l, Inc., 692 F. Supp. 568, 571 n.1 (E.D. Pa. 1988) (defendants made in limine motion to exclude animal studies and testimony of experts based on studies of the carcinogenicity of pesticide; court found evidence admissible; court noted that it was not deciding sufficiency issue and that if defendants could show at trial that "an overriding segment of the scientific community repudiates the value of the studies, it would be appropriate to seek their exclusion at that time").

152. 959 F.2d 1349 (6th Cir.), cert. denied, 113 S. Ct. 84 (1992).

153. Id. at 1359.

154. See discussion supra § III.C.1.b.

155. The court's rejection of the expert testimony of the medical doctor who testified that Bendectin did cause the defects in issue is discussed *infra* note 170.

defects; they did not assert that Bendectin had more probably than not caused birth defects. <sup>156</sup>

The admissibility of animal studies was considered at length in an extensive post-Daubert opinion by Judge Becker for the Third Circuit in *In re Paoli Rail-road Yard PCB Litigation* (Paoli II).<sup>157</sup> The court reviewed the cases decided prior to *Daubert* and concluded that the case law "is mixed."<sup>158</sup> It decided that the trial court had abused its discretion in excluding the particular animal studies at issue. The court distinguished other cases in which studies had been found inadmissible

because most involved the exclusion of animal studies in the face of *extensive* epidemiological data that failed to support causation, because none involved studies on animals particularly similar to humans in the way they react to the chemical in question, and because none involved studies the federal government had relied on as a basis for concluding the chemical was a probable health hazard.<sup>159</sup>

2. The need for probabilistic evidence: clinical medicine . The testimony of a physician that he or she is convinced to a reasonable degree of medical certainty that a plaintiff's problem was caused by a physician's negligence is frequently encountered in medical malpractice cases. Testimony that the plaintiff's problem was caused by the defendant's product is more problematic. Except in the case of signature diseases and certain known carcinogens and teratogens, some would maintain that a clinician's opinion alone is inadequate to link a particular birth injury or cancer diagnosis with exposure to a product in the absence of probabilistic evidence.<sup>160</sup> Cancers and most birth defects differ from diseases for which it can be demonstrated that exposure to a particular bacterium or virus obtained from one person will result in the development of the same disease in

<sup>156.</sup> Turpin, 959 F.2d at 1359–60. See also Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 830 (D.C. Cir. 1988) (plaintiff's expert himself acknowledged that "animal data alone would not be a sufficient bases [sic] for you to give an opinion with reasonable medical certainty that Bendectin causes birth defects in humans"), cert. denied, 493 U.S. 882 (1989).

<sup>157. 1994</sup> U.S. App. LEXIS 23722 (3d Cir. Aug. 31, 1994). The Third Circuit had previously, prior to the *Daubert* opinion, reversed a grant of summary judgment for the defendant. *In re* Paoli R.R. Yard PCB Litig. (Paoli I), 916 F.2d 829 (3d Cir. 1990), *cert. denied*, 499 U.S. 961 (1991). On remand, after five days of in limine hearings, the district court granted summary judgment again. 811 F. Supp. 1071 (E.D. Pa. 1992).

<sup>158.</sup> Paoli II, 1994 U.S. App. LEXIS at \*173.

<sup>159.</sup> Id. at \*175 (emphasis added). In applying the Daubert factors, the court noted that the studies themselves are testable, follow a generally accepted methodology, were published in peer-reviewed journals, and were used for purposes outside the litigation. Id. at \*178. "Finally," wrote the court, "although their 'fit' to proof of causation in humans is in dispute, all experts acknowledge they are of some use—at least in eliminating those chemicals not likely to cause disease in humans." Id. The court also found that the district court had abused its discretion in concluding that the studies could be excluded pursuant to Rule 403. Id. at \*178–79. See discussion infra § V.

<sup>160.</sup> See Troyen A. Brennan, Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous Substance Litigation, 73 Cornell L. Rev. 469 (1988), for a discussion of the need for probabilistic evidence to prove causation. See also infra § III.C.3.b, which considers the effect of a treating physician's testimony that specific facts about the injured party rule out causes for the disease other than the defendant's product and therefore affect the relative risk that would otherwise apply.

another person.<sup>161</sup> Because no validated theory as yet furnishes an adequate explanation about cancer formation or most birth defects, a claim by a physician that a particular product caused a plaintiff's injury based on the observation that the plaintiff developed a disease after exposure may amount to nothing more than a description of two events, exposure and disease, that are sequentially, but not causally, connected.<sup>162</sup>

Of course, the physician may have training in toxicology or epidemiology or possess specialized information about a particular controverted issue before the court.<sup>163</sup> Unlike the qualification issue discussed in section II, however, the question considered here is not whether the particular physician has enough specialized knowledge, but whether testimony by a physician relying on the methodology of clinical medicine will suffice to establish causation. *Ferebee v. Chevron Chemical Co.*, a case in which the manufacturer of paraquat, a herbicide, was sued for causing the decedent's death from pulmonary fibrosis, is often cited as holding that causation can be established by the testimony of treating physicians.<sup>164</sup> The *Ferebee* court stated:

[A] cause-effect relationship need not be clearly established by animal or epidemiological studies before a doctor can testify that, in his opinion, such a relationship exists. As long as the basic methodology employed to reach such a conclusion is sound, such as use of tissue samples, standard tests, and patient examination, products liability law does not preclude recovery until a "statistically significant" number of people have been injured or until science has had the time and resources to complete sophisticated laboratory studies of the chemical. In a courtroom, the test for allowing a plaintiff to recover in a tort suit of this type is not scientific certainty but legal sufficiency; if reasonable jurors could conclude from the expert testimony that paraquat more likely than not caused Ferebee's injury, the fact that another jury might reach the opposite conclusion or that science would require more evidence before conclusively considering the causation question resolved is irrelevant. That Ferebee's case may have been the first of its exact type, or that his doctors may have been the first alert enough to recognize such a case, does not mean that the testimony of those doctors, who are concededly well qualified in their fields, should not have been admitted.<sup>165</sup>

In a subsequent Bendectin case, *Richardson v. Richardson-Merrell, Inc.*,<sup>166</sup> the District of Columbia Circuit explained its *Ferebee* opinion as follows:

163. See discussion supra §§ II.C–D.

164. 736 F.2d 1529, 1535-36 (D.C. Cir.), cert. denied, 469 U.S. 1062 (1984).

165. Id. at 1535-36.

166. 857 F.2d 823 (D.C. Cir. 1988), cert. denied, 493 U.S. 882 (1989).

<sup>161.</sup> See discussion of Koch's postulates in Linda A. Bailey et al., Reference Guide on Epidemiology  $\$  IV.B, in this manual.

<sup>162.</sup> See, e.g., Porter v. Whitehall Lab., Inc., 9 F.3d 607, 611–16 (7th Cir. 1993) (court affirmed trial judge, who found that experts' conclusions were based on temporal relationship unsupported by studies or scientific methodology; appellate court found that trial court had anticipated *Daubert* in its analysis). See also In re Joint E. & S. Dist. Asbestos Litig. (Maiorana), 827 F. Supp. 1014, 1048–50 (S.D.N.Y. 1993) (finding that treating physician's differential diagnosis did not suffice to prove that plaintiff's colon cancer was caused by as bestos exposure in the absence of epidemiological proof).

Ferebee stands for the proposition that courts should be very reluctant to alter a jury's verdict when the causation issue is novel and "stand[s] at the frontier of current medical and epidemiological inquiry." If experts are willing to testify to causation in such situations and their methodology is sound, the jury's verdict should not be disturbed. <sup>167</sup>

Distinctions can be drawn between *Ferebee* and *Richardson*. Paraquat was known to be a toxic chemical; the particular injury to the lungs after chronic exposure <sup>168</sup> was biologically plausible; <sup>169</sup> and the physicians in question were experts on lung disease who relied on their examination of the patient as well on studies of the particular substance.

Recent cases suggest that courts may be unwilling to allow nonsignature cancer and birth injury claims to reach the jury solely on the basis of causation testimony by a clinical physician even in a case of first impression regarding the substance in question.<sup>170</sup>

2. Issues with regard to the methodology and reasoning of a particular scientific theory

Unlike the previous section, which concentrates on various issues that arise with the methodology of an entire discipline, this section examines expert testimony offered by a witness in an established field. The discussion focuses on a number of contexts in which courts have been confronted with challenges to a variety of theories on the ground that the expert's reasoning does not comport with the scientific method.

a. When does the expert's reasoning satisfy the Daubert test?

The Daubert opinion views science as an empirical enterprise and emphasizes the need for validation through testing. Does this mean that once an issue is labeled as "scientific," the parties must rely solely on theories that have been sub-

170. See Porter v. Whitehall Lab., Inc., 9 F.3d 607, 614–15 (7th Cir. 1993) (granting summary judgment post-Daubert); Chikovsky v. Ortho Pharmaceutical Corp., 832 F. Supp. 341, 345–46 (S.D. Fla. 1993) (same). Pre-Daubert cases: Turpin v. Merrell Dow Pharmaceuticals, Inc., 959 F.2d 1349, 1360 (6th Cir.) (in Bendectin case, court found affidavit by physician claiming that Bendectin caused plaintiff's birth defects insufficient to meet plaintiff's burden of proof on summary judgment motion: "Dr. Palmer does not testify on the basis of the collective view of his scientific discipline, nor does he take issue with his peers and explain the grounds for his differences."), cert. denied, 113 S. Ct. 84 (1992); See also Felgenhauer v. Texaco, Inc., No. 85-3671, 1987 U.S. Dist. LEXIS 11258, at \*4–9 (E.D. Pa. Nov. 30, 1987) (not officially reported) (plaintiff claimed that liver damage was caused by exposure to aromatic hydrocarbons in paints and solvents at his place of employment; court granted summary judgment after plaintiffs submitted affidavit of their expert physician (board certified in internal medicine) claiming a causal connection, although he had conceded in correspondence that he was not aware of any case reports or studies establishing such a connection). See also discussion of theories of cancer causation supra § III.C.1.b.2.

<sup>167.</sup> Id. at 832 (emphasis in original) (quoting Ferebee, 736 F.2d at 1534).

<sup>168.</sup> See In re "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1262 (E.D.N.Y. 1985), affd, 818 F.2d 187 (2d Cir. 1987), cert. denied, 487 U.S. 1234 (1988).

<sup>169.</sup> See Linda A. Bailey et al., Reference Guide on Epidemiology § IV.B.4, and Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § III.E, in this manual. See *also* Cella v. United States, 998 F.2d 418, 421 (7th Cir. 1993) (plaintiff's expert offered plausible explanation, discussed in medical literature, for why stress might cause disease from which plaintiff suffered).

jected to an empirical investigation? Or does Rule 702 still allow experts to rely on other types of "specialized knowledge" that do not satisfy the *Daubert* test for "scientific knowledge?"

1. Theories as to the cause of plaintiff's cancer . After Daubert, may a plaintiff establish causation in the absence of studies implicating a product as having a connection with the plaintiff's disease? The Thalidomide experience suggests that in some situations anecdotal observations will provide sufficient validation even in the absence of a formal study. But in the absence of a well-documented, strong association between a product and a disease, how does a party satisfy Rule 702's requirement of "a valid scientific connection to the pertinent inquiry as a precondition to admissibility?"<sup>171</sup>

For instance, in *Christophersen v. Allied-Signal Corp.*,<sup>172</sup> the majority held that the district court had not erred in excluding expert testimony that the decedent's death from a rare form of colon cancer was due to exposure to nickel and cadmium fumes at his place of work. The plaintiffs' expert conceded that he had never seen epidemiological, animal, or in vitro studies demonstrating a causal association between exposure to nickel or cadmium, or both, and colon cancer. Instead, he argued that nickel and cadmium had been implicated in small-cell carcinoma of the lungs, that the cells in the decedent's colon cancer were likewise small, and that one could conclude that small-cell carcinoma throughout the body had a similar pathogenesis.<sup>173</sup> The majority, invoking *Frye*, found that the witness had not "used a well-founded methodology or mode of reasoning."<sup>174</sup> It termed the expert's conclusion "a scientific hunch, which as far as the record shows, no one else shares."<sup>175</sup> The majority stressed that it was basing its conclusion on the lack of support in the scientific community for the expert's methodology and not on an evaluation of the correctness of the methodology.<sup>176</sup>

After *Daubert*, the admissibility of scientific evidence does not depend on the "general acceptance" of the expert's theory, although "general acceptance" is still a factor to be considered.<sup>177</sup> In the future, courts will have to determine whether reasoning by analogy, which the expert in *Christophersen* was attempting to do, might in some instances be scientifically sound.<sup>178</sup> *Daubert* suggests, however, that an expert who has not investigated the proposed analogy to the ex-

<sup>171.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993).

<sup>172. 939</sup> F.2d 1106 (5th Cir. 1991) (en banc), cert. denied, 112 S. Ct. 1280 (1992).

<sup>173.</sup> Id. at 1115-16.

<sup>174.</sup> Id. at 1111.

<sup>175.</sup> Id. at 1115.

<sup>176.</sup> Id. at 1116.

<sup>177.</sup> Cf. O'Connor v. Commonwealth Edison Co., 13 F.3d 1090, 1106 (7th Cir. 1994) (court found inadmissible expert's testimony that plaintiff's cataracts were radiation-induced; authorities on whom expert purported to rely agree that diagnosis cannot be made on observation alone and expert failed to use proper methodology for diagnosis).

<sup>178.</sup> See Edward J. Imwinkelried, The Admissibility of Expert Testimony in Christophersen v. Allied-Signal Corp.: The Neglected Issue of the Validity of Nonscientific Reasoning by Scientific Witnesses, 70 Denv. U. L. Rev. 473 (1993).

tent feasible is offering a hunch rather than an explanatory theory, and should not be permitted to offer an opinion.<sup>179</sup> A complicating problem in *Christophersen* was the weakness of the plaintiff's evidence on exposure, which came from a fellow worker's affidavit that provided no information about the composition of the fumes to which Christophersen was exposed or the physical facilities of the plant.<sup>180</sup>

2. Social science evidence . A fundamental question that the courts will have to address after *Daubert* is whether the Supreme Court's opinion applies to the social sciences or is limited to the physical and "hard" sciences. Are experts in such fields as psychology, <sup>181</sup> economics, <sup>182</sup> sociology, and political science testifying about "scientific knowledge" so that the *Daubert* standard of admissibility applies? If *Daubert* applies, how does a court determine whether an expert opinion grounded in the social sciences rests on a valid methodology? Although there has been a marked increase in proffers of social science evidence, especially psychological evidence, the federal courts rarely explored these issues in much depth prior to *Daubert*, in part because cases involving rape and child abuse in which syndrome evidence is prevalent were not usually brought in federal courts. <sup>183</sup>

179. See, e.g., Chikovsky v. Ortho Pharmaceutical Corp., 832 F. Supp. 341, 346 (S.D. Fla. 1993) (plaintiff claimed that defendant's product, Retin-A, which pregnant mother had used topically, had caused child's birth defects; Retin-A is a Vitamin A derivative, and there are no data linking Retin-A to birth defects; plaintiff's expert, an obstetrician-gynecologist with no specialized training in embryology or teratology (see *supra* § II.D.1.a), testified that high doses of other Vitamin A derivatives have been implicated in birth defects; how-ever, he did not know how much Vitamin A could have been absorbed through the skin, and "most significant" according to the trial court which excluded his testimony as not meeting *Daubert*, he had performed no comparisons between the dose of Vitamin A in the studies showing fetal harm and that found in Retin-A).

180. Cf. Peteet v. Dow Chem. Co., 868 F.2d 1428 (5th Cir.), cert. denied, 493 U.S. 935 (1989), in which plaintiffs claimed that decedent's death, probably of Hodgkin's disease, was due to exposure to defendant's herbicide. The plaintiffs' expert, a certified toxicologist, testified that numerous studies linked the herbicide and cancer. The opinion does not discuss to what extent, if any, these studies found a link to Hodgkin's disease. The expert relied on a "one-hit" theory of cancer that "suggests that one molecule of carcinogen, in the right place and at the right time, can cause cancer." *Id.* at 1433. The court quoted from an earlier case in which it had stated that "[w]hat is necessary is that the expert arrived at his causation opinion by relying upon *methods* that other experts in his field would reasonably rely upon in forming their own, possibly different opinions, about what caused the patient's disease." *Id.* at 1433 (quoting Osburn v. Anchor Lab., Inc., 825 F.2d 908, 915 (5th Cir. 1987)). Testimony had established that the decedent had worked on a weed control project and that the herbicide had frequently gotten on the workers' clothing and skin. *Id.* at 1430.

181. United States v. Amador-Galvan, 9 F.3d 1414, 1418 (9th Cir. 1993) (Daubert governs expert testimony about unreliability of eyewitness evidence).

182. See, e.g., Petruzzi's IGA Supermarkets, Inc. v. Darling-Delaware Co., 998 F.2d 1224, 1241 (3d Cir.) (court cited *Daubert* with regard to economists' testimony), *cert. denied*, 114 S. Ct. 554 (1993). See also Joy v. Bell Helicopter Textron, Inc., 999 F.2d 549, 569 (D.C. Cir. 1993); Scales v. George Washington Univ., No. 89-0796, 1993 U.S. Dist. LEXIS 10692, at \*25 (D.D.C. July 27, 1993).

183. Prior to the middle 1970s, courts almost never applied Frye to the social sciences. See David McCord, Syndromes, Profiles and Other Mental Exotica: A New Approach to the Admissibility of Nontraditional Psychological Evidence in Criminal Cases, 66 Or. L. Rev. 19, 77 (1987). Thereafter, some courts began extending Frye to social science evidence. See, e.g., United States v. Shorter, 809 F.2d 54, 59–61 (D.C. Cir.) (psychological testimony), cert. denied, 484 U.S. 817 (1987). United States v. Carmel, 801 F.2d 997, 998–99 (7th Cir. 1986); Spencer v. General Elec. Co., 688 F. Supp. 1072, 1076 (E.D. Va. 1988); Barrel of Fun, Inc. v. State Farm Fire & Casualty Co., 739 F.2d 1028, 1031–32 (5th Cir. 1984) (psychological stress evaluation). Finally, some courts did not use Frye as a test for "hard" or "soft" scientific evidence. The Third Circuit's approach, rejecting Frye and focusing on reliability and fit, upon which the Supreme Court relied in part in

Are there reasons why the "hard" and "soft" sciences should perhaps be handled differently? Two schools of thought about this issue can conveniently be compared by looking at the views of Professors David McCord<sup>184</sup> and David L. Faigman.<sup>185</sup> Even though McCord and Faigman wrote before *Daubert* and were primarily concerned with psychological syndrome evidence, their differing attitudes shed light on some underlying factors and assumptions. The crux of their disagreement centers on when evidence should be kept from a jury. Mc-Cord is much more willing to take the risk, which he thinks is low, that jurors will be swayed by worthless social science evidence if there is a chance that the evidence might be helpful. Faigman puts much greater stock in shielding against juror misuse of invalid evidence.<sup>186</sup>

According to McCord, the fundamental difference between hard and soft scientific evidence (at least of a psychological nature) makes a stringent test wholly inappropriate for the latter type of evidence. The justification for a stricter admissibility test—keeping from the jury evidence which "juries are not in a position to fairly and intelligently weigh" and which "appears to be unassailably 'scientific'"—does not apply to psychological evidence.<sup>187</sup>

The essence of such "soft" psychological evidence is not locked up in some mysterious nonhuman device or process, and the expert on the stand can be grilled regarding the foibles of psychological research. Further, and perhaps more important, most jurors do not conceive of psychological research as very, if at all, "scientific." It is not likely to elicit unquestioning juror acceptance. In short, the jury most likely has the ability to fairly and intelligently weigh the strengths and weaknesses of psychological evidence without being overwhelmed or overawed by it.<sup>188</sup>

The Supreme Court's opinion in *Barefoot v. Estelle*<sup>189</sup> might be read as consistent with McCord's position. In *Barefoot*, a death row inmate argued that the government should not have been permitted to call an expert psychiatrist at the guilt phase of his trial in order to predict the defendant's future dangerousness. The defendant claimed that psychiatrists are not capable of predicting future behavior, especially without interviewing the person.<sup>190</sup> The Court rejected this view, stating that if it is constitutionally permissible to base a death sentence

188. Id. at 85-86 (footnotes omitted).

189. 463 U.S. 880 (1983).

190. Id. at 896. An amicus brief filed by the American Psychiatric Association quoted by the dissent had concluded that "*two out of three* predictions of long-term future violence made by psychiatrists are wrong." *Id.* at 920 (citing Brief for American Psychiatric Association as Amicus Curiae at 9, 13).

Daubert (113 S. Ct. at 2797 n.12), was developed in United States v. Downing, 753 F.2d 1224 (3d Cir. 1985), a case dealing with the admissibility of a psychologist's testimony about eyewitness testimony.

<sup>184.</sup> McCord, supra note 183.

<sup>185.</sup> David L. Faigman, To Have and Have Not: Assessing the Value of Social Science to the Law as Science and Policy, 38 Emory L.J. 1005 (1989). See also American College of Trial Lawyers, supra note 5.

<sup>186.</sup> To perhaps a lesser extent, the same two camps exist with "hard" scientific evidence as well. As is discussed elsewhere, judges obviously differ in how willing they are to exclude evidence from a jury's consideration. One of the clearest reflections of this difference is the extent to which courts scrutinize the validity of the social science research underlying an expert's opinion.

<sup>187.</sup> McCord, supra note 183, at 85.

on the likelihood of future behavior, then an expert may give an opinion on that behavior.<sup>191</sup> The dissent agreed that future behavior is a permissible consideration, but objected strenuously to the expert testimony's lack of reliability.<sup>192</sup>

McCord concedes that "[e]ven with respect to 'soft' psychological evidence, some inquiry into reliability is still appropriate since the jury may well not be in the best position to completely understand the probative value of the evidence." <sup>193</sup> He suggests, however, that flexible, less stringent standards of reliability are appropriate, and that no one factor should be dispositive. "Even somewhat unreliable evidence may be admitted in certain circumstances, particularly where it is offered on a nondispositive issue in the case or offered by the defendant." <sup>194</sup> Ultimately, McCord endorses a four-factor balancing test that focuses on necessity, reliability, understandability, and importance. <sup>195</sup>

A very different view is expressed by Faigman. He endorses a test that sounds remarkably like the Supreme Court's language in *Daubert*. According to him, social science evidence should not be presented to jurors unless it rests on a scientific theory that has been empirically tested: "[f]alsifiability or testability represents the line of demarcation between science and pseudo-science, and the strength of particular scientific statements depends on the extent to which they have been tested appropriately." <sup>196</sup> He would insist on threshold screening by the judge of the methodology on which the social science evidence rests. <sup>197</sup> To Faigman, tying threshold admissibility determinations to "scientific" validity is as essential for "soft" evidence as it is for "hard" evidence. A restrictive test prevents scientific observation" and guards against "experts . . . [who] nullify legal rules themselves, by confusing jurors, or . . . call upon the jury to nullify a legal rule on the basis of policy considerations that the rule does not reflect."

191. Id. at 896-97.

192. Id. at 938 (Blackmun, J., dissenting). See Giannelli, supra note 137, at 113–17 (discussion of Barefoot to illustrate "junk science" in criminal cases).

193. McČord, supra note 183, at 86.

196. Faigman, *supra* note 185, at 1015. For a discussion of the empirical approach in the social sciences and references to more detailed discussions, see John Monahan & Laurens Walker, Social Science In Law: Cases and Materials 33–82 (2d ed. 1990). In a series of articles, Professors Walker and Monahan have also proposed an alternative concept for providing jurors with information stemming from social science research that treats such research more like legal precedents than factual findings. See John Monahan & Laurens Walker, *Empirical Questions Without Empirical Answers*, 1991 Wis. L. Rev. 569; Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 Cal. L. Rev. 877 (1988); Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 Va. L. Rev. 559 (1987); John Monahan & Laurens Walker, Social Authority: Obtaining, Evaluating and Establishing Social Science in Law, 134 U. Pa. L. Rev. 477 (1986).

197. Faigman, supra note 185, at 1090.

198. Id. at 1084, 1088.

Reference Manual on Scientific Evidence

<sup>194.</sup> Id. at 88.

<sup>195.</sup> Id. at 94.

Daubert, some courts reached the result Faigman advocates by relying on Rule 403 rather than Rule 702.<sup>199</sup>

Faigman would require courts to look for the hallmarks of scientific methodology before he would allow any expert to render an opinion based on the social sciences. He would require of the proffered expert "a cogent explanation of the methods and analyses that produced the scientific opinion."<sup>200</sup> Expert testi mony about the accuracy of eyewitness identifications meets a minimum threshold standard because it is based on a research design and statistical studies. <sup>201</sup>

3. *Psychological syndrome evidence*. The difference between the two approaches is apparent if one considers how courts have treated the admissibility of psychological syndrome, or profile, evidence. Expert testimony, usually by a psychologist, has been proffered in the federal courts concerning rape trauma syndrome (RTS).<sup>202</sup> Testimony about RTS or post-traumatic stress disorder (PTSD) is most often offered in a rape prosecution to counter the defendant's consent defense or to explain the victim's behavior. The absence of RTS has also been offered by the defense to show that the complainant was not raped, and the presence of RTS has been proffered in civil cases on a number of theories. The courts are divided on the admissibility of RTS expert testimony; some exclude all RTS evidence, <sup>203</sup> whereas others admit RTS evidence, although they differ on how the expert testimony may be used.<sup>204</sup>

A "*Daubert*" approach to social science evidence that insists on empirical validation might exclude expert testimony that the existence of certain symptoms proves that the alleged victim has been raped, but might admit testimony offered

203. See, e.g., People v. Bledsoe, 681 P.2d 291, 300–01 (Cal. 1984) (error, although not prejudicial, to admit testimony of rape counselor as expert testimony that victim was suffering from RTS; rape counselors nei-ther question the credibility of their clients nor probe inconsistencies, and therefore use of these opinions as expert testimony is problematic; scientific literature does not purport to claim that RTS is a scientifically reliable means of proving that a rape occurred); State v. Black, 745 P.2d 12, 15–19 (Wash. 1987); Commonwealth v. Gallagher, 547 A.2d 355, 358–59 (Pa. 1988).

204. See Infra notes 205–07. For a recent survey that considers the entire body of psychological research on rape, see Patricia A. Frazier & Eugene Borgida, Rape Trauma Syndrome: A Review of Case Law and Psychological Research, 16 Law & Hum. Behav. 293 (1992). Faigman has faulted much of the research in this area. See David L. Faigman, Note, The Battered Woman Syndrome and Self-Defense: A Legal and Empirical Dissent, 72 Va. L. Rev. 619 (1986).

<sup>199.</sup> See infra § V.

<sup>200.</sup> Faigman, supra note 185, at 1081.

<sup>201.</sup> Id. at 1089. Faigman concedes that some validity problems will remain for the jury because most eyewitness identification studies involve college students and are conducted under contrived circumstances. These are issues that he thinks jurors can comprehend and that can be explored adequately on cross-examination of the expert. Id.

<sup>202.</sup> See, e.g., Spencer v. General Elec. Co., 688 F. Supp. 1072, 1075–77 (E.D. Va. 1988) (court held that such evidence is not a scientifically reliable means of proving that a rape occurred and therefore does not satisfy *Frye* test; court noted that expert's methodology "bore little, if any, resemblance to traditional scientific or medical methodologies" and that probative value of such evidence is outweighed by its unfair prejudicial effect, citing Rule 403). See also United States v. Arcoren, 929 F.2d 1235, 1238–42 (8th Cir. 1991) (court upheld admission of evidence of battered woman syndrome to explain why witness recanted her testimony), *cert. denied*, 112 S. Ct. 312 (1991); United States v. Azure, 801 F.2d 336, 339–41 (8th Cir. 1986) (allowing expert testimony concerning post-traumatic stress reactions to child abuse but not permitting expert to give opinion as to truth of victim's story).

to rebut the defendant's defense that the complainant's behavior was inconsistent with the claim of rape.<sup>205</sup> A less rigid but still "scientific" view would permit experts who have interviewed or treated the victim to testify about the typical behavior of rape victims, and allow experts to state that the victim's behavior is consistent with that of rape victims.<sup>206</sup> Cross-examination of the experts could develop the limits of present scientific knowledge. Courts least inclined to take a rigid scientific approach to social science evidence, who believe that jurors are capable of evaluating soft evidence for what it is worth without being unduly swayed by the expert, allow RTS evidence as part of the prosecution's case in chief on the issue of whether a rape occurred in cases in which the defendant is claiming consent.<sup>207</sup>

#### b. Rejecting expert testimony because of skewed methodology

Courts may also be confronted with experts who purport to rely on a standard methodology. In the instant case, however, the opponent claims that this methodology is somehow skewed—nonconventional assumptions or irregular techniques were used, or errors have been found. Is this a Rule 702 problem under *Daubert*, a Rule 703 problem, a Rule 403 problem, or a problem of weight for the jury?

205. See discussion in Spencer, 688 F. Supp. at 1076–77 ("[T]he relevant issue is not whether rape victims may display certain symptoms, but 'whether the presence of various symptoms, denominated together as "rape trauma syndrome" [or PTSD], is a scientifically reliable method admissible in evidence and probative of the issue of whether an alleged victim was raped.'" (quoting State v. Black, 745 P.2d 12, 17 (Wash. 1987)); court excluded PTSD testimony by an expert to prove rape in tort case because expert's methodology "bore little, if any, resemblance to traditional scientific or medical methodologies"; but court allowed testimony to establish damages). See also People v. Taylor, 552 N.E.2d 131, 138 (N.Y. 1990) (allowing testimony concerning rape victim's lack of emotion after attack because RTS evidence is "relevant to dispel misconceptions that jurors might possess regarding the ordinary responses of rape victims in the first hours after their attack," but excluding RTS evidence to prove that rape occurred in a companion case; court stressed that "evidence of rape trauma syndrome is inadmissible when it inescapably bears solely on proving that a rape occurred"). When the evidence is being permitted to counter the defendant's defense, to prevent jurors from drawing the prohibited assumption—that the expert has concluded that the victim was raped and that the expert has a basis for this opinion—it has been suggested that the testimony about the rape victim's behavior be given by an expert who has not examined the victim. Deborah A. Dwyer, Note, *Expert Testimony on Rape Trauma Syndrome: An Argument for Limited Admissibility*, 63 Wash. L. Rev. 1063, 1084 (1988).

206. People v. Fasey, 829 P.2d 1314, 1315–17 (Colo. 1992) (expert in state's case in chief first described the symptoms of PTSD and stated that a sexual assault could be a traumatic experience that would cause the symptoms; he then described the symptoms exhibited by the victim and concluded that the victim did suffer from PTSD; he did not state that the syndrome was necessarily caused by a sexual assault; court found no error).

207. See, e.g., State v. Allewalt, 517 A.2d 741, 748 (Md. 1986) (defendant claimed consent in rape prosecution; psychiatrist permitted to state that victim's PTSD was caused by rape:

He [the expert] did not purport to have invented a scientific test for determining consent to sexual intercourse had months earlier. He did claim that he could use his special knowledge and the interviewing techniques of his profession to diagnose whether Mrs. Lemon, at the time of his examination of her, suffered from a medically recognized anxiety disorder. He did not claim that psychiatry could demonstrate conclusively that the cause of the PTSD was rape. He did claim the special knowledge and experience to be able to identify the cause of the PTSD by utilizing the history furnished by the patient . . . ).

The Third Circuit has dealt with this question in a number of contexts that illustrate the issues that may arise. Recently, in *In re Paoli Railroad Yard PCB Litigation (Paoli II)*, the court considered at length to what extent plaintiffs' experts, specialists in internal medicine, had to employ the technique of differential diagnosis in order for the court to find that their opinions were based on a reliable methodology that satisfied the standards of *Daubert*.<sup>208</sup> The court

agree[d] with the defendants that performance of physical examinations, taking of medical histories, and employment of reliable laboratory tests all provide significant evidence of a reliable differential diagnosis, and that their absence makes it much less likely that a differential diagnosis is reliable.<sup>209</sup>

#### But the court also

agree[d] with the plaintiffs that a doctor does not always have to employ all of these techniques in order for the doctor's differential diagnosis to be reliable.... [S]ometimes differential diagnosis can be reliable with less than full information, and to the extent that the district court concluded otherwise, we hold that it abused its discretion.<sup>210</sup>

The court then concluded that the district court could not exclude the opinions of the plaintiffs' physicians unless

(1) [they] engaged in very few standard diagnostic techniques by which doctors normally rule out alternative causes *and* the doctor offered no good explanation as to why his or her conclusion remained reliable, or (2) the defendants pointed to some likely cause of the plaintiff's illness other than the defendants' actions and [the plaintiff's physician] offered no reasonable explanation as to why he or she still believed that the defendants' actions were a substantial factor in bringing about that illness.<sup>211</sup>

In a previous review of the same case, *In re Paoli Railroad Yard PCB Litigation (Paoli I)*, the court discussed the admissibility of an expert's opinion based on a meta-analysis.<sup>212</sup> Meta-analysis is a statistical method for combining the

211. *Id.* at \*104–05 (emphasis in original). The court concluded that a court could exclude a physician's conclusions that were based solely on the "plaintiff's self-report of illness in preparation for litigation." *Id.* at \*111. One reliable source for the opinion, such as a physical examination or medical records, will ordinarily suffice. *Id.* The court ultimately found that the testimony of one of the plaintiff's physician experts was properly excluded and that the testimony of the other physician should have been admitted to a limited extent.

212. 916 F.2d 829, 856–59 (3d Cir. 1990), cert. denied, 111 S. Ct. 1584 (1991). Because the defendants were challenging the meta-analysis technique itself, the court invoked the standard it had announced in United States v. Downing, 753 F.2d 1224 (3d Cir. 1985), for analyzing expert testimony based on novel scientific techniques. *Paoli I*, 916 F.2d at 856. The majority opinion in *Daubert* acknowledged that its discussion of the reliability of scientific evidence was "draw[n] in part" from the Third Circuit's opinion in *Downing*, 753 F.2d at 1238–39. See Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2791 n.12 (1993). Judge Becker was the author of both *Downing* and In re Paoli. The admissibility of meta-analysis was not addressed in In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722 (3d Cir. Aug. 31, 1994).

<sup>208. 1994</sup> U.S. App. LEXIS 23722, at \*87-140 (3d Cir. Aug. 31, 1994).

<sup>209.</sup> Id. at \*100-01 (footnote omitted).

<sup>210.</sup> *Id.* at \*102–03 (footnote omitted). The court gives as an example a patient who comes to a physician with a medical record "demonstrating illness A known to be strongly associated with chemical X and evidence of exposure to that chemical." Even if the patient has been exposed to other chemicals that sometimes, though less frequently, cause the same illness, "[w]ith a basic understanding of probabilities, a physician might very well be able to reliably conclude that a person exposed to chemical X was more likely than not to have contracted illness A as a result of that exposure than as a result of any other cause." *Id.* at \*104 n.30.

results from separate published studies on a common scientific issue to see if all available data looked at collectively produce a result different from that obtained when small studies are looked at individually. Combining studies that measure different parameters may be controversial, although such a technique is often used by scientists.<sup>213</sup> Indeed, in *In re Paoli (Paoli I)*, the court noted that defendants' own experts did not question the reliability of all meta-analyses; they merely questioned the way in which plaintiffs' experts had applied meta-analysis in the instant case.<sup>214</sup> The court suggested, however, that a district court could exclude a particular meta-analysis pursuant to Rule 702 if it was "sufficiently unreliable."<sup>215</sup> The court declined "to define the exact level at which a district court can exclude a technique as sufficiently unreliable. Reliability indicia vary so much from case to case that any attempt to define such a level would most likely be pointless."<sup>216</sup>

A Bendectin case in the Third Circuit, *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, illustrates a situation in which the district court concluded after an in limine evidentiary hearing that the expert's methodology was so unreliable as to warrant exclusion.<sup>217</sup> The district court relied primarily on Rule 702, but also concluded that the testimony was excludable pursuant to Rule 703.<sup>218</sup> The district court, while recognizing that the defendant was attacking the plaintiff's expert's "methodology and not the underlying data he relied upon in making his calculations," stated that "[t]his is where Rules 702 and 703 intersect" and acknowledged that the expert had "used data upon which no epidemiologist would rely." <sup>219</sup>

In *DeLuca*, the plaintiff's principal expert acknowledged that published studies showed no statistically significant association between Bendectin and limb reduction defect, but claimed that his reanalysis of the studies established such

214. Paoli I, 916 F.2d at 857.

215. Id. at 858. "A reliable methodology . . . [that is] so altered as to skew the methodology itself" is properly subject to a Rule 702-based exclusion. Id.

216. *Id.* The court remanded for a fuller record and specific findings on reliability issues.

217. 911 F.2d 941, 944 (3d Cir. 1990) (reversing and remanding grant of summary judgment), on remand, 791 F. Supp. 1042 (D.N.J. 1992) (granting summary judgment), aff'd without op., 6 F.3d 778 (3d Cir. 1993), cert. denied, 114 S. Ct. 691 (1994).

218. 791 F. Supp. at 1058-59.

219. *Id.* at 1048 n.10, 1059. The district court quoted the court of appeals: "'If a study's method of data collection is faulty, it may be that no expert would rely upon the data generated as a basis for drawing any inference about the studied subject." *Id.* at 1059 (quoting *DeLuca*, 911 F.2d at 955 n.14). The district court also noted that the Third Circuit in *Paoli* made reference to its decision in *DeLuca*: "*DeLuca* announces an important rule by making clear that when it is a scientist's methodology that is being attacked, in contrast to the data relied on, the court must analyze the reliability of that methodology under *Downing* (and Rule 702)." *Id.* at 1047–48 n.10 (quoting *In* re Paoli R.R. Yard PCB Litig. (Paoli I), 916 F.2d 829, 856 (3d Cir. 1990), *cert. de-nied*, 111 S. Ct. 1584 (1991)).

<sup>213.</sup> See generally Frederic Marc Wolf, Meta-Analysis: Quantitative Methods for Research Synthesis (1986) (basic text on meta-analysis); Larry V. Hedges & Ingram Olkin, Statistical Methods for Meta-Analysis (1985) (discussion of strengths and weaknesses of statistical procedures); Handbook of Research Synthesis (Harris Cooper & Larry V. Hedges eds., 1994) (strengths and weaknesses of various approaches to meta-analysis); Thomas R. Einarson et al., A Method for Meta-Analysis of Epidemiological Studies, 22 Drug Intelligence & Clin. Pharmacy 813 (1988) (meta-analysis of Bendectin studies); Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers 254–55 (1990).

an association. At the in limine hearing, however, none of the epidemiologists testifying was able to replicate the expert's numbers, and his methodology was termed a "mystery."<sup>220</sup>

In a case such as *DeLuca*, the data collection problems that are discussed in section IV corroborate the weakness of the expert's methodology; separating the methodology and data issues is somewhat artificial, since both relate to the reliability of the expert's opinion. Whether courts will continue to make this distinction after *Daubert* is not clear;<sup>221</sup> the Supreme Court mentioned Rule 703 as a rule to consider but did not discuss its scope. Nevertheless, some of the problems with the expert's testimony in *DeLuca* are considered in section IV because, when screening expert testimony, many courts have considered the data on which the expert's opinion is based independently of methodological concerns.

The nondefinitive checklist the majority offers in *Daubert* of factors bearing on scientific validity all point to the flawed methodology of the expert's testimony in *DeLuca*. A more difficult case arises when no obvious errors emerge but the parties disagree about the research design of particular epidemiological studies. What must the judge do if one side argues, for instance, that the control group was improperly constituted, or that the classification scheme for identifying exposed individuals was faulty, or that confounding factors were not taken into account?<sup>222</sup> No epidemiological study can be perfect; will less blatant flaws

220. 791 F. Supp. at 1047. In addition to transposing numbers on his charts, relying on data that authors had corrected, and including anecdotal data which he conceded could not be used to show causation, the plaintiff's expert apparently calculated results for studies that did not specify odds ratios or relative risk and that did not contain data from which such calculations could easily be made. With respect to another study, he specified a relative risk far in excess of the confidence intervals specified in the study, although a relative risk must fall between these limits. For yet another study, he recalculated the odds, although he claimed to use the authors' calculations when possible. The expert also failed to consider the strength of the various studies in conjunction with their results, and he ignored two recent studies, one quite large, that found no causal relationship between Bendectin and birth defects. Although he testified that he included data from all studies he knew of, he did not include any data from post-1986 studies. Articles on two of these studies had been published in a peer-reviewed scientific journal. In addition, the court noted that he had failed to identify any literature or other expert who endorsed his technique. *Id.* at 1047–57.

221. In In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*52-53 (3d Cir. Aug. 31, 1994), the court stated:

we no longer think that the distinction between a methodology and its application is viable. To begin with, it is extremely elusive to attempt to ascertain which of an expert's steps constitute parts of a "basic" methodology and which constitute changes from that methodology. If a laboratory consistently fails to use certain quality controls so that its results are rendered unreliable, attempting to ascertain whether the lack of quality controls constitutes a failure of methodology or a failure of application of methodology may be an exercise in metaphysics. Moreover, any misapplication of a methodology that is significant enough to render it unreliable is likely to also be significant enough to skew the methodology.

222. See Linda A. Bailey et al., Reference Guide on Epidemiology § II, in this manual, for a discussion of factors to be considered in evaluating an epidemiology research design.

than those in *DeLuca* warrant exclusion of an expert's opinion?<sup>223</sup> Few cases have, as yet, considered methodological challenges.<sup>224</sup>

#### 3. Issues with regard to statistical estimates

The *Daubert* opinion does not discuss the statistical issues that frequently emerge in connection with scientific evidence.<sup>225</sup> The parties may, for instance, agree that an appropriate way to prove the controverted issue—does Substance A cause Disease B—is through an epidemiological study. They may even concur in finding no problems with the methodological design of the study.<sup>226</sup> But they may disagree strongly about the statistical significance of the study and the consequences with regard to admissibility or sufficiency. Or they may differ on what, if anything, the jury must be told about background statistical information. Finally, issues arise about the extent to which results of particular studies should be discounted by error rates. Each of these issues is discussed below.

223. See In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722 (3d Cir. Aug. 31, 1994). The majority concluded that

[t]he evidentiary requirement of reliability is lower than the merits standard of correctness.... A judge frequently should find an expert's methodology helpful even when the judge thinks that the expert's technique has flaws sufficient to render the conclusions in accurate. He or she will often still believe that hearing the expert's testimony and assessing its flaws was an important part of assessing what conclusion was correct and may certainly still believe that a jury attempting to reach an accurate result should consider the evidence. *Id.* at \*49–50.

In concurring, Judge Roth specifically declined to join this portion of the opinion, stating: "I do not believe that it is 'helpful' for the jury to receive information which the trial judge concludes is not accurate. In my opinion, the 'gatekeeper' function of the trial judge established by the Supreme Court in *Daubert* would not be fulfilled by permitting inaccurate information to go to the jury even though the trial judge may have determined that the methodology used to produce such results is reliable." *Id.* at \*245.

224. See discussion in Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1553 (D. Colo. 1990), *affd*, 972 F.2d 304 (10th Cir. 1992) (court appointed a geological expert to advise court as to whether it was methodologically proper to extrapolate all conclusions about exposure from a single water sample; court took into account expert's report and granted summary judgment). See also the various opinions rendered by the district court after remand in *In re* Paoli R.R. Yard PCB Litig., 790 F. Supp. 94 (E.D. Pa. 1992), *aff'd without op.*, 980 F.2d 724 (3d Cir. 1992), *on remand*, 1992 U.S. Dist. LEXIS 16287, 18427, 18428, 18429, 18430, 18431, 18432, 18433, 18434, 18435, 18436, 18437 (E.D. Pa. Oct. 21, 1992); 811 F. Supp. 1071 (E.D. Pa. 1992), *aff'd in part, rev'd in part, In re* Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722 (3d Cir. Aug. 31, 1994) (most recently the court of appeals affirmed exclusion of much of proffered expert testimony; it reversed with regard to some of the excluded evidence on exposure and harmful effects of PCBs so that the summary judgments entered by the district court on the thirty-eight plaintiffs' personal injury claims were reversed with regard to two of the plaintiffs).

225. Indeed, the parties at times seem unaware that statistical issues exist and that the expert who has specialized knowledge about the underlying physics, chemistry, or biology may lack adequate statistical training to explain the probabilities associated with his or her conclusion. See, e.g., United States v. Stifel, 433 F.2d 431, 435–41 (6th Cir. 1970) (participants seemingly failed to appreciate that neutron-activation-analysis testimony that tape samples came from same batch was misleading in the absence of testimony about the frequency with which such matches could be expected), cert. denied, 401 U.S. 994 (1971), conviction vacated, 594 F. Supp. 1525, 1537 (N.D. Ohio 1984) (conviction vacated primarily because of *Brady* violation, but evidentiary hearing also demonstrated that sample tape from bomb packing did not differ from other samples of tape from different batches).

226. See Linda A. Bailey et al., Reference Guide on Epidemiology § II, in this manual, for a discussion of methodological questions.

a. Statistical significance: An issue for scientists or for the court?

A threshold issue with regard to expert proof based on many different kinds of studies is whether courts ought to use the level of statistical significance that is conventionally used in the particular discipline to which the expert belongs and, if so, for what purpose. Although this problem has received some judicial attention, it has not been conclusively resolved. The problem has been discussed primarily in the context of epidemiological studies to prove causation in toxic tort cases. <sup>227</sup> Scientists customarily employ a 5% significance level in testing a hypothesis. In the context of an epidemiological study that reports a particular relative risk, this means that there is at most one chance in twenty of seeing such a big relative risk if the true relative risk is 1.0.<sup>228</sup>

A not-proven verdict in court, however, has very different consequences than a not-proven verdict in the context of scientific research.<sup>229</sup> A failure to satisfy the 5% significance level means only that more research is in order—it is not a statement of an established "truth."<sup>230</sup> Virtually no mechanisms exist for deferring judicial decisions until more proof is available or for correcting decisions erroneously made. The plaintiff or the defendant generally wins or loses at the moment the case is ripe for decision. The plaintiff, who has the burden of persuasion, bears the risk, and the loss, if the case is not proven.

Some commentators have suggested that the use of "confidence intervals" provides more meaningful information than statistical significance because a range of possible values is presented that is consistent with the observed data.<sup>231</sup> The use of confidence intervals does not eliminate the need to designate the

229. The Supreme Court explicitly recognized this difference between science and the law in Daubert v. Merrell Dow Pharmaceuticals, Inc.: "Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly." 113 S. Ct. 2786, 2798 (1993).

230. *Id. Cf.* Chemical Carcinogens, 50 Fed. Reg. 10371, 10377 (Office of Science & Technology Policy 1985) (final document) ("A high-quality negative epidemiological study, while useful, cannot prove the absence of an association between chemical exposure and human cancer."). *See also* Brief Amicus Curiae of Professor Kenneth Rothman et al. in Support of Petitioners, Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993) (No. 92-102), *reprinted in* 1 Shepard's Expert & Sci. Evidence Q. 75, 80 (1993) ("The result of using significance testing as a criterion for decision making is that the focus is changed from the information presented by the observations themselves to conjecture about the role chance *could* have played in bringing about those observations.") (emphasis in original).

231. See discussion in David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A.3, and Linda A. Bailey et al., Reference Guide on Epidemiology § III.C.1, in this manual.

<sup>227.</sup> See, e.g., Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 312 (5th Cir. 1989) (in granting judgment n.o.v., court referred to plaintiff's failure to provide a study with statistical significance that concludes that Bendectin is a human teratogen), *modified*, 884 F.2d 166 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990); DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 948–49 (3d Cir. 1990) (discusses issues arising from use of statistical significance concepts).

<sup>228.</sup> However, the significance level cannot be interpreted as the probability that the true relative risk differs from 1.0. See David H. Kaye & David A. Freedman, Reference Guide on Statistics §§ IV.B.1–2, in this manual. Social scientists as well adopt the .05 level of statistical significance. See discussion *infra* § III.C.2.a.2. See *also* Segar v. Smith, 738 F.2d 1249, 1282–83 (D.C. Cir. 1984) (Title VII action), *cert. denied*, 471 U.S. 1115 (1985).

confidence levels.<sup>232</sup> A court would still ultimately have to decide at what level it finds the evidence sufficiently probative.

In DeLuca v. Merrell Dow Pharmaceuticals, Inc., the Third Circuit summed up its extensive discussion of statistical significance by observing that "[t]he root issue . . . is what risk of what type of error the judicial system is willing to tolerate."<sup>233</sup> The court did not reach a conclusion because it found the record inadequate to resolve the issue. It expressed the hope that on remand, legal scholars and epidemiologists would assist the court with this problem, perhaps through amicus briefs.<sup>234</sup> On remand, as discussed in section III.C.2.b, the court excluded the plaintiffs' expert's reanalysis on the ground of unacceptable methodology and unreliability as well as on Rule 403 grounds. It never considered the level of statistical significance a study would have to satisfy in order to be admissible.<sup>235</sup>

Courts may consider that although the plaintiff ordinarily bears the burden of producing evidence, the plaintiff, particularly an individual plaintiff, often has no control over the amount of data that are available and no means of compelling anyone, including the defendant, to undertake additional research. Even if the evidence is admissible, however, whether it is sufficient is a separate issue. How these issues should be resolved may also rest more on substantive policy concerns than on the law of evidence. <sup>236</sup>

The use of confidence intervals, however, does not in any way eliminate the necessity for numerical standards. The most common and widely accepted standard is the use of a 95% confidence interval, which is precisely analogous to a *p* value of .05, which denotes 'statistical significance.' The critical issue is what level of  $\alpha$  is to be acceptable.... The choice of  $\alpha$  must be made whether statistical significance or confidence intervals are used.

233. 911 F.2d 941, 955 (3d Cir. 1990). Many opinions, even those that are sophisticated about statistical concepts, fail to consider this basic issue. See, e.g., Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 312 (5th Cir.) (makes statements about significance of confidence intervals without specifying the significance level used in constructing the confidence interval), *modified per curiam*, 884 F.2d 166 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990).

234. DeLuca, 911 F.2d at 956 & n.17.

235. DeLuca v. Merrell Dow Pharmaceuticals, Inc., 791 F. Supp. 1042 (D.N.J. 1992) (granting summary judgment on remand), aff'd without op., 6 F.3d 778 (3d Cir. 1993), cert. denied, 114 S. Ct. 691 (1994). The court did state in its findings of fact that "[i]n the analysis of Bendectin limb defect studies, the choice of a confidence interval of 90% or 95% does not change the result if that confidence interval contains the number 1.0." *Id.* at 1052. It is not clear whether this finding means that relative risk was calculated at a 90% confidence level in addition to a 95% level. The Third Circuit did not suggest that changing to a 90% level would resolve the issue of what type of error to tolerate.

236. For a discussion arguing that substantive state law issues must accordingly be taken into account, see Michael H. Gottesman, Should Federal Evidence Rules Trump State Tort Policy? The Federalism Values Daubert Ignored, 15 Cardozo L. Rev. 1837 (1994). See also In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*72–77 (3d Cir. Aug. 31, 1994) (holding Pennsylvania's requirement that experts testify that defendant's actions caused plaintiff's illness with a reasonable degree of medical certainty constitutes part of the plaintiff's burden of proof and, as a substantive rule, is not in conflict with the Federal Rules of Evidence and governs in a diversity action in federal court).

<sup>232.</sup> Brief Amicus Curiae of Professor Alvan R. Feinstein in Support of Respondent, Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993) (No. 92-102), *reprinted in* 1 Shepard's Expert & Sci. Evidence Q. 91, 104 (1993):

#### b. Correlation of statistical results with the burden of proof

A perplexing problem for the courts has been the interrelationship between an opinion couched in probabilistic terms and the applicable burden of proof. If, for instance, an expert testifies that epidemiological studies show that exposure to a defendant's product results in an increased risk of a particular form of cancer, how much of an increase in risk has to be demonstrated for plaintiffs to satisfy their burden of proof?<sup>237</sup> Epidemiological studies typically assign a relative risk ratio to a cohort study, or an odds ratio to a case-control study.<sup>238</sup> How much higher than 1.0 (which is the equivalent of no difference between the exposed and unexposed groups) must the relative risk or odds ratio be for the plaintiff to make out a prima facie case based on epidemiological proof?

As an abstract statistical proposition, a ratio under 2.0 does not comport with a preponderance-of-the-evidence standard. This conclusion was the basis for the district court's grant of summary judgment in *In re Joint Eastern & Southern District Asbestos Litigation*,<sup>239</sup> on the ground that plaintiff could not establish that her husband's colon cancer was caused by exposure to asbestos:

Only when the risk level exceeds 2.0 can it be said that the one risk factor is more likely to cause the disease than any other factor affecting the unexposed cohort.

As an example, if it is the case that in a random sample of 5000 people 100 are likely to contract colon cancer, and in a random sample of 5000 people who have been exposed to asbestos 150 are likely to develop the disease, then asbestos exposure would have a relative risk of 1.5 for this disease. However, only one third of the afflicted people in the exposed cohort could be said to have contracted colon cancer as a result of their exposure, because on average 100 would have developed it anyway. Epidemiology alone would offer no way to identify which 50 victims were attributable to asbestos. In the absence of any other evidence, the strongest conclusion which could be drawn would be that for each of the 150 afflicted individuals there was a one in three chance that the disease was caused by asbestos. As this probability is less than fifty percent, none of the victims could satisfy the legal standard of showing that it was more probable than not that the cancer was due to asbestos exposure.<sup>240</sup>

On reargument, the district court explained further that the plaintiff could not avoid summary judgment by adducing epidemiological studies indicating that some groups exposed to asbestos possess relative risks greater than 2.0.<sup>241</sup>

241. In re Joint E. & S. Dist. Asbestos Litig., 774 F. Supp. 113, 114–15 (S.D.N.Y. 1991) (court denied motion for reargument).

<sup>237.</sup> See Linda A. Bailey et al., Reference Guide on Epidemiology § V, in this manual.

<sup>238.</sup> See id. § III.A.

<sup>239. 758</sup> F. Supp. 199 (S.D.N.Y. 1991), rev'd, 964 F.2d 92 (2d Cir. 1992).

<sup>240.</sup> Id. at 202–03 (footnote omitted). See also In re "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 785 (E.D.N.Y. 1984) aff'd on other grounds, 818 F.2d 145 (2d Cir. 1987), cert. denied, 484 U.S. 1004 (1988):

A court . . . must observe the tort law requirement that a plaintiff establish a probability of more than fifty percent that the defendant's action injured him. This means that at least a two-fold increase in incidence of the disease attributable to Agent Orange exposure is required to permit recovery if epidemiological studies alone are relied upon.

The plaintiff also had to introduce evidence that her husband shared the heavy exposure that had been experienced by those in the exposed cohorts who had a relative risk over 2.0.<sup>242</sup>

The Second Circuit reversed, without, however, reaching the issues discussed above. The court found that the "plaintiff did not need to provide epidemiological evidence of a certain magnitude in order to defeat a summary judgment motion because she did not rely on epidemiological studies alone." <sup>243</sup> Plaintiff's experts had relied on the decedent's medical records and personal history as well as on epidemiological studies. On the basis of these, they had opined that other causes for the colon cancer, such as a diet high in fats, could be ruled out.<sup>244</sup>

In an individual case, it is unlikely that a plaintiff would ever rely solely on epidemiological studies. A failure by experts to consider medical records and personal history could lead a court to conclude that the expert was failing to consider evidence on which experts customarily rely and that the proffered opinion failed to satisfy Rule 703.<sup>245</sup>

At this time it appears that courts are reluctant to conclude that an epidemiological study will not be an adequate basis for an expert's conclusion about causation solely because there has been less than a twofold increase in risk, provided some positive correlation between exposure and disease is demonstrated.<sup>246</sup> This

243. 964 F.2d at 97. The court quoted with approval from a New Jersey case, Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991). *Id.* In *Grassis*, the court found that plaintiff's expert should not have been precluded from testifying even though she confirmed that most authoritative epidemio-logical studies linking asbestos and colon cancer were below the 2.0 level. The court explained:

[A] particular study might show a high correlation between asbestos and colon cancer, but it also might show a high correlation between the consumption of excessive alcohol and colon cancer. If there were also a very high correlation between those working with asbestos and the high consumption of alcohol, one could not tell whether the alcohol or asbestos or both actually were causative factors of the colon cancer, or even whether the presence of both were needed in order to be a producing factor of the disease. Each study must be analyzed to determine whether the absestos factor was really isolated. Where, however, study after study has shown some positive correlation, although not to the factor of 2.0, it might be said that asbestos is at least a producing factor in some colon cancers, even if the precise biological process has not yet been defined.

... The physician or other such qualified expert may view the epidemiological studies and factor out other known risk factors such as family history, diet, alcohol consumption, smoking (surprisingly, generally recognized as not being a risk in colon cancer, according to the testimony in this case), or other factors which might enhance the remaining recognized risks, even though the risk in the study fell short of the 2.0 correlation.

#### 591 A.2d at 675.

244. 964 F.2d at 96. After trial, resulting in a verdict for plaintiff of over \$4.5 million, the trial judge granted judgment n.o.v. on the ground that epidemiological studies failed to demonstrate a sufficiently strong and consistent association between asbestos exposure and colon cancer. *In re* Joint E. & S. Dist. Asbestos Litig., 827 F. Supp. 1014, 1037–43 (S.D.N.Y. 1993).

245. See discussion infra § IV.B.2.c.1.

246. In Landrigan v. Celotex Corp., 605 A.2d 1079 (N.J. 1992), the court reversed a directed verdict for defendant granted on the ground that epidemiological studies showed a relative risk smaller than 2.0. The court instructed the trial court to proceed as follows:

Without limiting the trial court on remand, its assessment of Dr. Sokolowski's testimony should include an evaluation of the validity both of the studies on which he relied and of

<sup>242.</sup> Id. at 115 & n.3. The court denied motion for rehearing of grant of summary judgment. In re Joint E. & S. Dist. Asbestos Litig., 774 F. Supp. 116 (S.D.N.Y. 1991).

issue is independent of problems with the study's underlying methodology or statistical significance.

c. Confusing the probability of a sample identification with a probability of guilt

The so-called "prosecutor's fallacy" occurs when a prosecutor presents statistical evidence to suggest that the evidence indicates the likelihood of the defendant's guilt rather than the odds of the evidence having been found in a randomly selected sample.<sup>247</sup> The danger that jurors will erroneously confuse the probability of a match with the probability of guilt exists whenever a test can reliably match two samples and the resulting match is being used to identify the defendant.

In United States v. Massey, for instance, an expert witness testified that three out of five hairs found on a ski mask worn by a bank robber matched one or more out of nine mutually dissimilar hairs taken from the defendant's scalp.<sup>248</sup> He further testified that in his work on more than 2,000 cases, there had only been a couple of occasions on which he had seen hair from two individuals that he couldn't distinguish. He also made reference to a Canadian study which concluded that for a hair that "matched in the manner which I have set forth, there's a chance of one in 4,500 these hairs could have come from another individual." <sup>249</sup> While the appellate court found this evidence somewhat confusing, it found reversible error because of comments the prosecutor made in closing argument which suggested that the hair evidence made the defendant's guilt 99.44% certain.<sup>250</sup>

his assumption that the decedent's asbestos exposure was like that of the members of the study populations. The court should also verify Dr. Sokolowski's assumption concerning the absence of other risk factors. Finally, the court should ascertain if the relevant scientific community accepts the process by which Dr. Sokolowski reasoned to the conclusion that the decedent's asbestos exposure had caused his cancer. Thus, to determine the admissibility of the witness's opinion, the court, without substituting its judgment for that of the expert, should examine each step in Dr. Sokolowski's reasoning.

#### Id. at 1088.

247. See, e.g., William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy, 11 Law & Hum. Behav. 167, 181–82 (1987); Jonathan J. Koehler, Error and Exaggeration in the Presentation of DNA Evidence at Trial, 34 Jurimetrics J. 21 (1993) (discussing instances of "prosecutor's fallacy" in the presentation of DNA evidence); see also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.2, in this manual.

248. 594 F.2d 676, 678–79 (8th Cir. 1979). See also United States v. Chischilly, 30 F.3d 1144, 1156 (9th Cir. 1994) (assessing the potential for prejudice arising from the possibility that the jury will accept the estimate of a coincidental match of a DNA profile "as a statement of source probability (i.e., the likelihood that the defendant is the source of the evidentiary sample)" rather than as an estimate of the rareness of the DNA profile).

249. Massey, 594 F.2d at 679. The appellate court found the expert's testimony confusing and the foundation for the witness's reference to the Canadian study insufficient because the witness "testified that he did not know the nature and extent of the studies conducted from which the statistics were gathered." *Id.* at 680.

250. Id. at 680. The prosecutor said:

Now in order to convict the defendant, you must find him guilty beyond a reasonable doubt.

... A handful [let's say that's] 3 to 5 out of 2,000. That's better than 99.44 percent; it's better than Ivory Soap, if you remember the commercial. It's very convincing.

Although trial judges can obviously prevent such a blatant misuse of statistical evidence, difficult problems remain. Should a court permit evidence of matching samples when no background rate is offered of the probability of a match or when there are disputes about the appropriate background rate?<sup>251</sup> The issue can arise with many varieties of trace evidence, such as fibers, soil, and tool marks.

A match without more undoubtedly satisfies the relevancy test set forth in Rule 401 of altering the probabilities, but when no background rate is offered, may the jury erroneously give the evidence far more weight than it actually has?<sup>252</sup> If, for instance, the samples that match are tape to which defendant had access at his place of work and tape used in manufacturing a bomb sent through the mails from an unknown location, the probative value of the evidence is virtually nonexistent if thousands of identical rolls of tape were distributed throughout the world.<sup>253</sup> *Daubert* contains a reminder of the trial judge's power to exclude pursuant to Rule 403 and quotes Judge Weinstein: "Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercises more control over experts than over lay witnesses."<sup>254</sup>

#### d. Reducing odds because of sampling uncertainties; DNA

One of the central issues in the debate about the admissibility of DNA evidence concerns the probability estimate that an expert may properly make when testing reveals a match.<sup>255</sup> Population geneticists have identified a number of problems

Now hair samples are not like fingerprints. It is not positive identification. There is a theoretical possibility (and it actually happened in the case of this examiner in 3 to 5 times out of say, 2,000) where the hairs of two different heads can look the same when you examine the whole range of their characteristics.

However, it is infinitesimally rare, and when we talk about the range of proof which we can use in deciding questions for us, these kinds of percentages are higher than the percentage we use in any other area I can think of in terms of making a decision.

I submit to you that if hair samples are found a known and an unknown and they are microscopically identical, that it is at the very least proof beyond a reasonable doubt that the unknown hair comes from the same head as the known hair.

ld. at 680.

251. Disputes about background rates are considered infra§III.C.3.d.

252. See, e.g., United States v. Bynum, 3 F.3d 769, 773 (4th Cir. 1993) (prosecution sought to link coconspirators by showing through gas chromatography that cocaine seized at different locations had identical composition; no evidence appears to have been offered about the extent to which batches of cocaine differ from each other), cert. denied, 114 S. Ct. 1105 (1994).

253. See United States v. Stifel, 433 F.2d 431, 435 (6th Cir. 1970) (expert testified that fragments of tape on bomb packing matched samples of tape taken from defendant's place of work and were "of the same manufacture" and from "the same batch"), cert. denied, 401 U.S. 994 (1971), conviction vacated, 594 F. Supp. 1525 (N.D. Ohio 1984) (conviction vacated primarily because of *Brady* violations, but evidentiary hearing also demonstrated that sample from bomb packing did not differ from other samples of tape that came from different batches).

254. Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2798 (1993) (quoting Jack B. Weinstein, Rule 702 of the Federal Rules of Evidence Is Sound; It Should Not Be Amended, 138 F.R.D. 631, 632 (1991)).

255. See Judith A. McKenna et al., Reference Guide on Forensic DNA Evidence § VII, in this manual.

Reference Manual on Scientific Evidence

that may cause serious underestimation of the probability of a coincidental match. Currently, there is considerable debate as to the frequency with which gene components known as alleles are found in particular populations. Furthermore, not enough may be known about whether specific alleles are independently inherited so as to warrant use of the product rule to multiply the frequency with which each allele is found.

Because of these as yet unresolved questions, the National Research Council (NRC) recommended using a "ceiling principle" in applying the multiplication rule for estimating the frequency of a particular DNA profile until more research is done. <sup>256</sup> This principle seeks to ensure that the assigned probability will always be greater than or equal to the true probability of a match despite our present lack of knowledge. <sup>257</sup>

The heated debate among population geneticists exemplifies the difficult issues that a court may face when an expert seeks to testify in probabilistic terms. How should a court deal with the proffered opinion if there is disagreement in the relevant scientific communities about the precise statistical conclusions that may validly be drawn, although a general consensus exists that the evidence on which the opinion is based does substantially alter probabilities with regard to an issue in controversy?

One approach, taken by the Second Circuit, is to treat this issue as one of weight. In *United States v. Jakobetz*, defense experts had challenged the statistical interpretation offered by the FBI on the ground that insufficient information was available about population substructures, making it "inappropriate to use one data base for all Caucasians and to use the product rule to calculate an allele pattern's frequency."<sup>258</sup> The Second Circuit found that the FBI's conclusion that the probability of a coincidental match was "one chance in 300 million" had properly been admitted.<sup>259</sup> Furthermore, the court disclaimed the need to conduct extensive hearings and findings thereafter:

[I]n future cases with a similar evidentiary issue, a court could properly take judicial notice of the general acceptability of the general theory and the use of these specific techniques... Beyond such judicial notice, the threshold for admissibility should require only a preliminary showing of reliability of the particular data to be offered, *i.e.*, some indication of how the laboratory work was done and what analysis and assumptions underlie the probability calculations. The probability data may well vary among different segments of the population. Affidavits should normally suffice to provide a sufficient basis for admissibility. DNA profiling evidence should be excluded only when the government cannot show this threshold level of reliability in its data. The district court should focus on whether [an] accepted protocol was adequately followed

257. Id.

<sup>256.</sup> Committee on DNA Technology in Forensic Science, National Research Council, DNA Technology in Forensic Science 13–14, 82–85 (1992) [hereinafter NRC Report]. See also Judith A. McKenna et al., Reference Guide on Forensic DNA Evidence § VII.B.2, in this manual.

<sup>258. 955</sup> F.2d 786, 799 (2d Cir.), cert. denied, 113 S. Ct. 104 (1992). 259. Id. at 789, 799.

in a specific case, but the court, in exercising its discretion, should be mindful that this issue should go more to the weight than to the admissibility of the evidence. Rarely should such a factual determination be excluded from jury consideration. With adequate cautionary instructions from the trial judge, vigorous cross-examination of the government's experts, and challenging testimony from defense experts, the jury should be allowed to make its own factual determination as to whether the evidence is reliable.<sup>260</sup>

Rather than admitting the expert's probability assessment or excluding the DNA evidence, a court could take the intermediate position of requiring a modification of the probability estimate. The NRC Report recommended that experts couch their opinions as follows in the interval before additional research furnishes needed information:

1) If no match is found with any sample in a total databank of N persons (as will usually be the case), that should be stated, thus indicating the rarity of a random match. 2) In applying the multiplication rule, the 95% upper confidence limit of the frequency of each allele should be calculated for separate U.S. "racial" groups and the highest of these values or 10% (whichever is the larger) should be used. Data on at least three major "races" (*e.g.*, Caucasians, blacks, Hispanics, Asians, and Native Americans) should be analyzed.<sup>261</sup>

Although no federal court has followed the specific recommendation of the NRC Report, recently the U.S. Court of Appeals for the Ninth Circuit upheld the admission of probability estimates of a coincidental matching DNA profile that included conservative adjustments similar to those recommended by the NRC.<sup>262</sup>

Other possible solutions are discussed in the NRC Report. <sup>263</sup> Experts could also be instructed to state a range of probabilities that take into account a variety of hypotheses, to use verbal formulations instead of numbers, <sup>264</sup> or to use more than the usual four probes in order to decrease the probability of a coincidental profile match.

262. United States v. Chischilly, 30 F.3d 1144, 1158 (9th Cir. 1994) (probability estimation employed conservative statistical estimates even though "not calculated pursuant to the NRC Report's controversial recommendation to adopt the ceiling principle").

263. NRC Report, supra note 256, at 84-85.

264. See, e.g., various suggestions for explaining significance of human leukocyte antigen (HLA) paternity testing in 1 Paul C. Giannelli & Edward J. Imwinkelried, Scientific Evidence § 17-9(A), at 578 (2d ed. 1993) (ABA and AMA approved guidelines provide for six steps, ranging from "no significance" to "paternity practically proven").

<sup>260.</sup> Id. at 799-800. See also United States v. Bonds, 12 F.3d 540 (6th Cir. 1993) (agreeing that substructure argument goes to weight; post-Daubert).

<sup>261.</sup> NRC Report, supra note 256, at 95. This approach is being questioned and being praised. Compare B. Devlin et al., Comments on the Statistical Aspects of the NRC's Report on DNA Typing, 39 J. Forensic Sci. 28, 34 (1994) ("no scientific motivation for the NRC panel's ceiling principle") with Richard Lempert, DNA, Science and the Law: Two Cheers for the Ceiling Principle, 34 Jurimetrics J. 41 (1993). See also David H. Kaye, The Forensic Debut of the National Research Council's DNA Report: Population Structure, Ceiling Frequencies and the Need for Numbers, 34 Jurimetrics J. 369 (1994) (reviewing judicial opinions that discuss the "ceiling principle" technique). The NRC has impaneled another committee to consider criticism of the recommendations of the report. Rorie Sherman, New Scrutiny for DNA Testing, Nat'l L.J., Oct. 18, 1993, at 3.

The court's decision on how to permit the probability estimate to be stated may in part turn on the quantum of other evidence in the case. In United States v. Jakobetz, for example, the victim positively identified the defendant, and the prosecution introduced an enormous amount of conventional circumstantial evidence linking the defendant to the kidnapping and rape with which he was charged.<sup>265</sup> It seems highly unlikely that a conviction obtained without the use of DNA evidence would have been overturned on insufficiency grounds. Consequently, an underestimation of the true probability would, at most, amount to harmless error. This conclusion suggests, however, that courts should perhaps hesitate in according judicial notice to the probabilistic underpinnings of a particular theory until a sufficient period has elapsed for the statistical assumptions to be thrashed out by the scientific community. Frequently, probability issues with regard to a particular form of evidence arise only in the context of forensic application; even though the underlying scientific theory is well grounded, as is the case with the theory of DNA typing, time is needed to consider the probabilistic implications.

## e. Incorporating proficiency test performance results; DNA

An additional contributor to uncertainty is that some risk of error with regard to test results stems from laboratory practices, such as improper handling of samples, and mistakes in interpreting and reporting results.<sup>266</sup>

The Daubert Court mentioned "the known or potential rate of error" and "the existence and maintenance of standards controlling the technique's operation" as methodological factors a court "should consider."<sup>267</sup> It cited two voiceprint evidence cases in which the courts found the evidence admissible.<sup>268</sup> Whether this means that once evidence passes a certain threshold with regard to error, the issue is solely one of weight to be considered by the trier of fact is not yet clear. In United States v. Bonds, the court reviewing the admissibility of DNA evidence after Daubert termed "troubling" deficiencies in calculating the rate of error and the failure to conduct external proficiency testing, or to specify the rate of error.<sup>269</sup> The court found, however, that when the district judge below affirmed the magistrate judge's finding under the then applicable *Frye* test that

269. 12 F.3d 540, 560 (6th Cir. 1993).

<sup>265. 955</sup> F.2d 786 (2d Cir.), cert. denied, 113 S. Ct. 104 (1992).

<sup>266.</sup> Whether errors of this type might cause a court to reject evidence as unreliable is discussed infra § IV.B.2.c.

<sup>267.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2797 (1993).

<sup>268.</sup> United States v. Smith, 869 F.2d 348, 353–54 (7th Cir. 1989); United States v. Williams, 583 F.2d 1194, 1198 (2d Cir. 1978), cert. denied, 439 U.S. 1117 (1979). Cf. John I. Thornton, Courts of Law v. Courts of Science: A Forensic Scientist's Reaction to Daubert, 1 Shepard's Expert & Sci. Evidence Q. 475, 481 (1994) (expressing dismay at court's reference to Smith because error rates testified to ranged from zero to 88.3%; author argues that "the types of examinations that are often conducted in crime laboratories do not lend them-selves to analysis by conventional statistics").

the FBI's procedures are generally accepted, this finding implicitly decided "that the rate of error is acceptable to the scientific community as well." <sup>270</sup>

The American Society of Crime Laboratory Directors (ASCLD) and the American Society of Crime Laboratory Directors—Laboratory Accreditation Board (ASCLD-LAB) have both recommended mandatory proficiency testing at regular intervals as a requirement for accreditation of forensic-science laboratories engaged in DNA typing.<sup>271</sup> If proficiency testing comes to pass, it will be possible to assign an error rate to each laboratory. Should the proficiency test performance rate then be somehow integrated with the estimation of the probability of a match?<sup>272</sup> Of course, the same issue may arise in connection with tests and studies relating to matters other than DNA. It may be that the discussion of proficiency testing with regard to DNA will have a spillover effect.<sup>273</sup>

270. Id.

271. See NRC Report, *supra* note 256, at 102–06, for information about these entities and their recommendations for laboratory accreditation.

272. See Michael J. Saks & Jonathan J. Koehler, What DNA "Fingerprinting" Can Teach the Law About the Rest of Forensic Science, 13 Cardozo L. Rev. 361, 368–69 (1991) (discusses a number of different models for how this could be done). But see Devlin et al., supra note 261, at 38 ("[A]n a priori estimate of a handling error is not sufficient to evaluate the probability of a handling error in any particular case.").

273. Saks & Koehler, supra note 272. See also discussion of errors in data leading to exclusion pursuant to Rule 703 infra § IV.B.2.c.

Reference Manual on Scientific Evidence

# IV. Is the Expert's Opinion Supported by Reliable Data?

## A. Rule 703

The objection that a witness is basing his or her opinion on evidence not "reasonably relied upon" is frequently encountered in judicial opinions treating a challenge to expert testimony. The phrase is derived from Federal Rule of Evidence 703, which provides:

The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.

The meaning of Rule 703 has always been the subject of controversy.<sup>274</sup> Although the Supreme Court's opinion in *Daubert* clarified the meaning of Rule 703 in some respects because Rule 703 no longer applies to issues the Court allocates to Rule 702, other issues remain about the two rules' interface that require resolution in the wake of *Daubert*. The *Daubert* opinion also contains a brief comment about Rule 703 itself, which while clearly dictum—the rule played no role in the majority's analysis—may nevertheless shed some light on Rule 703 issues. Furthermore, the Court's recognition that Rule 403 plays a role in the exclusion of expert testimony means that courts must also consider the boundary between Rule 703 and Rule 403.<sup>275</sup>

The discussion first examines how the Court's discussion of Rule 702 impacts on Rule 703 and then considers the Court's observation about Rule 703. It turns next to a variety of theoretical issues about the application of Rule 703 that the Court's opinion does not address. It concludes with a survey of contexts in which courts have relied on Rule 703 to exclude evidence.

<sup>274.</sup> See 3 Jack B. Weinstein & Margaret A. Berger, Weinstein's Evidence: Commentary on Rules of Evidence for the United States Courts and State Courts ¶ 703[03] (1993). 275. See infra § V.
- B. Rule 703; Scope of Rule
- 1. The impact of Daubert
- a. Reclassifying issues under Rule 702 that some courts had classified under Rule 703

1. *Fit.* In *Daubert*, the Court defined the scope of Rule 702 to encompass issues that some courts previously handled pursuant to Rule 703. Rule 702, rather than Rule 703, is now the proper vehicle for excluding expert opinions that do not "fit." By this term, the Court means that the court must make a pre-liminary assessment "of whether the reasoning or methodology underlying the testimony . . . can be applied to the facts in issue."<sup>276</sup>

2 Methodology. Daubert states that Rule 702 governs determinations about the experts' use of scientific reasoning in arriving at their conclusions. In making a preliminary inquiry into the admissibility of an opinion, the court is directed to examine its methodological underpinnings and not to rely solely on *Frye's* "general acceptance" approach. Consequently, issues concerning the reliability of a theory or discipline should be handled pursuant to Rule 702.<sup>277</sup> As the discussion in section III.C.2.b *supra* indicates, however, questions about data an expert used in applying a particular methodology may at times raise issues that straddle Rules 702 and 703.<sup>278</sup>

- b. Rule 703 reference
  - 1. Standard of proof. In Daubert, the Court stated:

Throughout, a judge assessing a proffer of expert scientific testimony under Rule 702 should also be mindful of other applicable rules. Rule 703 provides that expert opinions based on otherwise inadmissible hearsay are to be admitted only if the facts or data are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject."<sup>279</sup>

Does the mention of Rule 703 in connection with preliminary determinations pursuant to Rule 702 mean that inquiries under Rule 703, like those under Rule 702, are subject to a Rule 104(a) preponderance-of-the-evidence standard? Courts have rarely explicitly considered this issue. Furthermore, in a number of cases in which courts used a Rule 104(a) standard when excluding evidence pursuant to Rule 703, they were excluding scientific evidence on methodologi-

<sup>276.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993). See discussion supra  $\S$  I.B.

<sup>277.</sup> See discussion supra § III.

<sup>278.</sup> See, e.g., DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 955 n.14 (3d Cir. 1990) ("In this respect, Rules 702 and 703 intersect. If a study's method of data collection is faulty, it may well be that no expert would rely upon the data generated as a basis for drawing any inference about the studied subject."). See also discussion of *In re* Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722 (3d Cir. Aug. 31, 1994), supra note 221.

<sup>279.</sup> Daubert, 113 S. Ct. at 2797–98. The majority opinion also cites Rule 703 as well as Rule 702 after stating that "an expert is permitted wide latitude to offer opinions, including those that are not based on first-hand knowledge or observation." *Id.* at 2796.

cal grounds that according to *Daubert* raise Rule 702 issues.<sup>280</sup> This issue will be revisited in section IV.B.2 in the discussion of the various issues that courts have resolved pursuant to Rule 703.

2. Function of Rule 703 as a rule of admissibility. One function of Rule 703, which no one disputes, was to expand the common-law bases for an expert's opinion by authorizing experts to base their opinions on reliable inadmissible data. Some controversy exists over whether Rule 703 authorizes experts to testify on direct to the hearsay basis for their conclusions or whether the basis of an expert's opinion may only be brought out on the cross-examiner's option pursuant to Rule 705.<sup>281</sup> The Court's comment in *Daubert*—that expert opinions are to be admitted only if the test in Rule 703's second sentence is satisfied—seems to also acknowledge Rule 703's role as an independent source for excluding expert testimony. This approach is consistent with prior practice in the federal courts which construed Rule 703 as imposing conditions on admissibility, rather than as limited to expanding the bases of expert testimony and possibly the scope of expert testimony on direct.

- 2. Other theoretical issues about the function of Rule 703
- a. Does the second sentence of Rule 703 apply only when an expert relies on inadmissible evidence?

According to one view, the plain meaning of Rule 703 is that the "reasonably relied upon" language in the second sentence is a ground for exclusion only when an expert's opinion is based on otherwise inadmissible evidence. If the expert's opinion is based upon admissible evidence, Rule 703 does not apply. Consequently, a court must first determine whether the facts and data underlying the opinion could have been admitted into evidence.

Evidentiary Framework

<sup>280.</sup> See, e.g., Renaud v. Martin Marietta Corp., 972 F.2d 304, 308 (10th Cir. 1992) (excluding study based on only one sample of water pursuant to Rules 703 and 104(a)); Head v. Lithonia Corp., 881 F.2d 941, 944 (10th Cir. 1989) (court rejected evidence based on topographical brain mapping pursuant to *Frye* test; despite liberality of Rule 703, court must not abdicate its responsibility to assure minimum standards for admissi bility as required by Rule 104(a)) (citing *In re* "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1245 (E.D.N.Y. 1985), aff'd, 818 F.2d 187 (2d Cir. 1987), cert. denied, 487 U.S. 1234 (1988)).

<sup>281.</sup> See Ronald L. Carlson, Collision Course in Expert Testimony: Limitations on Affirmative Introduction of Underlying Data, 36 U. Fla. L. Rev. 234, 238, 251 (1984) (discusses objective of Federal Rules to sweep away cases in which, for instance, a physician was not permitted to base his or her opinion on nonrecord laboratory reports; objects to allowing examiner on direct to get inadmissible hearsay before the jury, particularly in criminal cases); James W. McElhaney, *Trial Notebook: Fixing the Expert Mess*, 20 Litigation 53, 56 (1993) (unfairness of allowing expert to get inadmissible hearsay before jury by mentioning basis of opinion on direct). See also University of R.I. v. A.W. Chesterton Co., 2 F.3d 1200, 1219 (1st Cir. 1993) ("we are given some pause by the district court's blanket statement that it 'always requires' the proponent to disclose on direct examination the factual basis for an expert opinion"; the court cites as a comparison example Lis v. Robert Packer Hosp., 579 F.2d 819, 822, 822–23 (3d Cir.) (expressed disapproval of such an invariable practice), *cert. denied*, 439 U.S. 955 (1978)). Cf. Datskow v. Teledyne Continental Motors Aircraft Prods., 826 F. Supp. 677, 684 (W.D.N.Y. 1993) (converts Rule 703 into hearsay exception by allowing letters to be admitted into evidence because they were the basis of expert's opinion).

Chief Judge Clark of the Fifth Circuit forcefully expressed this view in his concurring opinion in Christophersen v. Allied-Signal Corp.:<sup>282</sup>

If the facts or data are admissible, Rule 703 does not authorize exclusion of the expert opinion. If they are admissible, the inquiry ends, and nothing in Rule 703 authorizes exclusion of the expert's testimony. If they are not admissible, the district court must determine whether the reliability inquiry is satisfied. If it is satisfied, Rule 703 does not authorize exclusion. If it is not, the district court should exclude the testimony. No other reading is consistent with the plain language, history, and purpose of Rule 703.<sup>283</sup>

As the *Christophersen* en banc opinion itself demonstrates, however, a narrow view about the permissible ambit of Rule 703 does not mean that a court has no power to screen expert testimony. Despite his restrictive view of the scope of Rule 703, Chief Judge Clark concurred in upholding a grant of summary judgment for the defendant because he found the opinion of the plaintiff's expert excludable under Rule 403:

[I]f an opinion is fundamentally unsupported, then it offers no expert assistance to the jury; and that lack of reliable support can render an opinion substantially more prejudicial than probative, making it inadmissible under Rule 403.<sup>284</sup>

How courts apply Rule 403 to expert testimony is further discussed in section V.

The majority opinion in *Christophersen* takes a broader view that suggests that Rule 703 plays a role in screening expert testimony regardless of the evidentiary posture of the data on which the expert relies:

Although this rule is primarily directed toward permitting an expert to base his opinion on hearsay or otherwise inadmissible sources, the inquiry into the "types" of "facts and data" underlying an expert's testimony is not limited to the admissibility of that data. District judges may reject opinions founded on critical facts that are plainly untrustworthy, principally because such an opinion cannot be helpful to the jury.<sup>285</sup>

b. Determining what is "reasonably relied upon"

In situations in which courts agree that Rule 703 applies, appellate courts do not speak in unison about the trial court's role in determining whether an expert

282. 939 F.2d 1106, 1118 (5th Cir. 1991) (en banc), cert. denied, 112 S. Ct. 1280 (1992).

<sup>283.</sup> *Id.* at 1118. The judge suggested that facts and data might often be admissible pursuant to Rule 803(6) as records of regularly conducted activities, or Rule 803(18) under the learned treatises exception to the hearsay rule. *Id.* at 1119.

<sup>284.</sup> Id. at 1120.

<sup>285.</sup> *Id.* at 1114 (citation and footnote omitted). *See also* Soden v. Freightliner Corp., 714 F.2d 498, 505 (5th Cir. 1983) ("Though courts have afforded experts a wide latitude in picking and choosing the sources on which to base opinions, Rule 703 nevertheless requires courts to examine the reliability of those sources."); Head v. Lithonia Corp., 881 F.2d 941, 943 (10th Cir. 1989) (Rule 703 "provides a mechanism by which the court can evaluate the trustworthiness of the underlying data on which the expert relies."); Shatkin v. McDonnell Douglas Corp., 727 F.2d 202, 208 (2d Cir. 1984) (district judge had "the discretionary right under Fed. R. Evid. 703 to determine whether the expert acted reasonably in making assumptions of fact upon which he would base his testimony").

"reasonably relied." They disagree about the extent to which a court may peer beneath experts' averments that their testimony is based on data upon which experts in their field rely.

Before *Daubert*, courts espousing a "liberal" approach stressed that the Federal Rules of Evidence sought to expand the admissibility of expert testimony. Consequently, "Rule 703 is satisfied once there is a showing that an expert's testimony is based on the type of data a reasonable expert in the field would use in rendering an opinion on the subject at issue."<sup>286</sup> Courts advocating a more "restrictive" approach treated the reliability of expert testimony as a preliminary question of admissibility no different than other issues appropriate for determination under Rule 104(a). After *Daubert*, this distinction may no longer be tenable.<sup>287</sup> In any event, the disagreement between the two camps is one of emphasis that is perhaps reflected more in procedural distinctions than in evidentiary ones. Courts subscribing to the liberal view seemed more inclined to treat the proffered expert testimony as presumptively reliable unless and until the opponent made an adequate showing, and then to insist on a fully developed record before a judge will exclude the testimony.<sup>288</sup> Other courts have been willing to grant summary judgment without requiring motions in limine first.<sup>289</sup>

Because theoretical distinctions may fail to accord with what courts actually do, the discussion below concentrates on fact patterns of expert testimony that some courts have found problematic owing to the data on which the expert relied. The material is organized in terms of the most common categories that courts use when they screen testimony under Rule 703. The commentary also indicates other approaches that some courts use to deal with the problems that some of their judicial colleagues classify as falling within the ambit of Rule 703.

c. Circumstances in which courts use a "reasonably rely" test to exclude

1. Expert's failure to consider data that must be taken into account. Courts have at times relied on Rule 703 in excluding an opinion where the specific facts and data on

Evidentiary Framework

<sup>286.</sup> DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 953 (3d Cir. 1990), aff'd without op., 6 F.3d 778 (3d Cir. 1993), cert. denied, 114 S. Ct. 691 (1994).

<sup>287.</sup> See In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*62-63 (3d Cir. Aug. 31, 1994).

Court stated that its former view is no longer tenable in light of *Daubert*.... By requiring the judge to look to the views of other experts rather than allowing the judge to exercise independent judgment, current Third Circuit case law eviscerates the judge's gate-keeping role with respect to an expert's data and instead gives that role to other experts. The gatekeeping role is reduced even further by our *DeLuca* holding that the opinion of one expert that a type of data is reliable will generally be enough to render that data reliable.

<sup>288.</sup> See, e.g., DeLuca, 911 F.2d at 953 (remanding for record-supported factual findings); In re Paoli R.R. Yard PCB Litig. (Paoli I), 916 F.2d 829, 855 (3d Cir. 1990) (summary judgment would have had to be set aside solely on ground that plaintiffs were afforded insufficient process at the evidentiary stage), cert. denied, 499 U.S. 961 (1991).

<sup>289.</sup> See Christophersen v. Allied-Signal Corp., 939 F.2d 1106 (5th Cir. 1991), cert. denied, 112 S. Ct. 1280 (1992).

which the expert relies "are critically inaccurate or incomplete, as determined by what other experts would or would not be willing to base opinions upon."<sup>290</sup> The gist of this objection is that the expert has failed to consider data that must be taken into account in reaching the opinion that the expert is rendering.

An oft-cited case in the Fifth Circuit, *Viterbo v. Dow Chemical Co.*, is illustrative.<sup>291</sup> The plaintiff claimed that exposure to the defendant's pesticide had caused his nervousness, depression, renal failure, and hypertension. The district judge granted summary judgment on the ground that the testimony of the plaintiff's expert was excludable pursuant to Rule 703. The expert had reached his conclusion without considering the plaintiff's family history, even though a number of the plaintiff's relatives had been hospitalized for depression and hypertension. He failed to explain why the plaintiff had no reaction when he was exposed to the defendant's product in the expert's office. Furthermore, although a blood test of the plaintiff revealed a high level of another chemical that can cause depression, the expert ignored this result on the ground that the plaintiff had denied having had contact with that chemical, even though he failed to explain why the substance was found in the plaintiff's bloodstream. The appellate court affirmed, stating that the expert's "opinion simply lacks the foundation and reliability necessary to support expert testimony."<sup>292</sup>

2. Expert's reliance on data that should not be taken into account. Courts have also cited Rule 703 when faulting an expert for reaching a conclusion on the basis of facts or data that ought not to be taken into account.<sup>293</sup> A detailed illustration of problems considered pursuant to Rule 703 can be found in *DeLuca v. Merrell* 

293. See, e.g., United States v. Tran Trong Cuong, 18 F.3d 1132, 1143–44 (4th Cir. 1994) (although experts may consider hearsay, including reports of other experts, in reaching their opinions, the reports must qualify as data "of a type reasonably relied upon by experts in the particular field"; a physician in the field of family medicine would not usually rely upon forensic medical opinions "specifically prepared for purposes of litigation"; error in a criminal case for an expert testifying for the government to state that a prominent physician who had been a former president of the medical society agreed with him merely to convince the jury of the accuracy and reliability of the expert's opinions; this is unfair, as it denies the defendant his right to cross-examination "and is an improper use of expert testimony."). *See also* Marsee v. United States Tobacco Co., 866 F.2d 319, 323 (10th Cir. 1989) (excluding expert's testimony regarding conversations with other physicians about cases that supported his opinion).

<sup>290.</sup> Id. at 1115.

<sup>291. 826</sup> F.2d 420, 423 (5th Cir. 1987).

<sup>292.</sup> Id. at 424. Cf. Cella v. United States, 998 F.2d 418, 420–22 (7th Cir. 1993) (Jones Act action; plain tiff claimed that his disease, polymyositis, could have been caused by trauma on a ship; defendant attacked damage award to plaintiff as an example of "junk science" entering the courtroom, but court affirmed in an extensive opinion that explained the basis for plaintiff's conclusion about a possible link between plaintiff's disorder and stress; expert had conducted extensive neurological testing of plaintiff and had excluded all other possible factors, such as genetic defects, viral infections, vaccinations, and certain tropical diseases; although he conceded that the etiology was unknown in many cases, he pointed to references in the medical literature that discussed the possibility of stress as a precipitating cause, and a study that showed a link in some instances, and he articulated a plausible hypothesis for why stress would play a role in the etiology of the disease). See also In re "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1250–51 (E.D.N.Y. 1985) (court granted summary judgment for defendant; court found that testimony of experts was "insufficiently grounded in any reliable evidence"; for instance, one expert who concluded that plaintiffs' difficulties were caused by exposure to Agent Orange had failed to consider individual plaintiffs' past medical histories or their families' histories, smoking or drinking habits, or exposure to other substances and drugs), *affd*, 818 F.2d 187 (2d Cir. 1987). See also discussion of differential diagnosis in Paoli II, supra § III.C.2.b.

Dow Pharmaceuticals, Inc.<sup>294</sup> In DeLuca, the trial judge found that Rule 703 requires the exclusion of the testimony of the plaintiff's expert because he "specifically relied upon several types of data experts in the field would not use in forming their opinions."<sup>295</sup> The court concluded that epidemiologists would not rely on their own unpublished reanalyses of adverse drug reaction re ports (ADRs) and drug experience reports (DERs),<sup>296</sup> would not rely on preliminary drafts of studies that were later replaced by finalized published studies,<sup>297</sup> and would not rely on another expert's unpublished reanalysis of data.<sup>298</sup> The plaintiffs' expert conceded that the reporting of DERs is incomplete and may contain information from lawsuits and news accounts.<sup>299</sup>

Courts may be more hesitant to exclude testimony where experts make no such concessions. In *Mendes-Silva v. United States*, the court reversed a grant of summary judgment in an action under the Federal Torts Claim Act brought by a plaintiff who claimed that her encephalomyelitis was caused by having received yellow fever and smallpox vaccines on the same day.<sup>300</sup> The district court had rejected testimony relating to studies counseling against simultaneous administration of vaccines, on which the plaintiff's experts relied, because they did not involve adults and because they did not involve the same two vaccines. The court of appeals found that the district court's conclusion that such evidence is not of a type reasonably relied upon by experts in the field was "unsupported by the evidence available at the summary judgment stage of the proceedings below." <sup>301</sup> The court of appeals specifically noted that the experts did not concede that studies were not of a type reasonably relied upon. <sup>302</sup>

294. 791 F. Supp. 1042 (D.N.J. 1992), aff'd without op., 6 F.3d 778 (3d Cir. 1993), cert. denied, 114 S. Ct. 691 (1994). The trial court also relied on Rule 702 in excluding plaintiffs' expert's testimony. 791 F. Supp. at 1055. In concluding that the expert had used data on which experts in the field would not rely, the district court commented: "This is where Rules 702 and 703 intersect." *Id.* at 1059. See *also* discussion *supra* § III.C.2.b.

295. 791 F. Supp. at 1059.

296. Plaintiffs conceded that the expert's reanalyses of the FDA's ADR and DER data could be disregarded. 791 F. Supp. at 1059 n.20. In addition, the court noted in its "Factual Findings" that a review of DER data "cannot be used by itself to prove causation, but rather is merely a stimulus for further study." *Id.* at 1050.

297. One draft on which the expert relied had been labeled a preliminary draft by the author, who submitted an affidavit explaining that his subsequent draft corrected errors. The expert admitted the unreliability of the preliminary draft. 791 F. Supp. at 1050.

298. 791 F. Supp. at 1059.

299. *Id.* Furthermore, the expert admitted that he could not verify his DER data because he did not have a list of the DERs he had consulted. *Id.* at 1051. "ADRs have inherent biases as they are second-or-third hand reports, are affected by medical or mass media attention, and are subject to other distortions." *Id.* at 1050. See *also In re* Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*86–87, \*149 (court affirmed district court's holding that one expert could not rely on immunological test results from a chemical injury kit because results were not the type of data reasonably relied upon by experts in the field under Rule 703, and that another expert could not rely on recalculations that were "too rough to be considered reliable at all").

300. 980 F.2d 1482, 1486-87 (D.C. Cir. 1993).

301. Id. at 1486.

302. Id. Cf. Cantrell v. GAF Corp., 999 F.2d 1007, 1014 (6th Cir. 1993) (neither Rule 702 nor Rule 703 bars a physician from testifying "to confirmatory data, gained through his own clinical experience"; in action claiming personal injuries from asbestos exposure in the workplace, court found no error in allowing expert who testified about medical sources attesting to a link between asbestos and cancer to also state that 3 out of 150 employees had laryngeal cancer; court noted that expert was subject to cross-examination and conceded

Evidentiary Framework

3. Expert's reliance on data that are erroneous. The DeLuca case also illustrates an expert's reliance on data that are wrong.<sup>303</sup> In DeLuca, the plaintiff's expert could not account for some of the relative risk numbers he had entered on his charts. He seems to have transposed numbers, made arithmetical mistakes, changed numbers from an earlier draft chart without giving much of an explanation, and included the numbers from one study twice.<sup>304</sup> The court observed as part of its Rule 703 analysis that the "new data" that he used could not "in many instances be replicated by other experts in the field or even be explained." <sup>305</sup>

Exclusion on the ground that an error in data exists does not fit easily into the plain meaning of Rule 703, which speaks of the "type" of data.<sup>306</sup> Some commentators would argue that neither does Rule 702 apply, as Rule 702's concern is with the methodological reliability of the expert's theory in general<sup>307</sup> and not with its application in the particular case.<sup>308</sup> At some point, however, as *DeLuca* recognizes, an expert whose opinion is derived from faulty data combined with types of data not reasonably relied on is obviously using a skewed methodology, thereby implicating Rule 702 concerns. At other times, however, courts are willing to leave possible errors in data as questions of weight for the jury.<sup>309</sup>

In part the evidentiary issues may be defined by what is discoverable. In civil cases in which the mandatory expert disclosure provisions are in effect, experts must reveal data underlying their conclusions and are subject to deposition. <sup>310</sup> In a criminal case, the lessened opportunity for discovery undoubtedly decreases the likelihood of detecting actual errors in underlying data. In civil cases as well, however, discovery may not always produce the relevant data. When an expert

303. DeLuca v. Merrell Dow Pharmaceuticals, Inc., 791 F. Supp. 1042 (D.N.J. 1992), aff'd without op., 6 F.3d 778 (3d Cir. 1993), cert. denied, 114 S. Ct. 691 (1994).

304. 791 F. Supp. at 1048-50.

305. Id. at 1059.

306. See *In re* Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*149–54, \*152 n.39 (defendants argued that experts would not rely on a nationwide study of PCBs in fat to calculate background level of PCBs in blood; plaintiffs countered that defendants' argument was that experts would not reasonably rely on data from this particular study and not that experts would not reasonably rely on this *type* of data, and that the language of Rule 703 makes it permissible to rely on particular data "even if the particular data was imperfect"; court declined to "rest upon this difficult distinction, for defendants' argument is easily recharacterized as attacking expert reliance on fat data reported in broad ranges (a type of data) rather than as an attack on particular data"; court found that trial court had abused discretion in excluding testimony based on fat study data).

307. See discussion supra § III.

308. Professor Edward Imwinkelried has suggested that expert testimony has a syllogistic structure, the constituent parts of which are a major premise embodying the expert's explanatory theory, a minor premise constituting the case-specific data to which the expert applies the major premise, and a conclusion, which is the opinion the expert proffers. According to this analysis, Rule 702 addresses deficiencies in the major premise, and Rule 703 addresses deficiencies in the minor premise. See Edward J. Imwinkelried, *The "Bases" of Expert Testimony: The Syllogistic Structure of Scientific Testimony*, 67 N.C. L. Rev. 1, 2–3, 5 (1988). As the discussion below indicates, however, in practice it is difficult to discern a bright line between a theory and its application.

309. See discussion of DNA laboratory procedures supra § III.C.3.e.

310. See discussion supra § I.C.1.

that the three employees were smokers and that it was impossible to determine if their cancers were caused by asbestos).

relies on a study done by someone else, as Rule 703 clearly allows, the data underlying that study may not be readily available.<sup>311</sup>

4. Expert's opinion does not rest on a foundation that experts would generally find reliable. Prior to Daubert, the First Circuit excluded plaintiff's expert testimony in a Bendectin case pursuant to Rule 703,<sup>312</sup> and the District of Columbia Circuit suggested in two cases that the expert's testimony in a Bendectin case was inadmissible pursuant to Rule 703, although both cases arose in the context of rulings on the sufficiency of the evidence.<sup>313</sup> The courts seemed to conclude that a court was justified in excluding under Rule 703 testimony contrary to a scientific consensus.

In *Daubert*, the Supreme Court stated in the course of interpreting Rule 702: "The focus, of course, must be solely on the principles and methodology, not on the conclusions that they generate." <sup>314</sup> The Court also acknowledged that a directed verdict or a grant of summary judgment is appropriate "in the event the trial court concludes that the scintilla of evidence presented supporting a position is insufficient to allow a reasonable juror to conclude that the position more likely than not is true."<sup>315</sup> After a "*Cf.*, *e.g.*," cite, the Court referred to two Bendectin cases, *Turpin v. Merrell Dow Pharmaceuticals, Inc.*, 959 F.2d 1349 (6th Cir. 1987), cert. denied, 113 S. Ct. 84 (1992), and *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307 (5th Cir.), and modified, 884 F.2d 166 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990).<sup>316</sup> In *Turpin*, the appellate court affirmed the district court's grant of summary judgment; in *Brock*, the appellate court reversed a jury verdict for the plaintiffs.

The Supreme Court's opinion in *Daubert* raises but does not answer several valid questions: May a court rely on Rule 703 to exclude an expert's opinion that reaches a conclusion that is inconsistent with a scientific consensus or that lacks a scientific foundation? Does such a reading constitute a back-door resurrection of the *Frye* "general acceptance" test, which was rejected by the Court as incompatible with the Federal Rules of Evidence?<sup>317</sup> Should a court use a sufficiency analysis rather than an admissibility analysis when an expert uses an ap-

311. Cf. Fed. R. Civ. P. 45(c)(3)(B)(ii) (obtaining disclosure of unretained expert's opinion or information).

312. Lynch v. Merrell-National Lab., Inc., 830 F.2d 1190, 1196–97 (1st Cir. 1987) ("[T]he district court's firm rejection here of foundationless expert testimony was necessary, admirable, and entirely within the discretion of the court under Federal Rules of Evidence 403 and 703.").

314. Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2797 (1993).

315. Id. at 2798.

316. Id.

317. The Court stated: "Nor does respondent present any clear indication that Rule 702 or the Rules as a whole were intended to incorporate a 'general acceptance' standard." 113 S. Ct. at 2794 (emphasis added).

Evidentiary Framework

<sup>313.</sup> Ealy v. Richardson-Merrell, Inc., 897 F.2d 1159, 1162, 1164 (D.C. Cir.) ("[U]nder Rule 703, an opinion refuting this scientific consensus [that Bendectin is not teratogenic] is inadmissible for lack of an adequate foundation, in the absence of other substantial probative evidence on which to base this opinion"; court reversed trial judge's refusal to grant judgment n.o.v.), *cert. denied*, 498 U.S. 950 (1990); Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 829 (D.C. Cir. 1988) (court stated that Rule 703 "lays the foundation for our consideration of what constitutes adequate expert testimony"; case arose in the procedural posture of the trial court's grant of a judgment n.o.v.), *cert. denied*, 493 U.S. 882 (1989).

propriate methodology and relies on data that experts reasonably rely upon but nevertheless reaches an opinion at odds with the scientific community?

Two questions that surface with some regularity in toxic tort cases illustrate issues the appellate courts may have to address pursuant to Rule 703. First, may a court reject as inadmissible an opinion based on a study that fails to meet a certain level of statistical significance?<sup>318</sup> Second, may a court reject an expert's causation testimony based on animal studies?<sup>319</sup> Both of these issues have been discussed in connection with methodological concerns.<sup>320</sup> These questions indicate that the interrelationship between Rules 702 and 703 and between admissibility and sufficiency questions requires further consideration by the appellate courts.<sup>321</sup>

320. See supra § III.

321. See discussion of *In re* Paoli R.R. Yard PCB Litig. (Paoli II), *supra* note 221 and related text, questioning the viability of the distinction between a methodology and its application after *Daubert*.

<sup>318.</sup> See, e.g., Wade-Greaux v. Whitehall Lab., Inc., No. 30/1988, 1994 U.S. Dist. LEXIS 7649, at \*126 (D.V.I. Mar. 1, 1994) (rejecting testimony pursuant to Rule 703 based on epidemiological studies that do not show a statistically significant increase in the risk of limb reductions associated with the use of defendant's product because the experts "used data that experts in the field would not use in reaching conclusions on the subject").

<sup>319.</sup> *Id.* at \*12–13, \*117–18 (experts in teratology would not reasonably rely upon animal studies in the absence of positive epidemiological data); *In re* Paoli R.R. Yard PCB Litig., No. 86-2229, 1992 U.S. Dist. LEXIS 16287, at \*15–16 (E.D. Pa. Oct. 21, 1992) (same).

### V. Is the Expert's Opinion Subject to Exclusion Under Rule 403?

### A. The Interplay Between Rules 702, 703, and 403

In Daubert v. Merrell Dow Pharmaceuticals, Inc., the Supreme Court acknowledged in passing that Rule 403<sup>322</sup> may also be used to control scientific expert testimony. The Court stated:

Finally, Rule 403 permits the exclusion of relevant evidence "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury ...." Judge Weinstein has explained: "Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercises more control over experts than over lay witnesses."<sup>323</sup>

The Court's recognition that Rule 403 is a source for the exclusion of expert testimony states a proposition with which most judges have generally agreed. Nevertheless, the range of Rule 403's operation in connection with the Article VII rules was somewhat unclear before *Daubert*, when the circuits differed in their understanding of the scope of Rules 702 and 703. Now that *Daubert* has shed some light on the proper role of Rule 702, questions still remain about how Rule 403 fits into this analysis, and the appropriate boundary with Rule 703.

Potential uses for Rule 403 in excluding expert scientific testimony raise complex issues implicating the relationship between judge and jury. Particularly because the exclusion of the plaintiff's expert proof will often result in summary judgment for the defendant, courts will undoubtedly exercise sparingly their power to exclude scientific evidence that is sufficiently trustworthy to pass the test of Rule 702 but would nevertheless confuse or prejudice the jury.<sup>324</sup>

<sup>322.</sup> Fed. R. Evid. 403 provides: "Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence."

<sup>323. 113</sup> S. Ct. 2786, 2798 (quoting Jack B. Weinstein, Rule 702 of the Federal Rules of Evidence Is Sound; It Should Not Be Amended, 138 F.R.D. 631, 632 (1991)).

<sup>324.</sup> In re Paoli R.R. Yard PCB Litig. (Paoli II), 1994 U.S. App. LEXIS 23722, at \*58 n.16 (3d Cir. Aug. 31, 1994) ("exclusion under Rule 403 should be rare").

Whether some issues should be handled pursuant to Rule 702, 703, or 403 is not yet clear. For instance, may a court exclude evidence based on animal studies on the ground that extrapolating from the studies to humans is confusing or prejudicial? Is this an admissibility issue or a sufficiency issue? Does it matter whether the court relies on evidentiary principles or on rules governing sufficiency?

Prior lack of unanimity about the role of Rule 403 stemmed from some courts' view that the Rule 403 balancing test is built into Rules 702 and 703.<sup>325</sup> For instance, the "assist the trier" standard of Rule 702 led some courts to weigh the probative value of expert proof against the countervailing considerations of prejudice and confusion specified in Rule 403.<sup>326</sup> Similarly, the "reasonably rely" language in Rule 703 suggested to judges who take a broad view of the rule that evidence may be excluded without having to turn to Rule 403.<sup>327</sup>

### B. Examples of Situations in Which Courts Apply Rule 403

Some issues regarding expert testimony, such as the admissibility of cumulative testimony, raise questions precisely analogous to those that arise in the nonscientific evidence context. A few examples are discussed below.

#### 1. Prejudicial language

Courts rely on Rule 403 to exclude opinions which are couched in terms that a judge views as overly prejudicial even though the gist of the opinion is admissible.<sup>328</sup> A judge might, for instance, find the terms "voiceprint" or "DNA print" objectionable as suggesting an analogy to fingerprints that might cause a juror to overvalue the worth of the expert's opinion.

<sup>325.</sup> See Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1112, 1120–22 (5th Cir. 1991) (en banc) (per curiam) (while all the judges seemed to agree that expert witness testimony is subject to a Rule 403 analysis, the majority excluded the expert testimony in question without reaching Rule 403; the concurring opinion found that the testimony satisfied the expert rules but should have been excluded pursuant to Rule 403, and the dissent found that the testimony satisfied the expert rules and Rule 403), *cert. denied*, 112 S. Ct 1280 (1992). See also discussion supra § IV.B.2.a.

<sup>326.</sup> See United States v. Vance, 871 F.2d 572, 577 (6th Cir.) (listing as a factor that makes testimony admissible under Rule 702 that probative value outweighs prejudice), *cert. denied*, 110 S. Ct. 323 (1989); American Bearing Co. v. Litton Indus., Inc., 540 F. Supp. 1163, 1170–71 (E.D. Pa. 1982) ("It is apparent that when considering the admissibility of expert testimony, Rules 703 and 403 somewhat overlap, in that an opinion which is deemed inadmissible under one of the rules may also be deemed inadmissible on the basis of the other." Economist in antitrust action included figures from outside the defined market which thus could be misleading and speculative; court cited both Rule 403 and Rule 703).

<sup>327.</sup> See the dispute between the majority and concurring opinions in Christophersen v. Allied-Signal Corp., 939 F.2d 1106 (5th Cir. 1991), cert. denied, 112 S. Ct. 1280 (1992), as to the appropriateness of this approach. See also discussion supra § IV.B.2.a.

<sup>328.</sup> See, e.g., Scott v. Sears, Roebuck & Co., 789 F.2d 1052, 1055–56 (4th Cir. 1986) (not an abuse of discretion for plaintiff's expert, testifying about various elements of the defendant's grating that made it dangerous, to opine that a yellow curb causes human eye to fill in the discontinuities; court granted a new trial because the expert had also stated that the scene was an "accident waiting to happen"; testimony was rejected pursuant to Rule 403 as overly prejudicial).

### 2. "Aura of scientific infallibility" 329

Taken literally, the charge "aura of scientific infallibility" would lead to the exclusion of scientific evidence of the highest probative value. What courts mean when they use this phrase is that the "aura" is somewhat deceptive, but that jurors might be overwhelmed by the seeming "infallibility."<sup>330</sup>

Courts have relied on Rule 403 when they fear that statements of statistical probability might be overpersuasive and thus prejudice the jury. In *United States v. Massey*, <sup>331</sup> for example, the court reversed on the basis of plain error. The prosecution's expert witness who identified a hair sample as identical to one taken from the defendant testified to some statistical probabilities as to which no foundation had been established. In addition, the trial judge engaged in a colloquy with the expert concerning mathematical probabilities which was speculative and confusing. Prejudice was exacerbated by the prosecution's closing argument, which misstated what the expert had said and then dwelled on these misleading mathematical odds. <sup>332</sup>

### 3. In-court demonstrations or evidence of experiments

Courts will at times rely on Rule 403 to exclude visual evidence, such as videotaped demonstrations or computer-generated evidence. Evidence of this kind is so vivid and compelling that jurors may disregard its true value if it is at all misleading. Judges therefore scrutinize this type of evidence carefully to ensure that it is relevant and not improperly influential. Limiting instructions by the trial judge may help to obviate the dangers.

Two cases are illustrative of how courts analyze these cases, and how inextricably Rule 403 evaluations are tied to the particular facts of a case. In *Shipp v. General Motors Corp.*, the plaintiff claimed that a defect in the roof of a car manufactured by the defendant caused her more serious injuries than she would otherwise have suffered. Both the plaintiff and the defendant wanted to offer films into evidence.<sup>333</sup> The court admitted the plaintiff's film and photographs of a car similar to that involved in the accident being dropped on its roof, but excluded all tapes of drop tests performed on other model cars. The defendant

331. 594 F.2d 676, 680 (8th Cir. 1979).

332. See also discussion of statistical problems with regard to DNA evidence supra § III.C.3.d-e. 333. 750 F.2d 418, 422 n.4, 427 (5th Cir. 1985).

Evidentiary Framework

<sup>329.</sup> The quotation is from Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later*, 80 Colum. L. Rev. 1197, 1237 (1980). See Barefoot v. Estelle, 463 U.S. 880, 926 (1983) (Blackmun, J., dissenting); United States v. Anderson, 851 F.2d 384, 393 (D.C. Cir. 1988), *cert. denied*, 488 U.S. 1012 (1989) (prejudicial impact of expert testimony offered against the accused); United States v. Amaral, 488 F.2d 1148, 1152 (9th Cir. 1973) (prejudicial impact of expert testimony offered by criminal defendant in support of a defense).

<sup>330.</sup> Polygraph evidence has often been excluded on a Rule 403 analysis. See United States v. Alexander, 526 F.2d 161, 168 (8th Cir. 1975) ("polygraph evidence . . . is likely to be shrouded with an aura of near infallibility, akin to the ancient oracle of Delphi"), quoted in United States v. MacEntee, 713 F. Supp. 829, 831 (E.D. Pa. 1989). In *MacEntee*, the court relied on Rule 702 in excluding expert testimony that the government's witness was untruthful based on a failed polygraph but stated that it could also have excluded under Rule 403. *Id.* at 832.

sought to admit a film of rollover tests with dummies that showed how a body is tossed in an accident when seat belts are not worn. The defendant argued that this film was relevant to show general principles of occupant movement and was not being offered as a simulation of the accident.<sup>334</sup>

After expressing its distrust of demonstrations involving vehicles other than the model involved in the accident, the trial judge concluded that the jury "would likely consider it as more than a simple demonstration of general principles." <sup>335</sup> The appellate court found no abuse of discretion. <sup>336</sup>

In contrast, in *Harvey v. General Motors Corp.*, a case in which the plaintiff was seeking damages for injuries sustained when thrown through the roof of his car, the trial court admitted films of rollover tests offered to illustrate vehicle dynamics and not to re-create the accident.<sup>337</sup> The trial judge clearly and in detail instructed the jury not to ignore the distinctions in the model of cars.<sup>338</sup> The appellate court affirmed, noting that *Shipp* was not to the contrary: "Evidence properly excluded in one context is not automatically admitted erroneously in a separate context." <sup>339</sup>

Although some courts may continue to rely on Rule 403 in responding to fact patterns that other courts view as controlled solely by Rule 702 or 703, the dispute will probably not affect outcomes. Furthermore, it is often difficult to tell to what extent a particular decision rests on Rule 403, rather than on the expert rules, because courts frequently cite Rule 403 in addition to one of the expert rules.<sup>340</sup> If after *Daubert* the circuits insist on more uniformity in how trial judges must handle certain recurring issues pursuant to Rules 702 and 703, then Rule 403 may become correspondingly more important as a vehicle for the trial courts' exercise of discretion. The trial courts' resort to Rule 403 may also be af-

339. Id. at 1356. See also Swajian v. General Motors Corp., 916 F.2d 31, 36 & n.2 (1st Cir. 1990) (court affirmed the exclusion of videotaped testimony showing what occurs when an axle fractures, but allowed oral testimony about the experiments); Edwards v. Liz Claiborne, Inc., 17 Fed. R. Evid. Serv. (Callaghan) 1316, 1320 (E.D. Pa. 1984) (not officially reported) (defendant allowed to burn fiber in court which was used to show only a limited part of accident; jury would not be misled into believing it was an exact replication of the acci dent); Shekell v. Sturm, Ruger & Čo., 14 Fed. R. Evid. Serv. (Callaghan) 1634, 1637 (9th Cir. 1983) (unpublished opinion) (new trial ordered in product liability action where gun accidentally discharged; a live drop demonstration using a different gun was done only for effect and was probably too prejudicial); Raymond v. Riegel Textile Corp., 484 F.2d 1025, 1028 n.8 (1st Cir. 1973) (dicta) (may have been prejudicial to permit in-court exhibition of burning fabric in a jury trial, but not in a bench trial); Patterson v. F.W. Woolworth Co., 786 F.2d 874, 880 (8th Cir. 1986) (court admitted expert testimony concerning a demonstration that took place under different conditions; appellate court stated that test need not be conducted under exactly similar conditions and noted that trial court had limited prejudice by permitting the plaintiff's expert to remain in the courtroom and offer rebuttal); Wolf v. Procter & Gamble Co., 555 F. Supp. 613, 626-27 (D.N.J. 1982) (toxic shock syndrome case; court permitted plaintiff's expert to perform in-court experiment to explain the expert's testimony; any distinctions between the testing conditions and the human body could be explored on crossexamination).

340. United States v. Long, 917 F.2d 691 (2d Cir. 1990) (expert testimony on structure of crime family excluded; would not be helpful to jury; unclear as to which test was applied).

<sup>334.</sup> Id. at 427.

<sup>335.</sup> Id.

<sup>336.</sup> Id.

<sup>337. 873</sup> F.2d 1343, 1355 (10th Cir. 1989).

<sup>338.</sup> Id.

fected by the appellate courts' choice of standards for reviewing determinations pursuant to Rules 702 and 703.<sup>341</sup> If courts adopt stringent standards, then trial courts may tend to bolster their conclusions with a Rule 403 analysis that will be governed by an abuse-of-discretion standard of review.

341. See discussion supra § I.D.

Evidentiary Framework

### **Comment on the Use of the Reference Guides**

The reference guides that follow are intended to aid judges in litigation involving contested scientific issues. Their principal purpose is to facilitate the process of identifying and narrowing disputed issues concerning scientific evidence. Each guide sets forth the key issues implicated by posing a question that judges will be able to use in a dialogue with counsel to determine the specific areas of dispute. The questions are structured to give the judge an overview of the intellectual framework of the particular scientific discipline.

The guides are not intended to instruct judges concerning what scientific evidence should be admissible. Instead, they outline for judges the pivotal issues in the area of science that are often the subject of dispute between litigants. Reference to these outlines may be helpful in identifying issues and clarifying questions relevant to rulings on admissibility, such as relevance, reliability, and methodology of scientific evidence. Case citations are offered for illustrative purposes.

Each guide contains critical terms, the first occurrence of which is italicized in the text. A glossary of the key terms used in the text follows each reference guide. The use of certain critical terms, such as *causation*, may vary across reference guides just as it varies across scientific disciplines. Judges should keep in mind the manner in which a term is used in a particular reference guide.

Each reference guide was prepared by an author or authors with substantive knowledge of the scientific discipline and its use in litigation, and reviewed by attorneys, scientists, and others familiar with scientific evidence.

The guides are premised on the idea that management of issues in litigation is a joint effort by court and counsel. Judges are encouraged to make copies of relevant portions available to counsel and interested parties and to invite them to supplement or modify the guides as may be appropriate.

We invite comments and suggestions, as well as recommendations for additional areas of scientific evidence that should be addressed in this manual. Comments and suggestions should be addressed to: Reference Manual on Scientific Evidence - Comments, Federal Judicial Center, Research Division, One Columbus Circle, N.E., Washington, DC 20002-8003.

## **Reference Guide on Epidemiology**

Linda A. Bailey Leon Gordis Michael Green

With contribution of case citations by Paul Rothstein

Linda A. Bailey, M.H.S., J.D., is Program Director, Institute for Injury Reduction, Crofton, Maryland. Leon Gordis, M.D., Dr.P.H., is Professor, Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, Baltimore, Maryland, and Department of Pediatrics, Johns Hopkins School of Medicine. Michael Green, J.D., is Professor of Law, University of Iowa College of Law, Iowa City, Iowa. Paul Rothstein, J.D., is Professor of Law, Georgetown University Law Center, Washington, D.C.

### Contents

- I Introduction 125
- II. Trustworthiness of Research Methods 129
  - A. Was the Research Design Appropriate for Answering the Research Question? 131
    - 1. Cohort studies 134
    - 2. Case-control studies 136
  - B. Were the Study Populations Well Defined and Samples Adequately Selected So As to Allow for Meaningful Comparisons (Between Study Groups or Between Time Periods)? 138
    - 1. Did the researcher minimize the risk of selection bias? 138
      - a. How were the cases and controls (in case-control study) or exposed and unexposed subjects (in cohort study) identified and selected? 138
      - b. What percentage of those selected for the study agreed to participate? 139
      - c. What proportion of the subjects dropped out of the study before it was completed? 139
    - 2. Was the sample size adequate to draw a valid conclusion? 139
      - a. Sample size calculation 140
      - b. Power calculation 141
  - C. Was Exposure to the Putative Agent Measured Using a Standardized and Reliable Methodology? 143
    - 1. Were data collected from objective and reliable sources? 144
    - 2. What types of procedures were instituted to control the quality of measurements of exposure? 145
    - 3. Was information obtained from one group of the study population more accurate or complete than that obtained from the comparison group? 145
    - 4. Did the method of collecting data yield reliable information? 145
  - D. Were the Health Effects (i.e., Disease, Disability) Clearly Defined and Reliably Measured? 146

- III. Association Between Exposure and the Disease 147
  - A. What Is the Basis for Concluding That the Exposure Is Associated with an Increased Risk of Disease? 147
    - 1. Relative risk (RR) 147
    - 2. Odds ratio (OR) 149
    - 3. Attributable proportion of risk (APR) 149
  - B. What Categories of Error Might Have Produced a False Result? 150
  - C. What Statistical Methods Exist to Evaluate the Likelihood That the Result of an Epidemiological Study Was Due to Random Sampling Error? 151
    - 1. False positive error and statistical significance 152
    - 2. False negative error 155
    - 3. Power 156
  - D. What Biases May Have Existed That Would Result in an Erroneous Association? 156
- IV. General Causal Association Between Exposure and the Disease 157
  - A. Could a Confounding Factor Be Responsible for the Study Result? 158
    - 1. What techniques, if any, were used to identify confounding factors? 159
    - 2. What techniques, if any, were used to control confounding factors? 160
  - B. Overall, Does Application of the Guidelines for Causation Support a Finding of Causation? 160
    - 1. How strong is the association between the exposure and disease? 161
    - 2. Is there a temporal relationship? 162
    - 3. Is the association consistent with other research? 162
    - 4. Is the association biologically plausible (consistent with existing knowledge)? 163
    - 5. Have alternative explanations been ruled out? 163
    - 6. Does the association exhibit specificity? 163
    - 7. Is there a dose-response relationship? 164
  - C. What Type of Causal Association Has Been Demonstrated Between Exposure and Disease? 164

V. The Role of Epidemiology in Proving Individual Causation 167 Glossary of Terms 171 References on Epidemiology 179

### I. Introduction

*Epidemiology* is the field of public health that studies the *incidence*, distribution, and etiology of disease in human populations and applies the findings to alleviate health problems. The purpose of epidemiology is to better understand disease causation and to prevent disease in groups of individuals. Epidemiology assumes that disease is not distributed randomly in a group of individuals and that identifiable subgroups are at increased *risk* of contracting particular diseases.

Judges and juries increasingly are presented with epidemiological evidence as the basis of an expert's opinion. Judges determine whether such evidence, or the expert's opinion that relies on epidemiology, reaches the jury.<sup>1</sup> When judges are unclear about how to gauge the quality of the expert's science, and hence the validity of the expert's testimony, incorrect and inconsistent judgments may result.<sup>2</sup>

In the courtroom epidemiological research findings are offered to establish or dispute whether *exposure* to an agent caused a harmful effect or disease.<sup>3</sup>

In any case, an epidemiological study might be part of the basis of an expert's opinion and need not be independently admissible pursuant to Fed. R. Evid. 703. See In re "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1240 (E.D.N.Y. 1985), aff'd, 818 F.2d 187 (2d Cir. 1987), cert. denied, 487 U.S. 1234 (1988); cf. Grassis v. Johns-Manville Corp., 591 A.2d 671, 676 (N.J. Super. Ct. App. Div. 1991) (epidemiological study offered in evidence to support expert's opinion under New Jersey evidentiary rule equivalent to Fed. R. Evid. 703).

2. Compare Richardson v. Richardson-Merrell, Inc., 857 F.2d 823 (D.C. Cir. 1988) (affirming the grant of a judgment n.o.v. to defendant in a Bendectin case), *cert. denied*, 493 U.S. 882 (1989) *with* Oxendine v. Merrell Dow Pharmaceuticals, Inc., 506 A.2d 1100 (D.C. 1986) (reinstating a \$750,000 jury verdict for plaintiff in a Bendectin case), *cert. denied*, 493 U.S. 1074 (1990); *se also* Turpin v. Merrell Dow Pharmaceuticals, Inc., 959 F.2d 1349, 1349 (6th Cir.) ("the inconsistent results reached by courts and juries nationwide on the question of causation in Bendectin birth defect cases are of serious concern"), *cert. denied*, 113 S. Ct. 84 (1992).

3. DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 945–48, 953–59 (3d Cir. 1990) (litigation over morning sickness drug, Bendectin); Cook v. United States, 545 F. Supp. 306, 307–16 (N.D. Cal. 1982) (swine flu vaccine alleged to have caused plaintiff's Guillain-Barré disease); Allen v. United States, 588 F. Supp. 247, 416–25 (D. Utah 1984) (residents near atomic test site claimed exposure to radiation caused leukemia and other cancers), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987), cert. denied, 484 U.S. 1004 (1988); In re "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 780–90 (E.D.N.Y. 1984) (Vietnam

<sup>1.</sup> An epidemiological study, which often is published in a medical or scientific journal, is hearsay. An epidemiological study that is performed by the government, such as one performed by the Centers for Disease Control (CDC), may be admissible based on the hearsay exception for government records contained in Fed. R. Evid. 803(8)(C). See Ellis v. International Playtex, Inc., 745 F.2d 292, 300–01 (4th Cir. 1984); Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 899 (N.D. Iowa 1982), aff/d sub nom. Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1983). A study that is not conducted by the government might qualify for the learned treatise exception, Fed. R. Evid. 803(24) and 804(5).

Epidemiological evidence identifies agents that are associated with an increased risk of disease in groups of individuals, quantifies the amount of excess disease that is associated with an agent, and provides a profile of the type of individual who is likely to contract a disease after being exposed to an agent.<sup>4</sup> Epidemiology focuses on the question of general causation (i.e., is the agent capable of causing disease?) rather than that of specific causation (i.e., did it cause disease in this individual?).<sup>5</sup> For example, in the 1950s Doll and Hill published a series of articles about the increased risk of lung cancer in cigarette smokers. Their findings showed that smokers who smoked ten to twenty cigarettes a day had a lung cancer *mortality rate* that was about ten times higher than that for nonsmokers.<sup>6</sup> Doll and Hill's study identified an *association* between smoking cigarettes and death from lung cancer.

Association is not *causation*.<sup>7</sup> An association identified in an epidemiological study may or may not be causal.<sup>8</sup> Properly designed and executed studies enable epidemiologists to assess the existence (and strength) or absence of an association between an agent and a disease. Epidemiologists commonly use a measure called *relative risk* (RR) to indicate the strength of association between exposure and disease.<sup>9</sup> A strong association that is demonstrated consistently in a series of research projects leads a researcher to infer that a causal relationship exists. Even the best of studies do not demonstrate more than a high probability of a causal relationship between exposure to an agent and a disease. In the absence of an understanding of the biological and pathological mechanisms by which disease develops, epidemiological evidence is the most valid type of scientific evidence of toxic causation.<sup>10</sup>

4. An agent is a factor, such as a drug, a microorganism, a chemical substance, or a form of radiation, whose presence or absence can result in the occurrence of a disease. A disease can have a single agent, a number of independent alternative agents, or a complex of two or more factors whose combined presence is necessary for the development of the disease. Agents are also referred to as risk factors of a disease. A Dictionary of Epidemiology 4 (John M. Last ed., 1988).

5. See infra § V for a discussion of specific causation.

6. Richard Doll & A. Bradford Hill, Lung Cancer and Other Causes of Death in Relation to Smoking , 2 Brit. Med. J. 1071 (1956).

7. Association is more fully discussed *infra* § III. The term is used to describe the relationship between two events (e.g., exposure to a chemical agent and development of disease) that occur more frequently together than one would expect by chance. Association does not necessarily imply a causal effect. Causation is used to describe the association between two events when one event is a necessary link in a chain of events that results in the effect. Of course, alternative causal chains may exist that do not include the agent but that result in the same effect. Epidemiological methods cannot prove causation; however, scientific evidence can lead an epidemiologist to infer that a certain agent causes a disease.

8. See infra §§ IV.A-IV.B.

9. See infra § III.A for a discussion of relative risk and other measures of risk.

10. Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 830 (D.C. Cir. 1988) (epidemiology more probative than other forms of scientific studies), *cert. denied*, 493 U.S. 882 (1989); Conde v. Velsicol Chem.

veterans exposed to Agent Orange and dioxin contaminant brought suit for various diseases and birth defects in their offspring), aff d, 818 F.2d 145 (2d Cir. 1987); Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1115 (5th Cir. 1991) (cancer alleged to have resulted from exposure to nickel-cadmium fumes), cert. denied, 112 S. Ct. 1280 (1992); Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 898–902 (N.D. Iowa 1982) (toxic shock syndrome from use of Rely tampons), *aff'd sub nom.* Kehm v. Proctor & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1983).

An expert's opinion on causation in court is based on a series of epidemiological findings. It is important to note that often the expert testifying before the court is not the scientist who conducted the study or series of studies.<sup>11</sup> The epidemiological studies that form the basis of the expert's testimony should examine persons who represent the general population or the subgroup that is of concern to the court and should assess the risk of disease with a study methodology and statistical measures that limit the opportunity for invalid findings.<sup>12</sup> While the findings of epidemiology always involve a measure of uncertainty, systematic methods for assessing the characteristics of persons included in the study and their risk of disease can be used to help rule out known sources of *bias* and *error*.

The epidemiologist uses *sample size* calculations and inclusion and exclusion criteria for identifying exposed and unexposed study groups (or cases and controls) to reduce potential error and bias in a study. These methods and techniques of epidemiology provide a means of assessing the relationship between a disease and its causes. Unfortunately, these tools are incapable of discerning every association, and the absence of an association should not be interpreted to mean causation does not exist. Even in a well-designed and well-analyzed study, lack of an association may only mean that (1) the sample size was not large enough to detect a weak association, or (2) the disease has multiple causes (epidemiological methods are best able to identify a single cause of disease).

As a final caveat about the limitations of epidemiology, the precision of epidemiological methods is based on the stability of studying large numbers of people. There should be a sufficiently large number of subjects, so that a small change in the number of people with the disease does not appreciably affect the results of the study. Applying population-based results to an individual plaintiff is generally beyond the limits of epidemiology. Measurements of error and risk, the hallmarks of epidemiology, lose their meaning when they are applied to an

Corp., 804 F. Supp. 972, 1025–26 (S.D. Ohio 1992) ("Epidemiologic studies are the primary generally accepted methodology for demonstrating a causal relation between a chemical compound and a set of symptoms or a disease."), *aff'd*, No. 93-3092, 1994 U.S. App. LEXIS 10752 (6th Cir. May 16, 1994); DeLuca v. Merrell Dow Pharmaceuticals, Inc., 791 F. Supp. 1042, 1045 (D.N.J. 1992); *In re* "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1239 (E.D.N.Y. 1985) (epidemiology "assume[s] a role of critical importance"), *aff'd*, 818 F.2d 187 (2d Cir. 1987), *cert. denied*, 487 U.S. 1234 (1988).

<sup>11.</sup> See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 953 (3d Cir. 1990) (pediatric pharmacologist expert's credentials sufficient pursuant to Fed. R. Evid. 702 to interpret epidemiological studies and render an opinion based thereon); Rubanick v. Witco Chem. Corp., 593 A.2d 733, 736 (N.J. 1991) (expert with degree in biochemistry and experience as a cancer researcher permitted to testify about causation and in part rely on epidemiological evidence in support of opinion); *cf.* Landrigan v. Celotex Corp., 605 A.2d 1079, 1088 (N.J. 1992) (epidemiologist permitted to testify to both general causation and specific causation); Loudermill v. Dow Chem. Co., 863 F.2d 566, 569 (8th Cir. 1988) (toxicologist permitted to testify that chemical caused decedent's death).

<sup>12.</sup> In Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796–97 (1993), the Supreme Court addressed the standard for permitting an expert witness to testify to an opinion on a scientific matter. The ultimate issue in the Court's test is whether the methodology and reasoning that form the basis for the expert's opinion are scientifically valid. See generally Margaret A. Berger, Evidentiary Framework, in this manual.

individual. Nevertheless, a substantial body of legal precedent has developed that addresses the use of epidemiological evidence to prove causation for an individual litigant through probabilistic means.<sup>13</sup>

The following sets of questions address technical issues that arise in considering the admissibility of, and weight to be accorded to, epidemiological research findings. Over the past fifteen years, courts frequently have confronted the use of epidemiological studies as evidence and recognized their utility in proving causation. As the Third Circuit observed in *DeLuca v. Merrell Dow Pharmaceuti cals, Inc*.:

> The reliability of expert testimony founded on reasoning from epidemiological data is generally a fit subject for judicial notice; epidemiology is a well-established branch of science and medicine, and epidemiological evidence has been accepted in numerous cases.<sup>14</sup>

The use of epidemiology in legal disputes raises three issues for consideration:

- 1. Were the research methods trustworthy?
- 2. If so, is exposure to the agent associated with disease?
- 3. If the agent is associated with disease, is it a causal relationship?

There is an additional legal question that arises in most *toxic substances* cases. That issue is whether and how population-based epidemiological evidence can be used to infer specific causation. Sections II through V address these four questions. Section II examines research design and planning issues from the perspective of an epidemiologist planning a study and addresses concerns about designing a methodologically valid study. Section III looks at a completed study and explains the significance of a study's findings, statistical methods for assessing the possibility of sampling error, and methodological problems that may distort the outcome of a study.<sup>15</sup> Section IV discusses general causation, considering whether an agent is capable of causing disease. Section V examines issues of specific causation, considering whether an agent causation, considering whether an agent causation.

13. See infra§V.

<sup>14. 911</sup> F.2d 941, 954 (3d Cir. 1990); *se also* Smith v. Ortho Pharmaceutical Corp., 770 F. Supp. 1561, 1571 (N.D. Ga. 1991) (explaining increased reliance of courts on epidemiological evidence in toxic substances litigation).

<sup>15.</sup> For a more in-depth discussion of the statistical basis of epidemiology, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.D, in this manual, and two case studies: Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts,* 43 Hastings L.J. 301 (1992); Devra L. Davis et al., *Assessing the Power and Quality of Epidemiologic Studies of Asbestos-Exposed Populations*, 1 Toxic. & Indus. Health 93 (1985). See also References on Epidemiology and References on Law and Epidemiology at the end of this reference guide.

### II. Trustworthiness of Research Methods

Ethical constraints limit the research methods that the epidemiologist can use.<sup>16</sup> For example, to determine whether cigarette smoking is associated with lung cancer, the epidemiologist would like to compare two randomly selected groups, only one of which smokes cigarettes. Using true experimental methods, the epidemiologist would select a group of individuals and randomly assign half of them to cigarette smoke exposure and half to no cigarette smoke exposure to control for any differences that exist between smokers and nonsmokers. Thus, the epidemiologist could be relatively certain that any difference observed between the groups was caused by smoke exposure.

Since it is unethical to expose a group of human beings to a known harm, true experimental methods cannot be used. Instead, the epidemiologist uses observational methods. Observational methods are limited by the fact that researchers do not control the human subjects. Rather than randomly assign the study subjects to experimental groups (e.g., one group exposed to cigarette smoke and the other group not exposed), researchers identify a group of subjects who have voluntarily (or unknowingly) exposed themselves and compare the group's rate of disease with that of an unexposed group. Important factors that cannot be controlled directly by the epidemiologist include genetic background, lifestyle choices, and the amount and duration of exposure. These factors may be distributed differentially between the groups through random chance or some connection between exposure status and the other factors. The epidemiologist attempts to control and assess the influence of these factors through *research de* - *sign* and statistical analysis.<sup>17</sup>

In addition to observational epidemiology, *toxicology* models based on animal studies (*in vivo*) may be used to determine toxicity in humans. <sup>18</sup> Animal studies

17. True experimental studies require random assignment of subjects to groups. With the exception of controlled clinical trials (i.e., the type of studies used to test the effectiveness of new drug treatments), few epidemiological studies use true experimental methods. This reference guide focuses on observational studies.

18. For an in-depth discussion of toxicology, see Bernard D. Goldstein & Mary Sue Henifin, Reference

<sup>16.</sup> Experimental studies with human beings are ethically proscribed where the agent is known or thought to be toxic. Sæ Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 26 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976). Experimental studies can be used where the agent under investigation is believed to be beneficial, as is the case in the development and testing of new pharmaceutical drugs. Sæ, e.g., E.R. Squibb & Sons, Inc. v. Stuart Pharmaceuticals, No. 90-1178, 1990 U.S. Dist. LEXIS 15788 (D.N.J. Oct. 16, 1990); Gordon H. Guyatt, Using Randomized Trials in Pharmacepidemiology, in Drug Epidemiology and Post-Marketing Surveillance 59 (Brian L. Strom & Giampaolo Velo eds., 1992).

have a number of advantages. They can be conducted as experiments, and researchers control all aspects of the animals' lives. This avoids the problem of confounding, <sup>19</sup> which epidemiology often confronts. Exposure can be carefully controlled and measured. Ethical limitations are diminished and animals can be sacrificed, which may improve the accuracy of disease assessment. Animal studies often provide useful information about pathological mechanisms and play a complementary role to epidemiology by assisting in framing hypotheses and in developing study designs for epidemiological studies.

Animal studies, however, have two significant disadvantages. First, animal study results must be extrapolated to another species—human beings—where differences in absorption, metabolism, and other factors may result in interspecies variation in responses. For example, one powerful human *teratogen*, thalidomide, does not cause birth defects in most rodent species.<sup>20</sup> The second difficulty with inferring human causation from animal studies is that the high *doses* customarily used in animal studies require consideration of the *dose-response relationship* and whether a threshold no-effect dose exists.<sup>21</sup> Those matters are almost always fraught with considerable, and currently unresolvable, uncertainty.

Toxicologists also use *in vitro* methods, in which human or animal cells or tissue are grown in laboratories and exposed to certain substances. The problem with this approach is also extrapolation—whether one can generalize the findings from the tissues in laboratories to whole human beings.<sup>22</sup>

Often toxicological studies are the only or best available evidence of toxicity. Epidemiological studies are difficult, time-consuming, and expensive and consequently do not exist for a large array of environmental agents. Where both animal toxicology and epidemiological studies are available, no universal rules exist for how to interpret or reconcile them. <sup>23</sup> Careful assessment of the method -

Guide on Toxicology, in this manual.

- 20. Phillip Knightley et al., Suffer the Children: The Story of Thalidomide 271-72 (1979).
- 21. See infra § IV.B.7 and accompanying note.

22. For a further discussion of these issues, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § III.A, in this manual.

23. A number of courts have grappled with the role of animal studies in proving causation in a toxic substance case. One line of cases takes a very dim view of their probative value. For example, in Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 313 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990), the court noted the "very limited usefulness of animal studies when confronted with questions of toxicity." A similar view is reflected in Richardson v. Richardson-Merrell, Inc., 857 F.2d 823, 830 (D.C. Cir. 1988), *cert. denied*, 493 U.S. 882 (1989); Bell v. Swift Adhesives, Inc., 804 F. Supp. 1577, 1579–80 (S.D. Ga. 1992); and Cadarian v. Merrell Dow Pharmaceuticals, Inc., 745 F. Supp. 409, 412 (E.D. Mich. 1989). Other courts have been more amenable to the use of animal toxicology in proving causation. Thus, in Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1094 (D. Md. 1986), *aff'd sub nom*. Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987), the court observed: "There is a range of scientific methods for investigating questions of causation—for example, toxicology and animal studies, clinical research, and epidemiology—which all have distinct advantages and disadvantages." *See also* Villari v. Terminix Int'l, Inc., 692 F. Supp. 568, 571 (E.D. Pa. 1988); Peterson v. Sealed Air Corp., Nos. 86-C3498, 88-C9859 Consol., 1991 U.S. Dist. LEXIS 5333, at \*27 – 29 (N.D. Ill. Apr. 23, 1991); *cf. In re* Paoli R.R. Yard PCB Litig., 916 F.2d 829, 853–54 (3d Cir. 1990) (questioning the exclusion of animal studies by the lower court), *cert. denied*, 499 U.S. 961 (1991).

<sup>19.</sup> See infra § IV.A.

ological *validity* and power<sup>24</sup> of the epidemiological evidence must be undertaken as well as consideration of the quality of the toxicology studies and the questions of interspecies extrapolation and dose-response relationship.

When reviewing the methodological validity of an epidemiological study, four issues should be considered:

- 1. Was the research design appropriate for answering the research question?
- 2. Were the study populations well defined and *samples* adequately selected so as to allow for meaningful comparisons (between study groups or between time periods)?
- 3. Was exposure to the putative agent measured using a standardized and reliable methodology?
- 4. Were the health effects (i.e., disease, disability) clearly defined and reliably measured? <sup>25</sup>

# A. Was the Research Design Appropriate for Answering the Research Question?

Research begins with formulation of the research question. This question should be stated clearly by the researcher before the data collection begins, since the researcher cannot measure the uncertainty or potential for error when the findings are unrelated to the research question. Unrelated findings may have some validity and therefore be relevant to a disputed issue in court. However, such findings should be carefully examined for bias. <sup>26</sup>

In reviewing the research methods used to conduct an epidemiological study, the potential for bias should be considered. When scientists use the term bias, it does not necessarily carry an imputation of prejudice or other subjective factors, such as the researcher's desire for a particular outcome. This differs from con-

One explanation for these conflicting lines of cases may be that animal toxicology has much less probative value where a substantial body of epidemiological evidence that addresses the causal issue is available. That was the case, for example, in the Bendectin cases of *Richardson*, *Brock*, and *Cadarian*. Where epidemiological evidence is not available, animal toxicology may be thought to play a more prominent role in resolving a causal dispute. Sæ Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation, 86 Nw. U. L. Rev. 643, 680–82 (1992) (arguing that plaintiffs should be required to prove causation by a preponderance of the available evidence); Turpin v. Merrell Dow Pharmaceuticals, Inc., 959 F.2d 1349, 1359 (6th Cir.), *cert. denied*, 113 S. Ct. 84 (1992); Inre Paoli R.R. Yard PCB Litig, No. 86-2229, 1992 U.S. Dist. LEXIS 16287, at \*16 (E.D. Pa. Oct. 21, 1992). For another explanation of these cases, see Gerald W. Boston, A Mass-Exposure Model of Toxic Causation: The Control of Scientific Proof and the Regulatory Experience , 18 Colum. J. Envtl. L. 181 (1993) (arguing that epidemiological evidence should be required in mass exposure cases but not in isolated exposure cases). See also Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § I.F, in this manual.

<sup>24.</sup> See infra §§ II.B.2.b, III.C.3.

<sup>25.</sup> These questions are based on guidelines for assessing epidemiological studies: 3 U.S. Envtl. Protection Agency, Air Quality Criteria for Particulate Matter and Sulfur Oxides § 14.1.2 (1982).

<sup>26.</sup> See infra § III.

ventional (and legal) usage in which bias refers to a partisan point of view.<sup>27</sup> Bias refers to anything (other than random sampling error)<sup>28</sup> that results in error in a study and thereby compromises its validity. Bias in research can result from a defect in the design or conduct of a study. Although dozens of biases have been catalogued,<sup>29</sup> the two main classes of bias are selection bias (differences in the characteristics between the individuals who are selected for study and those who are not) <sup>30</sup> and information bias (a flaw in measuring expo sure or disease between study groups).

No epidemiological study is perfect; all have some degree of bias that may affect the outcome. Some studies may be so flawed as to be virtually worthless. Finding the bias, however, can be difficult if not impossible. In reviewing the validity of an epidemiological study, the epidemiologist must identify potential biases and analyze (or use educated estimates of) the amount of error that might have been induced by the existence of the bias. Moreover, the direction of error can often be determined; depending on the specific type of bias, it may exaggerate the real association, dilute it, or even completely mask it.

A type of bias that occurs in the formulation of the research question is conceptual bias. Conceptual bias usually means that the research question and hypothesis are biologically implausible as a result of faulty logic, faulty premises, or mistaken beliefs on the part of the researcher. For example, if the researcher defines the disease of interest as all birth defects, rather than a specific birth defect, he or she must have a scientific basis to hypothesize that the effects of the agent being investigated could be so varied.<sup>31</sup> Failure to have such a basis raises concerns about conceptual bias.<sup>32</sup>

Once the research question has been identified, the researcher designs a study that elicits information directly relevant to the question. There are two classes of epidemiological research designs for studying human populations: (1) studies that collect data about the group as a whole and (2) studies that collect data about individuals within the group.

Studies that collect data about the group as a whole are called ecological studies. <sup>33</sup> Such studies are useful for identifying associations but generally are re-

30. See infra note 46.

<sup>27.</sup> A Dictionary of Epidemiology, *supra* note 4, at 14; Edmond A. Murphy, The Logic of Medicine 239 – 62 (1976).

<sup>28.</sup> See infra§III.B.

<sup>29.</sup> See David L. Sackett, Bias in Analytic Research , 32 J. Chron. Dis. 51 (1979).

<sup>31.</sup> In Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 312 (5th Cir. 1989), *cert. denied*, 494 U.S. 1046 (1990), the court discussed a reanalysis of a study in which the effect was narrowed from all congenital malformations to limb reduction defects. The magnitude of the association changed by 50% when the effect was defined in this narrower fashion.

<sup>32.</sup> The effect of this conceptual bias would be to dilute or mask any real effect that the agent might have on a specific type of birth defect. Sæ Kenneth J. Rothman, Modern Epidemiology 88 (1986) ("[u]nwarranted assurances of a lack of effect can easily emerge from studies in which a wide range of etiologically unrelated outcomes are grouped.").

<sup>33.</sup> In Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1551 (D. Colo. 1990), affd, 972 F.2d 304 (10th Cir. 1992), plaintiffs attempted to rely on an excess incidence of cancers in their neighborhood to prove

garded by epidemiologists as "weak." An example of an ecological study follows.

If the researcher is interested in determining whether a high dietary fat intake is associated with breast cancer, he or she can compare different countries on the basis of their average fat intakes and their average rates of breast cancer. If a country with a high average fat intake also tends to have a high rate of breast cancer, the findings would suggest an association between dietary fat and breast cancer. However, such a finding would be far from conclusive because it lacks particularized information about an individual's exposure and disease status (i.e., whether an individual with high fat intake is more likely to have breast cancer). In addition to the lack of information about an individual's intake of fat, the researcher does not know about alternative individual exposures to other agents (or family history) that may also be responsible for the increased risk of breast cancer. This lack of particularized information about an individual's exposure to an agent and disease status detracts from the usefulness of the study and can lead to an erroneous inference about the relationship between fat intake and breast cancer, known as an ecological fallacy. However, the study is useful in that it identi fies an area for further research: the fat intake of individuals who have breast cancer as compared with the fat intake of those who do not.

Another type of group study compares disease rates over time. Secular trend studies (also called time-line studies) focus on disease rates before and after a point in time when some event of interest took place.<sup>34</sup> Thalidomide's terato-genicity (capacity to cause birth defects) was discovered after Dr. Widukind Lenz found a dramatic increase in the incidence of limb reduction birth defects in Germany beginning in 1960. Yet other than with such powerful agents as thalidomide, which increased the incidence of limb reduction defects by several orders of magnitude, time-line studies are less powerful than the studies described below. Other variables associated with the disease, such as improved diagnostic techniques and changes in lifestyle or age demographics, may change over time. If those variables can be identified and measured, it may be possible to control for them with statistical methods. Of course, unknown factors cannot be controlled for in these studies.

Observational studies, which are research designs that collect data about individuals within a group, allow the researcher to draw stronger inferences about associations between risk factors and disease. For example, in an observational study conducted in the population described above (with a high average fat intake and an increased rate of breast cancer), the researcher gathers information

causation. Unfortunately, the court confused the role of epidemiology in proving causation with the issue of plaintiffs exposure to the alleged carcinogen and never addressed the evidentiary value of plaintiffs' evidence of a disease cluster (i.e., the aggregation of a particular disease in a neighborhood or community). *Id.* at 1554.

<sup>34.</sup> In Wilson v. Merrell Dow Pharmaceuticals, Inc., 893 F.2d 1149, 1152–53 (10th Cir. 1990), defendant introduced evidence showing total sales of Bendectin and the incidence of birth defects during the 1970–1984 period. In 1983, Bendectin was removed from the market, but the rate of birth defects did not change. The Tenth Circuit affirmed the lower court's ruling that the time-line data were admissible and that defendant's expert witnesses could rely on it in rendering their opinions.

about how much dietary fat each individual consumes and whether she has breast cancer. The researcher then compares the dietary fat intake of individuals who have breast cancer with that of those who do not to determine if fat intake is associated with breast cancer.

There are two main types of observational studies: *cohort studies* and *case-con*-*trol studies*.<sup>35</sup> The difference between these two is the use of exposure or disease as the independent variable. Cohort studies, which use exposure as the independent variable, compare two groups: one group that is exposed to the agent, and a *control group* that consists of persons with similar characteristics who have not been exposed. Case-control studies compare a *case group*, those who have the disease or outcome being studied, and a control group, those who do not have the disease in question. The researcher compares the odds of having the disease when exposed to suspected agents and when not exposed. Case-control studies use disease as the independent variable.

The goal of both types of studies is to determine if there is an association between exposure to an agent and a disease and the strength (magnitude) of that association.

#### 1. Cohort studies

In cohort studies (also called prospective studies, concurrent studies, follow-up studies, incidence studies, or longitudinal studies), the researcher identifies two groups of individuals: (1) individuals who have been exposed to a substance that is thought might cause a disease and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion of each group that develops the disease is compared.<sup>36</sup> If the exposure is associated with or causes the disease, the researcher would expect a greater proportion of the exposed individuals to develop the disease (see Figure 1).

<sup>35.</sup> Case-control studies also are referred to as case history studies, case-comparison studies, and retrospective studies, because researchers gather historical information about rates of exposure to an agent in the case and control groups.

<sup>36.</sup> Sometimes retrospective cohort studies (also known as historical cohort or retrospective follow-up studies) are conducted, in which the researcher gathers historical data about exposure and disease outcome of the exposed cohort. Harold A. Kahn, An Introduction to Epidemiologic Methods 39–41 (1983). Irving Selikoff, in his seminal study of asbestotic disease in insulation workers, included several hundred workers who had died before he began his study. Selikoff was able to obtain information about exposure from union records and information about disease from hospital and autopsy records. Irving J. Selikoff et al., *The Occurrence of Asbestosis Among Insulation Workers in the United States*, 132 Annals N.Y. Acad. Sci. 139, 143 (1965).

### Figure 1 Design of a Cohort Study



An advantage of the cohort study design is that the temporal relationship between exposure and disease can be established. By tracking the exposed and unexposed groups over time, the researcher can determine the time of disease onset. This temporal relationship is relevant to the question of causation, since exposure must precede disease onset if exposure caused the disease.

As an example, in 1950 a cohort study was begun to determine whether uranium miners exposed to radon were at increased risk for lung cancer as compared with nonminers. The study group (also referred to as the exposed *cohort*) consisted of 3,400 white, underground miners. The control group comprised white nonminers from the same geographic area. Members of both groups were examined every three years, and the degree of exposure of the exposed cohort to radon was measured from samples taken in the mines. The ongoing testing of rock samples for radioactivity and the periodic medical monitoring of lungs permitted the researchers to examine whether disease was linked to prior work exposure to radiation and to discern the relationship between exposure to radiation and disease. Exposure to radiation was associated with the development of lung cancer in uranium miners.<sup>37</sup>

The cohort design is often used in occupational studies. A weakness of this design is that an increased risk of disease among the exposed group may be caused by agents other than the exposure. A cohort study of workers in a certain industry that pays below average wages might find a higher risk of cancer in those workers. This may be because they work in that industry, or it may be because low-wage groups are exposed to other harmful agents, such as environmental toxins present in higher concentrations in their neighborhood. The researcher must attempt in the study design to identify factors other than the exposure that may be responsible for the increased risk of disease. If data are gathered

<sup>37.</sup> This example is based on a study description in Abraham M. Lilienfeld & David E. Lilienfeld, Foundations of Epidemiology 237–39 (2d ed. 1980). The original study is Joseph K. Wagoner et al., *Radiation as the Cause of Lung Cancer Among Uranium Miners*, 273 New Eng. J. Med. 181 (1965).

on other possible *etiologic factors*, the researcher can sometimes use statistical methods<sup>38</sup> to assess whether a true association exists between working in the industry and cancer. Evaluating whether the association is causal involves additional analysis, as discussed in sections IV–IV.B.

2. Case-control studies

In case-control studies, the researcher begins with a group of individuals who have the disease (cases) and then selects a group of individuals who do not have the disease (controls). Instead of observing each group, as is done in a cohort study, the researcher compares past exposures. If a past exposure is associated with or caused the disease, the researcher expects to find a higher proportion of past exposure among the cases. For example, we expect a higher proportion of past cigarette smoking among lung cancer cases than among controls who do not have lung cancer (see Figure 2).

#### Figure 2

Design of a Case-Control Study



An advantage of the case-control study is that the study can be completed in less time and with less expense than a cohort study. Case-control studies also are often more powerful and therefore reveal weaker associations than cohort studies, especially when the disease or outcome is rare.<sup>39</sup>

The case-control research design poses a number of potential methodological problems. However, the researcher can prevent or diminish these problems with careful attention to the design and conduct of the study. For instance, the researcher depends on information from the past to determine exposure and disease and their temporal relationship. In some situations the researcher is required to interview the subject about past exposures, thus relying on his or her memory. Research has shown that individuals with disease (cases) may more readily recall past exposures than individuals with no disease (controls);<sup>40</sup> this

<sup>38.</sup> See Daniel L. Rubinfeld, Reference Guide on Multiple Regression § II.B, in this manual.

<sup>39.</sup> Thus, for example, to detect a doubling of disease caused by exposure to an agent where the incidence of disease is 1 in 100 in the unexposed population would require sample sizes of 3,100 each for a cohort study, but only 177 each for a case-control study. Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology 66 (1989).

<sup>40.</sup> Steven S. Coughlin, Recall Bias in Epidemiologic Studies, 43 J. Clin. Epidemiol. 87 (1990); Rothman, supra note 32, at 85.

creates a potential for biased data.

For example, consider a case-control study conducted to examine the cause of congenital malformations. The epidemiologist is interested in whether the malformation was caused by an infection during the mother's pregnancy. A group of mothers of malformed infants (cases) and a group of mothers of infants with no malformation (controls) are interviewed regarding infections during pregnancy. Mothers of children with malformations may recall an inconsequential fever or runny nose during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if the infection rate in mothers of malformed chil dren is no different from the rate in mothers of normal children, the result in this study would be an apparently higher rate of infection in the mothers of the children with the malformations solely on the basis of differential recall between the two groups. The problem of recall bias can sometimes be overcome by finding a second source of data to validate the mother's response (e.g., blood test results from prenatal visits or medical records that document symptoms of infection).<sup>41</sup> A cohort study would not be feasible, because malformations occur so rarely, and cohort studies may not be powerful enough to detect outcomes that are rare.

Selecting members of the control group (those without disease) also may be problematic in case-control studies, especially if these individuals differ in many of their characteristics from members of the case group (those with disease).<sup>42</sup> The selection of an appropriate control group has been described as the Achilles' heel of a case-control study.<sup>43</sup> One key to a valid control group is to ensure that the controls were selected independently of their exposure status, as illustrated below.<sup>44</sup>

Since many researchers are located in medical centers, they often select hospital patients as study participants. However, the selection of controls from a hospital's inpatient population can introduce selection bias into a study. For example, suppose an association is found between coffee drinking and coronary heart disease using hospital patients as a control group. However, the hospitalized control group may include individuals who had been advised against drink-

<sup>41.</sup> Two researchers who used a case-control study to examine the association between congenital heart disease and the mother's use of drugs during pregnancy corroborated interview data with the mother's medical records. See Sally Zierler & Kenneth J. Rothman, *Congenital Heart Disease in Relation to Maternal Use of Bendectin and Other Drugs in Early Pregnancy*, 313 New Eng. J. Med. 347, 347–48 (1985).

<sup>42.</sup> The types of characteristics most commonly considered when selecting the study population include age, race, socioeconomic status, years of education, and occupation.

<sup>43.</sup> William B. Kannel & Thomas R. Dawber, Coffee and Coronary Disease (editorial), 289 New Eng. J. Med. 100 (1973).

<sup>44.</sup> Another important criterion for selecting controls occurs where the cases are a sample of a given population rather than all cases within the population. In that situation, care must be taken to select controls who, if they had developed the disease, would have been included as cases in the study. Thus, if the cases consist of the leukemia patients at a hematology-oncology clinic, it is important that the control group be limited to persons who, if they had contracted leukemia, would be patients (and therefore cases) at the same clinic. For additional explanation on selecting controls in case-control studies, see Brian MacMahon & Thomas F. Pugh, Epidemiology: Principles and Methods 244–56 (1970); Rothman, *supra* note 32, at 62–68.

ing coffee for medical reasons, such as a peptic ulcer (the reason for lower consumption is not important, but the unrepresentativeness of the control group is). If this is true, the amount of coffee drinking in the control group would understate the extent of coffee drinking in the general population. Understating the exposure to coffee in the population without disease would result in inflating the impact of exposure to coffee on heart disease. This bias must be considered when the study's findings and implications are being made. Extrapolation of the findings to the general population may not be possible (or should be done cautiously), since the control group differs in an important way (exposure to the agent).<sup>45</sup>

Examining a study for potential sources of bias is an important task that helps determine the accuracy of a study's conclusions. In addition, when a source of bias is identified, it may be possible to identify whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

B. Were the Study Populations Well Defined and Samples Adequately Selected So As to Allow for Meaningful Comparisons (Between Study Groups or Between Time Periods)?

As stated above, the two main types of bias are selection bias, in which there is a systematic difference between those individuals included in the study and those who are not, and information bias, which involves error in measuring disease or exposure among those included in the study.

- 1. Did the researcher minimize the risk of selection bias?<sup>46</sup>
- a. How were the cases and controls (in case-control study) or exposed and unexposed subjects (in cohort study) identified and selected?

A list of criteria for inclusion in and exclusion from the study must be articulated by the researcher. These criteria should be documented clearly before the subjects are recruited for the study to ensure that no overt or covert biases enter into the selection process. Such biases could lead to erroneous inferences regarding causation. For example, in a *prospective study* of cervical cancer, those who are

<sup>45.</sup> Hershel Jick et al., Coffee and Myocardial Infarction, 289 New Eng. J. Med. 63 (1973).

<sup>46.</sup> Selection bias is defined as "[e]rror due to systematic differences in characteristics between those who are selected for study and those who are not. Examples include hospital cases or cases under a physician's care, excluding those who die before admission to hospital because the course of their disease is so acute, those not sick enough to require hospital care, or those excluded by distance, cost, or other factors." A Dictionary of Epidemiology, *supra* note 4, at 15.

In *Inre* "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 783 (E.D.N.Y. 1985), *aff'd*, 818 F.2d 145 (2d Cir. 1987), the court expressed concern about selection bias. The exposed cohort consisted of young, healthy males who served in Vietnam. Comparing mortality rates between that exposed cohort and a control group made up of civilians might have resulted in error due to selection bias. Failing to account for health status as an independent variable would tend to understate any association between exposure and disease where the exposed cohort is healthier.

not at risk for the disease—women who have had their cervices removed and men—should be excluded from the study population. Inclusion of such individuals as controls in a case-control study could result in erroneous findings by overstating the association between the agent and the disease. If the study population is not described precisely by the researcher, the ability to generalize the results is compromised.

b. What percentage of those selected for the study agreed to participate?

Even when a study is well designed, bias may be introduced if a large proportion of those selected as subjects refuse to participate. Many studies have shown that individuals who participate in studies differ significantly from those who do not. Consequently, if only a small proportion of selected subjects agree to participate, the findings may not apply to the general population.

If a significant portion of either study group refuses to participate in the study, the researcher should investigate reasons for refusal and whether those who refused are different from those who agreed. The researcher can show that those in the study are not a biased sample by interviewing those who refused to participate or by comparing the relevant characteristics of those who refused to participate with those who participated.

c. What proportion of the subjects dropped out of the study before it was completed?

Many of the same issues discussed above apply here. If, for example, a significant number of subjects drop out of a study before completion, it may be impossible to extrapolate the findings from a small number of subjects to the general population. The researcher should examine whether the study group is still representative of the general population.

2. Was the sample size adequate to draw a valid conclusion?

Common sense leads one to believe that researchers who do not study a large enough sample of individuals may not be able to discern the relationship between exposure to a substance and a disease. Common sense also leads one to believe that by enlarging the sample size (size of the study group), researchers can form a more accurate conclusion and reduce the chance of error in their results. Both statements are correct: researchers can increase the accuracy of the measurement of the risk of disease by enlarging the sample size. This commonsense intuition is illustrated by a test to determine if a two-sided coin is fair. A test in which the coin is flipped 500 times is much more helpful than a test in which the coin is flipped 10 times. Both common sense and statistics reveal that it is far more likely that 80% of the flips in the latter test will result in heads than in the former test if the coin is fair. The concern with the design of the coin test, as with epidemiology, is that both involve sampling techniques to draw an inference. Estimates based on samples are subject to random error, which can be reduced by increasing the size of the samples.

Sample size calculations are important in two circumstances. The first circumstance occurs during the planning of a study, when the researcher estimates the size and expense of the study. At this point, the researcher must determine the number of subjects that will have to participate to obtain research findings of acceptable precision. Since enlarging the study size increases the time, cost, and complexity of conducting the study, a balance must be maintained between the scientific precision of the findings and the cost of the project.

The second circumstance occurs after a study has been completed. The outcome of a study can be incorrect or inaccurate because of sampling error. Thus, a study erroneously can find no association between exposure to an agent and a disease if the sample size is too small to detect the association that existed.<sup>47</sup> If the researchers suspect that this is the case after a study has been completed, they can determine the likelihood that the size of the sample (i.e., the number of participants) was sufficient to permit detection of an association of a given magnitude, if there actually was one. Similarly, a study can find an association that is spurious—the result of random error. This is similar to the example mentioned above, in which a fair coin flipped ten times results in eight heads. Statistical techniques can be used to estimate the likelihood that the association is due to sampling error.<sup>48</sup>

Researchers use a variety of approaches to determine an appropriate size for a study population, including a sample size calculation and a power calculation. Although these calculations generally are completed before the study begins, they require an estimation of the study's findings. In general, the calculations help determine whether a study is feasible (i.e., whether the researcher can recruit enough subjects and finance the project adequately). It should be recognized that sample size calculations are based on public health considerations and costs, which may not coincide with the level of precision that would be optimal for legal standards of proof.

a. Sample size calculation

The sample size calculation provides researchers with an estimate of the number of individuals they should study to detect whether exposure to a substance increases the risk of disease. The calculation is based on four factors:

1. the specified level of statistical significance, or alpha, that is desired;<sup>49</sup>

<sup>47.</sup> Junius C. McElveen, Jr. & Pamela S. Eddy, *Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology*, 33 Clev. St. L. Rev. 29, 40–41 (1984) (detecting an increase of 200 cancers per population of 100,000, with a significance level of .05, where the background rate of cancer is 20,000 per 100,000 individuals, would require exposed and control cohorts of 700,000 persons each).

<sup>48.</sup> See infra§III.C.1.

<sup>49.</sup> The specified level of statistical significance, also called alpha or type I error, is the probability of observing an association of the magnitude found in the study or greater when there is no association (i.e., false positive). See *infra* § III.C.1.

- 2. the chance of missing a real effect, or beta, that the researcher selects; <sup>50</sup>
- 3. the estimated magnitude of the increased risk of disease, or *effect size*;<sup>51</sup> and
- 4. the background risk of disease or exposure.<sup>52</sup>

Factors (1) and (2) are set by convention and generally do not change from study to study. Changes in these values should be based on logical and defensible scientific needs. For example, it may be important to increase beta or alpha to examine carefully the effect of a specific risk factor.

Factors (3) and (4) are critical in determining an adequate sample size. For example, in a study of exposure to video display terminals (VDTs) and spontaneous abortion, the researcher must consider two questions. First, what is the background rate of spontaneous abortion in the group? Second, how many excess spontaneous abortions are thought to be related to VDT exposure? In general, when there is a high background rate of a particular disease and when the increased risk of disease is small, the researcher needs a larger sample size. In the example given above, the researcher would need a fairly large sample size, since the background rate of spontaneous abortions is high and other studies suggest that the risk associated with VDT exposure is small.

Since the sample size calculation generally is performed before the study is initiated, the researcher often estimates values for the increased risk of disease and the standard deviation. <sup>53</sup> Researchers can rely on values from similar studies conducted, or if no such studies are available, they can rely on educated guesswork. When little is known about a particular disease, researchers can calculate the sample size and then increase it to allow for the uncertainty.

This method of calculating sample size has been criticized for being subject to manipulation, because the researcher must estimate values that will not be known until the study has been completed. Nonetheless, it still is used for determining the size of a study population.

#### b. Power calculation

The results of a power calculation, often displayed as a diagram, present the probability that a researcher will be able to find a hypothetical increased risk of disease for specified sample sizes. After reviewing the diagram, a researcher

<sup>50.</sup> The chance of missing a real effect, or beta, is the probability that an association that exists will be missed by the study. See infra § III.C.2.

<sup>51.</sup> The magnitude of the increased risk in disease, or effect size, is best thought of as the amount of disease that is caused by exposure to a toxic substance. For example, the risk of contracting lung cancer may be ten times higher for cigarette smokers than for nonsmokers. The magnitude of the increased risk is therefore a factor of 10. See infra § III.A.

<sup>52.</sup> The background rate, or background risk, in a population is the amount of disease that occurs in individuals who have no known exposures to an alleged risk factor for the disease. For example, the background rate for all birth defects is 3%-5% of live births.

<sup>53</sup>. Standard deviation is a summary statistic that describes how widely dispersed the data are around the mean (average) value.
chooses the appropriate sample size. Figure 3 shows sample power curves, with *power* plotted against sample size for several anticipated levels of relative risk (RR, a measure of association discussed below). Many statistical computer programs used in epidemiology can generate power curves.





The power curves in Figure 3 are drawn on the assumption that the researcher will conduct a case-control study that will have an equal number of cases (those with disease) and controls (those without disease). The estimated exposure rate in the control group is 0.3, which means that 30% of the controls are predicted to have been exposed to the agent. In addition, the curves are calculated and drawn based on a level of statistical significance of 0.1. The y-axis displays the power, or probability (with 1.0 being 100%), of the study being able to detect an association of the magnitude shown for each of the curves based on the sample size on the x-axis. Thus, if the researcher wants to detect a relative risk of at least 3.0 with 80% probability, there should be approximately 75 subjects in each of the case and control groups.

After a study is completed, a power curve is used to determine the likelihood that the study would have detected an association of a given magnitude. Thus, a case-control study with 100 subjects in each group has a slightly better than 60% probability (determined from reading the curve) of detecting a relative risk of 2.0

Reference Manual on Scientific Evidence

that will be statistically significant. Put another way, there is a 40% chance that the study failed to detect a relative risk of up to 2.0 at a statistical significance level of 0.1.

Power calculations performed in advance of a study have also been criticized as being subject to manipulation. They can be manipulated because the researcher must estimate study variables in advance of the study. Nevertheless, power calculations are becoming a more common method for determining study size, in part because a power diagram provides more information than a sample size calculation. <sup>54</sup> This criticism is not applicable to power calculations performed after a study is completed, because at that time, exposure rates (for case-control studies) and disease incidence (for cohort studies) are known.

Both the power calculation and the sample size calculation require some estimation or educated guesswork by the researcher and are subject to uncertainty. Although there is no absolute right number of participants, the researcher usually uses one of these two approaches to determine the minimum number of participants needed. The assumptions that underlie a researcher's estimations can be examined for soundness. The researcher should be able to articulate a reasonable and scientific basis for estimates of the magnitude of the increased risk of disease and the background risk. If the study size is too small, or if the actual risk from exposure is less than the researcher estimated, the study may be inconclusive.

## C. Was Exposure to the Putative Agent Measured Using a Standardized and Reliable Methodology?

One of the most difficult areas in epidemiology concerns exposure: determining whether a person was exposed to an agent in the past and, if so, measuring the intensity and length of such an exposure.<sup>55</sup> Exposure can be measured directly or indirectly.<sup>56</sup> Sometimes researchers use a *biological marker* as a direct measure of exposure—an alteration in tissue or body fluids that occurs as a result of

<sup>54.</sup> Rothman, supra note 32, at 79-81.

<sup>55.</sup> Dose generally refers to the intensity or magnitude of exposure multiplied by the time exposed. For a discussion of the difficulties of determining dosage from atomic fallout, see Allen v. United States, 588 F. Supp. 247, 425–26 (D. Utah 1984), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987), cert. denied, 484 U.S. 1004 (1988).

A different, but related, problem often arises in court. Determining plaintiff's exposure to the alleged toxic substance always involves a retrospective determination and may involve difficulties similar to those faced by an epidemiologist planning a study. Thus, in Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1113 (5th Cir. 1991), *cert. denied*, 112 S. Ct. 1280 (1992), the court criticized the plaintiff's expert who relied on an affidavit of a co-worker to determine the dose of nickel and cadmium to which the decedent had been exposed.

In asbestos litigation, a number of courts have adopted a requirement that plaintiff demonstrate (1) regular use by an employer of defendant's asbestos-containing product; (2) plaintiff's proximity to those products; and (3) exposure over an extended period of time. *See, e.g.*, Lohrmann v. Pittsburgh Corning Corp., 782 F.2d 1156, 1162–64 (4th Cir. 1986).

<sup>56.</sup> See In re Paoli R.R. Yard PCB Litig., No. 86-2229, 1992 U.S. Dist LEXIS 18430, at \*9-11 (E.D. Pa. Oct. 21, 1992) (discussing valid methods of determining exposure to chemicals).

an exposure and that can be detected in the laboratory. Biological markers are only available for a small number of toxins and only reveal whether or not a person was exposed. Biological markers rarely help determine the intensity or duration of exposure.<sup>57</sup>

Monitoring devices also can be used to measure exposure directly but often are not available for exposures that occurred in the past. For past exposures, epidemiologists often use indirect means of measuring exposure, such as interviewing workers and reviewing employment records. Thus, all those employed installing asbestos insulation may be treated as having been exposed to asbestos during the period that they were employed. However, there may be a wide variation of exposure within any job, and these measures may have limited applicability to a given individual. Where the agent of interest is a drug, medical or hospital records can be used to determine exposure. Thus, retrospective occupational or environmental measurements of exposure are usually less accurate than prospective or follow-up studies, especially ones where drugs or medical intervention is the independent variable being measured.

The route (e.g., inhalation or absorption), duration, and intensity of exposure are important factors in assessing disease causation. Even with environmental monitoring, the dose measured in the environment generally is not the same as the dose that reaches internal target organs. If the researcher has calculated the internal dose of exposure, the scientific basis for this calculation should be examined for soundness.<sup>58</sup>

#### 1. Were data collected from objective and reliable sources?

Medical records, government documents, employment records, death certificates, and interviews are examples of data sources that are used by epidemiologists.<sup>59</sup> The accuracy of a particular source may affect the validity of a research finding. If different data sources are used to collect information about a study group, differences in the accuracy of those sources may affect the validity of the findings.

<sup>57.</sup> The timing of exposure may also be critical, especially where the disease of interest is birth defects. In Smith v. Ortho Pharmaceutical Corp., 770 F. Supp. 1561, 1577 (N.D. Ga. 1991), the court criticized a study for its inadequate measure of exposure to spermicides. The researchers had defined exposure as receipt of a prescription for spermicide within 600 days of delivery: This definition of exposure is too broad because environmental agents are only likely to cause birth defects during a narrow band of time.

<sup>58.</sup> See also Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology § I.C, in this manual.

<sup>59.</sup> Even these sources may produce unanticipated error. Identifying the causal connection between asbestos and mesothelioma, a rare form of cancer, was complicated and delayed because doctors who were unfamiliar with mesothelioma erroneously identified other causes of death in death certificates. Sæ David E. Lilienfeld & Paul D. Gunderson, The "Missing Cases" of Pleural Malignant Mesothelioma in Minnesota, 1979–81: Preliminary Report, 101 Pub. Health Rep. 395, 397–98 (1986).

For example, using employment records to gather information about exposure to narcotics probably would lead to inaccurate results, since employees tend to keep such information private. If the researcher uses an unreliable source of data, the study may not be useful to the court.

2. What types of procedures were instituted to control the quality of measurements of exposure?

The types of quality control procedures used depend on the source of data. For data collected by interview, quality control procedures should probe the reliability of the individual and whether the information is verified by other sources. For data collected and analyzed in the laboratory, quality control procedures should probe the reliability of the laboratory test.

3. Was information obtained from one group of the study population more accurate or complete than that obtained from the comparison group?

Error can be introduced into a study if there are differences in the accuracy or completeness of the subjects' recollection of past events or experiences. This type of bias, known as recall bias, is a special concern in case-control studies.

For example, a researcher may be interested in whether fetal malformation is caused by a mother's exposure to a virus during pregnancy. A group of mothers of malformed infants (cases) and a group of mothers of infants with no malformation (controls) are interviewed regarding infections during pregnancy. Mothers of the malformed infants may tend to recall inconsequential fevers or runny noses during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if the true viral infection rate in mothers of malformed infants is no different from the rate in mothers of normal infants, the results of this study would indicate a false association between infection during pregnancy and birth defects because of differential recall between the two groups.<sup>60</sup>

#### 4. Did the method of collecting data yield reliable information?

Errors in data collection can compromise the validity of the research findings. Evidence of staff training and data collection guidelines may be available for review by opposing experts. If the data were coded by members of the research team, reliability checks and coefficients may be reported on the coding and data

<sup>60.</sup> See Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 311–12 (5th Cir. 1989) (discussion of recall bias among women who bear children with birth defects), *cert. denied*, 494 U.S. 1046 (1990). It should be noted that the court was mistaken in its assertion that a confidence interval could correct for recall bias, or for any bias for that matter. Confidence intervals are a statistical device for analyzing error that may result from random sampling. Systematic errors (bias) in the design or data collection are not addressed by statistical methods, such as confidence intervals or statistical significance. See Green, supra note 23, at 667–68; Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases*, 12 Risk Analysis 343, 344–45 (1992); *infra* § III.B.

entry processes. Further, if the data were collected through interview protocol, survey form, or code sheet, a pilot test may have been conducted using the data collection instrument. The results of the pilot test would indicate whether the data collection instruments (questionnaires, forms, etc.) posed problems for data collection staff.

### D. Were the Health Effects (i.e., Disease, Disability) Clearly Defined and Reliably Measured?

The outcome or health effects being studied should be clearly defined by the researcher in the study design. Precise definition of the disease ensures that the same variable is consistently measured throughout the study. For example, if a researcher is studying birth defects, it is necessary to define the age at which defects will be measured, as some birth defects are not apparent at birth and only are diagnosed later in childhood.

The quality and sophistication of the diagnostic methods used to detect a disease should be assessed. The proportion of subjects who were examined also should be questioned. If, for example, many of the subjects refused to be tested, the fact that the test used was of high quality would be of relatively little value.

The scientific validity of the research findings is influenced by the reliability of the diagnosis of disease.<sup>61</sup> For example, a researcher interested in studying spontaneous abortion in the first trimester needs to test women for pregnancy. Diagnostic criteria that are accepted by the medical community should be used to make the diagnosis. If a diagnosis is made using an unreliable home pregnancy kit known to have a high rate of false positives (indicating pregnancy when the woman is not pregnant), the study will overestimate the number of spontaneous abortions.

<sup>61.</sup> In *Inre* Swine Flu Immunization Prod. Liab. Litig., 508 F. Supp. 897, 903 (D. Colo. 1981), *aff'd sub nom*. Lima v. United States, 708 F.2d 502 (10th Cir. 1983), the court critically evaluated a study relied on by an expert whose testimony was stricken. In that study, determination of whether a patient had Guillain-Barré syndrome was made by medical clerks, not physicians who were familiar with diagnostic criteria.

### **III.** Association Between Exposure and the Disease

Exposure to an agent and disease are said to be associated when they occur more frequently together than one would expect by chance.<sup>62</sup> The term association implies a range of possible relationships, but it does not necessarily imply a cause-effect relationship between exposure and disease. Of course, a causal relationship is one possible explanation for the association, which is of ultimate concern to epidemiologists.

This section begins with a description of the epidemiological methods for expressing the strength of an association between exposure and disease. It goes on to review ways in which an incorrect result can be produced and then examines statistical methods for evaluating whether an association is real or due to sampling error.

# A. What Is the Basis for Concluding That the Exposure Is Associated with an Increased Risk of Disease?

The strength of an association between exposure and disease can be stated as a relative risk (RR), odds ratio (OR), or attributable proportion of risk (APR). Each of these measurements of association examines the degree to which the risk of disease increases when individuals are exposed to an agent.

#### 1. Relative risk (RR)

A commonly used approach for expressing the association between an agent and disease is relative risk. It is defined as the ratio of the incidence of disease in exposed individuals compared to the incidence in unexposed individuals.<sup>63</sup> Thus, it can be expressed algebraically as:

$$RR = \frac{I_{e}}{I_{c}}$$

<sup>62.</sup> A negative association may imply that the agent has a protective or curative effect. Because the concern in toxic substances litigation is whether an agent has caused disease, this reference guide focuses on positive associations.

<sup>63.</sup> This definition of relative risk assumes that the researcher is conducting a cohort study (examining the risk of disease in an exposed and an unexposed population). In a case-control study, the equivalent of the relative risk, the odds ratio, compares the odds of having disease when exposed to a suspected agent and when not exposed.

In the formula above, *RR* is the relative risk,  $I_e$  is the incidence of disease in the exposed population, and  $I_c$  is the incidence of disease in the control population.

For example, a researcher studies 100 individuals who are exposed to an agent and 100 who are not exposed. After several years, 40 of the exposed individuals are diagnosed as having a disease, and 10 of the unexposed individuals also are diagnosed as having disease. The relative risk of contracting the disease is calculated as follows:

- The incidence of disease in the exposed individuals is 40 cases per 100 persons (40/100), or 0.4.
- The incidence of disease in the unexposed individuals is 10 cases per 100 persons (10/100), or 0.1.
- The relative risk is calculated as the incidence in the exposed group (0.4) divided by the incidence in the unexposed group (0.1), or 4.0.

A relative risk of 4.0 indicates that the risk of disease in the exposed group is four times higher than the risk of disease in the unexposed group. <sup>64</sup>

In general, the relative risk can be interpreted as follows:

- If the relative risk equals 1.0, the risk in exposed individuals is the same as the risk in unexposed individuals. There is no association between exposure to the agent and disease.
- If the relative risk is greater than 1.0, the risk in exposed individuals is greater than the risk in unexposed individuals. There is a positive association between the agent and the disease.
- If the relative risk is less than 1.0, the risk in exposed individuals is less than the risk in unexposed individuals. There is a negative association, which could reflect a protective or curative effect of the agent on risk of disease. For example, immunizations lower the risk of disease. The results suggest that immunization is associated with a decrease in disease and may have a protective effect on the risk of disease.

Researchers should scrutinize their results for error. Error in the design of a

64. The court in Gaul v. United States, 582 F. Supp. 1122, 1125 n.9 (D. Del. 1984), defined relative risk as follows:

Relative risk, or relative risk ratio, describes the relationship between the risk of an occurrence, such as contracting a disease, in a population exposed to a certain stimulus, and the risk of the occurrence in a population not exposed to the stimulus. It is the ratio of the former risk to the latter. It is another way of explaining how much more likely a person exposed to the stimulus is to get a disease than an unexposed person. For example, using hypothetical numbers and facts, if one in every 100,000 vegetarians contracts stomach cancer while five in every 100,000 meat eaters contract this disease, the relative risk of contracting cancer among meat eaters would be 5/1, or 5. In other words, the risk of getting stomach cancer would be five times greater for meat eaters than vegetarians, assuming all other factors are held constant.

See also DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 947 (3d Cir. 1990) ("[i]n the context of an epidemiological study of Bendectin's relationship to birth defects, the relative risk is the ratio of the incidence rate of birth defects in the study group exposed to Bendectin divided by the rate in the control group not exposed to Bendectin.").

study could yield an incorrect relative risk. Sources of bias should be examined. Whenever a positive association is uncovered, further analysis should be conducted to determine if the association is causal.<sup>65</sup>

#### 2. Odds ratio (OR)

The odds ratio is similar to a risk ratio. An odds ratio is used for case-control studies and is based on a comparison of the odds of having a disease when exposed to a suspected agent and when not exposed. For all practical purposes, the odds ratio is comparable to the relative risk when the disease is rare. <sup>66</sup> However, as the disease becomes more common, these measures diverge.

The odds ratio is expressed algebraically as:

$$OR = \frac{D_{\rm e} \times C_{\rm u}}{D_{\rm u} \times C_{\rm e}}$$

In the formula above, OR is the odds ratio,  $D_e$  is the number of cases (those with the disease) who were exposed to the agent,  $C_u$  is the number of controls (those without the disease) who were not exposed to the agent,  $D_u$  is the number of cases who were not exposed to the agent, and  $C_e$  is the number of controls who were exposed.

Consider the following hypothetical study: A researcher finds 10 individuals with a disease. Four of those individuals were exposed to the agent and 6 were not. The control group consists of 100 persons, none of whom have the disease "by definition." Among the control group, 20 have been exposed and 80 have not. The calculation of the odds ratio would be:

$$OR = \frac{4 \times 80}{6 \times 20} = 2.67$$

If the disease is relatively rare in the general population (about 5% or less), the odds ratio is close to a relative risk of 2.67, which means that there is almost a tripling of the disease in those exposed to the agent.

#### 3. Attributable proportion of risk (APR)

Perhaps the most useful measurement of risk, the attributable proportion of risk (also called etiologic fraction and attributable risk percent) represents the proportion of the disease among exposed individuals that is associated with the exposure. The attributable proportion reflects the maximal amount of the disease that could be prevented by blocking the effect of the exposure or by eliminating the exposure.<sup>67</sup> In other words, if the association is causal, the attributable proportion of risk is the amount of disease in an exposed population caused by

<sup>65.</sup> See infra §§ IV-IV.B.

<sup>66.</sup> For further detail about the odds ratio and its calculation, see Kahn & Sempos, *supra* note 39, at 47–56. 67. Rothman, *supra* note 32, at 38–39. *See also* Landrigan v. Celotex Corp., 605 A.2d 1079, 1086 (N.J.

<sup>1992) (</sup>illustrating that relative risk of 1.55 conforms to attributable risk of 35%).

the agent (see Figure 4).

To determine the proportion of a disease that is attributable to an exposure, a researcher would need to know the incidence of the disease in the exposed group and the incidence of disease in the nonexposed group. With that information, the attributable proportion of risk can be stated algebraically as:

$$APR = \frac{I_e - I_c}{I_e}$$

In the above formula, APR is the attributable proportion of risk,  $I_e$  is the incidence of disease in the exposed group, and  $I_c$  is the incidence of disease in the control group.

#### Figure 4 Risks in Exposed and Not Exposed Groups



The attributable proportion of risk can be calculated using the example described in section III.A.1. Suppose a researcher studies 100 individuals who are exposed to a substance and 100 who are not exposed. After several years, 40 of the exposed individuals are diagnosed as having a disease, and 10 of the unexposed individuals are also diagnosed as having a disease.

- The incidence of disease in the exposed group is 40 persons in 100.
- The incidence of disease in the unexposed group is 10 persons in 100.
- The maximum proportion of disease that is attributable to the exposure is 30 persons out of 40, or 75%.

This means that up to 75% of the disease in the exposed group is attributable to the exposure.

#### B. What Categories of Error Might Have Produced a False Result?

Incorrect study results occur in a variety of ways. A study may find a positive association (relative risk greater than 1.0) when there is no association. Or a study erroneously may conclude that there is no association. Finally, a study may find an association when one truly exists, but the association found may be greater or less than the real association.

Two categories of error in an epidemiological study can produce these incor-

rect results. <sup>68</sup> The first, known as sampling error, occurs because all epidemiological studies are based on sampling a small proportion of the relevant population. As stated in section II.B.2, the size of the sample can be adjusted to reduce (but not eliminate) the likelihood of sampling error. Statistical techniques<sup>69</sup> permit an assessment of the plausibility that the results of a study represent a true association or random error.

Systematic error or bias also can produce error in the outcome of a study. Many of the potential sources of bias were described in section II in connection with the planning or conduct of an epidemiological study. However, even the best designed and conducted studies still can have biases. Thus, after a study is completed (and this is the time when most lawyers and judges confront an epidemiological study), it should be evaluated for potential sources of bias. Sometimes, after bias is identified, the epidemiologist can determine whether the bias would tend to inflate or dilute any association that may exist. Identification of the bias may enable an assessment of whether the study's results should be adjusted, and if so, the direction of such an adjustment and the range of error that is indicated. Sometimes, epidemiologists conduct reanalyses of a study's underlying data to correct for a bias identified in a completed study.<sup>70</sup>

# C. What Statistical Methods Exist to Evaluate the Likelihood That the Result of an Epidemiological Study Was Due to Random Sampling Error?<sup>n</sup>

Before detailing the statistical methods used to assess random error, two concepts are explained that are central to epidemiology and statistical analysis. Understanding these concepts should facilitate comprehension of the statistical methods.

Epidemiologists often refer to the *true association* (also called real association), which is the association that really exists between agent and exposure and that might be found by a perfect (but nonetheless nonexistent) study. The true association is a concept that is used in evaluating the results of a given study.

Epidemiologists begin each study with a hypothesis that they seek to dis-

<sup>68.</sup> In DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 955 (3d Cir. 1990), the court recognized and discussed random sampling error. It then went on to refer to other errors (i.e., systematic bias) that create as much or more error in the outcome of a study. For a similar description of error in study procedure and random sampling, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV, in this manual.

<sup>69.</sup> See infra § III.C.

<sup>70.</sup> E.g., Richard A. Kronmal et al., *The Intrauterine Device and Pelvic Inflammatory Disease: The Women's Health Study Reanalyzed*, 44 J. Clin. Epidemiol. 109 (1991) (reanalysis of a study that found an association between use of IUDs and pelvic inflammatory disease concluded that IUDs do not increase the risk of pelvic inflammatory disease).

<sup>71.</sup> For a bibliography on the role of statistical significance in legal proceedings, see Sanders, *supra* note 15, at 329 n.138.

prove.<sup>72</sup> The hypothesis most often used is called the *null hypothesis*, which posits that there is no true association between agent and exposure; thus, the epidemiologist begins by assuming that the relative risk is 1.0 and seeks to develop data that disprove the hypothesis.<sup>73</sup>

#### 1. False positive error and statistical significance

When a study results in a positive association (i.e., a relative risk greater than 1.0), epidemiologists try to determine whether that outcome represents a true association or whether it is the result of random error. Random error is similar to the error that occurs when a fair coin yields five heads out of five tosses.<sup>74</sup> Thus, even though the true association is a relative risk of 1.0, an epidemiological study may find a positive association because of random error. An erroneous conclusion that the null hypothesis is false (due to random error) is a *false positive* (also, *alpha error* or type *l error*).

The essential concern is with the numerical stability of the sampling conducted by the epidemiologist. A researcher who compares two coins and finds a 50% incidence of heads in one coin and a 75% incidence of heads in the second might conclude that the second coin is biased and the first is fair. However, if each test consists of only four flips, the results are highly unstable, because if the next flip for each coin results in a tail, each one will have resulted in a 60% incidence of a head or a tail. Nothing, then, could be said about which coin is a biased one. If the test is conducted with larger numbers (1,000 flips each), the stability of the outcome is less likely to be influenced by random error, and the researcher would have greater confidence in the inferences drawn from data that found 75% heads in one coin and 50% in the other.<sup>75</sup>

One means for evaluating the possibility that an effect is due to random error is by calculating a *p-value*. <sup>76</sup> A *p*-value represents the probability that a positive association like that found would result due to random error if no association is in fact present. <sup>77</sup> Thus, a *p*-value of .1 means that there is a 10% chance that an effect at least as large as that found is due solely to random error.<sup>78</sup>

76. Sæ also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B, in this manual ( p value reflects the implausibility of the null hypothesis).

77. Technically, a *p*-value represents the probability that the study's association or a larger one would occur due to sampling error where no association (or, equivalently, the null hypothesis) is the true situation.

78. Technically, a p-value of .1 means that 10% of all similar studies would be expected to yield the same

<sup>72.</sup> Sæ, e.g., Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993) (scientific methodology involves generating and testing hypotheses).

<sup>73.</sup> See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 945 (3d Cir. 1990).

<sup>74.</sup> See id. at 946-47.

<sup>75.</sup> This explanation of numerical stability was drawn from Brief Amicus Curiae of Professor Alvan R. Feinstein in Support of Respondent at 12–13, Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993) (No. 92-102). Sæ also Allen v. United States, 588 F. Supp. 247, 417–18 (D. Utah 1984), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987), cert. denied, 484 U.S. 1004 (1988). The Allen court observed that while "[s]mall communities or groups of people are deemed 'statistically unstable'" and "data from small populations must be handled with care does not mean that it cannot provide substantial evidence in aid of our effort to describe and understand events."

To avoid false positive error, epidemiologists use a convention that the *p*-value must fall below some selected level known as alpha for the results of the study to be considered statistically significant.<sup>79</sup> The most common level of statistical significance or alpha used is .05. Using this value for significance testing accepts as statistically significant those studies where a positive association erroneously is found (because the true situation is that there is no association) no more than 5 times out of 100 due to random error.<sup>80</sup> Although .05 is often the level of alpha selected, other defensible levels can and have been used legitimately. Thus, in its study of the effects of secondhand smoke, the Environmental Protection Agency (EPA) used a .10 standard for statistical significance.<sup>81</sup>

There is some controversy among epidemiologists about the appropriate role of significance testing.<sup>82</sup> To the strictest significance testers, any study whose *p*-

or larger association due solely to random error where, in fact, there is no association.

This claim is incorrect, although the reasons are a bit complex. First, alpha does not address the likelihood that plaintiff's disease was caused by exposure to the agent; the magnitude of the association bears on that question. See infra § V. Second, significance testing only bears on whether the magnitude of association found was due to random chance, not on whether the null hypothesis is true. Third, using stringent significance testing to avoid false positive error comes at a complementary cost of inducing false negative error. See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 947 (3d Cir. 1990). Fourth, using an alpha of 0.5 would not be equivalent to saying that the probability the association found is real is 50%, and the probability it is a result of random error is 50%. Statistical methodology does not permit assessments of those probabilities. See Green, supra note 23, at 686; see also David H. Kaye, Apples and Oranges: Confidence Coefficients and the Burden of Persuasion, 73 Cornell L. Rev. 54, 66 (1987); David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.2, in this manual; Allen v. United States, 588 F. Supp. 247, 417 (D. Utah 1984) ("whether a correlation between a cause and group of effects is more likely than not, particularly in a legal sense, is a different question from that answered by tests of statistical significance"), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987), cert. denied, 484 U.S. 1004 (1988); Turpin v. Merrell Dow Pharmaceuticals, Inc., 959 F.2d 1349, 1357 n.2 (6th Cir.), cert. denied , 113 S. Ct. 84 (1992); cf. DeLuca, 911 F.2d at 959 n.24 ("The relationship between confidence levels and the more likely than not standard of proof is a very complex one . . . and in the absence of more education than can be found in this record, we decline to comment further on it.").

81. U.S. Envtl. Protection Agency, Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders (1992); *et also Turpin*, 959 F.2d at 1353–54 n.1 (alpha frequently set at 95%, though 85% and 90% are also used; selection of the value is "somewhat arbitrary").

82. Similar controversy exists among the courts that have confronted the issue of whether statistically significant studies are required to satisfy the burden of production. The leading case advocating statistically significant studies is Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 312 (5th Cir.), amended, 884 F.2d 167, 167 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990). Overturning a jury verdict for the plaintiff in a Bendectin case, the court observed that no statistically significant study had been published that found an increased relative risk for birth defects in children whose mothers had taken Bendectin. The court concluded:

[W]e do not wish this case to stand as a bar to future Bendectin cases in the event that new and statistically significant studies emerge which would give a jury a firmer basis on

which to determine the issue of causation. 884 F.2d at 167.

A number of courts have followed the Brock decision or have indicated strong support for significance testing

Epidemiology

<sup>79.</sup> Allen, 588 F. Supp at 416–17 (discussing statistical significance and selection of a level of alpha); see also Sanders, supra note 15, at 343–44 (explaining alpha, beta, and their relationship to sample size).

<sup>80.</sup> A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof. Thus, one will often see a statement that using an alpha of .05 for statistical significance imposes a burden of proof on plaintiff far higher than the civil burden of a preponderance of the evidence (i.e., greater than 50%). See, e.g., Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 28 n.58 (D.C. Cir.), cert. denied, 426 U.S. 941 (1976); Hodges v. Secretary Dep't Health & Human Servs., No. 92-5089, 1993 U.S. App. LEXIS 29590, at \*29, 41 (Fed. Cir. Nov. 15, 1993) (Newman, J., dissenting); Edward J. Imwinkelried, The Admissibility of Expert Testimony in Christophersen v. Allied-Signal Corp.: The Neglected Issue of the Validity of Nonscientific Reasoning by Scientific Witnesses , 70 Denv. U. L. Rev. 473, 478 (1993).

value does not exceed the level chosen for statistical significance should be rejected as inadequate to disprove the null hypothesis. In the past, authors of a study simply would report whether or not the results were statistically significant, without providing any information about the p-value.<sup>83</sup>

For others, a statistical device known as a *confidence interval* permits a more refined assessment of appropriate inferences about the association found in an epidemiological study.<sup>84</sup> The advantage of a confidence interval is that it displays more information than a *p*-value. What a *p*-value does not provide is the magnitude of the association found in the study or an indication of how numerically stable that association is. A confidence interval for any study shows the relative risk determined in the study as a point on an axis. It also displays the boundaries of relative risk consistent with the data found in the study based on one or several selected levels of alpha or statistical significance. A sample confidence interval is displayed in Figure 5. The confidence interval represents a study that found a relative risk of 1.5, with boundaries of .8 to 3.4 for alpha equal to .05 and boundaries of 1.1 to 2.2 for alpha equal to .1. Because the boundaries of the confidence limits with alpha set at .05 encompass a relative risk of 1.0, the study

Although the trial court had relied in part on the absence of statistically significant epidemiological studies, the Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993), did not explicitly address the matter. The Court did, however, in identifying factors relevant to the scientific validity of an expert's methodology, refer to "the known or potential rate of error." *Id.* at 2797. The Court did not address any specific rate of error, although two cases that it cited affirmed the admissibility of voice spectrograph results that the courts reported were subject to a 2%–6% chance of error due to either false matches or false eliminations.

83. Epidemiological studies have become increasingly more statistically sophisticated in their treatment of random error. See Sanders, *supra* note 15, at 342 (describing the improved handling and reporting of statistical analysis in studies of Bendectin after 1980).

84. Kenneth Rothman, Professor of Public Health at Boston University and Adjunct Professor of Epidemiology at the Harvard School of Public Health, is one of the leaders in advocating use of confidence intervals and rejecting strict significance testing. In *DeLuca*, 911 F.2d at 947, the Third Circuit discussed Rothman's views on the appropriate level of alpha and the use of confidence intervals. In *Turpin*, 959 F.2d at 1353–54 n.1, the court discussed the relationship among confidence intervals, alpha, and power. The use of confidence intervals in evaluating sampling error more generally than in the epidemiological context is discussed in David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

as a screening device. See Renaud v. Martin Marietta Corp., 749 F. Supp. 1545, 1555 (D. Colo. 1990) (quoting *Brock* approvingly), *affd*, 972 F.2d 304 (10th Cir. 1992); Thomas v. Hoffman-LaRoche, Inc., 731 F. Supp. 224, 228 (N.D. Miss. 1989) (granting judgment n.o.v. and observing that "there is a total absence of any statistically significant study to assist the jury in its determination of the issue of causation"), *affd on other grounds*, 949 F.2d 806 (5th Cir.), *cert. denied*, 112 S. Ct. 2304 (1992); Daubert v. Merrell Dow Pharmaceuticals, Inc., 727 F. Supp. 570, 575 (S.D. Cal. 1989), *affd on other grounds*, 951 F.2d 1128 (9th Cir. 1991), *vacated*, 113 S. Ct. 2786 (1993).

By contrast, a number of courts appear more cautious about using significance testing as a necessary condition, instead recognizing that assessing the likelihood of random error is important in determining the probative value of a study. In *Allen*, 588 F. Supp. at 417, the court stated: "The cold statement that a given relationship is not 'statistically significant' cannot be read to mean there is no probability of a relationship." The Third Circuit described confidence intervals (i.e., the range of values within which the true value is thought to lie, with a specified level of confidence) and their use as an alternative to statistical significance in *DeLuca*, 911 F.2d at 948–49. *See also Turpin*, 959 F.2d at 1357 ("The defendant's claim overstates the persuasive power of these statistical studies. An analysis of this evidence demonstrates that it is possible that Bendectin causes birth defects even though these studies do not detect a significant association."); *In re* Bendectin Prod. Liab. Litig., 732 F. Supp. 744, 748–49 (E.D. Mich. 1990) (rejecting defendant's claim that plaintiff could not prevail without statistically significant epidemiological evidence).

would not be statistically significant at that level. By contrast, since the confidence boundaries for alpha equal to .1 do not overlap with a relative risk of 1.0, the study does have a positive finding that is statistically significant at that level of alpha. The larger the sample size in a study (all other things being equal), the narrower the confidence boundaries will be (indicating greater numerical stability), reflecting the decreased likelihood that the association found in the study would occur if the true association is 1.0.<sup>85</sup>

#### Figure 5 Confidence Intervals



2. False negative error

False positives can be reduced by adopting more stringent values for alpha. Using a level of .01 or .001 will result in fewer false positives than with alpha at .05. The trade-off for reducing false positives is an increase in *false negatives* (also, *beta error* or *type II error*). This concept reflects the possibility that a study will be interpreted not to disprove the null hypothesis when in fact there is a true association of a specified magnitude.<sup>86</sup> The beta for any study can be calculated only based on an alternative hypothesis about a given positive relative risk and

Epidemiology

<sup>85.</sup> Where multiple epidemiological studies are available, a technique known as *meta-analysis* (see infra § IV.B.3) can be used to combine the results of the studies to reduce the numerical instability of all. *See generally* Frederic M. Wolf, Meta-Analysis: Quantitative Methods for Research Synthesis (1986). Meta-analysis is better suited to pooling results from randomly controlled experimental studies, but if carefully performed it may also be helpful for observational studies, such as in the epidemiological field. *See* Zachary B. Gerbarg & Ralph I. Horwitz, *Resolving Conflicting Clinical Trials: Guidelines for Meta-Analysis* , 41 J. Clin. Epidemiol. 503 (1988).

In *In* re Paoli R.R. Yard PCB Litig., 916 F.2d 829, 856–57 (3d Cir. 1990), *cert. denied*, 499 U.S. 461 (1991), the court discussed the use and admissibility of meta-analysis as a scientific technique. Overturning the district court's exclusion of a report using meta-analysis, the Third Circuit observed that meta-analysis is a regularly used scientific technique. The court recognized that the technique might be poorly performed and required the district court to reconsider the validity of the expert's work in performing the meta-analysis. *See also* E.R. Squibb & Sons, Inc. v. Stuart Pharmaceuticals, No. 90-1178, 1990 U.S. Dist. LEXIS 15788, at \*41 (D.N.J. Oct. 16, 1990) (acknowledging the utility of meta-analysis but rejecting its use in that case because one of the two studies included was poorly performed); Tobin v. Astra Pharmaceutical Prods., Inc., 993 F.2d 528, 538–39 (6th Cir. 1992) (identifying an error in the performance of a meta-analysis, in which the Food and Drug Administration (FDA) pooled data from control groups in different studies in which some gave the control a placebo and others gave the control an alternative treatment), *cert. denied*, 114 S. Ct. 304 (1993).

<sup>86.</sup> See also DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 947 (3d Cir. 1990).

the level of alpha selected.<sup>87</sup> That is, beta, or the likelihood of erroneously failing to reject the null hypothesis, depends on the selection of an alternative hypothesis about the magnitude of association and the level of alpha chosen.

#### 3. Power

The power of a study expresses the likelihood of detecting a postulated level of effect, assuming such an effect exists.<sup>88</sup> The power of a study is the complement of beta  $(1 - \beta)$ . Thus, a study with a likelihood of .25 of failing to detect a true relative risk of 2.0<sup>89</sup> or greater has a power of .75. This means the study has a 75% chance of detecting a true relative risk of 2.0. If the power of a negative study to find a relative risk of 2.0 or greater is low, it has significantly less probative value than a study with similar results but a higher power.<sup>90</sup>

### D. What Biases May Have Existed That Would Result in an Erroneous Association?

Systematic error or bias can produce an erroneous association in an epidemiological study.<sup>91</sup> Major sources of bias in the context of planning an epidemiological study were discussed previously in section II. After a study is completed, similar inquiries can be made about whether the study design, data collection, or analysis are flawed and therefore create error. Such an inquiry would be informed by the same concerns described in section II.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, another inquiry remains. An association, as repeatedly noted, does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert must consider the possibility of confounding factors and use several criteria to determine whether an inference of causation is appropriate. These matters are discussed in section IV.

Reference Manual on Scientific Evidence

<sup>87.</sup> See Green, supra note 23, at 684-89.

<sup>88.</sup> For clarification, see supra § II.B.2.b and Figure 3.

<sup>89.</sup> The use of a relative risk of 2.0 for illustrative purposes is because of the legal significance of this magnitude of association. See infra  $\S$  V.

<sup>90.</sup> See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.3.a, in this manual.

<sup>91.</sup> See supra § III.B.

# IV. General Causal Association Between Exposure and the Disease

Once an association has been found between exposure to a substance and a disease, researchers consider whether the association reflects a true cause-effect relationship or, alternatively, a spurious finding.<sup>92</sup> As mentioned in section I, epidemiology cannot prove causation; causation is a judgment issue for epidemiologists and others interpreting the epidemiological data.<sup>93</sup>

Researchers first look for alternative explanations for the association, such as bias or confounding factors, the latter of which is discussed below. Once this process is completed, researchers consider the guidelines for causation. These guidelines consist of seven inquiries that assist researchers in making a judgment about causation.<sup>94</sup> As a final step, researchers interpret the data and draw a conclusion about the existence of a cause-effect relationship. Most researchers are conservative when it comes to assessing causal relationships, often calling for stronger evidence and more research before a conclusion of causation is drawn.<sup>95</sup>

<sup>92.</sup> When epidemiologists evaluate whether a cause-effect relationship exists between an agent and disease, they are using the term causation in a way similar to, but not identical with, the familiar "but for" or sine qua non test used in law for cause in fact. "An act or an omission is not regarded as a cause of an event if the particular event would have occurred without it." W. Page Keeton et al., Prosser & Keeton on Torts 265 (5th ed. 1984); *se also* Restatement (Second) of Torts § 432(1) (1965). Epidemiologists use the term to mean that increase in disease among the exposed group would not have occurred in the group had they not been exposed to the agent. Thus, exposure is a necessary condition for the increase in the incidence of disease among those exposed. See Rothman, *supra* note 32, at 11 ("We can define a cause of a disease as an event, condition or characteristic that plays an essential role in producing an occurrence of the disease."); Allen v. United States, 588 F. Supp. 247, 405 (D. Utah 1984) (quoting a physician on the meaning of the statement that radiation causes cancer), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988). Translating the epidemiological concept of cause to the legal question of whether exposure to an agent caused an individual's disease is addressed *infra* § V.

<sup>93.</sup> In epidemiology, the practice of drawing inferences about causation is extremely controversial. On one side of this controversy, Professor Kenneth Rothman and his supporters argue that drawing conclusions about causation is not part of science at all, but the domain of public policy. They suggest that scientists should provide policy makers with information but should not advocate a particular interpretation. On the other side of this controversy are more traditional epidemiologists who contend that the researcher is often in the best position to interpret the results, and ought to do so when possible. See Stephan F. Lanes, Causal Inference Is Not a Matter of Science (abstract), 122 Am. J. Epidemiol. 550 (1985).

<sup>94.</sup> The guidelines, referred to as "Koch's postulates" (see infra § IV.B), were used first in the field of infectious diseases. See Mervyn Susser, Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology (1973).

<sup>95.</sup> In Cadarian v. Merrell Dow Pharmaceuticals, Inc., 745 F. Supp. 409, 412 (E.D. Mich. 1989), the court refused to permit an expert to rely on a study that the authors had concluded should not be used to

This section of the reference guide is organized around the following three topics:

- 1. identification and adjustment for potential confounding factors;
- 2. application of guidelines for causation; and
- 3. interpretation of the results.

#### A. Could a Confounding Factor Be Responsible for the Study Result<sup>296</sup>

Even when an association exists, researchers must determine whether the exposure causes the disease or whether the exposure and disease are caused by some other confounding factor. A *confounding factor* is both a risk factor for the disease and associated with the exposure of interest. For example, researchers may conduct a study that finds individuals with gray hair have a higher rate of death than those with hair of another color. Instead of hair color having an impact on death, the results might be explained by the confounding factor of age. If old age is associated differentially with the gray hair group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death. <sup>97</sup> Researchers must separate the relationship between gray hair and risk of death and old age and risk of death. When researchers find an association between an agent and a disease, it is critical to determine whether the association is causal or the result of confounding.<sup>98</sup>

In 1981, Dr. Brian MacMahon, Professor and Chairman of the Department of Epidemiology at the Harvard School of Public Health, reported an association between coffee drinking and cancer of the pancreas in the *New England Journal of Medicine*. <sup>99</sup> This observation caused a great stir, and in fact, one coffee distributor ran a large advertisement in the *New York Times* refuting the findings of the study. What could MacMahon's findings mean? The first possibility is that the association is causal and that drinking coffee causes an increased risk of cancer of the pancreas. However, there is also another possibility. It is known that smoking is an important risk factor for cancer of the pancreas. It also is known that it is difficult to find a smoker who does not drink coffee. Thus, drinking coffee and smoking are associated. An observed association between coffee consumption and an increased risk of cancer of the pancreas could reflect the fact

96. Sæ Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991) (discussing the possibility that confounders may lead to an erroneous inference of a causal relationship).

97. This example is drawn from Kahn, supra note 36, at 63.

98. Similarly, a finding of no association may be erroneous because a confounding factor with a protective effect is differentially associated with those exposed to the agent. One example of a confounding factor with a protective effect is vaccination.

99. Brian MacMahon et al., Coffee and Cancer of the Pancreas, 304 New Eng. J. Med. 630 (1981).

support an inference of causation in the absence of independent confirmatory studies. The court did not address the question of whether the degree of certainty employed by epidemiologists before making a conclusion of cause was consistent with the legal standard. See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 957 (3d Cir. 1990) (standard of proof for scientific community is not necessarily appropriate standard for expert opinion in civil litigation); Wells v. Ortho Pharmaceutical Corp., 788 F.2d 741, 745 (11th Cir.), cert. denied, 479 U.S. 950 (1986).

that smoking causes cancer of the pancreas and that smoking also is associated closely with coffee consumption. The association MacMahon found between drinking coffee and pancreatic cancer could be due to the confounding factor of smoking. To consider the possible confounding role of cigarettes, MacMahon examined smokers and nonsmokers separately to determine whether the relationship between coffee and cancer of the pancreas held in both groups. When smoking was held constant, he still found an increasing risk of pancreatic cancer with increasing consumption of coffee, particularly in women.

The main problem in many observational studies such as MacMahon's is that the individuals are not assigned randomly to the exposed cohort and the control group.<sup>100</sup> Instead, individuals self-select themselves for that exposure (or in many studies someone else selects them), a feature of virtually all observational human population studies without *randomization*. The lack of randomization leads to the potential problem of confounding. Thus, for example, the exposed cohort might consist of those who are exposed at work to an agent suspected of being an industrial toxin. The members of this cohort may have been "selected"—by themselves or by others—based on residence, socioeconomic status, age, or other factors.<sup>101</sup> These other selection factors may be causing the disease, but because of the selection an apparent (yet false) association of the disease with exposure to the agent may appear.

Confounding factors that are known in advance can be controlled during the study design and through study group selection. Unanticipated confounding factors that can be identified can sometimes be controlled during data analysis if data are gathered about them. There is always a risk, however, that an undiscovered confounding factor is responsible for a study's findings.<sup>102</sup>

1. What techniques, if any, were used to identify confounding factors?

Care in the design of a research project (e.g., methods to select the subjects, diagnose disease, and assess exposure) can prevent confounding. To identify potential confounding factors, the researcher must assess a range of factors that could influence risk. This procedure often involves complex statistical manipulation to compare the overall risk of exposure with the risk when identified potential confounding factors have been removed from the calculation.

Using MacMahon's study as an example, the researcher would test whether smoking is a confounding factor by comparing the risk of pancreatic cancer in all coffee drinkers (including smokers) with the risk in nonsmoking coffee

Epidemiology

<sup>100.</sup> Randomization assumes that the presence of a characteristic, such as coffee drinking, is governed by chance, as opposed to being determined by the presence of an underlying medical condition. See the Glossary of Terms for additional comments on randomization and confounding.

<sup>101.</sup> See , e.g. , In re "Agent Orange" Prod. Liab. Littig., 597 F. Supp. 740, 783 (E.D.N.Y. 1984) (discussing the problem of confounding that might result in a study of the effect of exposure to Agent Orange on Vietnam servicemen), aff'd, 818 F.2d 145 (2d Cir. 1987).

<sup>102.</sup> Rothman, supra note 32, at 125.

drinkers. If the risk is the same, smoking is not a confounding factor (e.g., smoking does not distort the relationship between coffee drinking and the development of pancreatic cancer).

2. What techniques, if any, were used to control confounding factors?

To control for confounding factors during data analysis researchers can use one of two techniques: stratification or multivariate analysis.

Stratification reduces or eliminates confounding by evaluating the effect of an exposure at different levels (strata) of exposure of the confounding variable. Statistical methods then can be applied to combine the different results of each stratum into an overall single estimate of risk. For example, in MacMahon's study of smoking and pancreatic cancer, if smoking had been a confounding factor, the researchers could have stratified the data by creating subgroups based on how many cigarettes each subject smoked a day (e.g., a nonsmoking group, a light smoking group, a medium smoking group, and a heavy smoking group). By comparing the different rates of pancreatic cancer for people in each group who drink the same amount of coffee, the effect of smoking on pancreatic cancer is revealed. The effect of the confounding factor can then be removed from the study results.

Multivariate analysis controls the confounding factor through mathematical modeling. *Models* are developed to describe the simultaneous effect of exposure and confounding factors on the increase in risk. This technique relies on building a series of mathematical models to predict who will get the disease.<sup>103</sup> For instance, MacMahon might have begun a multivariate analysis with a simple model to determine how well the individual's daily intake of coffee predicts whether he or she will contract pancreatic cancer. In the next model, he could add the number of years the person had been a coffee drinker. If the second model better predicts who would contract cancer, MacMahon would continue to create more complex models (including variables such as age, gender, and ethnic group) until he found a model that best predicts who will contract cancer.

If the association between exposure and disease remains after completing the assessment and adjustment for confounding factors, the researcher applies the guidelines described in section IV.B to determine whether an inference of causation is warranted.

## B. Overall, Does Application of the Guidelines for Causation Support a Finding of Causation?

Seven factors should be considered when an epidemiologist determines whether

103. For a more complete discussion, see Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual.

the association between an agent and a disease is causal.<sup>104</sup> These factors guide the epidemiologist in making a judgment about causation. They are

- 1. strength of the association;
- 2. temporal relationship;
- 3. consistency of the association;
- 4. biologic plausibility (coherence with existing knowledge);
- 5. consideration of alternative explanations; <sup>105</sup>
- 6. specificity of the association; and
- 7. dose-response relationship.

These guidelines, known as Koch's postulates, were proposed first about 100 years ago by two infectious disease researchers, Koch and Henle.<sup>106</sup> Each factor is considered in the following subsections.

1. How strong is the association between the exposure and disease?<sup>107</sup>

The relative risk is one of the cornerstones for causal inferences.<sup>108</sup> Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.<sup>109</sup> For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.<sup>110</sup> That is, the risk of lung cancer in smokers is approximately nine to ten times the risk in nonsmokers.

A relative risk of 9 to 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any kind of error in the study that would have produced it. The higher the relative risk, the stronger the association, and the more likely an epidemiologist will consider it causal. Although lower relative risks can reflect causality, the epidemiologist will scrutinize the association more closely.

Attributable risk, another measure of excess risk, is particularly important in the legal arena, because it measures the excess risk caused by exposure to the

110. See Doll & Hill, supra note 6.

Epidemiology

<sup>104.</sup> See Smith v. Ortho Pharmaceutical Corp., 770 F. Supp. 1561, 1575-76 (N.D. Ga. 1991).

<sup>105.</sup> See supra note 94 and accompanying text.

<sup>106.</sup> The two factors, dose-response relationship and specificity of the association, are not always used. See infra note 119 and accompanying text.

<sup>107.</sup> Assuming that an association is determined to be causal, the strength of the association plays an important role legally in determining the specific causation question—whether the agent caused the individual plaintiffs injury. See infra  $\S$  V.

<sup>108.</sup> See supra § III.A.1.

<sup>109.</sup> The reason that a higher relative risk is more likely to indicate a true causal relationship is because such a strong effect is unlikely to be the result of bias or random sampling error. Findings of small relative risks are much more susceptible to these errors. Sæ Cook v. United States, 545 F. Supp. 306, 316 n.4 (N.D. Cal. 1982); Landrigan v. Celotex Corp., 605 A.2d 1079, 1085 (N.J. 1992). The use of the strength of the association as a factor does not reflect a belief that weaker effects are rarer phenomena than stronger effects. Sæ Green, supra note 23, at 652–53 n.39. Indeed, the apparent strength of a given agent is dependent on the prevalence of the other necessary elements that must occur with the agent to produce the disease, rather than on some inherent characteristic of the agent itself. Sæ Rothman, supra note 32, at 12–13.

agent.<sup>111</sup> For example, if a group of individuals is exposed to PCBs and has a high risk of cancer, attributable risk permits the epidemiologist to subtract the background risk of disease from the exposed group's total risk of disease. In doing so, the epidemiologist measures the increased risk of disease that can be attributed to a specific exposure, which can then be used to determine the benefit that would be gained by eliminating a particular exposure.<sup>112</sup>

#### 2. Is there a temporal relationship?

A temporal or chronological relationship must exist for causation. If an exposure causes disease, the exposure must occur before the disease develops.<sup>113</sup> If the exposure occurs after the disease develops, it cannot cause the disease.

#### 3. Is the association consistent with other research?<sup>114</sup>

The need to replicate research findings permeates most fields of science. In epidemiology, research findings often are replicated in different populations.<sup>115</sup> Consistency in these findings is an extremely important factor in making a judgment about causation. Different studies that examine the same exposuredisease relationship should yield similar results. Any inconsistencies signal a need to question whether the relationship is causal.

*Meta-analysis* is an analytic technique that allows epidemiologists to combine the results of several research studies to better understand the relationship between exposure to an agent and a disease. <sup>116</sup> The combined data are analyzed to determine if they render different results from those in the individual studies performed with smaller sample sizes. <sup>117</sup> Particular concern must be paid to the

116. See In re Paoli R.R. Yard PCB Littig., 916 F.2d 829, 856–57 (3d Cir. 1990), cert. denied, 499 U.S. 961 (1991); Hines v. Consolidated Rail Corp., 926 F.2d 262, 273 (3d Cir. 1991).

117. See supra note 85.

<sup>111.</sup> Risk is not zero among the control group (those not exposed) when there are other causal chains that cause the disease that do not require exposure to the agent. For example, a proportion of birth defects are the result of genetic sources, which do not require the presence of any environmental agent. Also, some degree of risk in the control group may be the result of background exposure to the agent being studied. For example, nonsmokers in a control group may have been exposed to passive cigarette smoke, which is responsible for some cases of lung cancer and other diseases. *Sæ also* Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 25 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976). There are some diseases that do not occur without exposure to an agent; these are known as signature diseases. *See infra* note 122.

<sup>112.</sup> The benefit gained by eliminating a particular exposure would be equivalent to the amount of disease that could be prevented by eliminating that exposure. See *supra* § III.A for an example of how to calculate this amount.

<sup>113.</sup> See Carroll v. Litton Sys., Inc., No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at \*29 (W.D.N.C. Oct. 29, 1990) ("[I]t is essential for . . . [the plaintiffs' medical experts opining on causation] to know that exposure preceded plaintiffs' alleged symptoms in order for the exposure to be considered as a possible cause of those symptoms. . . . ").

<sup>114.</sup> In Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 901 (N.D. Iowa 1982), aff'd sub nom. Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1983), the court remarked on the persuasive power of multiple independent studies, each of which reached the same finding of an association between toxic shock syndrome and tampon use.

<sup>115.</sup> Sæ Cadarian v. Merrell Dow Pharmaceuticals, Inc., 745 F. Supp. 409, 412 (E.D. Mich. 1989) (holding a study on Bendectin insufficient to support an expert's opinion, because "the study's authors themselves concluded that the results could not be interpreted without independent confirmatory evidence").

propriety of combining different study populations and to the appropriate inferences to be drawn from the meta-analysis.

 Is the association biologically plausible (consistent with existing knowledge)?<sup>118</sup>

*Biological plausibility* is not a simple criterion to use. When an association is biologically plausible, the plausibility is appealing and provides supporting evidence. For example, the conclusion that high cholesterol is a cause of coronary heart disease is plausible because cholesterol is found in atherosclerotic plaques. However, observations have been made in epidemiological studies that were not biologically plausible at the time but subsequently were shown to be correct. When an observation is inconsistent with current biological knowledge, it should not be discarded, but the observation should be confirmed before significance is attached to it. The saliency of this factor varies depending on the extent of scientific knowledge about the cellular and subcellular mechanisms through which the disease process works. The mechanisms of some diseases are understood better than others.

5. Have alternative explanations been ruled out?

Alternative explanations and confounding factors should be examined and ruled out to avoid reaching an erroneous conclusion. However, it is never possible to rule out every alternative explanation. Epidemiology cannot prove causation. It is an inference for the scientist to make and usually is not made lightly.

The last two factors, specificity of the association and dose-response relationship, differ in significant ways from the five factors mentioned above. Although the presence of specificity and dose-response strengthens the inference of causation, the absence of either does not weaken the inference. Epidemiologists have begun to question the use of these two factors as guidelines for causation in noninfectious diseases.<sup>119</sup>

#### 6. Does the association exhibit specificity?

An association exhibits specificity if the exposure is associated only with a single disease or type of disease. As mentioned above, epidemiologists no longer require that the effect of exposure to an agent be specific for a single disease. For example, cigarette manufacturers have long claimed that since cigarettes have

<sup>118.</sup> A number of courts have adverted to this criterion in the course of their discussions of causation in toxic substances cases. E.g., Cook v. United States, 545 F. Supp. 306, 314–15 (N.D. Cal. 1982) (discussing biological implausibility of a two-peak increase of disease when plotted against time); Landrigan v. Celotex Corp., 605 A.2d 1079, 1085–86 (N.J. 1992) (discussing the existence vel non of biological plausibility). See also Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, in this manual.

<sup>119.</sup> Koch's postulates were originally formulated for determining causation of infectious diseases. Specificity and dose-response remain important factors in infectious disease epidemiology. See supra § IV.B and note 106.

been linked to lung cancer, emphysema, bladder cancer, heart disease, pancreatic cancer, and other conditions, there is no specificity and the relationships are not causal. The scientific bases that have undermined the guideline include the following: (1) Human cells and tissues share many common features. They all have a basic structure, including nuclei, DNA, and other characteristics. There is every reason to expect that a certain agent will act on certain cellular components and structures even if they are in different tissues and different organs; and (2) Tobacco and cigarette smoke are not single agents but mixtures of harmful agents. Smoking represents exposure to multiple agents and specificity would not be expected. However, most known teratogens cause a specific birth defect or a related pattern of birth defects.

#### 7. Is there a dose-response relationship?

A dose-response relationship assumes that the more intense the exposure, the greater the risk of disease. However, the researcher may not observe a dose-response relationship when there is a *threshold phenomenon* (i.e., a low dose exposure may not cause disease until the exposure exceeds a certain dose).<sup>120</sup> Evidence of a dose-response relationship strengthens the conclusion that the relationship between an agent and disease is causal; however, a dose-response relationship is not necessary to infer causation.

## C. What Type of Causal Association Has Been Demonstrated Between Exposure and Disease?

Assuming an association is not due to confounding factors and that the epidemiologist has decided that the scientific findings overwhelmingly support an inference of causation, the epidemiologist next determines which type of causal relationship exists between the agent and the disease in the exposed population. Epidemiologists divide causes into four categories (see Figure 6). It should be noted that the terms applied to the four categories of causation are not consistent

<sup>120.</sup> The question of whether there is a no-effect threshold dose is a controversial one in a variety of toxic substances areas. *See*, e.g., Irving J. Selikoff, Disability Compensation for Asbestos-Associated Disease in the United States: Report to the U.S. Dep't of Labor 181–220 (1981); Paul Kotin, *Dose-Response Relationships and Threshold Concepts*, 271 Annals N.Y. Acad. Sci. 22 (1976); K. Robock, *Based on Available Data*, *Can We Project an Acceptable Standard for Industrial Use of Asbestos? Absolutely*, 330 Annals N.Y. Acad. Sci. 205 (1979); Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir.) (dose-response relationship for low doses is "one of the most sharply contested questions currently being debated in the medical community"), cert. denied, 469 U.S. 1062 (1984).

Moreover, good evidence to support or refute the threshold dose hypothesis is exceedingly unlikely because of the inability of epidemiology or animal toxicology to ascertain very small effects. *Cf.* Arnold L. Brown, *The Meaning of Risk Assessment*, 37 Oncology 302, 303 (1980). Even the question of the shape of the dose-response curve—whether linear or curvilinear, and if the latter, the shape of the curve—is a matter of hypothesis and speculation. *Sæ* Allen v. United States, 588 F. Supp. 247, 419–24 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988); Troyen A. Brennan & Robert F. Carter, *Legal and Scientific Probability of Causation for Cancer and Other Environmental Disease in Individuals*, 10 J. Health Pol'y & L. 33, 43–44 (1985).

with legal terminology. Nevertheless, these terms may be useful in understanding them when they appear in a published study or when used by an epidemiologist. <sup>121</sup>

Figure 6 Four Categories of Causation

- 2. Necessary but not sufficient

 $A_1 + A_2 + A_3 \dots$  Disease

(Causal chain may also involve a specific temporal sequence)

3. Sufficient but not necessary



 Neither necessary nor sufficient (probably true for most of the diseases we study)



- 1. Exposure to an agent may be a *necessary and sufficient cause* of the disease. This type of causal relationship assumes that the disease will not result unless an individual is exposed. Nothing but the agent is needed to cause the disease.
- 2. Exposure can be necessary but not a sufficient cause of the disease. In

121. See Mervyn Susser, What Is a Cause and How Do We Know One? A Grammar for Pragmatic Epidemiology , 133 Am. J. Epidemiol. 635, 637 (1991).

#### Epidemiology

this causal relationship the disease will not result unless an individual is exposed, but exposure in and of itself is not enough to cause the disease.

- 3. An exposure may be a *sufficient but not necessary cause* of the disease when the disease occurs not only in the presence of exposure but also in the presence of exposures to other agents. Leukemia is an example of this relationship; exposure to radiation or benzene can result in the occurrence of disease.
- 4. The last possibility is that exposure is *neither a necessary nor sufficient cause* of the disease. This takes place when the disease occurs in the absence of exposure and does not always occur in its presence. This complicated relationship is probably the one that most faithfully represents the causal relationships in the majority of diseases encountered. The disease can occur through a variety of combinations of different exposures.

### V. The Role of Epidemiology in Proving Individual Causation

Epidemiology is concerned with the incidence of disease in populations and does not address the question of the cause of an individual's disease.<sup>122</sup> This question, sometimes referred to as specific causation, is beyond the domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiology addresses whether an agent causa a disease, not whether an agent did cause a plaintiff's disease.<sup>123</sup>

Nevertheless, the specific causation issue is a necessary element in a toxic substance case. The plaintiff must establish not only that the defendant's agent is capable of causing disease but also that it did cause the plaintiff's disease. Thus, a number of courts have confronted the legal question of what is acceptable proof of specific causation and the role that epidemiological evidence plays in answering that question. This question is not a question about which an epidemiologist would have any expertise to contribute. Rather it is a legal question with which a number of courts have grappled. An explanation of how these courts have resolved this question follows.

There are two legal issues that arise with regard to the role of epidemiology in proving individual causation: admissibility and sufficiency of evidence to meet the burden of production. The first issue tends to receive less attention by the courts but nevertheless deserves mention. An epidemiological study that is suffi-

123. Cf. In re "Agent Orange" Prod. Liab. Litig., 597 F. Supp. at 780.

<sup>122.</sup> See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 945 & n.6 (3d Cir. 1990) ("Epidemiological studies do not provide direct evidence that a particular plaintiff was injured by exposure to a substance."); Smith v. Ortho Pharmaceutical Corp., 770 F. Supp. 1561, 1577 (N.D. Ga. 1991); Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991); Michael Dore, A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact , 7 Harv. Envtl. L. Rev. 429, 436 (1983).

There are some diseases that do not occur without exposure to a given toxic agent. This is the same as saying that the toxic agent is a necessary cause for the disease and sometimes referred to as a "signature disease," because the existence of the disease necessarily implies the causal role of the agent. See Kenneth S. Abraham & Richard A. Merrill, *Scientific Uncertainty in the Courts,* Issues Sci. & Tech., Winter 1986, at 93, 101. Asbestosis is a *signature disease* for asbestos, and adenocarcinoma (in young adult women) is a signature disease for *in utero* DES exposure. See In re "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 834 (E.D.N.Y. 1984) (Agent Orange allegedly caused a wide variety of diseases in Vietnam veterans and their offspring), *aff'd*, 818 F.2d 145 (2d Cir. 1987).

ciently rigorous to justify a conclusion that it is scientifically valid should be admissible, <sup>124</sup> as it tends to make an issue in dispute more or less likely.<sup>125</sup>

Far more courts have confronted the role that epidemiology plays with regard to the sufficiency of the evidence and the burden of production. The civil burden of proof is described most often as requiring the fact finder to "believe that what is sought to be proved . . . is more likely true than not true." <sup>126</sup> The relative risk from an epidemiological study can be adapted to this 50% plus standard to yield a probability or likelihood that an agent caused an individual's disease. <sup>127</sup> The threshold for concluding that an agent was more likely the cause of a disease than not is a relative risk greater than 2.0. Recall that a relative risk of 1.0 means that the agent has no effect on the incidence of disease. When the relative risk reaches 2.0, the agent is responsible for an equal number of cases of disease as all other background causes. Thus, a relative risk of 2.0 implies a 50% likelihood that an exposed individual's disease was caused by the agent. A rela-

124. See *DeLuca*, 911 F.2d at 958; *cf*. Kehm v. Procter & Gamble Co., 580 F. Supp. 890, 902 (N.D. Iowa 1982) ("These [epidemiological] studies were highly probative on the issue of causation—they all concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists."), *aff'd sub nom*. Kehm v. Procter & Gamble Mfg. Co., 724 F.2d 613 (8th Cir. 1984).

Hearsay concerns may limit the independent admissibility of the study (*see supra* note 1); but the study could be relied on by an expert in forming an opinion and may be admissible pursuant to Fed. R. Evid. 703 as part of the underlying facts or data relied on by the expert.

In Ellis v. International Playtex, Inc., 745 F.2d 292, 303 (4th Cir. 1984), the court concluded that certain epidemiological studies were admissible despite criticism of the methodology used in the studies. The court held that the claims of bias went to the weight rather than the admissibility of the studies. *Cf.* Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1109 (5th Cir. 1991) ("As a general rule, questions relating to the bases and sources of an expert's opinion affect the weight to be assigned that opinion rather than its admissibility..."), *cert. denied*, 112 S. Ct. 1280 (1992). 125. Even if evidence is relevant, it may be excluded if its probative value is substantially outweighed by

125. Even if evidence is relevant, it may be excluded if its probative value is substantially outweighed by prejudice, confusion, or inefficiency. Fed. R. Evid. 403. Exclusion of an otherwise relevant epidemiological study on Rule 403 grounds is unlikely.

In Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2796 (1993), the Court invoked the concept of "fit," which addresses the relationship of an expert's scientific opinion with the facts of the case and the issues in dispute. In a toxic substance case in which cause in fact is disputed, an epidemiological study of the same agent to which the plaintiff was exposed that examined the association with the same disease from which the plaintiff suffers would almost surely have sufficient "fit" to be a part of the basis of an expert's opinion. The Court's concept of "fit," borrowed from United States v. Downing, 753 F.2d 1224, 1242 (3d Cir. 1985), appears equivalent to the more familiar evidentiary concept of probative value.

126. 2 Edward J. Devitt & Charles B. Blackmar, Federal Jury Practice and Instruction §71.13 (3d ed. 1977); *also* United States v. Fatico, 458 F. Supp. 388, 403 (E.D.N.Y. 1978) ("Quantified the preponderance standard would be 50+% probable."), *aff'd*, 603 F.2d 1053 (2d Cir. 1979), *cert. denied*, 444 U.S. 1073 (1980).

127. An adherent of the frequentist school of statistics would resist this adaptation, which may explain why so many epidemiologists and toxicologists also resist it. To take the step identified in the text requires a shift from a frequentist approach, which involves sampling or frequency data from an empirical test, to a subjective probability about a discrete event. Thus, a frequentist might assert, after conducting a sampling test, that 60% of the balls in an opaque container are blue. The same frequentist would resist the statement: "The probability that a single ball removed from the box and hidden behind a screen is blue is 60%." The ball is either blue or not, and no frequentist data would permit the latter statement. "[T]here is no logically rigorous definition of what a statement of probability means with reference to an individual instance. . . ." Lee Loevinger, *On Logic and Sociology*, 32 Jurimetrics J. 527, 530 (1992); see also Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence*, 96 Yale L.J. 376, 382–92 (1986). Subjective probabilities about discrete events are the product of adherents to Bayes Theorem. See Kaye, *supra* note 80, at 54–62; David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B, in this manual.

tive risk greater than 2.0 would permit an inference that an individual plaintiff's disease was more likely than not caused by the implicated agent. A substantial number of courts in a variety of toxic substances cases have accepted this reasoning.<sup>128</sup>

An alternative, yet similar means to address probabilities in individual cases is by use of the attributable proportion of risk parameter.<sup>129</sup> The attributable risk is a measurement of the excess risk that can be attributed to an agent, above and beyond the background risk due to other causes. When the attributable risk exceeds 50% (equivalent to a relative risk greater than 2.0), this logically might be converted to a belief that the agent was more likely than not the cause of the plaintiff's disease.

The discussion above assumes that the only evidence bearing on cause in fact is epidemiological. Such an assumption is unlikely, and a variety of additional pieces of evidence, although less quantifiable, affect a fact finder's assessment. Biases in the epidemiological studies might justify a conclusion that the real magnitude of increased risk is greater or lower than that revealed in the studies. The dose to which the plaintiff was exposed may be greater or lesser than those in the epidemiological study, thereby requiring some extrapolation.<sup>130</sup> In addition, there may be factors peculiar to the plaintiff—excess exposure to another known cause, pathological mechanism,<sup>131</sup> family history of disease, or conflicting diagnoses—that modify any probability based solely on the available epidemiological evidence.<sup>132</sup>

129. See supra § III.A.3.

130. See supra § IV.B.5; se also Ferebee v. Chevron Chem. Co., 736 F.2d 1529, 1536 (D.C. Cir.) ("The dose-response relationship at low levels of exposure for admittedly toxic chemicals like paraquat is one of the most sharply contested questions currently being debated in the medical community."), cert. denied , 469 U.S. 1062 (1984); In re Joint E. & S. Dists. Asbestos Litig., 774 F. Supp. 113, 115 (S.D.N.Y. 1991) (discussing different relative risks associated with different dose levels), rev'd on other grounds , 964 F.2d 92 (2d Cir. 1992).

131. Sæ Tobin v. Astra Pharmaceutical Prods., Inc., 993 F.2d 528 (6th Cir.) (plaintiff's expert relied predominantly on pathogenic evidence), cert. denied, 114 S. Ct. 304 (1993).

132. An example of a judge sitting as fact finder and considering individualistic factors for a number of plaintiffs in deciding cause in fact is contained in Allen v. United States, 588 F. Supp. 247, 429–43 (D. Utah 1984), rev'd on other grounds, 816 F.2d 1417 (10th Cir. 1987), cert. denied, 484 U.S. 1004 (1988); see also

<sup>128.</sup> See DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941, 958-59 (3d Cir. 1990) (Bendectin allegedly caused limb reduction birth defects); In re Joint E. & S. Dists. Asbestos Litig., 758 F. Supp. 199, 203 (S.D.N.Y. 1991) (asbestos allegedly caused colon cancer), rev'd, 964 F.2d 92 (2d Cir. 1992) (relative risk less than 2.0 may still be sufficient to prove causation); Manko v. United States, 636 F. Supp. 1419, 1434 (W.D. Mo. 1986) (swine flu vaccine allegedly caused Guillain-Barré syndrome), aff'd in part, 830 F.2d 831 (8th Cir. 1987); Marder v. G.D. Searle & Co., 630 F. Supp. 1087, 1092 (D. Md. 1986) (pelvic inflammatory disease allegedly caused by Copper 7 IUD), aff'd without op. sub nom. Wheelahan v. G.D. Searle & Co., 814 F.2d 655 (4th Cir. 1987); In re "Agent Orange" Prod. Liab. Litig., 597 F. Supp. 740, 835-37 (E.D.N.Y. 1984) (Agent Orange allegedly caused a wide variety of diseases in Vietnam veterans and their offspring), aff'd, 818 F.2d 145 (2d Cir. 1987); Cook v. United States, 545 F. Supp. 306, 308 (N.D. Cal. 1982) (swine flu vaccine allegedly caused Guillain-Barré syndrome); Landrigan v. Celotex Corp., 605 A.2d 1079, 1087 (N.J. 1992) (relative risk greater than 2.0 "support[s] an inference that the exposure was the probable cause of the disease in a specific member of the exposed population"). But cf. In re Fibreboard Corp., 893 F.2d 706, 711-12 (5th Cir. 1990) (The court disapproved a trial in which several representative cases would be tried and the results extrapolated to a class of some 3,000 asbestos victims, without consideration of any evidence about the individual victims. The court remarked that general causation, which ignores any proof particularistic to the individual plaintiff, could not substitute under Texas law for cause in fact.).

This additional evidence bearing on causation has led a few courts to conclude that a plaintiff may satisfy his or her burden of production even if a relative risk less than 2.0 emerges from the epidemiological evidence.<sup>133</sup> For example, genetics might be known to be responsible for 50% of the incidence of a disease. If genetics can be ruled out in an individual's case, then a relative risk greater than 1.5 might be sufficient to support an inference that the agent was more likely than not responsible for the plaintiff's disease.<sup>134</sup>

Manko v. United States, 636 F. Supp. 1419, 1437 (W.D. Mo. 1986), *affd*, 830 F.2d 831 (8th Cir. 1987). 133. See, e.g., Grassis v. Johns-Manville Corp., 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991): The physician or other qualified expert may view the epidemiological studies and factor

out other known risk factors such as family history, diet, alcohol consumption, smoking... or other factors which might enhance the remaining risks, even though the risk in the study fell short of the 2.0 correlation.

See also Inre Joint E. & S. Dists. Asbestos Litig., 964 F.2d 92, 97 (2d Cir. 1992), rev'g, 758 F. Supp. 199, 202–03 (S.D.N.Y. 1991) (requiring relative risk in excess of 2.0 for plaintiff to meet burden of production).

134. The use of probabilities in excess of .50 to support a verdict results in an all-or-nothing approach to damages that some commentators have criticized. The criticism reflects the fact that defendants responsible for toxic agents with a relative risk just above 2.0 may be required to pay damages not only for the disease that their agents caused, but also for all instances of the disease. Similarly, those defendants whose agents increase the risk of disease by less than double may not be required to pay for any of the disease that their agents caused. See, e.g., 2 American Law Inst., Reporter's Study on Enterprise Responsibility for Personal Injury: Approaches to Legal and Institutional Change 369–75 (1991). To date, courts have not adopted a rule that would apportion damages based on the probability of cause in fact in toxic substances cases.

Reference Manual on Scientific Evidence

### **Glossary of Terms**

The following terms and definitions were adapted from a variety of sources, including: A Dictionary of Epidemiology (John M. Last ed., 1988); 1 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988); James K. Brewer, Everything You Always Wanted To Know About Statistics, But Didn't Know How To Ask (1978); and R. A. Fisher, Statistical Methods for Research Workers (1973).

- Agent. Also, risk factor. A factor, such as a drug, microorganism, chemical substance, or form of radiation, whose presence or absence can result in the occurrence of a disease. A disease may have a single agent, a number of independent alternative agents, or a complex of two or more factors whose combined presence is necessary for the development of the disease (e.g., a virus is the agent of measles).
- Alpha. The level of statistical significance chosen by the researcher to determine if any association found in the study is sufficiently unlikely to have occurred by chance (due to random sampling error) if the null hypothesis (no association) is true. Researchers commonly adopt an alpha of .05, but the choice is arbitrary and other values can be justified.
- Alpha Error. Alpha error, also called type I error, occurs when the researcher rejects a null hypothesis when it is actually true (i.e., when there is no association). This can occur when an apparent difference is observed between the control and experimental groups, but the difference is not real (i.e., it occurred by chance). A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof.
- Association. The degree of statistical dependence between two or more events or variables. Events are said to be associated when they occur more or less frequently together than one would expect by chance. Association does not necessarily imply a causal relationship. Events are said not to have an association when the agent (or independent variable) has no apparent effect on the incidence of a disease (the dependent variable). This corresponds to a relative risk of 1.0. A negative association means that the events occur less frequently together than one would expect by chance, thereby implying a

preventive or protective role for the agent (e.g., a vaccine).

- Attributable Proportion of Risk (APR). This term has been used to denote the fraction of risk that is attributable to exposure to a substance (e.g., X% of lung cancer is attributable to cigarettes).
- Background Risk of Disease . Background risk of disease (or background rate of disease) is the amount of disease in a population that occurs in individuals who have no known exposures to an alleged risk factor for the disease. For example, the background risk for all birth defects is 3%–5% of live births.
- Beta Error. Beta error, also called type II error or false negative, occurs when the researcher fails to reject a null hypothesis when it is incorrect (i.e., when there is an association). This can occur when no statistically significant difference is detected between the control and experimental groups, but a difference does exist.
- *Bias.* Any effect at any stage of investigation or inference tending to produce results that depart systematically from the true values. The term bias does not necessarily carry an imputation of prejudice or other subjective factor, such as the experimenter's desire for a particular outcome. This differs from conventional usage in which bias refers to a partisan point of view.
- Biological Marker. A biological marker is an alteration in tissue or body fluids that occurs as a result of an exposure and that can be detected in the laboratory. Biological markers are only available for a small number of toxins.
- Biological Plausibility. This factor considers existing knowledge about human biology and disease pathology to provide a judgment about the plausibility that an agent causes a disease.
- Case-Comparison Study. See Case-Control Study.
- *Case-Control Study.* Also, case-comparison study, case history study, case referent study, retrospective study. A study that starts with the identification of persons with the disease (or other outcome variable) and a suitable control (comparison, reference) group of persons without the disease. Such a study is called retrospective because it starts after the onset of disease and looks back to the postulated causal factors.
- *Case Group.* A group of individuals who have been exposed to the disease, intervention, procedure, or other variable whose influence is being studied.
- *Causation.* Causation, as we use the term, denotes an event, condition, characteristic, or agent that is a necessary element of a set of other events that produce an outcome, such as a disease. Thus, a cause may be thought of as a necessary link in some causal chain that results in an outcome of interest.
- *Cohort.* Any designated group of persons followed or traced over a period of time to examine health or mortality experience.

- *Cohort Study.* The method of epidemiologic study in which groups of individuals can be identified who are, have been, or in the future may be differentially exposed to a factor or factors hypothesized to influence the probability of occurrence of a disease or other outcome. The groups are observed to find out if the exposed group is more likely to develop disease. The alternative terms for a cohort study (concurrent study, follow-up study, incidence study, longitudinal study, prospective study) describe an essential feature of the method, which is observation of the population for a sufficient number of person-years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both.
- Confidence Interval. A range of values within which the results of a study sample would be likely to fall if the study were repeated numerous times. Thus, if a *p*-value of .05 is selected, a confidence interval would indicate the range of relative risk values that would result 95% of the time if the study were repeated. The width of the confidence interval provides an indication of the precision of the point estimate or relative risk found in the study; the narrower the confidence interval, the greater the confidence in the relative risk estimate found in the study. Where the confidence interval contains a relative risk of 1.0, the results of the study are not statistically significant.
- Confounding Factor. A confounding factor is both a risk factor for the disease and associated with the exposure of interest. Confounding refers to a situation in which the effects of two processes are not separated. The distortion can lead to an erroneous result.
- Control Group. A comparison group (identified as a rule before a study is begun) comprising individuals who have not been exposed to the disease, intervention, procedure, or other variable whose influence is being studied. In statistics, control procedures try to filter out the effects of confounding variables on nonexperimental data, typically by "adjusting" through statistical procedures (like multiple regression).
- *Dose.* Dose generally refers to the intensity or magnitude of exposure multiplied by the duration of exposure.
- Dose-Response Relationship. A relationship in which a change in amount, intensity, or duration of exposure is associated with a change—either an increase or a decrease—in risk of disease.
- *Ecological Fallacy.* An error that occurs when a correlation between an agent and disease in a group (ecological) is not reproduced when individuals are studied. For example, at the ecological (group) level, a correlation has been found in several studies between the quality of drinking water and mortality rates from heart disease; it would be an ecological fallacy to infer from this alone that exposure to water of a particular level of hardness necessarily in-

fluences the individual's chances of contracting or dying of heart disease.

- *Effect Size.* The effect size, or magnitude of the increased risk in disease, is best thought of as the amount of disease that is caused by exposure to a toxic substance.
- *Epidemiology.* The study of the distribution and determinants of health-related states and events in populations and the application of this study to control of health problems.
- *Error.* Random error (sampling error) is that due to chance when the result obtained in the sample differs from the result that would be obtained if the entire population (universe) were studied. Two varieties of sampling error are type I error, or alpha error, and type II error, or beta error.

When hypotheses testing is used, rejecting a null hypothesis when it is actually true is called type I error. Failing to reject a null hypothesis when it is incorrect is called type II error.

Etiologic Factor. An agent that plays a role in causing a disease.

- *Exposed, Exposure.* In epidemiology, the exposed group (or the exposed) is used to describe a group whose members have been exposed to an agent that may be a cause of a disease or health effect of interest, or possess a characteristic that is a determinant of a health outcome.
- False Negative Error. See Beta Error.
- False Positive Error. See Alpha Error.
- Follow-Up Study. See Cohort Study.
- In Vitro. Within an artificial environment such as a test tube (e.g., the cultivation of tissue in vitro).
- In Vivo. Within a living organism (e.g., the cultivation of tissue in vivo).
- *Incidence.* The number of people in a specified population falling ill from a particular disease during a given period. More generally, the number of new events (e.g., new cases of a disease in a defined population) within a specified period of time.
- Incidence Study. See Cohort Study.
- Inference. The intellectual process of making generalizations from observations. In statistics, the development of generalization from sample data, usually with calculated degrees of uncertainty.
- Meta-Analysis. A technique used to combine the results of several studies to enhance the precision of the estimate of the effect size and reduce the plausibility that the association found is due to random sampling error. Meta-analysis is better suited to pooling results from randomly controlled experimental studies, but if carefully performed, it also may be used for

observational studies.

- Morbidity Rate. Morbidity is the state of illness or disease. Morbidity rate may refer to the incidence rate or prevalence rate of disease.
- Mortality Rate. Mortality refers to death. The mortality rate expresses the proportion of a population that dies of a disease or of all causes. The numerator is the number of individuals dying; the denominator is the total population in which the deaths occurred. The unit of time is usually a calendar year.
- *Model.* A representation or simulation of an actual situation. This may be either (1) a mathematical representation of characteristics of a situation that can be used to examine consequences of various actions, or (2) a representation of a country's situation through an "average region" with characteristics resembling those of the whole country.
- Multivariate Analysis. A set of techniques used when the variation in several variables has to be studied simultaneously. In statistics, any analytic method that allows the simultaneous study of two or more factors or variables.
- Null Hypothesis. At the outset of any observational or experimental study, the researcher must state a principle or proposition that will be tested in the study. In epidemiology, this principle typically addresses the existence of a causal relation between an agent and a disease. Most often, the null hypothesis is a statement that Agent A does not cause Disease D. The results of the study may justify a conclusion that the null hypothesis has been disproved (e.g., a study that finds a strong association between smoking and lung cancer). A study may fail to disprove the null hypothesis, but that does not justify a conclusion that the null hypothesis has been proved.
- Observational Study. An observational study is an epidemiological study in situations where nature is allowed to take its course, without intervention from the investigator. For example, in an observational study the subjects of the study are permitted to determine their level of exposure to an agent.
- Odds Ratio (OR). Also, cross-product ratio, relative odds. The ratio of two odds. For most purposes the odds ratio from a case-control study is quite similar to a risk ratio from a cohort study.
- *P* (*Probability*), *p*-Value. The *p*-value is the probability of getting a value of the test statistic equal to or more extreme than the result observed, given that the null hypothesis is true.

The letter p, followed by the abbreviation *n.s.* (not significant) or by the symbol for less than (<) and a decimal notation such as .01 or .05, is a statement of the probability that the difference observed could have occurred by chance.

Investigators may arbitrarily set their significance levels, but in most biomedical and epidemiological work, a study result whose probability value is less than 5% (p < .05) or less than 1% (p < .01) is considered sufficiently unlikely to have occurred by chance to justify the designation statistically significant.

- *Power.* The probability that a difference of a specified amount will be detected by the statistical hypothesis test, given that a difference exists. In less formal terms, power is like the strength of a magnifying lens in its capability to identify an association that truly exists. Power is equivalent to one minus type II error.
- Prospective Study. In a prospective study, two groups of individuals are identified: (1) individuals who have been exposed to a risk factor; and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion that develops disease in each group is compared. See Cohort Study.
- Random. The term implies that an event is governed by chance. See Randomization.
- Randomization. Allocation of individuals to groups (e.g., for experimental and control regimens) by chance. Within the limits of chance variation, randomization should make the control and experimental groups similar at the start of an investigation and ensure that personal judgment and prejudices of the investigator do not influence allocation.

Randomization should not be confused with haphazard assignment. Random assignment follows a predetermined plan that usually is devised with the aid of a table of random numbers. Randomization cannot be used where the exposure is known to cause harm (e.g., cigarette smoking).

- *Relative Risk (RR).* The ratio of the risk of disease or death among the exposed to the risk among the unexposed. For instance, if 10% of all people exposed to a chemical develop a disease, compared with 5% of people who are not exposed, the disease occurs twice as frequently among the exposed people: The relative risk is 10%/5% = 2. A relative risk of 1 indicates no association.
- *Research Design.* The procedures and methods, predetermined by an investigator, to be adhered to in conducting a research project.
- *Risk.* A probability that an event will occur (e.g., that an individual will become ill or die within a stated period of time or by a certain age).
- Sample. A selected subset of a population. A sample may be random or nonrandom and may be representative or nonrepresentative.
- Sample Size. The number of subjects who participate in a study.
- Secular Trend Study. Also, time-line study. This type of study examines changes over a period of time, generally years or decades. Examples include the decline of tuberculosis mortality and the rise, followed by a decline, in coro-

nary heart disease mortality in the United States in the past fifty years.

Sensitivity, Specificity. Sensitivity measures the accuracy of a diagnostic or screening test or device in identifying disease (or some other outcome) when it truly exists. For example, assume that we know that 20 women in a group of 1,000 women have cervical cancer. If the entire group of 1,000 women is tested for cervical cancer and the screening test only identifies 15 (of the known 20) cases of cervical cancer, the screening test has a sensitivity of 15/20, or 75%.

Specificity measures the accuracy of a diagnostic or screening test in identifying those who are disease free. Once again, assume that 980 women out of a group of 1,000 women do not have cervical cancer. If the entire group of 1,000 women is screened for cervical cancer and the screening test only identifies 900 women as without cervical cancer, then the screening test has a specificity of 900/980, or 92%.

- Signature Disease. A disease that is associated uniquely with exposure to an agent (e.g., asbestosis and exposure to asbestos).
- Statistical Significance. This term is used to describe a study result or difference that exceeds the type I error rate (or *p*-value) that was selected by the researcher at the outset of the study. In formal significance testing, a statistically significant result is unlikely to be the result of random sampling error and justifies rejection of the null hypothesis. Some epidemiologists believe that formal significance testing is inferior to using a confidence interval to express the results of a study.

Statistical significance, which addresses the role of random sampling error in producing the results found in the study, should not be confused with the importance (for public health or public policy) of a research finding.

- Stratification. The process of or result of separating a sample into several subsamples according to specified criteria, such as age, socioeconomic status, and so forth. The effect of confounding variables may be controlled by stratifying the analysis of results. For example, lung cancer is known to be associated with smoking. To examine the possible association between urban atmospheric pollution and lung cancer, the population may be divided into strata according to smoking status, thus controlling for smoking. The association between air pollution and cancer then can be appraised separately within each stratum.
- *Teratogen.* An agent that produces abnormalities in the embryo or fetus by disturbing maternal health or by acting directly on the fetus in utero.
- *Teratogenicity.* The capacity for an agent to produce abnormalities in the embryo or fetus.
- Threshold Phenomenon. A certain level of exposure to an agent below which disease does not occur and above which disease does occur.
- *Toxicology.* The science of the nature and effects of poisons, their detection, and the treatment of their effects.
- Toxic Substance. A substance that is poisonous.
- *True Association.* Also, real association. The association that really exists between agent and exposure and that might be found by a perfect (but nonetheless nonexistent) study.
- Type I Error. See Alpha Error and Error.
- Type II Error. See Beta Error and Error.
- Validity. The degree to which a measurement measures what it purports to measure.
- *Variable.* Any attribute, condition, or other item in a study that can have different numerical characteristics. In a study of the causes of heart disease, blood pressure and dietary fat intake are variables that might be measured.

# **References on Epidemiology**

Causal Inferences (Kenneth J. Rothman ed., 1988).

William G. Cochran, Sampling Techniques (1977).

A Dictionary of Epidemiology (John M. Last ed., 1988).

Joseph L. Fleiss, Statistical Methods for Rates and Proportions (1981).

Harold A. Kahn, An Introduction to Epidemiologic Methods (1983).

Harold A. Kahn & Christopher T. Sempos, Statistical Methods in Epidemiology (1989).

Abraham M. Lilienfeld & David E. Lilienfeld, Foundations of Epidemiology (1980).

Judith S. Mausner & Anita K. Bahn, Epidemiology: An Introductory Text (1974).

Kenneth J. Rothman, Modern Epidemiology (1986).

Mervyn Susser, Epidemiology, Health and Society: Selected Papers (1987).

### **References on Law and Epidemiology**

- 2 American Law Institute, Reporters' Study on Enterprise Responsibility for Personal Injury (1991).
- Bert Black & David Lilienfeld, Epidemiologic Proof in Toxic Tort Litigation , 52 Fordham L. Rev. 732 (1984).

Gerald Boston, A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience, 18 Colum. J. Envtl. L. 181 (1993).

Vincent M. Brannigan et al., Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases, 12 Risk Analysis 343 (1992).

Troyen Brennan, Causal Chains and Statistical Links: The Role of Scientific

Uncertainty in Hazardous-Substance Litigation, 73 Cornell L. Rev. 469 (1988).

- Troyen Brennan, Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts, 51 U. Pitt. L. Rev. 1 (1989).
- Comment, Epidemiologic Proof of Probability: Implementing the Proportional Recovery Approach in Toxic Exposure Torts , 89 Dick. L. Rev. 233 (1984).
- Richard Delgado, Beyond Sindell: Relaxation of Cause-in-Fact Rules for Indeterminate Plaintiffs, 70 Cal. L. Rev. 881 (1982).
- Michael Dore, A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact, 7 Harv. Envtl. L. Rev. 429 (1983).
- Daniel Farber, Toxic Causation, 71 Minn. L. Rev. 1219 (1987).
- Steve Gold, Note, Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence, 96 Yale L.J. 376 (1986).
- Leon Gordis, Epidemiologic Approaches for Studying Human Disease in Relation to Hazardous Waste Disposal Sites, 25 Hous. L. Rev. 837 (1988).
- Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation, 86 Nw. U. L. Rev. 643 (1992).
- Khristine L. Hall & Ellen Silbergeld, Reappraising Epidemiology: A Response to Mr. Dore, 7 Harv. Envtl. L. Rev. 441 (1983).
- Constantine Kokkoris, Comment, DeLuca v. Merrell Dow Pharmaceuticals, Inc.: Statistical Significance and the Novel Scientific Technique , 58 Brook. L. Rev. 219 (1992).
- James P. Leape, Quantitative Risk Assessment in Regulation of Environmental Carcinogens, 4 Harv. Envtl. L. Rev. 86 (1980).
- Junius McElveen, Jr. & Pamela Eddy, Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology, 33 Clev. St. L. Rev. 29 (1984).
- Note, The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation , 35 Stan. L. Rev. 575 (1983).
- David Rosenberg, The Causal Connection in Mass Exposure Cases: A Public Law Vision of the Tort System, 97 Harv. L. Rev. 849 (1984).
- Joseph Sanders, The Bendectin Litigation: A Case Study in the Life-Cycle of Mass Torts, 43 Hastings L.J. 301 (1992).

# **Reference Guide on Toxicology**

Bernard D. Goldstein Mary Sue Henifin

**Bernard D. Goldstein** M.D., is Director, Environmental & Occupational Health Sciences Institute, Piscataway, New Jersey, and Chairman, Department of Environmental & Community Medicine, UMDNJ - Robert Wood Johnson Medical School, Piscataway, New Jersey. **Mary Sue Henifin**, J.D., M.P.H., is an attorney with Goldshore, Wolf & Lewis, Lawrenceville, New Jersey, and an adjunct professor of public health law, Department of Environmental & Community Medicine, UMDNJ - Robert Wood Johnson Medical School, Piscataway, New Jersey, and School, Piscataway, New Jersey, New Jersey.

# Contents

- I Introduction 185
  - A. Toxicology and the Law 186
  - B. Purpose of the Reference Guide on Toxicology 186
  - C. Toxicological Research Design 186
    - 1. In vivo research 187
      - a. Dose-response relationships 188
      - b. Acute toxicity testing—lethal dose 50 (LD50) 188
      - c. No observable effect level 188
      - d. No threshold model and determination of cancer risk 189
      - e. Maximum tolerated dose (MTD) and chronic toxicity tests 189
    - 2. In vitro research 191
  - D. Extrapolation from Animal and Cell Research to Humans 191
  - E. Safety and Risk Assessment 192
  - F. Toxicology and Epidemiology 194
- II. Expert Qualifications 197
  - A. Does the Proposed Expert Have an Advanced Degree in Toxicology, Pharmacology, or a Related Field? If the Expert Is a Physician, Is He or She Board Certified in a Field Such As Occupational Medicine? 197
  - B. Has the Proposed Expert Been Certified by the American Board of Toxicology, Inc., or Does He or She Belong to a Professional Organization, Such As the Academy of Toxicological Sciences or the Society of Toxicology? 198
  - C. What Other Indicia of Expertise Does the Proposed Expert Possess? 199
- III. Demonstrating an Association Between Exposure and Risk of Disease 201
  - A. On What Species of Animals Was the Compound Tested? What Is Known About the Biological Similarities and Differences Between the Test Animals and Humans? How Do These Similarities and Differences Affect the Extrapolation from Animal Data in Assessing the Risk to Humans? 201

- B. Does Research Show That the Compound Affects a Specific Target Organ? Will Humans Be Affected Similarly? 202
- C. Has the Compound Been the Subject of In Vitro Research, and If So, Can the Findings Be Related to What Occurs In Vivo? 203
- D. What Is Known About the Chemical Structure of the Compound and Its Relationship to Toxicity? 203
- E. Is the Association Between Exposure and Disease Biologically Plausible? 204
- IV. Specific Causal Association Between an Individual's Exposure and the Onset of Disease 205
  - A. Was the Plaintiff Exposed to the Substance, and If So, Did the Exposure Occur in a Manner That Can Result in Absorption into the Body? 206
  - B. Were Other Factors Present That Can Affect the Distribution of the Compound Within the Body? 206
  - C. What Is Known About How Metabolism in the Human Body Alters the Toxic Effects of the Compound? 207
  - D. What Excretory Route Does the Compound Take, and How Does This Effect Its Toxicity? 207
  - E. Does the Temporal Relationship Between Exposure and the Onset of Disease Support or Contradict Causation? 207
  - F. If Exposure to the Substance Is Associated with the Disease, Is There a No Observable Effect or Threshold Level, and If So, Was the Individual Exposed Above the No Observable Effect Level? 208
- V. Medical History 209
  - A. Is the Medical History of the Individual Consistent with the Toxicologist's Expert Opinion Concerning the Injury? 209
  - B. Are the Complaints Specific or Nonspecific? 209
  - C. Do Laboratory Tests Indicate Exposure to the Compound? 210
  - D. What Other Causes Could Lead to the Given Complaint? 210
  - E. Is There Evidence of Interaction with Other Chemicals? 211
  - F. Do Humans Differ in the Extent of Susceptibility to the Particular Compound in Question? Are These Differences Relevant in This Case? 211
  - G. Has the Expert Considered Data That Contradict His or Her Opinion? 212

Glossary of Terms 213

References on Toxicology 219

## I. Introduction

*Toxicology* classically is known as the science of poisons. A modern definition is "the study of the adverse effects of chemical agents on biological systems."<sup>1</sup> While an age-old science, toxicology is still struggling to become a discipline distinct from pharmacology, biochemistry, cell biology, and related fields.

There are three central tenets of toxicology. First, "the dose makes the poison"; this implies that all chemical agents are harmful—it is only a question of *dose*.<sup>2</sup> Even water, if consumed in large quantities, can be *toxic*. Second, many chemical agents produce a specific pattern of toxic effects that are used to establish disease causation.<sup>3</sup> Third, the responses of laboratory animals are useful predictors of toxic responses in humans. Each of these tenets, and their exceptions, are discussed in greater detail below.

The science of toxicology attempts to determine at what doses foreign agents produce their effects. The foreign agents of interest to toxicologists are all chemicals (including foods) and physical agents in the form of radiation, but not living organisms that cause infectious diseases.<sup>4</sup>

The discipline of toxicology provides scientific information relevant to the following questions:

- 1. What hazards, if any, does a chemical or physical agent present to human populations or the environment?
- 2. What degree of risk is associated with chemical *exposure* at any given dose?

Toxicological studies, by themselves, rarely offer direct evidence that a disease in an individual was caused by a chemical exposure. However, toxicology can provide scientific information regarding the increased risk of contracting a disease at any given dose and helps rule out other risk factors for the disease. Toxi-

<sup>1.</sup> Louis J. Casarett & John Doull, Casarett and Doull's Toxicology: The Basic Science of Poisons 3 (Mary O. Amdur et al. eds., 4th ed. 1991).

<sup>2.</sup> A discussion of more modern formulations of this principle, which was articulated by Paracelsus in the sixteenth century, may be found in Ellen K. Silbergeld, *The Role of Toxicology in Causation: A Scientific Perspective,* 1 Cts. Health Sci. & L. 374, 378 (1991).

<sup>3.</sup> Some substances, such as central nervous system toxicants, can produce complex and nonspecific symptoms, such as headaches, nausea, and fatigue.

<sup>4.</sup> Forensic toxicology, a subset of toxicology generally concerned with criminal matters, is not addressed in this reference guide, since it is a highly specialized field with its own literature and methodologies which do not relate directly to toxic tort or regulatory issues.

cological evidence also explains how a chemical causes a disease by describing metabolic, cellular, and other physiological effects of exposure.

### A. Toxicology and the Law

The growing concern about chemical causation of disease is reflected in the public attention devoted to lawsuits alleging toxic torts, as well as litigation concerning the many federal and state regulations related to the release of potentially toxic compounds into the environment. These lawsuits inevitably involve toxicological evidence.

Toxicological evidence frequently is offered in two types of litigation: tort and regulatory proceedings. In tort litigation toxicologists offer evidence that either supports or refutes plaintiffs' claims that their diseases or injuries were caused by chemical exposures. <sup>5</sup> In regulatory litigation toxicological evidence is used to either support or challenge government regulations concerning a chemical or a class of chemicals. In this situation toxicological evidence addresses the question of how exposure affects populations rather than specific causation, and agency determinations are usually subject to deference. <sup>6</sup>

### B. Purpose of the Reference Guide on Toxicology

This reference guide focuses on scientific issues that arise most frequently in toxic tort cases. Where it is appropriate, the reference guide explores the use of regulatory data and how the courts treat such data. This reference guide provides an overview of the basic principles and methodologies of toxicology and offers a scientific context for proffered expert opinion based on toxicological data.<sup>7</sup> The reference guide describes research methods in toxicology and the relationship between toxicology and *epidemiology*, and provides model questions for evaluating the admissibility and strength of an expert's opinion. Following each question is an explanation of the type of information or toxicological data that is offered in response to the question, as well as a discussion of its significance.

### C. Toxicological Research Design

Toxicological research usually involves exposing laboratory animals (*in vivo* research) or cells or tissues (*in vitro* research) to chemicals, monitoring their out-

<sup>5.</sup> See, e.g., Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993).

<sup>6.</sup> See, e.g., Simpson v. Young, 854 F.2d 1429, 1435 (D.C. Cir. 1988) (toxicology research methods approved by the Food and Drug Administration (FDA) given deference by the court).

<sup>7.</sup> The use of toxicological evidence in tort litigation is discussed at length in Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation, 86 Nw. U. L. Rev. 643 (1992). See also Joan E. Bertin & Mary S. Henifin, Science, Law, and the Search for Truth in the Courtroom: Lessons from Daubert v. Merrell Dow, 22 J.L. Med. & Ethics 6 (1994). For a more general discussion of issues that arise in considering expert testimony, see Margaret A. Berger, Evidentiary Framework, in this manual.

comes, such as cellular abnormalities or tumor formation, and comparing them to unexposed control groups. As explained below,<sup>8</sup> the extent to which animal and cell experiments accurately predict human responses to chemical exposures is subject to debate.<sup>9</sup> However, because it is almost always unethical to experiment on humans by exposing them to known doses of suspected poisons, animal toxicological evidence often provides the best scientific information about the risk of disease from a chemical exposure.<sup>10</sup>

Only rarely are humans exposed to chemicals in a manner that permits a quantitative determination of adverse outcomes. This area of toxicological research, known as *clinical toxicology*, may consist of case series, case reports, or even experimental studies in which individuals or groups of individuals have been exposed under circumstances that permit analysis of *dose-response* relationships, mechanisms of action, or other aspects of toxicology. For example, individuals occupationally and environmentally exposed to PCBs prior to prohibitions on their use have been studied to determine the routes of absorption, *distribution*, *metabolism*, and *excretion* for this chemical. Human exposure occurs most frequently in occupational settings where workers are exposed to industrial chemicals like lead or asbestos; however, even under these circumstances, it is usually difficult, if not impossible, to quantify the amount of exposure. Moreover, human populations are exposed to many other chemicals and risk factors, making it difficult to isolate the increased risk of a disease due to any one chemical.<sup>11</sup>

Toxicologists use a relatively wide range of experimental techniques, depending in part on their area of specialization. Some of the more active areas of toxicological research are classes of chemical *compounds*, such as metals; body system effects, such as *neurotoxicology* and *immunotoxicology*; and effects on physiological process, including inhalation toxicology and molecular biology (the study of how chemicals interact with cell molecules). Each of these areas of research include both in vivo and in vitro research.<sup>12</sup>

#### 1. In vivo research

Animal research in toxicology generally falls under two headings: *safety assessment* and classic laboratory science, with a continuum in between. As explained in section I.E., safety assessment is a relatively formal approach in which a chemical's potential for toxicity is tested in vivo or in vitro using standardized tech-

<sup>8.</sup> See infra §§ I.D, III.A.

<sup>9.</sup> The controversy over the use of toxicological evidence in tort cases is described in Silbergeld, *supra* note 2

<sup>10.</sup> See, e.g., Office of Technology Assessment, U.S. Congress, Reproductive Health Hazards in the Workplace 8 (1985).

<sup>12.</sup> See infra §§ I.C.1, I.C.2.

niques often prescribed by regulatory agencies, such as the Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA).

Basic toxicological laboratory research focuses on the mechanisms of action of exogenous agents. It is based on the standard elements of scientific studies, including appropriate experimental design using controls and statistical evaluation. In general, toxicological research attempts to hold all variables constant except for that of the chemical exposure.<sup>13</sup> Any change in the experimental group not found in the control group is assumed to be perturbation caused by the chemical. An important component of toxicological research is dose-response. Thus, most toxicological studies generally use a range of doses for a chemical.<sup>14</sup>

#### a. Dose-response relationships

Animal experiments are conducted to determine the dose-response relationships of a compound by measuring the extent of any observed effect at various doses and diligently searching for a dose that has no measurable physiological effect. This information is useful in understanding the mechanisms of toxicity and extrapolating data from animals to humans.<sup>15</sup>

#### b. Acute toxicity testing—lethal dose 50 (LD50)

To determine the dose-response relationship for a compound, a short-term *lethal dose* 50 (LD50) is derived experimentally. The LD50 is the dose at which a compound kills 50% of laboratory animals within a period of a few days. This easily measured endpoint gradually is being abandoned, in part because recent advances in toxicology have provided more pertinent endpoints, and also because of pressure from animal rights activists to reduce or replace the use of animals in laboratory research.

#### c. No observable effect level (NOEL)

A dose-response study also permits determination of another important characteristic of a chemical—the *no observable effect level* (NOEL).<sup>16</sup> The NOEL sometimes is called a *threshold*, since it is the level above which observable effects in test animals are believed to occur and below which no toxicity is ob-

<sup>13.</sup> Alan Poole & George B. Leslie, A Practical Approach to Toxicological Investigations (1989); Principles and Methods of Toxicology (A. Wallace Hayes ed., 2d ed. 1989); see also discussion on acute, short-term, and long-term toxicity studies and acquisition of data in Frank C. Lu, Basic Toxicology: Fundamentals, Target Organs, and Risk Assessment 77–92 (2d ed. 1991).

<sup>14.</sup> Rolf Hartung, Dose-Response Relationships, in Toxic Substances and Human Risk: Principles of Data Interpretation 29 (Robert G. Tardiff & Joseph V. Rodricks eds., 1987).

<sup>15.</sup> See infra §§ I.D, III.A.

<sup>16.</sup> For example, undiluted acid on the skin can cause a horrible burn. As the acid is diluted to lower and lower concentrations, less and less of an effect occurs until there is a concentration sufficiently low (e.g., one drop in a bathtub of water, or a sample with less than the acidity of vinegar) that no effect occurs. This no observable effect concentration differs from person to person. For example, a baby's skin is more sensitive than that of an adult, and skin that is irritated or broken responds to the effects of an acid at a lower concentration. However, the key point is that there is some concentration that is completely harmless to the skin. See, e.g., Paul Kotin, Dose-Response Relationships and Threshold Concepts, 271 Annals N.Y. Acad. Sci. 22 (1976).

served.<sup>17</sup> Of course, since the NOEL is dependent on the ability to observe the effect, the level is sometimes lowered once more sophisticated methods of detection are developed, particularly for central nervous system effects.

#### d. No threshold model and determination of cancer risk

Certain mutational events, such as those leading to cancer and some inherited disorders, are believed to occur without any threshold. In theory, the cancer-specific alteration in the genetic material of the cell can be produced by any one molecule of the mutational agent. The *no threshold model* led to the development of the *one hit theory* of cancer risk, in which each molecule of a chemical has some finite possibility of producing the mutation that leads to cancer. This risk is very small, since it is unlikely that any one molecule of a potentially cancer-causing agent will reach that one particular spot in a specific cell and result in the change that then eludes the body's defenses and leads to a clinical case of cancer. However, the risk is not zero. The same model also can be used to predict the risk of inheritable mutational events.<sup>18</sup>

#### e. Maximum tolerated dose (MTD) and chronic toxicity tests

Another type of study uses different doses of a chemical agent to establish what is known as the *maximum tolerated dose* (MTD) (the highest dose that does not cause death or significant overt toxicity). The MTD is important because it enables researchers to calculate the dose of a chemical that an animal can be exposed to without reducing its life span, thus permitting evaluation of the *chronic* 

18. For further discussion of the no threshold model of carcinogenesis, see Office of Technology Assessment, U.S. Congress, Assessment of Technologies for Determining the Cancer Risks from the Environment (1981); Gary M. Williams & John H. Weisburger, *Chemical Carcinogenesis, in Casarett and Doull's Toxicology: The Basic Science of Poisons, supra note 1, at 127.* 

The no threshold model, as adopted by the Occupational Safety and Health Administration (OSHA) in its regulation of workplace carcinogens, has been upheld. Public Citizen Health Research Group v. Tyson, 796 F.2d 1479, 1498 (D.C. Cir. 1986) (as set forth in 29 C.F.R. § 1990.143(h) (1985), "no determination will be made that a 'threshold' or 'no effect' level of exposure can be established for a human population exposed to carcinogens in general, or to any specific substance"), *clarified sub nom*. Public Citizen Health Research Group v. Brock, 823 F.2d 626 (D.C. Cir. 1987).

While the one hit model explains the response to most carcinogens, there is accumulating evidence that for certain cancers there is in fact a multistage process, and that some cancer-causing agents act through nonmutational processes, so-called *epigenetic* or nongenotoxic agents. Committee on Risk Assessment Methodology, National Research Council, Issues in Risk Assessment 34–35, 187, 198–201 (1993). For example, the multistage cancer process may explain the carcinogenicity of benzo(a)pyrene (produced by the combustion of hydrocarbons such as oil) and chlordane (a termite pesticide). On the other hand, nonmutational responses to asbestos cause its carcinogenic effect. What the appropriate mathematical model is to depict the dose-response relationship for such an agent is still a matter of debate. *Id.* at 197–201.

<sup>17.</sup> The significance of the NOEL was relied on by the court in Graham v. Canadian Nat'l Ry. Co., 749 F. Supp. 1300 (D. Vt. 1990), in granting judgment for defendants. The court found the defendant's expert, a medical toxicologist, persuasive. The expert testified that plaintiffs' injuries could not have been caused by herbicides, since their exposure was well below the reference dose, which he calculated by taking the NOEL and decreasing it by a safety factor to ensure no human effect. For additional background on the concept of NOEL, see Robert G. Tardiff & Joseph V. Rodricks, *Comprehensive Risk Assessment, in* Toxic Substances and Human Risk: Principles of Data Interpretation, *supra* note 14, at 391.

effects of exposure.<sup>19</sup> These studies last about two years depending on the species.

Chronic toxicity tests evaluate carcinogenicity or other types of toxic effects. Federal regulatory agencies frequently require lifetime carcinogenicity studies on both sexes of two species, usually rats and mice. A standard pathological evaluation is done on the tissues of animals that died during the study and those that are sacrificed at the conclusion of the study.

The rationale for using the MTD in chronic toxicity tests, such as *carcinogenicity bioassays*, often is misunderstood. It is preferable to use realistic doses of *carcinogens* in all animal studies. However, this leads to a significant loss of statistical power, thereby limiting the ability of the test to detect carcinogens or other toxic compounds. Consider the possibility of a chemical in which a realistic dose causes a tumor in 1 in 100 laboratory animals. If the lifetime background incidence without exposure to the chemical is 6 in 100 animals, a toxicological test involving 100 control animals and 100 exposed animals who were fed the realistic dose would reveal 6 control animals and 7 exposed animals with the cancer. A researcher may not detect this difference using conventional statistical tests. However, if the study started with ten times the realistic dose, the researcher would get 16 cases in the exposed group and 6 cases in the control group, a significant difference that is unlikely to be overlooked.

Unfortunately, even this example does not demonstrate the difficulties of determining risk.<sup>20</sup> Regulators are responding to public concern about cancer by regulating risks of 1 in 1 million—not 1 in 100 as in the example given above. To test risks of 1 in 1 million, a researcher would have to either increase the lifetime dose from 10 times to 100,000 times the realistic dose or expand the numbers of animals under study into the millions. However, increases of this magnitude are beyond the world's testing capabilities and are also prohibitively expensive. Inevitably, then, animal studies must trade statistical power for *extrapolation* from higher doses to lower doses.

Accordingly, proffered toxicological expert opinion on potentially cancercausing chemicals almost always is based on review of research studies that extrapolate from animal experiments involving doses significantly higher than that to which humans are exposed.<sup>21</sup> Such extrapolation is accepted in the regulatory

<sup>19.</sup> Even the determination of the MTD can be fraught with controversy. *See*, *e.g.*, Simpson v. Young, 854 F.2d 1429, 1431 (D.C. Cir. 1988) (petitioners unsuccessfully argued that the FDA improperly certified color additive blue number two dye as safe because researchers failed to administer the MTD to research animals, as required by FDA protocols). *See also* David P. Rall, *Laboratory and Animal Toxicity and Carcinogenesis Testing: Underlying Concepts, Advantages and Constraints*, 534 Annals N.Y. Acad. Sci. 78 (1988); Frank B. Cross, Environmentally Induced Cancer and the Law: Risks, Regulation, and Victim Compensation 54–57 (1989).

<sup>20.</sup> See, e.g., Committee on Risk Assessment Methodology, National Research Council, supra note 18, at 43–51.

<sup>21.</sup> See, e.g., Human Risk Assessment: The Role of Animal Selection and Extrapolation (M. Val Roloff ed., 1987).

arena. However, in toxic tort cases, experts use additional background information<sup>22</sup> to offer opinions about disease causation and risk.<sup>23</sup>

#### 2. In vitro research

In vitro research concerns the effects of a chemical on cells, bacteria, body organs, or embryos. Thousands of in vitro toxicological tests have been described in the scientific literature. Many tests are for *mutagenesis* in bacterial or mammalian systems. There are short-term in vitro tests proposed for just about every physiological response and every organ system, such as perfusion tests and DNA studies. Relatively few of the tests described in the research literature have been validated by many different laboratories or compared with outcomes in animal studies to determine if they are predictive of whole animal toxicity.<sup>24</sup>

Criteria of reliability for in vitro tests include the following: (1) whether the test has come through a published protocol in which many laboratories used the same in vitro method on a series of unknown compounds prepared by a reputable organization (such as the National Institutes of Health (NIH) or the International Agency for Research on Cancer (IARC)) to determine if the test consistently and accurately measures toxicity; (2) whether the test has been adopted by a U.S. or international regulatory body; and (3) whether it is predictive of in vivo outcomes related to the same cell or *target organ* system.

#### D. Extrapolation from Animal and Cell Research to Humans

Two types of extrapolation must be considered: from animal data to humans and from higher doses to lower doses. In qualitative extrapolation one can usually rely on the fact that a compound causing an effect in one mammalian species will cause it in another species. If a heavy metal such as mercury causes kidney toxicity in laboratory animals, it will almost certainly do so at some dose in humans. However, the dose at which mercury causes this effect in laboratory animals is modified by many internal factors, and the exact *dose-response curve* may be different from that of humans. Through the study of factors that modify the toxic effects of chemicals, including absorption, distribution, metabolism, and excretion, researchers can improve the ability to extrapolate from laboratory animals to humans and from higher to lower doses.<sup>25</sup>

<sup>22.</sup> See infra §§ IV, V.

<sup>23.</sup> Policy arguments concerning extrapolation from low doses to high doses are explored in Troyen A. Brennan & Robert F. Carter, Legal and Scientific Probability of Causation of Cancer and Other Environmental Disease in Individuals, 10 J. Health Pol. Pol'y & L. 33 (1985).

<sup>24.</sup> See generally In Vitro Toxicity Testing: Applications to Safety Evaluation (John M. Frazier ed., 1992); In Vitro Methods in Toxicology (C. K. Atterwill & C. E. Steele eds., 1987) (discussion of the strengths and weaknesses of specific in vitro tests).

<sup>25.</sup> For example, benzene undergoes a complex metabolic sequence that results in toxicity to the bone marrow in all species, including humans. Robert Snyder et al., *The Toxicology of Benzene*, 100 Envtl. Health Persp. 293 (1990). The exact metabolites responsible for this bone marrow toxicity are the subject of much interest but remain incompletely known. Mice are more susceptible to benzene than rats. If researchers could

Mathematical depiction of the process by which an external dose moves through various compartments in the body until it reaches the target organ is often called physiologically based *pharmacokinetics*. Regulatory agencies are using research into factors causing differences in *target organ doses* for laboratory animals and humans after exposure to the same external doses to improve extrapolation in the risk-assessment process.<sup>26</sup>

Extrapolation from studies in nonmammalian species requires sufficient information on similarities in absorption, distribution, metabolism, and excretion; quantitative determinations of human toxicity based on in vitro studies usually are not considered appropriate. As discussed in section I.F, reliance on in vitro data for elucidating mechanisms of toxicity is more persuasive where positive human epidemiological data also exist.

### E. Safety and Risk Assessment

Toxicological expert opinion also relies on formal safety and *risk assessments*. Safety assessment is the area of toxicology relating to the testing of chemicals and drugs for toxicity. It is a relatively formal approach in which the potential for toxicity of a chemical is tested in vivo or in vitro using standardized techniques. The protocols for such studies usually are developed through scientific consensus and are subject to oversight by governmental regulators or other watchdog groups.

After a number of bad experiences, including outright fraud, the government imposed a code on industrial and contract laboratories involved in safety assessment. Known as *Good Laboratory Practice* (GLP), this code governs many aspects of laboratory standards, including such details as the number of animals per cage and the handling of tissue specimens.<sup>27</sup> Although both the FDA and

determine the differences in metabolism of benzene between mice and rats, they would have a useful clue into which portion of the metabolic scheme is responsible for benzene toxicity to the bone marrow. See, e.g., Curtis D. Klaassen & Karl Rozman, Absorption, Distribution, and Excretion of Toxicants, in Casarett and Doull's Toxicology: The Basic Science of Poisons, supra note 1, at 50; I. Glenn Sipes & A. Jay Gandolfi, Biotransformation of Toxicants, in Casarett and Doull's Toxicology: The Basic Science of Poisons, supra note 1, at 88.

<sup>26.</sup> For an analysis of methods used to extrapolate from animal toxicity data to human health effects, see, e.g., Robert E. Menzer, Selection of Animal Models for Data Interpretation, in Toxic Substances and Human Risk: Principles of Data Interpretation, supra note 14, at 133; Thomas J. Slaga, Interspecies Comparisons of Tissue DNA Damage, Repair, Fixation and Replication, 77 Envtl. Health Persp. 73 (1988); Lorenzo Tomatis, The Predictive Value of Rodent Carcinogenicity Tests in the Evaluation of Human Risks, 19 Ann. Rev. Pharmacol. & Toxicol. 511 (1979); Willard J. Visek, Issues and Current Applications of Interspecies Extrapolation of Carcinogenic Potency as a Component of Risk Assessment, 77 Envtl. Health Persp. 49 (1988); Gary P. Carlson, Factors Modifying Toxicity, in Toxic Substances and Human Risk: Principles of Data Interpretation, supra note 14, at 47; Michael D. Hogan & David G. Hoel, Extrapolation to Man, in Principles and Methods of Toxicology, supra note 13, at 879; James P. Leape, Quantitative Risk Assessment in Regulation of Environmental Carcinogens, 4 Harv. Envtl. L. Rev. 86 (1980).

<sup>27.</sup> A dramatic case of fraud involving a toxicology laboratory that performed tests to assess the safety of consumer products is described in United States v. Keplinger, 776 F.2d 678 (7th Cir. 1985), *cert. denied*, 476 U.S. 1183 (1986). Keplinger and the other defendants in this case were toxicologists who were convicted of falsifying data on product safety by underreporting animal morbidity and mortality and omitting negative data and conclusions from their reports.

the EPA also have published good laboratory practice standards, <sup>28</sup> major differences exist in the required procedures for testing drugs and environmental chemicals. Federal law requires and specifies both efficacy and safety testing of drugs in humans and animals. Carefully controlled clinical trials using doses within the expected therapeutic range are required for premarket testing of drugs. This is because exposures to prescription drugs are carefully controlled and do not exceed specified ranges. However, in the case of environmental chemicals and agents, no premarket testing in humans is required. Moreover, since exposures are less predictable, a wider range of doses usually is given in the animal tests. Finally, since exposures to environmental chemicals may continue over the lifetime and affect both young and old, test designs called *lifetime bioas*says have been developed in which relatively high doses are given to experimental animals. Interpretation of results requires extrapolation from animals to humans, from high to low doses, and from short exposures to multiyear estimates. It must be emphasized that less than 1% of the 60,000-75,000 chemicals in commerce have been subjected to a full safety assessment, and only 10%-20% have any toxicological data at all.

Risk assessment is an approach increasingly used by regulatory agencies to estimate and compare the risks of hazardous chemicals and to assign priority for avoiding their adverse effects.<sup>29</sup> The National Academy of Sciences defines four components of risk assessment: *hazard identification*, dose-response estimation, exposure assessment, and *risk characterization*.<sup>30</sup>

Although risk assessment is not an exact measurement, it should be viewed as a useful estimate on which policy decision making can be based. In recent years, codification of the methodology used to assess risk has increased confidence that the process can be reasonably free of bias; however, significant controversy remains, particularly when generally conservative default assumptions are used where limited actual data are available.<sup>31</sup>

While risk assessment information about a chemical can be somewhat useful in a toxic tort situation, at least in terms of setting reasonable boundaries as to the likelihood of causation, the impetus for the development of risk assessment has been the regulatory process, which has different goals.<sup>32</sup> Because of the

<sup>28.</sup> See, e.g., 40 C.F.R. § 160 (1989); Lu, supra note 13, at 89.

<sup>29.</sup> Committee on Risk Assessment Methodology, National Research Council, supra note 18, at 1.

<sup>30.</sup> National Research Council, Risk Assessment in the Federal Government: Managing the Process (1983). See also Bernard D. Goldstein, Risk Assessment/Risk Management Is a Three-Step Process: In Defense of EPA's Risk Assessment Guidelines, 7 J. Am. C. Toxicol. 543 (1988); Bernard D. Goldstein, Risk Assessment and the Interface Between Science and Law, 14 Colum. J. Envtl. L. 343 (1989).

<sup>31.</sup> An example of conservative default assumptions can be found in Superfund risk assessment. The EPA has determined that Superfund sites should be cleaned up to reduce cancer risk from between 1 in 10,000 to 1 in 1,000,000. A number of assumptions can go into this calculation, including conservative assumptions about intake, exposure frequency and duration, and cancer potency factors for the chemicals at the site. *See, e.g.*, Robert H. Harris & David E. Burmaster, *Restoring Science to Superfund Risk Assessment*, 6 Toxics L. Rep. (BNA) 1318 (March 25, 1992).

<sup>32.</sup> See, e.g., Steven Shavell, Liability for Harm Versus Regulation of Safety, 13 J. Legal Stud. 357 (1984). Risk assessment has been heavily criticized on a number of grounds. The major argument of industry has been

necessarily conservative assumptions in areas of uncertainty and the use of default assumptions where there are limited data, risk assessments intentionally encompass the upper range of possible risks.

### F. Toxicology and Epidemiology

Epidemiology is the study of the incidence and distribution of disease in human populations. Clearly, both epidemiology and toxicology have much to offer in elucidating the causal relationship between chemical exposure and disease.<sup>33</sup> These sciences often go hand in hand in assessing the risks of chemical exposure without artificial distinctions being drawn between the two fields. However, while courts generally rule epidemiological expert opinion admissible, admissibility of toxicological expert opinion has been more controversial because of uncertainties regarding extrapolation from animal and in vitro data to humans. This particularly has been the case where relevant epidemiological research data exist. However, since animal and cell studies permit researchers to isolate the effects of exposure to a single chemical or to known mixtures, toxicological evidence offers unique information concerning dose-response relationships, mechanisms of action, specificity of response, and other information relevant to the assessment of causation.<sup>34</sup>

Even though there is little toxicological data on many of the 75,000 compounds in general commerce, there is far more information from toxicological studies than from epidemiological studies.<sup>35</sup> It is much easier, and more economical, to expose an animal to a chemical or to perform in vitro studies than it

33. See Linda A. Bailey et al., Reference Guide on Epidemiology § IV, in this manual.

that it is overly conservative, and thus greatly overstates the actual risk. The rationale for conservatism in part is the prudent public health approach of "above all, do no harm." In other cases, including cancer risk, the conservative approach is used because it is sometimes more feasible to extrapolate to a plausible upper boundary for a risk estimate than it is to estimate a point of maximum likelihood. For a sample of the debate over risk assessment, see, e.g., Bruce N. Ames & Lois S. Gold, *Too Many Rodent Carcinogens: Mitogenesis Increases Mutagenesis*, 249 Science 970 (1990); Jean Marx, *Animal Carcinogen Testing Challenged*, 250 Science 743 (1990); Philip H. Abelson, *Incorporation of a New Science into Risk Assessment*, 250 Science 1497 (1990); Frederica P. Perera, *Letter to the Editor: Carcinogens and Human Health*, *Part* 1, 250 Science 1644 (1990); Bruce N. Ames & Lois S. Gold, *Response*, 250 Science 1645 (1990); David P. Rall, *Letter to the Editor: Carcinogens and Human Health*, *Part* 2, 251 Science 10 (1991); Bruce N. Ames & Lois S. Gold, *Response*, 251 Science 12 (1991); John C. Bailar III et al., *One-Hit Models of Carcinogenesis: Conservative or Not?*, 8 Risk Analysis 485 (1988).

<sup>34.</sup> Both commonalities and differences between animal and human responses to chemical exposures were recognized by the court in International Union, United Auto., Aerospace & Agric. Implement Workers of Am. v. Pendergrass, 878 F.2d 389, 394 (D.C. Cir. 1989). In reviewing the results of both epidemiological and animal studies on formaldehyde, the court stated: "humans are not rats, and it is far from clear how readily one may generalize from one mammalian species to another. In light of the epidemiological evidence [of carcinogenicity] that was not the main problem. Rather it was the absence of data at low levels." The court remanded the matter to OSHA to reconsider its findings that formaldehyde presented no specific carcinogenic risk to workers at exposure levels of 1 part per 1,000,000 or less.

<sup>35.</sup> National Research Council, supra note 30. See also Lorenzo Tomatis et al., Evaluation of the Carcinogenicity of Chemicals: A Review of the Monograph Program of the International Agency for Research on Cancer, 38 Cancer Res. 877, 881 (1978); National Research Council, Toxicity Testing: Strategies to Determine Needs and Priorities (1984); Myra Karstadt & Renee Bobal, Availability of Epidemiologic Data on Humans Exposed to Animal Carcinogens, 2 Teratogenesis, Carcinogenesis & Mutagenesis 151 (1982).

is to perform epidemiological studies.<sup>36</sup> This difference in data availability is evident even for cancer-causation, for which toxicological study is particularly expensive and time-consuming. Of the perhaps two dozen chemicals that reputable international authorities agree are known human carcinogens based on positive epidemiological studies, arsenic is the only one not known to be an animal carcinogen. Yet, there are more than 100 known animal carcinogens for which there is no valid epidemiological database, in addition to a handful of others for which the epidemiological database is equivocal (e.g., butadiene).<sup>37</sup> To clarify any findings, regulators can require a repeat of an equivocal two-year animal toxicological study or the performance of additional laboratory studies in which animals deliberately are exposed to the chemical. Such deliberate exposure is not possible in humans. As a general rule, equivocally positive epidemiological studies reflect prior workplace practices leading to relatively high levels of exposure to a limited number of individuals that, fortunately, in most cases no longer occur. Thus, an additional prospective epidemiological study often is not possible, and even the ability to do retrospective studies is constrained by the passage of time.

<sup>36.</sup> See Linda A. Bailey et al., Reference Guide on Epidemiology § II, in this manual.

<sup>37.</sup> Rall, supra note 32.

## II. Expert Qualifications

The basis of the toxicologist's expert opinion is a thorough review of the research literature and treatises concerning effects of exposure to the chemical at issue, applied to the specific case. To arrive at an opinion, the expert assesses the strengths and weaknesses of the research studies. The expert also bases an opinion on fundamental concepts of toxicology relevant to understanding the actions of chemicals in biological systems.

As the following series of questions indicates, no single academic degree, research specialty, or career path qualifies an individual as an expert in toxicology. Toxicology is a heterogeneous field. A number of indicia of expertise, however, can be explored, relevant to both admissibility and weight of the proffered expert opinion.

### A. Does the Proposed Expert Have an Advanced Degree in Toxicology, Pharmacology, or a Related Field? If the Expert Is a Physician, Is He or She Board Certified in a Field Such As Occupational Medicine?

A graduate degree in toxicology demonstrates that the proposed expert has a substantial background in the basic issues and tenets of toxicology. Many universities have established graduate programs in toxicology only recently. These programs are administered by the faculties of medicine, pharmacology, pharmacy, or public health.

However, given the relatively recent establishment of toxicology programs, a number of highly qualified toxicologists are physicians or hold doctoral degrees in related disciplines (e.g., pharmacology, biochemistry, environmental health, or industrial hygiene). For a person with this type of background, a single course in toxicology is unlikely to provide sufficient background to develop an expertise in the field.

A proposed expert should be able to demonstrate an understanding of the discipline of toxicology, including statistics, toxicological research methods, and disease processes. A physician without particular training or experience in toxicology is unlikely to have sufficient background to evaluate the strengths and weaknesses of toxicological research. Most practicing physicians have little knowledge of environmental and occupational medicine. Generally, physicians are quite knowledgeable as to identification of effects, and subspecialty physicians may have particular knowledge of a cause-and-effect relationship (e.g., pulmonary physicians have knowledge of the relationship between asbestos exposure and asbestosis). However, most physicians have little training in chemical toxicology and lack an understanding of exposure assessment and dose-response relationships. An exception is physicians who are certified in medical toxicology by the American Board of Medical Toxicology based on their substantial training in toxicology and successful completion of rigorous examinations.

Some physicians who are occupational health specialists also have training in toxicology. Of the occupational physicians practicing today, only a small group, perhaps 1,000, has successfully completed the board examination in occupational medicine, which contains some questions about chemical toxicology. <sup>38</sup>

B. Has the Proposed Expert Been Certified by the American Board of Toxicology, Inc., or Does He or She Belong to a Professional Organization, Such As the Academy of Toxicological Sciences or the Society of Toxicology?

As of December 1989, 991 individuals from nine countries have received board certification from the American Board of Toxicology, Inc. To sit for the examination, which has a pass rate of 67%, the candidate must be involved full-time in the practice of toxicology, including designing and managing toxicological experiments or interpreting results and translating them to identify and solve human and animal health problems. To become certified, the candidate must pass all three parts of the examination within two years. Diplomats must be recertified through examination every five years.

The Academy of Toxicological Sciences (ATS) was formed to provide credentials in toxicology through peer review only. They do not administer examinations for certification.

The Society of Toxicology (SOT), the major professional organization for the field of toxicology, was formed in 1960 and has grown dramatically in recent

<sup>38.</sup> Another group of physicians, known as *clinical ecologists*, has offered opinions regarding *multiple chemical hypersensitivity* and immune system responses to chemical exposures. These physicians generally have a background in the field of allergy, not toxicology, and their theoretical approach is derived in part from classic concepts of allergic responses and immunology. Clinical ecologists often belong to the American Academy of Environmental Medicine.

In Sterling v. Velsicol Chem. Corp., 855 F.2d 1188, 1208–09 (6th Cir. 1988), the court considered the admissibility of expert opinions based on clinical ecology theories. The court ruled the opinions inadmissible, finding that the experts "never personally examined or interviewed plaintiffs, nor performed the requisite medical tests." But see Elam v. Alcolac, Inc., 765 S.W.2d 42, 86 (Mo. Ct. App. 1988), cert. denied, 493 U.S. 817 (1989) (expert opinion based on clinical ecology theories admissible). See also Gregg L. Spyridon, Scientific Evidence vs. "Junk Science"—Proof of Medical Causation in Toxic Tort Litigation: The Fifth Circuit "Fryes" a New Test (Christophersen v. Allied Signal Corp.), 61 Miss. L.J. 287, 295–96 (1991); California Medical Ass'n Scientific Bd. Task Force on Clinical Ecology, Clinical Ecology—A Critical Appraisal, 144 W.J. Med. 239 (1986).

years; it currently has 2,944 members.<sup>39</sup> It has reasonably strict criteria for membership. Qualified people must have conducted and published original research in some phase of toxicology (excluding graduate work) or be generally recognized as expert in some phase of toxicology and be approved by a majority vote of the board of directors. Many environmental toxicologists who meet these qualifications belong to SOT.

Physician toxicologists can join the American College of Medical Toxicology and the American Academy of Clinical Toxicologists. Other organizations in the field include the American College of Toxicology, which has less stringent criteria for membership, the International Society of Regulatory Toxicology and Pharmacology, and the Society of Occupational and Environmental Health. The last two organizations require only the payment of dues for membership.

## C. What Other Indicia of Expertise Does the Proposed Expert Possess?

The success of academic scientists in toxicology, as in other biomedical sciences, usually is measured by the following types of criteria: the quality and number of peer-reviewed publications, the ability to compete for grants, service on scientific advisory panels, and university appointments.

Publication of articles in peer-reviewed journals indicates an expertise in toxicology. The number of articles, their topics, and whether the individual is the principal author are important factors in determining the expertise of a toxicologist.<sup>40</sup>

Most grants from government agencies and private foundations are highly competitive. Successful competition for funding and publication of the findings indicate competence in an area.

Selection for local, national, and international regulatory advisory panels usually implies a degree of recognition in the field. Examples include panels convened by the EPA, the FDA, the World Health Organization (WHO), and the International Agency for Research on Cancer (IARC). Recognized industrial organizations, including the American Petroleum Institute, Electric Power Research Institute, and Chemical Industry Institute of Toxicology, and public interest groups, such as the Environmental Defense Fund and the Natural Resources Defense Council, employ toxicologists directly and as consultants and enlist academic toxicologists to serve on advisory panels. Because of a growing interest in environmental issues, the demand for scientific advice has outgrown

<sup>39.</sup> There are currently six specialty sections of SOT that represent the different types of research needed to understand the wide range of toxic effects associated with chemical exposures. These sections are mecha nisms, molecular biology, inhalation toxicology, metals, neurotoxicology, and immunotoxicology.

<sup>40.</sup> Examples of reputable, peer-reviewed journals are Journal of Toxicology and Environmental Health; Toxicology and Applied Pharmacology; Science; British Journal of Industrial Medicine; Clinical Toxicology; Archives of Environmental Health; Journal of Occupational Medicine; Annual Review of Pharmacology and Toxicology; Teratogenesis, Carcinogenesis and Mutagenesis; Fundamental and Applied Toxicology; Inhalation Toxicology; Biochemical Pharmacology; Toxicology Letters; Environmental Research; Environmental Health Perspectives; and American Journal of Industrial Medicine.

the supply of available toxicologists. It is thus common for reputable toxicologists

to serve on advisory panels. Finally, a faculty appointment in toxicology, risk assessment, or a related field signifies an expertise in that area.

# III. Demonstrating an Association Between Exposure and Risk of Disease

Once the expert has been qualified, he or she is expected to offer an opinion on whether the plaintiff's disease was caused by exposure to a chemical. To do so, the expert relies on the principles of toxicology to provide a scientifically valid methodology for establishing causation and then applies the methodology to the facts of the case.

An opinion on causation should be premised on three preliminary assessments. First, the toxicologist should analyze whether the disease can be related to chemical exposure by a *biologically plausible* theory. Second, the expert should examine if the plaintiff was exposed to the chemical in a manner that can lead to absorption into the body. Finally, the expert should offer an opinion as to whether the dose to which the plaintiff was exposed is sufficient to cause the disease.

The following questions help evaluate the strengths and weaknesses of toxicological evidence.

A. On What Species of Animals Was the Compound Tested? What Is Known About the Biological Similarities and Differences Between the Test Animals and Humans? How Do These Similarities and Differences Affect the Extrapolation from Animal Data in Assessing the Risk to Humans?

All living organisms share a common biology that leads to marked similarities in the responsiveness of subcellular structures to toxic agents. Among mammals, more than sufficient common organ structure and function readily permits the extrapolation from one species to another in most cases. Through the study of factors that modify the toxic effects of chemicals, including absorption, distribution, metabolism, and excretion, the ability to extrapolate from laboratory animals to humans can improve.<sup>41</sup>

<sup>41.</sup> See, e.g., supra notes 25–26 and accompanying text; Edward J. Calabrese, Principles of Animal Extrapolation (1983); Human Risk Assessment: The Role of Animal Selection and Extrapolation, *supra* note 21.

The expert should review similarities and differences in absorption, distribution, metabolism, and excretion in the animal species in which the compound has been tested and in humans. This should form the basis of the opinion as to whether extrapolation between animals and humans is warranted.

In general, there is an overwhelming similarity in the biology of all living things and a particularly good relationship among mammals. Of course, laboratory animals differ from humans in many ways. For example, rats do not have gall bladders. Thus, rat data would not be pertinent to the possibility that a compound produces human gall bladder toxicity.<sup>42</sup>

# B. Does Research Show That the Compound Affects a Specific Target Organ? Will Humans Be Affected Similarly?

Some chemical and physical agents demonstrate specific effects at a particular dose. The organ specificity of a toxic chemical may be due to absorption, distribution, metabolism, excretion, or organ dysfunction.<sup>43</sup> For example, specificity may reflect the relatively high level in an organ of an enzyme system capable of metabolizing a parent compound to a toxic metabolite, or it may reflect the relatively low level of an enzyme system capable of detoxifying a compound. An example of the former is liver toxicity caused by inhaled carbon tetrachloride, for which there is extensive metabolism to a toxic intermediate within the liver but relatively little such metabolism in the lung.<sup>44</sup>

Some chemicals, on the other hand, may cause nonspecific effects or even multiple effects. Liver toxins may interfere with the role of red blood cells in the metabolism of certain drugs and release cellular enzymes into blood, leading to a number of nonspecific effects. Lead is an example of a *toxic agent* that affects many organ systems, including red blood cells, the central and peripheral nervous systems, reproductive systems, and the kidneys, leading to cardiovascular effects.

The basis of specificity usually reflects the function of individual organs. For example, the thyroid is particularly susceptible to radioactive iodine in atomic fallout because thyroid hormone is unique within the body in that it requires iodine. Through evolution a very efficient and specific mechanism has developed

44. Brian Jay Day et al., Potentiation of Carbon Tetrachloride-Induced Hepatotoxicity and Pneumotoxicity by Pyridine, 8 J. Biochemical Toxicol. 11 (1993).

<sup>42.</sup> See, e.g., Table 14-1: Some Biochemical/Physiological/Morphological Differences of Potential Toxicological Significance Between Rats and Humans, in Human Risk Assessment: The Role of Animal Selection and Extrapolation, supra note 21, at 583–89. Species differences producing a qualitative difference in response to xenobiotics are well known. Sometimes understanding the mechanism underlying the species difference can allow prediction of whether the effect will occur in humans. Thus, carbaryl, an insecticide commonly used, among other things, for gypsy moth control, produces fetal abnormalities in dogs but not in hamsters, mice, rats, and monkeys. Dogs lack the specific enzyme involved to metabolize carbaryl; the other species tested all have this enzyme, as do humans. On this basis, it has been reasoned that humans are not at risk for fetal malformations produced by carbaryl.

<sup>43.</sup> See infra § IV.

which concentrates any absorbed iodine preferentially within the thyroid, thus rendering the thyroid particularly at risk from radioactive iodine. In a test tube the radiation from radioactive iodine can affect the genetic material obtained from any cell in the body, but in the intact laboratory animal or human, only the thyroid is at risk.

### C. Has the Compound Been the Subject of In Vitro Research, and If So, Can the Findings Be Related to What Occurs In Vivo?

Cellular and tissue culture research can be particularly helpful in identifying mechanisms of toxic action and potential target organ toxicity. The major barrier to use of in vitro results is the frequent inability to relate dosages that cause cellular toxicity to whole animal toxicity. In many critical areas, knowledge that permits such extrapolation is lacking.<sup>45</sup> Nevertheless, the ability to quickly test new products through in vitro tests, using human cells, makes these tests invaluable "early warning systems" for toxicity.

# D. What Is Known About the Chemical Structure of the Compound and Its Relationship to Toxicity?

Understanding the structural aspects of chemical toxicology has led to the use of *structure activity relationships* (SAR) as a formal method of predicting toxicity of new chemicals. This technique compares the chemical structure of compounds with known toxicity to the chemical structure of compounds with unknown toxicity. Toxicity then is estimated based on molecular similarities between the two compounds. While SAR is used extensively by the EPA in testing many new chemicals required to be tested under the registration requirements of the Toxic Substances Control Act (TSCA), its reliability has a number of limitations.<sup>46</sup>

<sup>45.</sup> In Vitro Toxicity Testing: Applications to Safety Evaluation, supra note 24, at 8.

<sup>46.</sup> For example, benzene and alkyl benzenes, which include toluene, xylene, and ethyl benzene, share a similar chemical structure and are common bulk chemicals and constituents of gasoline. SAR works exceptionally well in predicting the acute central nervous system anesthetic-like effects of these compounds; the slight difference in dose-response is readily explainable by the interrelated factors of chemical structure, vapor pressure, and lipid solubility (the brain is highly lipid). National Research Council, The Alkyl Benzenes (1981). However, among these closely related compounds it is only benzene that produces damage to the bone marrow and leukemia. This is because of the specific metabolic products of benzene, a specificity so great that when the closely related compound toluene is administered with benzene to laboratory animals it actually protects against bone marrow toxicity.

Expert opinion based on SAR has been proffered in a number of cases alleging that fetal exposure to the pregnancy antinausea drug Bendectin resulted in birth defects. Lower courts, applying varying standards, have accepted and rejected expert opinion based on SAR. These cases are analyzed in Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 Hastings L.J. 301 (1992); Ernest J. Getto et al., *The Artification of Science: The Problem of Unscientific "Scientific" Evidence*, 23 Envtl. L. Rep. 10435 (1993); Green, *supra* note 7. See also Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993), which rejected a per se exclusion of SAR, animal data, and reanalyses of previously published epidemiological data where there was negative epidemiological data, and remanded the issue of admissibility to the trial court for reconsideration.

E. Is the Association Between Exposure and Disease Biologically Plausible?

No matter how strong the temporal relationship between exposure and development of disease, or the supporting epidemiological evidence, it is difficult to accept an association between a compound and a health effect where no mechanism can be ascribed by which the chemical exposure leads to the putative effect.

# IV. Specific Causal Association Between an Individual's Exposure and the Onset of Disease

An expert who opines that exposure to a compound caused a person's disease engages in deductive clinical reasoning.<sup>47</sup> In most instances, cancers and other diseases do not wear labels documenting their causation.<sup>48</sup> The opinion is based on an assessment of the individual's exposure, including the amount, the temporal relationship between the exposure and disease, and exposure to other disease-causing factors. This information is then compared to research data on the relationship between exposure and disease. The certainty of the expert's opinion depends on the strength of the research data demonstrating a relation-ship between exposure and the disease at the dose in question and the absence of other disease-causing factors (also known as *confounding factors*).<sup>49</sup>

Particularly problematic are generalizations made in personal injury litigation from regulatory positions. For example, if regulatory standards are discussed in toxic tort cases to provide a reference point for assessing exposure levels, it must be recognized that there is a great deal of variability in the extent of evidence required to support different regulations. <sup>50</sup> The extent of certainty required for regulatory activity for primary pollutants on adverse health consequences to sensitive populations with an adequate margin of safety and with no consideration of economic consequences, while regulatory activity under TSCA clearly asks

47. For an example of deductive clinical reasoning based on known facts about the toxic effects of a chemical and the individual's pattern of exposure, see Bernard D. Goldstein, *Is Exposure to Benzene a Cause of Human Multiple Myeloma?*, 609 Annals N.Y. Acad. Sci. 225 (1990).

48. Research, which is still in the preliminary stages, shows that certain cancers do "wear labels" in the form of DNA adducts and mutational spectra. National Research Council, Biologic Markers in Reproductive Toxicology (1989).

49. Causation issues are discussed in Joseph Sanders, From Science to Evidence: The Testimony on Causation in the Bendectin Cases, 46 Stan. L. Rev. 1 (1993); Troyen A. Brennan, Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation, 73 Cornell L. Rev. 469 (1988); Daniel A. Farber, Toxic Causation, 71 Minn. L. Rev. 1219 (1987); Steve Gold, Note, Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence, 96 Yale L.J. 376 (1986); Orrin E. Tilevitz, Judicial Attitudes Towards Legal and Scientific Proof of Cancer Causation, 3 Colum. J. Envtl. L. 344, 381 (1977); David L. Bazelon, Science and Uncertainty: A Jurist's View, 5 Harv. Envtl. L. Rev. 209 (1981); and William V. Dunlap & E. Michael Thomas, Tort Actions for Cancer: Deterrence, Carcinogenesis, 90 Yale L.J. 840 (1981). See also In re Joint E. & S. Dists. Asbestos Litig., 827 F. Supp. 1014, 1026 (S.D.N.Y. 1993) (under Daubert standards worker's causation evidence insufficient to support jury finding that asbestos exposure caused colorectal cancer).

50. The relevance of regulatory standards to toxic tort litigation is explored in Silbergeld, supra note 2.

for some balance between the societal benefits and risk of new chemicals); (2) the specific endpoint of concern (e.g., consider the concern caused by cancer and adverse reproductive outcomes versus almost anything else); and (3) the societal impact, as evidenced by the different degree of public support for control of an industry versus altering personal automobile use patterns. These three concerns, as well as others, including costs, politics, and the virtual certainty of litigation challenging the regulation, impact on the level of scientific proof required by the regulatory decision maker.

### A. Was the Plaintiff Exposed to the Substance, and If So, Did the Exposure Occur in a Manner That Can Result in Absorption into the Body?

Evidence of exposure is essential in determining the effects of harmful substances. Basically, potential human exposure is measured in one of three ways. First, where direct measurements cannot be made, exposure can be measured by mathematical modeling, in which one uses a variety of physical factors to estimate the transport of the pollutant from the source to the receptor. For example, mathematical models take into account such factors as wind variations to allow calculation of the transport of pollutants (e.g., radioactive iodine from a federal atomic research facility to nearby residential areas). Second, exposure can be measured using direct measurements of the medium in question-air, water, food, or soil. Where the medium of exposure is water, soil, or air, exposure calculations frequently draw on the expertise of hydrogeologists or meteorologists. The third approach directly measures human receptors through some form of biological monitoring, such as blood lead levels or a urinary metabolite, which shows pollutant exposure. Ideally, both environmental testing and biological monitoring are performed; however, this is not always possible, particularly in instances of historical exposure.

The toxicologist, on the other hand, must determine if the individual was exposed to the compound in a manner that can result in absorption into the body. The absorption of the compound is a function of its physiochemical properties, its concentration, and the presence of other agents or conditions that assist or interfere with its uptake. For example, inhaled lead is absorbed almost totally, while ingested lead is taken up only partially into the body. An iron deficiency or low nutritional calcium intake, both common conditions among inner-city children, increases the amount of ingested lead that is absorbed in the gastrointestinal tract and passes into the bloodstream.

# B. Were Other Factors Present That Can Affect the Distribution of the Compound Within the Body?

Once a compound is absorbed into the body through the skin, lungs, or gastrointestinal tract, it is distributed throughout the body through the bloodstream. Thus, the rate of distribution depends on the rate of blood flow to various organs and tissues. Distribution and resulting toxicity are also influenced by other factors, including the dose, route of entry, tissue solubility, lymphatic supplies to the organ, metabolism, and the presence of specific receptors or uptake mechanisms within body tissues.

# C. What Is Known About How Metabolism in the Human Body Alters the Toxic Effects of the Compound?

Metabolism is the alteration of a chemical by bodily processes. It does not necessarily result in less toxic compounds being formed. In fact, many of the organic chemicals that are known human cancer-causing agents require metabolic transformation before they can cause cancer. A distinction often is made between *direct-acting agents*, which cause toxicity without any metabolic conversion, and *indirect-acting agents*, which require metabolic activation before they can produce adverse effects. Metabolism is complex, since a variety of pathways compete for the same agent; some produce harmless metabolites, and others produce toxic agents.<sup>51</sup>

# D. What Excretory Route Does the Compound Take, and How Does This Affect Its Toxicity?

Excretory routes are urine, feces, sweat, saliva, expired air, and lactation. Many inhaled volatile agents are eliminated primarily by exhalation. The excretion of small water soluble compounds is usually through urine. Higher molecular weight compounds are often excreted through the biliary tract into the feces. Certain fat-soluble, poorly metabolized compounds, such as PCBs, may persist in the body for decades, although they can be excreted in the milk fat of lactating women.

# E. Does the Temporal Relationship Between Exposure and the Onset of Disease Support or Contradict Causation?

In most *acute* injuries, there is a short time period between cause and effect. However, in some situations, the length of basic biological processes necessitates a longer period of time between initial exposure and the onset of observable disease. For example, acute myelogenous leukemia, the adult form of acute leukemia, requires one to two years from initial exposure to radiation, benzene, or cancer chemotherapy until the manifestation of a clinically recognizable case of leukemia. A toxic tort claim alleging a shorter time period between cause and

<sup>51.</sup> Courts have explored the relationship between metabolic transformation and carcinogenesis. See, e.g., Stites v. Sundstrand Heat Transfer, Inc., 660 F. Supp. 1516 (W.D. Mich. 1987).

effect is scientifically untenable. Much longer time periods are necessary for the manifestation of solid tumors caused by asbestos.

F. If Exposure to the Substance Is Associated with the Disease, Is There a No Observable Effect or Threshold Level, and If So, Was the Individual Exposed Above the No Observable Effect Level?

Even if an individual was exposed to a chemical, if the level of exposure was below the no observable effect or threshold level, a relationship between the exposure and disease cannot be established. The NOEL is extrapolated from animals to humans by calculating the animal NOEL based on experimental data and decreasing it by a safety factor to ensure no human effect. <sup>52</sup> This analysis, however, is not applied to substances that exert toxicity by causing mutations leading to cancer. Theoretically, any exposure at all to *mutagens* may increase the risk of cancer, although the risk may be very slight. <sup>53</sup>

<sup>52.</sup> See, e.g., supra notes 17–18 and accompanying text. Joseph V. Rodricks & Robert G. Tardiff, *Comprehensive Risk Assessment, in Toxic Substances and Human Risk: Principles of Data Interpretation, supra* note 14, at 391. Joseph V. Rodricks, Calculated Risks 165–70, 193–96 (1992); Lu, *supra* note 13, at 84. 53. See sources cited *supra* note 18.

## V. Medical History

### A. Is the Medical History of the Individual Consistent with the Toxicologist's Expert Opinion Concerning the Injury?

One of the basic and most useful tools in diagnosis and treatment of disease is the patient's medical history. While a thorough, standardized patient information questionnaire would be particularly useful for recognizing the *etiology* or causation of illnesses related to toxic exposures, there is currently no validated or widely used questionnaire that gathers all pertinent information.<sup>54</sup> Nevertheless, it is widely recognized that a thorough medical history involves the questioning and examination of the patient as well as appropriate medical testing. The patient's written medical records should also be examined.

The following information is relevant to a patient's medical history: past and present occupational and environmental history and exposure to toxic agents; lifestyle characteristics (e.g., use of nicotine and alcohol); family medical history (e.g., medical conditions, diseases of relatives); and personal medical history (e.g., present symptoms and results of medical tests as well as past injuries, medical conditions, diseases, surgical procedures, and medical test results).

In some instances, the reporting of symptoms can be in itself diagnostic of exposure to a specific substance, particularly where evaluating acute effects. For example, individuals acutely exposed to organophosphate pesticides report headaches, nausea, and dizziness accompanied by anxiety and restlessness. Other reported symptoms include muscle twitching, weakness, and hypersecretion with sweating, salivation, and tearing.<sup>55</sup>

### B. Are the Complaints Specific or Nonspecific?

Acute exposure to many toxic agents produces a constellation of nonspecific symptoms, such as headaches, nausea, lightheadedness, and fatigue. These types of symptoms are part of human experience and can be triggered by a host of medical and psychological conditions. They are almost impossible to quantify or

<sup>54.</sup> Office of Technology Assessment, U.S. Congress, supra note 10, at 365–89.

<sup>55.</sup> Environmental Protection Agency, Recognition and Management of Pesticide Poisonings (4th ed. 1989).

document beyond the patient's report. Thus, these symptoms can be attributed mistakenly to an exposure to a toxic agent or discounted as unimportant when in fact they reflect a significant exposure.

A careful medical history focuses on the time pattern of symptoms in relation to any exposure and on the constellation of symptoms to determine causation. It is easier to establish causation when a symptom is unusual and rarely is caused by anything other than the suspect chemical (e.g., such rare cancers as hemangiosarcoma, associated with vinyl chloride exposure, and mesothelioma, associated with asbestos exposure). However, many cancers and other conditions are associated with several causative factors, thus complicating proof of causation.

### C. Do Laboratory Tests Indicate Exposure to the Compound?

There are two types of tests: routine tests, which are used in medicine to detect changes in normal body status, and relatively specialized tests, which are used to detect the presence of the chemical or physical agent. For the most part, tests used to demonstrate the presence of a toxic agent are frequently unavailable from clinical laboratories. Even when available from a hospital or a clinical laboratory, a test such as that for carbon monoxide combined to hemoglobin is done so rarely that it may raise concerns as to its accuracy. Other tests, such as the test for blood lead levels, are required for routine surveillance of potentially exposed workers. However, just because a laboratory is certified for testing of blood lead in workers, for which the OSHA action level is 40 micrograms per deciliter ( $\mu$ g/dl), does not necessarily mean that it will give reliable data on blood lead levels at the much lower Centers for Disease Control (CDC) action level of 10  $\mu$ g/dl.

### D. What Other Causes Could Lead to the Given Complaint?

With few exceptions, acute and chronic diseases, including cancer, are either caused by a toxic agent or other agents or conditions. A careful medical history examines the possibility of competing causes or confounding factors for any disease, leading to a *differential diagnosis*. The failure of a physician to elicit such history, or of a toxicologist to pay attention to such a history, leaves open the possibility of competing causes of the injury.<sup>56</sup>

<sup>56.</sup> See, e.g., Bell v. Swift Adhesives, Inc., 804 F. Supp. 157 (S.D. Ga. 1992) (expert's opinion that workplace exposure to methylene chloride caused plaintiff's liver cancer, without ruling out plaintiff's infection with hepatitis B virus, a known liver carcinogen, was insufficient to withstand motion for summary judgment for defendant).

### E. Is There Evidence of Interaction with Other Chemicals?

Simultaneous exposure to different compounds may change the response from that which would be expected from exposure to only one of the compounds.<sup>57</sup> When the effect of multiple agents is that which would be predicted by the sum of the effects of individual agents, it is called an *additive effect*; when it is greater than this sum, it is known as a *synergistic effect*; when one agent causes a decrease in the effect produced by another, the result is termed *antagonism*; and when an agent that by itself produces no effect leads to an enhancement of the effect of another agent, the response is termed *potentiation*.<sup>58</sup>

Three types of toxicological approaches are pertinent to understanding the effects of mixtures of agents. One is based on the standard toxicological evaluation of common commercial mixtures, such as gasoline; the second is from studies in which the known toxicological effect of one agent is used to explore the mechanism of action of another agent, such as using a known specific inhibitor of a metabolic pathway to determine whether the toxicity of a second agent depends on this pathway; and the third is based on an understanding of the basic mechanism of action of the individual components of the mixture, thereby allowing prediction of the combined effect, which can then be tested in an animal model.<sup>59</sup>

### F. Do Humans Differ in the Extent of Susceptibility to the Particular Compound in Question? Are These Differences Relevant in This Case?

Individuals who exercise inhale more than sedentary individuals and therefore are exposed to higher doses of airborne environmental toxins. Similarly, differences in metabolism, which are inherited or caused by external factors, such as the levels of carbohydrates in a person's diet, may result in differences in the delivery of a toxic product to the target organ. <sup>60</sup>

Moreover, for any given level of a toxic agent that reaches a target organ, damage may be greater because of differing responses to allergens. In addition, for any given level of target organ damage, there may be a greater impact on particular individuals. For example, an elderly individual or someone with preexist-

60. Id.

<sup>57.</sup> See, e.g., Edward J. Calabrese, Multiple Chemical Interactions (1991).

<sup>58.</sup> Courts have been called on to consider the issue of synergy. In International Union, United Auto., Aerospace & Agric. Implement Workers of Am. v. Pendergrass, 878 F.2d 389 (D.C. Cir. 1989), the court found that OSHA failed to sufficiently explain its findings that formaldehyde presented no significant carcinogenic risk to workers at exposure levels of 1 part per 1,000,000 or less. The court particularly criticized OSHA's use of a linear low-dose risk curve rather than a risk adverse model, after the agency had described evidence of synergy between formaldehyde and other substances that workers would be exposed to, especially wood dust.

<sup>59.</sup> See, e.g., Calabrese, supra note 57.

ing lung disease is less likely to tolerate a small decline in lung function caused by an air pollutant than is a healthy individual with normal lung function.

A person's level of physical activity, age, sex, and genetic makeup, as well as exposure to therapeutic agents (such as prescription or over-the-counter drugs), affect the metabolism of the compound and hence its toxicity.<sup>61</sup>

### G. Has the Expert Considered Data That Contradict His or Her Opinion?

Multiple avenues of deductive reasoning based on research data lead to scientific acceptance of causation in any field, particularly in toxicology. However, it is also one of the most difficult aspects of causation to describe quantitatively. For example, if animal studies, pharmacological research on mechanisms of toxicity, in vitro tissue studies, and epidemiological research all document toxic effects of exposure to a compound, an expert's opinion about causation in a particular case is much more likely to be true.<sup>62</sup>

The more difficult problem is how to evaluate conflicting research results. Where different research studies reach different conclusions regarding toxicity, the expert must be asked to explain how those results have been taken into account in the formulation of the expert's opinion.

61. The problem of differences in chemical sensitivity was addressed by the court in Gulf S. Insulation v. United States Consumer Prods. Safety Comm'n, 701 F.2d 1137 (5th Cir. 1983). The court overturned the Commission's ban on urea-formaldehyde foam insulation because the Commission failed to document in sufficient detail the level at which segments of the population were affected and whether their response was slight or severe: "[P]redicting how likely an injury is to occur, at least in general terms, is essential to a determination of whether the risk of that injury is unreasonable." *Id.* at 1148.

There was no dispute that both nitrosamines and polonium 210 are present in defendant's snuff products. Further, defendant conceded that animal studies have accurately and consistently demonstrated that these substances cause cancer in test animals. Finally, the Court found evidence based on experiments with animals particularly valuable and important in this litigation since such experiments with humans are impossible. Under all these circumstances, the Court found this evidence probative on the issue of causation.

See also sources cited supra note 7.

<sup>62.</sup> Consistency of research results was considered by the court in Marsee v. United States Tobacco Co., 639 F. Supp. 466, 469–70 (W.D. Okla. 1986). The defendant, the manufacturer of snuff alleged to cause oral cancer, moved to exclude epidemiological studies conducted among the populations of Asia that demonstrate a link between smokeless tobacco and oral cancer. Defendant also moved to exclude evidence demonstrating that the nitrosamines and polonium 210 contained in the snuff are cancer-causing agents in some forty different species of laboratory animals. The court denied both motions, finding:

## **Glossary of Terms**

The following terms and definitions were adapted from a variety of sources, including: Office of Technology Assessment, U.S. Congress, Reproductive Health Hazards in the Workplace (1985); Louis J. Casarett & John Doull, Casarett and Doull's Toxicology: The Basic Science of Poisons (Mary O. Amdur et al. eds., 4th ed. 1991); National Research Council, Biologic Markers in Reproductive Toxicology (1989); Committee on Risk Assessment Methodology, National Research Council, Issues in Risk Assessment (1993); M. Alice Ottoboni, The Dose Makes the Poison: A Plain-Language Guide to Toxicology (2d ed. 1991); Environmental and Occupational Health Sciences Inst., Glossary of Environment Health Terms (1989).

- Acute. Extremely severe or sharp, as in acute pain. Or, with an acute disease, the symptoms develop suddenly and quickly. An acute disease lasts only a short time (a few days).
- Additive Effect. When exposure to more than one toxic agent results in the same response as would be predicted by the sum of the effects of exposure to individual agents.
- Antagonism. When exposure to one agent causes a decrease in the effect produced by another toxic agent.
- *Bioassay.* A test for measuring the toxicity of an agent by exposing laboratory animals to the substance and observing the effects.
- Biological Monitoring. Measurement of toxic agents or the results of their metabolism in biological materials, such as blood, urine, expired air, or biopsied tissue, to test for exposure to toxic agents or the detection of physio-logical changes due to exposure.
- Biologically Plausible. A biological explanation for the relationship between exposure to an agent and adverse health outcomes.
- Carcinogen. A chemical substance or other agent that causes cancer.
- Carcinogenicity Bioassay. Limited or long-term tests using laboratory animals to evaluate the potential carcinogenicity of a chemical.
- *Chronic.* A condition that lasts a long time and frequently recurs. Unlike acute conditions, the symptoms develop slowly but continue for a long time and often can go away, only to repeatedly return.
- Clinical Ecologists. Physicians who believe that exposure to certain chemical agents can result in damage to the immune system, causing multiple chemical hypersensitivity. Clinical ecologists have a background in the field of allergy, not toxicology, and their theoretical approach is derived in part from classic concepts of allergic responses and immunology.
- *Clinical Toxicology.* The study and treatment of humans exposed to chemicals and the quantification of resulting adverse health effects. Clinical toxicology includes the application of pharmacological principles to the treatment of chemically exposed individuals and research on measures to enhance elimination of toxic agents.
- *Compound*. In chemistry, the combination of two or more different substances in definite proportions that, when combined, acquire differing properties than the original substances.
- Confounding Factors. A variable that is related to both the exposure and the outcome. A confounding factor can obscure the relationship between the toxic agent and the adverse health outcome associated with that agent.
- Differential Diagnosis. The method by which a physician determines what disease process has caused a patient's symptoms. The physician considers all relevant potential causes of the symptoms and then eliminates alternative causes based on a physical examination, clinical tests, and a thorough case history.
- Direct-Acting Agents. Agents that cause toxic effects without metabolic activation or conversion.
- Distribution. Movement of the toxic agent throughout the organ systems of the body (e.g., the liver, kidney, bone, fat, and central nervous system). The rate of distribution is usually determined by the blood flow through the organ and the ability of the chemical to pass the cell membranes of the various tissues.
- *Dose, Dosage.* The measured amount of a chemical that is administered at one time, or that an organism is exposed to in a defined period of time.
- Dose-Response. The way a living organism responds to a toxic substance. The more time spent in contact with a toxic substance, or the higher the dose, the greater the organism's response. For example, a small dose of carbon monoxide will cause drowsiness; a large dose can be fatal.
- Dose-Response Curve. A graphic representation of the relationship between the dose administered and the effect produced.

- *Epidemiology*. The study of the occurrence and distribution of disease among people. Epidemiologists study groups of people to discover the cause of a disease, or where, when, and why disease occurs.
- *Epigenetic*. Pertaining to nongenetic mechanisms by which certain agents cause diseases such as cancer.
- Etiology. A branch of medical science concerned with the causation of diseases.
- *Excretion.* The process by which toxicants are eliminated from the body, including the kidney and urinary excretion, the liver and biliary system and fecal excretor, and processes involving the lungs, sweat, saliva, and lactation.
- *Exposure*. The intake into the body of a hazardous material. The main routes of exposure to substances are through the skin, mouth, and lungs.
- Extrapolation. The process of estimating unknown values from known values.
- Good Laboratory Practice (GLP). A code developed by the federal government in consultation with the laboratory-testing industry that governs many aspects of laboratory standards.
- Hazard Identification. In risk assessment, the qualitative analysis of all available experimental animal and human data to determine whether and at what dose an agent is likely to cause toxic effects.
- Hydrogeologists, Hydrologists. Scientists that specialize in the movement of ground and surface waters and the distribution and movement of contaminants in waters.
- *Immunotoxicology.* A branch of toxicology concerned with the effects of toxic agents on the immune system.
- Indirect-Acting Agents. Agents that require metabolic activation or conversion before they exhibit toxic effects on living organisms.
- *In Vitro.* A research or testing methodology that employs an artificial or test tube system, or is otherwise outside of a living organism.
- In Vivo. A research or testing methodology that employs living organisms.
- Lethal Dose 50 (LD50). The dose at which 50% of laboratory animals die within a few days.
- Lifetime Bioassay. See Bioassay.
- Maximum Tolerated Dose (MTD). The highest dose that an organism can be exposed to without causing death or significant overt toxicity.
- *Metabolism.* The sum total of the biochemical reactions that a chemical undergoes in an organism.
- Multiple Chemical Hypersensitivity. A physical condition whereby individuals react to many different chemicals at extremely low exposure levels.

- Mutagen. A substance that causes physical changes in chromosomes or biochemical changes in genes.
- Mutagenesis. The process by which agents cause changes in chromosomes and genes.
- *Neurotoxicology.* A branch of toxicology concerned with the effects of exposure to toxic agents on the central nervous system.
- No Observable Effect Level (NOEL). The level above which observable effects are believed to occur and below which no toxicity is observed.
- No Threshold Model. A model for understanding disease causation which postulates that any exposure to a harmful chemical (such as a mutagen) may increase the risk of disease.
- One Hit Theory. A theory of cancer risk in which each molecule of a chemical mutagen may mutate or change a gene in a manner that may lead to tumor formation or cancer.
- *Pharmacokinetics.* A mathematical model that expresses the movement of a toxic agent through the organ systems of the body to the target organ.
- Potentiation. The process by which the addition of one substance, which by itself has no toxic effect, increases the toxicity of another chemical when exposure to both substances occurs simultaneously.
- *Risk* Assessment. The use of scientific evidence to estimate the likelihood of adverse effects on the health of individuals or populations from exposure to hazardous materials and conditions.
- Risk Characterization. The final step of risk assessment, which summarizes information about the agent and evaluates it in order to estimate risk.
- Safety Assessment. Toxicological research that tests the toxic potential of a chemical in vivo or in vitro using standardized techniques required by governmental regulatory agencies.
- Structure Activity Relationships (SAR). A method used by toxicologists to predict the toxicity of new chemicals by comparing their molecular similarities and differences to compounds with known toxic effects.
- Synergistic Effect. The effect that occurs when one agent enhances the effect of another agent.
- Target Organ. The organ system that is affected by a particular toxic agent.
- Target Organ Dose. The dose at which a specific organ is affected.
- *Teratogen.* A substance or agent that changes eggs, sperm, or embryos, thereby increasing the risk of birth defects.
- *Teratogenic.* Pertaining to the ability to produce birth defects. (Teratogenic effects do not pass on to future generations.) See Teratogen.

*Threshold.* The level above which observable effects occur and below which no observable effects occur. See No Observable Effect Level.

Toxic. Of, relating to, or caused by a poison—or a poison itself.

Toxic Agent. An agent or substance that is toxic.

*Toxicology.* The science of the nature and effects of poisons, their detection, and the treatment of their effects.

### **References on Toxicology**

- Edward J. Calabrese, Multiple Chemical Interactions (1991).
- Edward J. Calabrese, Principles of Animal Extrapolation (1983).
- Louis J. Casarett & John Doull, Casarett and Doull's Toxicology: The Basic Science of Poisons (Mary O. Amdur et al. eds., 4th ed. 1991).
- Committee on Risk Assessment Methodology, National Research Council, Issues in Risk Assessment (1993).
- Genetic Toxicology of Complex Mixtures (Michael D. Waters et al. eds., 1990).
- Human Risk Assessment: The Role of Animal Selection and Extrapolation (M. Val Roloff ed., 1987).
- In Vitro Toxicity Testing: Applications to Safety Evaluation (John M. Frazier ed., 1992).
- Michael A. Kamrin, Toxicology: A Primer on Toxicology Principles and Applications (1988).
- Frank C. Lu, Basic Toxicology: Fundamentals, Target Organs, and Risk Assessment (2d ed. 1991).
- Methods for Biological Monitoring (T. J. Kneip & J. V. Crable eds., 1988).
- National Research Council, Biologic Markers in Reproductive Toxicology (1989).
- M. Alice Ottoboni, The Dose Makes the Poison: A Plain-Language Guide to Toxicology (2d ed. 1991).
- Alan Poole & George B. Leslie, A Practical Approach to Toxicological Investigations (1989).
- Principles and Methods of Toxicology (A. Wallace Hayes ed., 2d ed. 1989).
- Joseph V. Rodricks, Calculated Risks (1992).
- Short-Term Toxicity Tests for Non-genotoxic Effects (Philippe Bourdeau et al. eds., 1990).

Statistical Methods in Toxicology: Proceedings of a Workshop During Eurotox '90, Leipzig, Germany, September 12–14, 1990 (L. Hutnom ed., 1990).

Toxic Interactions (Robin S. Goldstein et al. eds., 1990).

Toxic Substances and Human Risk: Principles of Data Interpretation (Robert G. Tardiff & Joseph V. Rodricks eds., 1987).

### **Reference Guide on Survey Research**

Shari Seidman Diamond

**Shari Seidman Diamond**, J.D., Ph.D., is Professor of Psychology, University of Illinois at Chicago, Senior Research Fellow, American Bar Foundation, and Lecturer in Law, University of Chicago Law School.

### Contents

- I Introduction 225
  - A. Use of Surveys in Court 227
  - B. A Comparison of Survey Evidence and Individual Testimony 228
- II. Purpose and Design of the Survey 231
  - A. Was the Survey Designed to Address Relevant Questions? 231
  - B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey? 232
  - C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced? 232
  - D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced? 233
- III. Population Definition and Sampling 235
  - A. Was an Appropriate Universe or Population Identified? 235
  - B. Did the Sampling Frame Approximate the Population? 235
  - C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population? 237
  - D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey? 239
  - E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample? 240
  - F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey? 241
- IV. Survey Questions and Structure 243
  - A. Were Questions on the Survey Framed to Be Clear, Precise, and Unbiased? 243
  - B. Were Filter Questions Provided to Reduce Guessing? 244
  - C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified? 246

- D. If Probes Were Used to Clarify Ambiguous or Incomplete Answers, What Steps Were Taken to Ensure That the Probes Were Not Leading and Were Administered in a Consistent Fashion? 248
- E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects? 248
- F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question? 249
- G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey? 252
  - 1. In-person interviews 252
  - 2. Telephone surveys 253
  - 3. Mail surveys 254
- V. Surveys Involving Interviewers 257
  - A. Were the Interviewers Appropriately Selected and Trained? 257
  - B. What Did the Interviewers Know About the Survey and Its Sponsorship? 258
  - C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias? 259
- VI. Data Entry and Grouping of Responses 261
  - A. What Was Done to Ensure That the Data Were Recorded Accurately? 261
  - B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately? 261
- VII. Disclosure and Reporting 263
  - A. When Was Information About the Survey Methodology and Results Disclosed? 263
  - B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics? 264
  - C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents? 265

Glossary of Terms 267

References on Survey Research 271

### I. Introduction

Surveys are used to describe or enumerate objects or the beliefs, attitudes, or behavior of persons or other social units.<sup>1</sup> Surveys typically are offered in legal proceedings to establish or refute claims about the characteristics of those objects, individuals, or social units. Although surveys may count or measure every member of the relevant *population* (e.g., all plaintiffs eligible to join in a suit, all employees currently working for a corporation, all trees in a forest), *sample sur veys* count or measure only a portion of the objects, individuals, or social organisms that the survey is intended to describe.

Some statistical and sampling experts apply the phrase "sample survey" only to a survey in which probability sampling techniques are used to select the sample.<sup>2</sup> Although probability sampling offers important advantages over nonprobability sampling,<sup>3</sup> experts in some fields (e.g., marketing) regularly rely on various forms of nonprobability sampling when conducting surveys. Consistent with Federal Rule of Evidence 703, courts generally have accepted such evidence.<sup>4</sup> Thus, in this reference guide, both the *probability sample* and the *nonprobability sample* are discussed. The strengths of probability sampling and the weaknesses of various types of nonprobability sampling are described so that the trier of fact can consider these features in deciding what weight to give to a particular sample survey.

As a method of data collection, surveys have several crucial potential advantages over less systematic approaches.<sup>5</sup> When properly designed, executed, and

<sup>1.</sup> Social scientists describe surveys as "conducted for the purpose of collecting data from individuals about themselves, about their households, or about other larger social units." Peter H. Rossi et al., *Sample Surveys: History, Current Practice, and Future Prospects, in* Handbook of Survey Research 1, 2 (Peter H. Rossi et al. eds., 1983). Used in its broader sense, however, the term survey applies to any description or enumeration, whether or not an individual is the source of this information. Thus, a report on the number of trees destroyed in a forest fire might require a survey of the trees and stumps in the damaged area.

<sup>2.</sup> E.g., Leslie Kish, Survey Sampling 26 (1965).

<sup>3.</sup> See infra § III.C.

<sup>4.</sup> Fed. R. Evid. 703 recognizes facts or data "of a type reasonably relied upon by experts in the particular field . . . ."

<sup>5.</sup> This does not mean that surveys are perfect measuring devices that can be relied on to address all types of questions. For example, some respondents may not be able to predict accurately whether they would volunteer for military service if Washington, D.C., were to be bombed. Their inaccuracy may arise not because they are unwilling to answer the question or to say they don't know, but because they believe they can predict accurately, and they are simply wrong. Thus, the availability of a "don't know" option cannot cure the inaccuracy. Although such a survey is suitable for assessing their predictions, it does not provide useful information about what their actual responses would be.

described, surveys (1) economically present the characteristics of a large group of objects or respondents and (2) permit an assessment of the extent to which the measured objects or respondents are likely to adequately represent a relevant group of objects, individuals, or social organisms.<sup>6</sup> All questions asked of respondents and all other measuring devices used can be examined by the court and the opposing party for objectivity, clarity, and relevance, and all answers or other measures obtained can be analyzed for completeness and consistency. In order to permit the court and the opposing party to closely scrutinize the survey so that its relevance, objectivity, and representativeness can be evaluated, the design and execution of the survey are described in detail by the party proposing to offer it as evidence.

The questions listed in this reference guide are intended to assist judges in identifying, narrowing, and addressing issues bearing on the adequacy of surveys either offered as evidence or proposed as a method for developing information.<sup>7</sup> These questions can be (1) raised from the bench during a pretrial proceeding to determine the admissibility of the survey evidence; (2) presented to the contending experts before trial for their joint identification of disputed and undisputed issues; (3) presented to counsel with the expectation that the issues will be addressed during the examination of the experts at trial; or (4) raised in bench trials when a motion for a preliminary injunction is made to help the judge evaluate what weight, if any, the survey should be given.<sup>8</sup> These questions are intended to improve the utility of cross-examination by counsel, where appropriate, not to replace it.

All sample surveys, whether they measure objects, individuals, or other social organisms, should address the issues concerning purpose and design (section II), population definition and sampling (section III), accuracy of data entry (section VI), and disclosure and reporting (section VII). Questionnaire and interview surveys raise methodological issues involving survey questions and structure (section IV) and confidentiality (section VII.C), and interview surveys introduce additional issues (e.g., interviewer training and qualifications) (section V). The sections of the reference guide are labeled to immediately identify those topics that are relevant to the type of survey being considered. The scope of this reference guide is limited necessarily, and additional issues might arise in particular cases.

<sup>6.</sup> The ability to quantitatively assess the limits of the likely margin of error is unique to probability sample surveys.

<sup>7.</sup> See infra text accompanying note 24.

<sup>8.</sup> Lanham Act cases involving trademark infringement or deceptive advertising frequently require expedited hearings that request injunctive relief, so that judges may need to be more familiar with survey methodology than if these cases were being submitted to a jury.

#### A. Use of Surveys in Court

Thirty years ago the question whether surveys constituted acceptable evidence still was unsettled.<sup>9</sup> Early doubts about the admissibility of surveys centered on their use of sampling techniques<sup>10</sup> and their status as hearsay evidence.<sup>11</sup> Federal Rule of Evidence 703 settled both matters by redirecting attention to the "validity of the techniques employed." <sup>12</sup> The inquiry under Rule 703 focuses on whether facts or data are "of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject."<sup>13</sup> In the case of a survey, the question becomes, "Was the poll or survey conducted in accordance with generally accepted survey principles, and were the results used in a statistically correct way?"<sup>14</sup>

Because the survey method provides an economical and systematic way to gather information about a large number of individuals or social units, surveys are used widely in business, government, and, increasingly, administrative settings and judicial proceedings. Both federal and state courts have accepted survey evidence on a variety of issues.<sup>15</sup> In a case involving allegations of discrimination in jury panel composition, the defense team surveyed prospective jurors to obtain age, race, education, ethnicity, and income distribution.<sup>16</sup> Surveys of employees or prospective employees are used to support or refute claims of employment discrimination.<sup>17</sup> Requests for a change of venue on grounds of jury pool bias often are backed by evidence from a survey of jury-eligible respondents

11. Judge Wilfred Feinberg's thoughtful analysis in Zippo Mfg. Co. v. Rogers Imports, Inc., 216 F. Supp. 670, 682–83 (S.D.N.Y. 1963), provides two alternative grounds for admitting opinion surveys: (1) surveys are not hearsay because they are not offered in evidence to prove the truth of the matter asserted; and (2) even if they are hearsay, they fall under one of the exceptions as a "present sense impression."

12. Fed. R. Evid. 703 advisory committee's note.

13. Fed. R. Evid. 703.

15. Some surveys are so well accepted that they even may not be recognized as surveys. For example, U.S. Census Bureau data are based on sample surveys. Similarly, the Standard Table of Mortality, which is accepted as proof of the average life expectancy of an individual of a particular age and gender, is based on survey data. 16. People v. Harris, 679 P.2d 433 (Cal.), cert. denied, 469 U.S. 965 (1984).

17. Richardson v. Quik Trip Corp., 591 F. Supp. 1151, 1153 (S.D. Iowa 1984); EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1308 (N.D. Ill. 1986), affd, 839 F.2d 302 (7th Cir. 1988).

<sup>9.</sup> Hans Zeisel, The Uniqueness of Survey Evidence , 45 Cornell L.Q. 322, 345 (1960).

<sup>10.</sup> In an early use of sampling, Sears, Roebuck & Co. claimed a tax refund based on sales made to individuals living outside city limits. Sears randomly sampled 33 of the 826 working days in the relevant working period, computed the proportion of sales to out-of-city individuals during those days, and projected the sample result to the entire period. The court refused to accept the estimate based on the sample. When a complete audit was made, the result was almost identical to that obtained from the sample. Sears, Roebuck & Co. v. City of Inglewood, described in R. Sprowls, *The Admissibility of Sample Data into a Court of Law: A Case History*, 4 UCLA L. Rev. 222, 226–29 (1956–57).

<sup>14.</sup> Manual for Complex Litigation, Second, § 2.712. Survey research also is addressed in the Manual for Complex Litigation, Third, § 21.493 (forthcoming 1995) [hereinafter MCL 3d]. Note, however, that experts who collect survey data, along with the professions that rely on those surveys, may differ in some of their methodological standards and principles. The required precision of sample estimates and an evaluation of the sources and magnitude of likely bias are required to distinguish methods that are acceptable from methods that are not.

in the area of the original venue.<sup>18</sup> The plaintiff in an antitrust suit conducted a survey to assess what characteristics, including price, affected consumers' preferences. The survey was offered as one way to estimate damages.<sup>19</sup> A routine use of surveys in federal courts occurs in Lanham Act<sup>20</sup> cases, where the plaintiff alleges trademark infringement<sup>21</sup> or claims that false advertising<sup>22</sup> has confused or deceived consumers. The pivotal legal question in such cases virtually demands survey research because it centers on consumer perception (i.e., is the consumer likely to be confused about the source of a product, or does the advertisement imply an inaccurate message?).<sup>23</sup> In addition, survey methodology has been used creatively to assist federal courts in managing mass torts litigation. Faced with the prospect of conducting discovery concerning 10,000 plaintiffs, the plaintiffs and defendants in *Wilhoite v. Olin Corp.*<sup>24</sup> jointly drafted a discovery survey that was administered in person by neutral third parties, thus replacing interrogatories and depositions. It resulted in substantial savings in both time and cost.

#### B. A Comparison of Survey Evidence and Individual Testimony

To illustrate the value of a survey, it is useful to compare the information that can be obtained from a competently done survey with the information obtained by other means. A survey is presented by a survey expert who testifies about the responses of a substantial number of individuals who have been selected according to an explicit sampling plan and asked the same set of questions by interviewers who were not told who sponsored the survey or what answers were predicted or preferred. Although parties presumably are not obliged to present a

18. E.g., United States v. Eagle, 586 F.2d 1193, 1195 (8th Cir. 1978); Powell v. Superior Court, 283 Cal. Rptr. 777, 783 (Cal. Ct. App. 2d Dist. 1991).

19. Dolphin Tours, Inc. v. Pacifico Creative Servs., Inc., 773 F.2d 1506, 1508 (9th Cir. 1985). See also Benjamin F. King, Statistics in Antitrust Litigation, in Statistics and the Law 49 (Morris H. DeGroot et al. eds., 1986). Surveys also are used in litigation to help define relevant markets. In United States v. E.I. DuPont de Nemours & Co., 118 F. Supp. 41, 60 (D. Del. 1953), *aff'd*, 351 U.S. 377 (1956), a survey was used to develop the "market setting" for the sale of cellophane.

20. Lanham Act § 43(a), 15 U.S.C. § 1125(a) (1946) (amended 1992).

21. E.g., Union Carbide Corp. v. Ever-Ready, Inc., 531 F.2d 366 (7th Cir.), cert. denied, 429 U.S. 830 (1976). According to Neal Miller, Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation, 40 Rutgers L. Rev. 101, 137 (1987), trademark law has relied on the institutionalized use of statistical evidence more than any other area of the law.

22. E.g., American Home Prods. Corp. v. Johnson & Johnson, 577 F.2d 160 (2d Cir. 1978).

23. Courts have observed that "the court's reaction is at best not determinative and at worst irrelevant. The question in such cases is—what does the person to whom the advertisement is addressed find to be the message?" American Brands, Inc. v. R.J. Reynolds Tobacco Co., 413 F. Supp. 1352, 1357 (S.D.N.Y. 1976). The wide use of surveys in recent years was foreshadowed in Triangle Publications, Inc. v. Rohrlich, 167 F.2d 969, 974 (2d Cir. 1948) (Frank, J., dissenting). Called on to determine whether a manufacturer of girdles labeled "Miss Seventeen" infringed the trademark of the magazine, *Seventeen*, Judge Frank suggested in the absence of a test of the reactions of "numerous girls and women," the trial court judge's finding as to what was likely to confuse was "nothing but a surmise, a conjecture, a guess," noting that "neither the trial judge nor any member of this court is (or resembles) a teen-age girl or the mother or sister of such a girl." *Id.* at 976.

24. No. CV-83-C-5021-NE (N.D. Ala. filed Jan. 11, 1983). The case ultimately settled before trial. See Francis E. McGovern & E. Allan Lind, *The Discovery Survey*, Law & Contemp. Probs., Autumn 1988, at 41.

survey conducted in anticipation of litigation by a nontestifying expert if it produced unfavorable results, <sup>25</sup> the court can and should scrutinize the method of respondent selection for any survey that is presented.

A party using a nonsurvey method generally identifies several witnesses who testify about their own characteristics, experiences, or impressions. While the party has no obligation to select these witnesses in any particular way or to report on how they were chosen, the party is not likely to select witnesses whose attributes conflict with the party's interests. The witnesses who testify are aware of the parties involved in the case and have discussed the case before testifying.

Although surveys are not the only means to demonstrate particular facts, the testimony of an expert describing the results of a well-done survey is an efficient way to inform the trier of fact about a large and representative group of potential witnesses. In some cases, courts have described surveys as the most direct form of evidence that can be offered.<sup>26</sup> Indeed, several courts have drawn negative inferences from the absence of a survey, taking the position that failure to undertake a survey may strongly suggest that a properly done survey would not support the plaintiff's position.<sup>27</sup>

Survey Research

<sup>25.</sup> Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 205 (S.D.N.Y. 1981) (distinguishing between surveys conducted in anticipation of litigation and surveys conducted for nonlitigation purposes which cannot be reproduced because of the passage of time, concluding that parties should not be compelled to introduce the former at trial, but may be required to provide the latter).

<sup>26.</sup> E.g., Charles Jacquin et Cie, Inc. v. Destileria Serralles, Inc., 921 F.2d 467, 475 (3d Cir. 1990). See also Brunswick Corp. v. Spinit Reel Co., 832 F.2d 513, 522 (10th Cir. 1987).

<sup>27.</sup> E.S. Originals, Inc. v. Stride Rite Corp., 656 F. Supp. 484, 490 (S.D.N.Y. 1987); see also Information Clearing House, Inc. v. Find Magazine, 492 F. Supp. 147, 160 (S.D.N.Y. 1980); Henri's Food Prods. Co. v. Kraft, Inc., 717 F.2d 352, 357 (7th Cir. 1983).

### II. Purpose and Design of the Survey

#### A. Was the Survey Designed to Address Relevant Questions?

The report describing the results of a survey should include a statement describing the purpose or purposes of the survey. One indication that a survey offers probative evidence is that it was designed to collect information relevant to the legal controversy (e.g., to estimate damages in an antitrust suit or to assess consumer confusion in a trademark case).<sup>28</sup> Surveys not conducted specifically in preparation for, or in response to, litigation may provide important information,<sup>29</sup> but they frequently ask irrelevant questions<sup>30</sup> or select inappropriate samples of respondents for study.<sup>31</sup> Nonetheless, surveys do not always achieve their stated goals. Thus, the content and execution of a survey must be scrutinized even if the survey was designed to provide relevant data on the issue before the court.

29. See, e.g., Wright v. Jeep Corp., 547 F. Supp. 871, 874 (E.D. Mich. 1982). Indeed, as courts increasingly have been faced with scientific issues, parties have requested in a number of recent cases that the courts compel production of research data and testimony by unretained experts. The circumstances under which an unretained expert can be compelled to testify or to disclose research data and opinions, as well as the extent of disclosure that can be required when the research conducted by the expert has a bearing on the issues in the case, are the subject of considerable current debate. See, e.g., Richard L. Marcus, Discovery Along the Litigation/Science Interface, 57 Brook. L. Rev. 381 (1991); Joe S. Cecil, Judicially Compelled Disclosure of Research Data, 1 Cts. Health Sci. & L. 434 (1991).

30. Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 206 (S.D.N.Y. 1981) (marketing surveys conducted before litigation were designed to test for brand awareness, whereas the "single issue at hand . . . [was] whether consumers understood the term 'Super Glue' to designate glue from a single source").

31. In Craig v. Boren, 429 U.S. 190 (1976), the state unsuccessfully attempted to use its annual roadside survey of the blood alcohol level, drinking habits, and preferences of drivers to justify prohibiting the sale of 3.2% beer to males under the age of 21 and to females under the age of 18. The Court suggested that the data were biased because it was likely that the male would be driving if both the male and female occupants of the car had been drinking. As pointed out in 2 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy: Tort Law, Evidence, and Health 527 (1988), the roadside survey would have provided more relevant data if all occupants of the cars had been included in the survey (and if the type and amount of alcohol most recently consumed had been requested so that the consumption of 3.2% beer could have been isolated).

<sup>28.</sup> Note, however, that if a survey was not designed for purposes of litigation, one source of bias is less likely: the party presenting the survey is less likely to have designed and constructed the survey to prove its side of the issue in controversy.

B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey?

An early handbook for judges recommended that interviews be "conducted independently of the attorneys in the case."<sup>32</sup> Some courts have interpreted this to mean that any evidence of attorney participation is objectionable.<sup>33</sup> A better interpretation is that the attorney should have no part in carrying out the survey. However, some attorney involvement in the survey design is necessary to ensure that relevant questions are directed to a relevant population.<sup>34</sup> The trier of fact evaluates the objectivity and relevance of the questions on the survey and the appropriateness of the definition of the population used to guide sample selection. These aspects of the survey are visible to the trier of fact and can be judged on their quality, irrespective of who suggested them. In contrast, the interviews themselves are not directly visible, and any potential bias is minimized by having interviewers and respondents blind to the purpose and sponsorship of the survey and by excluding attorneys from any part in conducting interviews and tabulating results.

# C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced?

Experts prepared to design, conduct, and analyze a survey generally should have graduate training in psychology, sociology, marketing, communication sciences, statistics, or a related discipline; that training should include courses in survey research methods, sampling, measurement, interviewing, and statistics. In some cases, professional experience in conducting and publishing survey research may provide the requisite background. In all cases, the expert must demonstrate an understanding of survey methodology, including sampling, <sup>35</sup> instrument design (questionnaire and interview construction), and statistical analysis. <sup>36</sup> Publication in peer-reviewed journals, authored books, membership in professional organizations, faculty appointments, consulting experience, and membership on scientific advisory panels for government agencies or private foundations are indications of a professional's area and level of expertise. In addition, if the survey involves highly technical subject matter (e.g., the particular preferences held by electrical engineers for various pieces of electrical equipment and the

<sup>32.</sup> Judicial Conference of the U.S., Handbook of Recommended Procedures for the Trial of Protracted Cases 75 (1960).

<sup>33.</sup> E.g., Boehringer Ingelheim G.m.b.H. v. Pharmadyne Lab., 532 F. Supp. 1040, 1058 (D.N.J. 1980).

<sup>34. 3</sup> J. Thomas McCarthy, McCarthy on Trademarks and Unfair Competition § 32.48(3) (3d ed. 1992). 35. The one exception is that sampling expertise is unnecessary if the survey is administered to all members

of the relevant population. See, e.g., McGovern & Lind, supra note 24.

<sup>36.</sup> If survey expertise is being provided by several experts, a single expert may have general familiarity but not special expertise in all these areas.

bases for those preferences) or involves a special population (e.g., developmentally disabled adults with limited cognitive skills), the survey expert also should be able to demonstrate sufficient familiarity with the topic or population (or assistance from an individual on the research team with suitable expertise) to design a survey instrument that will communicate clearly with relevant respondents.

# D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced?

Parties often call on an expert to testify about a survey conducted by someone else. The secondary expert's role is to offer support for a survey commissioned by the party who calls the expert, to critique a survey presented by the opposing party, or to introduce findings or conclusions from a survey not conducted in preparation for litigation or by any of the parties to the litigation. The trial court should take into account the exact issue that the expert seeks to testify about and the nature of the expert's field of expertise.<sup>37</sup> The secondary expert who gives an opinion about the adequacy and interpretation of a survey not only should have general skills and experience with surveys and be familiar with all of the issues addressed in this reference guide, but also should demonstrate familiarity with the following properties of the survey being discussed:

- 1. the purpose of the survey;
- 2. the survey methodology, including
  - a. the target population,
  - b. the sampling design used in conducting the survey,
  - c. the survey instrument (questionnaire or interview schedule), and
  - d. (for interview surveys) interviewer training and instruction;
- 3. the results, including rates and patterns of missing data; and
- 4. the statistical analyses used to interpret the results.

37. See Margaret A. Berger, Evidentiary Framework § II.C, in this manual.

### **III.** Population Definition and Sampling

#### A. Was an Appropriate Universe or Population Identified?

One of the first steps in designing a survey or in deciding whether an existing survey is relevant is to identify the *target population* (or *universe*).<sup>38</sup> The target population consists of all elements (i.e., objects, individuals, or other social organisms) whose characteristics or perceptions the survey is intended to represent. Thus, in trademark litigation, the relevant population in some disputes may include all prospective and actual purchasers of the plaintiff's goods or services. Similarly, the population for a discovery survey may include all potential plaintiffs or all employees who worked for Company A between two specific dates. The definition of the relevant population is crucial because there may be systematic differences in the responses of members of the population and nonmembers. (For example, consumers who are prospective purchasers may know more about the product category than consumers who are not considering making a purchase.)

The universe must be defined carefully. For example, a commercial for a toy or breakfast cereal may be aimed at children, who in turn influence their parents' purchases. If a survey assessing the commercial's tendency to mislead were conducted based on the universe of prospective and actual adult purchasers, it would exclude a crucial group of eligible respondents. Thus, the appropriate population in this instance would include children as well as parents.

#### B. Did the Sampling Frame Approximate the Population?

The target population consists of all the individuals or units that the researcher would like to study. The *sampling frame* is the source (or sources) from which the sample actually is drawn. The surveyor's job generally is easier if a complete list of every eligible member of the population is available (e.g., all plaintiffs in a discovery survey), so that the sampling frame lists the identity of all members of

<sup>38.</sup> Identification of the proper universe is recognized uniformly as a key element in the development of a survey. See, e.g., Judicial Conference of the U.S., *supra* note 32; MCL 3d, *supra* note 14, § 21.493. See *also* 3 McCarthy, *supra* note 34, § 32.47; and Council of Am. Survey Res. Orgs., Code of Standards for Survey Research § III.B.4.

the target population. Frequently, however, the target population includes members who are inaccessible or who cannot be identified in advance. As a result, some compromises are required in developing the sampling frame. The survey report should contain a description of the target population, a description of the *survey population* actually sampled, a discussion of the difference be tween the two populations, and an evaluation of the likely consequences of that difference.

A survey that provides information about a wholly irrelevant universe of respondents is itself irrelevant.<sup>39</sup> More commonly, however, either the target population or the sampling frame is underinclusive or overinclusive. If either is underinclusive, the survey's value depends on the extent to which the excluded population is likely to react differently from the included population. Thus, a survey of spectators and participants at running events would be sampling a sophisticated subset of those likely to purchase running shoes. Because this subset probably would consist of the consumers most knowledgeable about the trade dress used by companies that sell running shoes, a survey based on this population would be likely to substantially overrepresent the strength of a particular design as a trademark, and the extent of that overrepresentation would be unknown and not susceptible to any reasonable estimation.<sup>40</sup>

Similarly, in a survey designed to project demand for cellular phones, the assumption that businesses would be the primary users of cellular service led surveyors to exclude potential nonbusiness users from the survey. The Federal Communications Commission (FCC) found the assumption unwarranted and concluded that the research was flawed, in part because of this underinclusive universe.<sup>41</sup>

In some cases, it is difficult to determine whether an underinclusive universe distorts the results of the survey and, if so, the extent and likely direction of the bias. For example, a trademark survey was designed to test the likelihood of confusing an analgesic currently on the market with a new product that was similar in appearance.<sup>42</sup> The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the universe should have included users of other analgesics, "so that the full range of potential customers

<sup>39.</sup> A survey assessing response to an advertisement made for presentation to persons in the trade should not be evaluated on a sample of consumers. Home Box Office v. Showtime/The Movie Channel, 665 F. Supp. 1079, 1083 (S.D.N.Y.), aff d in part & vacated in part, 832 F.2d 1311 (2d Cir. 1987). But see Lon Tai Shing Co. v. Koch + Lowy, No. 90-C4464, 1990 U.S. Dist. LEXIS 19123, at \*50 (S.D.N.Y. Dec. 14, 1990), in which the judge was willing to find likelihood of consumer confusion from a survey of lighting store salespersons questioned by a survey researcher posing as a customer. The court was persuaded that the salespersons who were misstating the source of the lamp, whether consciously or not, must have believed reasonably that the consuming public would be misled.

<sup>40.</sup> Brooks Shoe Mfg. Co. v. Suave Shoe Corp., 533 F. Supp. 75, 80 (S.D. Fla. 1981), affd, 716 F.2d 854 (11th Cir. 1983).

<sup>41.</sup> Gencom, Inc., 56 Rad. Reg. 2d (P&F) 1597, 1604 (1984). This position was affirmed on appeal. Gencom, Inc. v. FCC, 832 F.2d 171, 186 (D.C. Cir. 1987).

<sup>42.</sup> American Home Prods. Corp. v. Barr Lab., 656 F. Supp. 1058 (D.N.J. 1987), affd, 834 F.2d 368 (3d Cir. 1987).

for whom plaintiff and defendants would compete could be studied."<sup>43</sup> In this instance, it is unclear whether users of the plaintiff's product would be more or less likely to be confused than users of the defendant's product or users of a third analgesic.<sup>44</sup>

An overinclusive universe generally presents less of a problem in interpretation than does an underinclusive universe. If the survey expert can demonstrate that a sufficiently large (and representative) subset of respondents in the survey was drawn from the appropriate universe, the responses obtained from that subset can be examined, and inferences about the relevant universe can be drawn based on that subset. <sup>45</sup> If the sample is drawn from an underinclusive universe, there is no way to know how the unrepresented members would have responded.

### C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population?

Identification of a survey population must be followed by selection of a sample that accurately represents that population.<sup>46</sup> The use of probability sampling techniques maximizes both the representativeness of the survey results and the ability to assess the accuracy of estimates obtained from the survey.

Probability samples range from simple random samples to complex *multistage* sample designs that use stratification, clustering of population elements into var - ious groupings, or both. In *simple random sampling*, the most basic type of prob - ability sampling, every element in the population has a known, equal probability of being included in the sample, and all possible samples of a given size are equally likely to be selected.<sup>47</sup> In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample, <sup>48</sup> which gives probability sampling two important advantages. First, the sample can provide an unbiased estimate of the responses of all persons in the population from which the sample was drawn; that is, the results from the sample are projectable. Second, the researcher can calculate a *confi*-

<sup>43.</sup> Id. at 1070.

<sup>44.</sup> See also Craig v. Boren, 429 U.S. 190 (1976).

<sup>45.</sup> This occurred in National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc., 532 F. Supp. 651, 657–58 (W.D. Wash. 1982).

<sup>46.</sup> MCL 3d, supra note 14, § 21.493. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.B, in this manual.

<sup>47.</sup> Systematic sampling, in which every *n*th unit in the population is sampled and the starting point is selected randomly, fulfills the first of these conditions. It does not fulfill the second because no systematic sample can include elements adjacent to one another on the list of population members from which the sample is drawn. Except in very unusual situations when periodicities occur, systematic samples and simple random samples generally produce the same results. Seymour Sudman, *Applied Sampling*, *in* Handbook of Survey Research, *supra* note 1, at 145, 169.

<sup>48.</sup> Other probability sampling techniques include (1) stratified random sampling, in which the researcher subdivides the population into mutually exclusive and exhaustive subpopulations, or strata, and then randomly selects samples from within these strata; and (2) *cluster sampling*, in which cases are sampled in groups or clusters, rather than on an individual basis. Martin Finkel, *Sampling Theory*, in Handbook of Survey Research, supra note 1, at 21, 37, 47.

dence interval that describes explicitly how reliable the sample estimate of the population is likely to be. Thus, suppose a survey tested a sample of 400 dentists randomly selected from the population of all dentists licensed to practice in the United States and found that 80, or 20%, of them mistakenly believed that a new toothpaste, Goldgate, was manufactured by the makers of Colgate. A survey expert properly could compute a confidence interval around the 20% estimate obtained from this sample. If the survey were repeated a large number of times, and a 95% confidence interval was computed each time, 95% of the confidence intervals would include the actual percentage of dentists in the entire population who would believe that Goldgate was manufactured by the makers of Colgate.<sup>49</sup> In this example, the confidence interval, or margin of error, is the estimate (20%) plus or minus 4%, or the distance between 16% and 24%.

All sample surveys produce estimates of *population values*, not exact measures of those values. Strictly speaking, the margin of *sampling error* associated with the sample estimate assumes probability sampling. Assuming a probability sample, a confidence interval describes how stable the mean response in the sample is likely to be. The width of the confidence interval depends on three characteristics:

- 1. the size of the sample (the larger the sample, the narrower the interval);
- 2. the variability of the response being measured; and
- 3. the confidence level the researcher wants to have.

Traditionally, scientists adopt the 95% level of confidence, which means that if 100 samples of the same size were drawn, the confidence interval expected for at least 95 of the samples would include the true population value.<sup>50</sup>

Although probability sample surveys often are conducted in organizational settings and are the recommended sampling approach in academic and government publications on surveys, probability sample surveys can be expensive when in-person interviews are required, the target population is dispersed widely, or qualified respondents are scarce. A majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability con - venience samples .<sup>51</sup> They are admitted into evidence based on the argument that nonprobability sampling is used widely in marketing research and that "results of these studies are used by major American companies in making decisions of

<sup>49.</sup> Actually, since survey interviewers would be unable to locate some dentists and some dentists would be unwilling to participate in the survey, technically the population to which this sample would be projectable would be all dentists with current addresses who would be willing to participate in the survey if they were asked.

<sup>50.</sup> To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%), the width of the confidence interval can be expanded. An increase in the confidence interval brings an increase in the confidence level. For further discussion of confidence intervals, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

<sup>51.</sup> Jacob Jacoby & Amy H. Handlin, Non-Probability Sampling Designs for Litigation Surveys, 81 Trademark Rep. 169 (1991). For probability surveys conducted in trademark cases, see National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc., 532 F. Supp. 651 (W.D. Wash. 1982); James Burrough, Ltd. v. Sign of Beefeater, Inc., 540 F.2d 266 (7th Cir. 1976).

considerable consequence."<sup>52</sup> Nonetheless, when respondents are not selected randomly from the relevant population, the expert should be prepared to justify the method used to select respondents. Special precautions are required to reduce the likelihood of biased samples.<sup>53</sup> In addition, quantitative values computed from such samples (e.g., percentage of respondents indicating confusion) should be viewed as rough indicators rather than as precise quantitative estimates. Confidence intervals should not be computed.

# D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey?

Even when a sample is drawn randomly from a complete list of elements in the target population, responses or measures may be obtained on only part of the selected sample. If this lack of response were distributed randomly, valid inferences about the population could be drawn from the characteristics of the available elements in the sample. The difficulty is that nonresponse often is not random, so that, for example, persons who are single typically have three times the "not at home" rate in U.S. Census Bureau surveys as do family members.<sup>54</sup> Efforts to increase response rates include making several attempts to contact potential respondents and providing financial incentives for participating in the survey.

One suggested formula for quantifying a tolerable level of nonresponse in a probability sample is based on the guidelines for statistical surveys issued by the former U.S. Office of Statistical Standards.<sup>55</sup> According to these guidelines, response rates of 90% or more are reliable and generally can be treated as random samples of the overall population. Response rates between 75% and 90% usually yield reliable results, but the researcher should conduct some check on the representativeness of the sample. Potential bias should receive greater scrutiny when the response rate drops below 75%. If the response rate drops below 50%, the survey should be regarded with significant caution as a basis for precise quantitative statements about the population from which the sample was drawn.<sup>56</sup>

54. 2 Gastwirth, *supra* note 31, at 501. This volume contains a useful discussion of sampling, along with a set of examples. 2 *ld.* at 467.

56. For thoughtful examples of judges closely scrutinizing potential sample bias when response rates were below 75%, see Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224 (N.D. Tex. 1980); Rosado v. Wyman, 322 F. Supp. 1173 (E.D.N.Y.), affd, 437 F.2d 619 (2d Cir. 1970), affd, 402 U.S. 991 (1971).

<sup>52.</sup> National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 515 (D.N.J. 1986). A survey of the 130 members of the Council of American Survey Research Organizations, the national trade association for commercial survey research firms in the United States, revealed that 95% of the in-person interview studies done in 1985 took place in malls or shopping centers. Jacoby & Handlin, *supra* note 51, at 172–73, 176.

<sup>53.</sup> See infra § III.E.

<sup>55.</sup> This standard is cited with approval by Gastwirth. 2 Id. at 502.

Determining whether the level of nonresponse in a survey is critical generally requires an analysis of the determinants of nonresponse. For example, even a survey with a high response rate may seriously underrepresent some portions of the population, such as the unemployed or the poor. If a general population sample was used to chart changes in the proportion of the population that knows someone with the HIV virus, the survey would underestimate the population value if some groups more likely to know someone with HIV (e.g., intravenous drug users) were underrepresented in the sample. The survey expert should be prepared to provide evidence on the potential impact of nonresponse on the survey results.

In surveys that include sensitive or difficult questions, particularly those that are self-administered, some respondents may refuse to provide answers or may provide incomplete answers. To assess the impact of nonresponse to a particular question, the survey expert should analyze the differences between those who answered and those who did not answer. Procedures to address the problem of missing data include recontacting respondents to obtain the missing answers and using the respondent's other answers to predict the missing response.<sup>57</sup>

### E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample?

If it is impractical for a survey researcher to sample randomly from the entire target population, the researcher still can apply probability sampling to some aspects of respondent selection, even in a mall intercept study, to reduce the like-lihood of biased selection. For example, mall locations can be sampled randomly from a list of possible sites. By administering the survey at several different malls, the expert can test for and report on any differences observed across sites. To the extent that similar results are obtained in different locations using different on-site interview operations, it is less likely that idiosyncrasies of sample selection or administration can account for the results.<sup>58</sup> Similarly, since the characteristics of persons visiting a shopping center vary by day of the week and time of day, bias in sampling can be reduced if the survey design calls for sampling time segments as well as mall locations.<sup>59</sup>

In *mall intercept surveys*, the organization that manages the on-site interview facility generally employs recruiters who approach potential survey respondents in the mall and ascertain if they are qualified and willing to participate in the survey. If a potential respondent agrees to answer the questions and meets the specified criteria, he or she is escorted to the facility where the survey interview

<sup>57.</sup> Andy B. Anderson et al., Missing Data: A Review of the Literature, in Handbook of Survey Research, supra note 1, at 415.

<sup>58.</sup> Note, however, that differences across sites may be due to genuine differences in respondents across geographic locations.

<sup>59.</sup> Seymour Sudman, Improving the Quality of Shopping Center Sampling , 17 J. Mktg. Res. 423 (1980).

takes place. If recruiters are free to approach potential respondents without controls on how an individual is to be selected for screening, shoppers who spend more time in the mall are more likely to be approached than shoppers who visit the mall only briefly. Moreover, recruiters naturally prefer to approach friendly looking potential respondents, so that it is more likely that certain types of individuals will be selected. These potential biases in selection can be reduced by providing appropriate selection instructions and training recruiters effectively. Training that reduces the interviewer's discretion in selecting a potential respondent is likely to reduce bias in selection, as are instructions to approach every third person entering the facility through a particular door.

# F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey?

In a carefully executed survey, each potential respondent is questioned or measured on the attributes that determine his or her eligibility to participate in the survey. Thus, the initial questions screen potential respondents to determine if they are within the target universe of the survey (e.g., Is she at least 14 years old? Does she own a dog? Does she live within 10 miles?). The screening questions must be drafted so that they do not convey information that will influence the respondent's answers on the main survey. For example, if respondents must be prospective and recent purchasers of Sunshine orange juice in a trademark survey designed to assess consumer confusion with Sun Time orange juice, potential respondents might be asked to name the brands of orange juice they have purchased recently or expect to purchase in the next six months. They should not be asked specifically if they recently have purchased, or expect to purchase, Sunshine orange juice, because this may affect their responses on the survey either by implying who is conducting the survey or by supplying them with a brand name that otherwise would not occur to them.

The criteria for determining whether to include a potential respondent in the survey should be objective and clearly conveyed, preferably using written instructions addressed to those who administer the screening questions. These instructions and the completed screening questionnaire should be made available to the court and the opposing party along with the interview form for each respondent.

### **IV.** Survey Questions and Structure

#### A. Were Questions on the Survey Framed to Be Clear, Precise, and **Unbiased**?

Although it seems obvious that questions on a survey should be clear and precise, phrasing questions to reach that goal is often difficult. Even questions that appear clear can convey unexpected meanings and ambiguities to potential respondents. For example, the question "What is the average number of days each week you have butter?" appears to be straightforward. Yet some respondents wondered whether margarine counted as butter, and when the question was revised to include the introductory phrase "Not including margarine," the reported frequency of butter use dropped dramatically.<sup>60</sup> When unclear questions are included in a survey, they may threaten the validity of the survey by systematically distorting responses if respondents are misled in a particular direction, or by inflating random error if respondents guess because they do not understand the question.<sup>61</sup>

Texts on survey research generally recommend pretests as a way to increase the likelihood that questions are clear and unambiguous, 62 and some courts have recognized the value of pretests.<sup>63</sup> In a pretest, the proposed survey is administered to a small sample (usually between twenty-five and seventy-five)<sup>64</sup> of the same type of respondents who would be eligible to participate in the fullscale survey.<sup>65</sup> The interviewers observe the respondents for any difficulties they may have with the questions and probe for the source of any such difficulties so that the questions can be rephrased if confusion or other difficulties arise. Attorneys who commission surveys for litigation sometimes are reluctant to ap-

60. Floyd J. Fowler, Jr., How Unclear Terms Affect Survey Data, 56 Pub. Opinion Q. 218, 225-26 (1992). 61. Id. at 219.

62. For a thorough treatment of pretesting methods, see Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire 51 (1986). See also Fred W. Morgan, Judicial Standards for Survey Research: An Update and Guidelines , 54 J. Mktg. 59, 64 (1990). 63. E.g., Zippo Mfg. Co. v. Rogers Imports, Inc., 216 F. Supp. 670 (S.D.N.Y. 1963).

64. Converse & Presser, supra note 62, at 69. Converse and Presser suggest that a pretest with twenty-five respondents is appropriate when the survey uses professional interviewers.

65. The terms pretest and pilot test are sometimes used interchangeably. When they are distinguished, the difference is that a pretest tests the questionnaire, while a pilot test generally tests proposed collection procedures as well.

prove pilot work or to reveal that pilot work has taken place because they are concerned that if a pretest leads to revised wording of the questions, the trier of fact may believe that the survey has been manipulated and is biased or unfair. A more appropriate reaction is to recognize that pilot work can improve the quality of a survey and to anticipate that it often results in word changes that increase clarity and correct misunderstandings. Thus, changes may indicate informed survey construction rather than flawed survey design.<sup>66</sup>

#### B. Were Filter Questions Provided to Reduce Guessing?

Some survey respondents may have no opinion on an issue under investigation, either because they have never thought about it before, or because the question mistakenly assumes a familiarity with the issue. For example, survey respondents may not have noticed that the commercial they are being questioned about guaranteed the quality of the product being advertised and thus may have no opinion on the kind of guarantee it indicated. Likewise, in an employee survey, respondents may not be familiar with the parental leave policy at their company and thus may have no opinion on whether they would consider taking advantage of the parental leave policy if they became parents. The following three alternative question structures will affect how those respondents answer and how their responses are counted.

First, the survey can ask all respondents to answer the question (e.g., "Did you understand the guarantee offered by Clover to be a one-year guarantee, a sixtyday guarantee, or a thirty-day guarantee?"). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an answer even if the respondent did not notice the guarantee (or is unfamiliar with the parental leave policy). Such answers will reflect only what the respondent can glean from the question, or they may reflect pure guessing. The size of the random element that this approach introduces will increase with the proportion of respondents who are unfamiliar with the topic at issue.

Second, the survey can use a *quasi-filter question* to reduce guessing by providing "don't know" or "no opinion" options as part of the question (e.g., "Did you understand the guarantee offered by Clover to be for more than a year, a year, or less than a year, or don't you have an opinion?").<sup>67</sup> By signaling to the respondent that it is appropriate not to have an opinion, the question reduces the demand for an answer and, as a result, the inclination to hazard a guess just to comply. Respondents are more likely to endorse a "no opinion" option if it is mentioned explicitly by the interviewer than if it is merely accepted when the respondent spontaneously offers it as a response. The consequence of this change in format is substantial. Studies indicate that, although the relative pro-

<sup>66.</sup> See infra § VII.B for a discussion of obligations to disclose pilot work.

<sup>67.</sup> Norbert Schwarz & Hans-Jürgen Hippler, Response Alternatives: The Impact of Their Choice and Presentation Order, in Measurement Errors in Surveys 41 (Paul P. Biemer et al. eds., 1991).

portions of the listed choices are unlikely to change dramatically, presentation of an explicit "don't know" or "no opinion" alternative commonly leads to an increase in that category of about 20% to 25%.<sup>68</sup>

Finally, the survey can include *full-filter questions*, that is, questions that lay the groundwork for the substantive question by first asking the respondent if he or she has an opinion about the issue or happened to notice the feature that the interviewer is preparing to ask about (e.g., "Based on the commercial you just saw, do you have an opinion about how long Clover stated or implied that its guarantee lasts?"). The interviewer then asks the substantive question only to those respondents who have indicated that they have an opinion on the issue.

The choice among these three approaches and the way they are used can affect the rate of "no opinion" responses that the question will evoke.<sup>69</sup> Respondents are more likely to say they do not have an opinion on an issue if a full-filter is used than if a quasi-filter is used.<sup>70</sup> However, in maximizing respondent expressions of "no opinion," full filters may produce an underreporting of opinions. There is some evidence that full-filter questions discourage respondents who actually have opinions from offering them by conveying the implicit suggestion that the respondent can avoid difficult follow-up questions by saying that he or she has no opinion.<sup>71</sup>

In general, then, a survey that uses full filters tends to provide a conservative estimate of the number of respondents holding an opinion, while a survey that uses neither full filters nor quasi-filters tends to overestimate the number of respondents with opinions, because some respondents offering opinions are guessing. The strategy of including "no opinion" or "don't know" as a quasi-filter avoids both of these extremes. Thus, rather than asking, "Based on the commercial, do you believe that the two products are made in the same way, or are they made differently?" or prefacing the question with a preliminary, "Do you have an opinion, based on the commercial, concerning the way that the two products are made?" the question could be phrased, "Based on the commercial, do you believe that the two products are made in the same way, that they are made differently, or don't you have an opinion about the way they are made?" <sup>72</sup>

<sup>68.</sup> Howard Schuman & Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context 113–46 (1981).

<sup>69.</sup> Considerable research has been conducted on the effects of filters. For a review, see George F. Bishop et al., Effects of Filter Questions in Public Opinion Surveys, 47 Pub. Opinion Q. 528 (1983).

<sup>70.</sup> Schwarz & Hippler, supra note 67, at 45–46.

<sup>71.</sup> Id. at 46.

<sup>72.</sup> The question in the example without the "no opinion" alternative was based on a question rejected by the court in Coors Brewing Co. v. Anheuser-Busch Cos., 802 F. Supp. 965, 972–73 (S.D.N.Y. 1992).

# C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified?

The questions that make up a survey instrument may be open-ended, closedended, or a combination of both. *Open-ended questions* require the respondent to formulate and express an answer in his or her own words (e.g., "What was the main point of the commercial?" "Where did you catch the fish you caught in these waters?" <sup>73</sup>). *Closed-ended questions* may provide the respondent with an explicit set of responses from which to choose, yes or no (e.g., "Is Colby College coeducational?" <sup>74</sup>), or they may offer respondents the choice among a number of specific alternatives (e.g., The two pain relievers have (1) the same likelihood of causing gastric ulcers; (2) about the same likelihood of causing gastric ulcers; (3) a somewhat different likelihood of causing gastric ulcers; (4) a very different likelihood of causing gastric ulcers; or (5) none of the above. <sup>75</sup>).

Open-ended and closed-ended questions may elicit very different responses.<sup>76</sup> Most responses are less likely to be volunteered in answering an open-ended question than to be endorsed in answering a closed-ended question. The response alternatives in a closed-ended question may remind respondents of options that they would not otherwise consider or which simply do not come to mind as easily.<sup>77</sup>

The advantage of open-ended questions is that they give the respondent fewer hints about the answer that is expected or preferred. Precoded responses on a closed-ended question, in addition to reminding respondents of options that they might not otherwise consider,<sup>78</sup> also may direct the respondent away from or toward a particular response. For example, a commercial reported that in shampoo tests with more than 900 women the sponsor's product received higher rat-

<sup>73.</sup> A relevant example from *Wilhoite v. Olin Corp.* is described in McGovern & Lind, *supra* note 24, at 76.

<sup>74.</sup> Presidents & Trustees v. Colby College, 508 F.2d 804, 809 (1st Cir. 1975).

<sup>75.</sup> This question is based on one asked in American Home Prods. Corp. v. Johnson & Johnson, 654 F. Supp. 568, 581 (S.D.N.Y. 1987) that was found to be a leading question by the court, primarily because the choices suggested that the respondent had learned about aspirin's and ibuprofen's relative likelihood of causing gastric ulcers. In contrast, in McNeilab, Inc. v. American Home Prods. Corp., 501 F. Supp. 517, 525 (S.D.N.Y. 1980), the court accepted as nonleading the question: "Based only on what the commercial said, would Maximum Strength Anacin contain more pain reliever, the same amount of pain reliever, or less pain reliever than the brand you, yourself, currently use most often?"

<sup>76.</sup> Howard Schuman & Stanley Presser, Question Wording as an Independent Variable in Survey Analysis, 6 Soc. Methods & Res. 151 (1977); Schuman & Presser, supra note 68, at 79–112; Converse & Presser, supra note 62, at 33.

<sup>77.</sup> For example, when respondents in one survey were asked, what is the most important thing for children to learn to prepare them for life, 62% picked "to think for themselves" from a list of five options, but only 5% spontaneously offered that answer when the question was open-ended. Schuman & Presser, *supra* note 68, at 104–07. An open-ended question presents the respondent with a free recall task, while a closed-ended question is a recognition task. Recognition tasks in general reveal higher performance levels than recall tasks. Mary M. Smyth et al., Cognition in Action 25 (1987). In addition, there is evidence that respondents answering open-ended questions may be less likely to report some information that they would reveal in response to a closed-ended question when that information seems self-evident or irrelevant.

<sup>78.</sup> Schwarz & Hippler, supra note 67, at 43.

ings than other brands. <sup>79</sup> According to a competitor, the commercial deceptively implied that each woman in the test rated more than one shampoo, when in fact each woman rated only one. To test consumer impressions, a survey might have shown the commercial and asked an open-ended question: "How many different brands mentioned in the commercial did each of the 900 women try?"<sup>80</sup> Instead, the survey asked a closed-ended question; respondents were given the choice between "one," "two," "three," "four," or "five or more." The choices in the closed-ended question (four of the five) implied that the correct answer was probably more than one.<sup>81</sup> Note, however, that the open-ended question also may suggest that the answer is more than one. By asking "how many different brands," the question suggests (1) that the viewer should have received some message from the commercial about the number of brands each woman tried and (2) that different brands were tried. Thus, the wording of a question, open-or closed-ended, can be leading, and the degree of suggestiveness of each question must be considered in evaluating the objectivity of a survey.

Closed-ended questions have some additional potential weaknesses that arise if the choices are not constructed properly. If the respondent is asked to choose one response from among several choices, the response only will be meaningful if the list of choices is exhaustive, that is, if the choices cover all possible positions a respondent might take on the question. If the list of possible choices is incomplete, a respondent may be forced to choose one that does not express his or her opinion.<sup>82</sup> Moreover, even if respondents are told explicitly that they are not limited to the choices presented, most respondents nevertheless will select an answer from among the listed choices.<sup>83</sup>

Although courts prefer open-ended questions on the grounds that they tend to be less leading, the value of any open- or closed-ended question depends on the information it is intended to elicit. Open-ended questions are more appropriate when the survey is attempting to gauge what comes first to a respondent's mind, but closed-ended questions are suitable for assessing choices between well-identified options or obtaining ratings on a clear set of alternatives.

<sup>79.</sup> See Vidal Sassoon, Inc. v. Bristol-Myers Co., 661 F.2d 272, 273 (2d Cir. 1981).

<sup>80.</sup> This was the wording of the stem of the closed-ended question in the survey discussed in Vidal Sassoon, 661 F.2d at 275–76.

<sup>81.</sup> Ninety-five percent of the respondents who answered the closed-ended question in the plaintiff's survey said that each woman had tried two or more brands. The open-ended question was never asked. *Vidal Sassoon*, 661 F.2d at 276. Norbert Schwarz, Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction, in Research Methods in Personality and Social Psychology 98 (Clyde Hendrick & Margaret S. Clark eds., 1990), suggests that respondents often rely on the range of response alternatives as a frame of reference when they are asked for frequency judgments.

See, e.g., American Home Prods. Corp. v. Johnson & Johnson, 654 F. Supp. 568, 581 (S.D.N.Y. 1987).
See Howard Schuman, Ordinary Questions, Survey Questions, and Policy Questions, 50 Pub. Opinion Q. 432 (1986).

D. If Probes Were Used to Clarify Ambiguous or Incomplete Answers, What Steps Were Taken to Ensure That the Probes Were Not Leading and Were Administered in a Consistent Fashion?

When questions allow respondents to express their opinions in their own words, some of the respondents may give ambiguous or incomplete answers. In such cases, interviewers may be instructed to record any answer that the respondent gives and move on to the next question, or they may be instructed to probe to obtain a more complete response or clarify the meaning of the ambiguous response. In either case, interviewers should record verbatim both what the respondent says and what the interviewer says in the attempt to get clarification. Failure to record the entire exchange in the order in which it occurs raises questions about the reliability of the survey, because neither the court nor the opposing party can evaluate whether the *probe* affected the views expressed by the respondent.

If the survey is designed to allow for probes, interviewers must be given explicit instructions on when they should probe and what they should say in probing. Standard probes used to draw out all that the respondent has to say (e.g., "Any further thoughts?" "Anything else?" "Can you explain that a little more?") are relatively innocuous and noncontroversial in content, although they may convey the idea to the respondent that he or she has not yet produced the "right" answer. Interviewers should be trained in delivering probes to maintain (as they should during the rest of the interview) a professional and neutral relationship with the respondent, which minimizes any sense of passing judgment on the content of the answers offered. Moreover, interviewers should be given explicit instructions on when to probe, so that probes are administered consistently.

A more difficult type of probe to construct and deliver reliably is one that requires a substantive question tailored to the answer given by the respondent. The survey designer must provide sufficient instruction to avoid giving directive probes that suggest one answer over another. Those instructions, along with all other aspects of interviewer training, should be made available for evaluation by the court and the opposing party.

# E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects?

The order in which questions are asked on a survey and the order in which response alternatives are provided in a closed-ended question can influence the answers.<sup>84</sup> Thus, although asking a general question before a more specific

<sup>84.</sup> See Schuman & Presser, supra note 68, at 23, 56–74; Norman M. Bradburn, Response Effects, in Handbook of Survey Research, supra note 1, at 289, 302. In R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc., 511 F. Supp. 867, 875 (S.D.N.Y. 1980), the court recognized the biased structure of a survey which dis-

question on the same topic is unlikely to affect the response to the specific question, reversing the order of the questions may influence responses to the general question. As a rule, then, surveys are less likely to be subject to *order effects* if the questions go from the general (e.g., "What do you recall being discussed in the commercial?") to the specific (e.g., "Based on your reading of the advertisement, what companies do you think the ad is referring to when it talks about rental trucks that average five miles per gallon?").<sup>85</sup>

The mode of questioning can influence the form that an order effect takes. In mail surveys, respondents are more likely to select the first choice offered (a *primacy* effect), while in telephone surveys, respondents are more likely to choose the last choice offered (a *recency* effect). Although these effects are typically small, no general formula is available that can adjust values to correct for order effects, because the size and even the direction of the order effects may be affected by the nature of the question being asked and the choices being offered. Moreover, it may be unclear which order is most appropriate. For example, if the respondent is asked to choose between two different products, and there is a tendency for response level the most accurate?

To control for order effects, the order of the questions and the order of the response choices should be rotated, so that, for example, one-third of the respondents have Product A listed first, one-third of the respondents have Product B listed first, and one-third of the respondents have Product C listed first. If the three different orders are distributed randomly among respondents, no response alternative will have an inflated chance of being selected due to position, and the average of the three will provide the most appropriate estimate of response level. <sup>87</sup>

# F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question?

Most surveys that are designed to provide evidence of trademark infringement or deceptive advertising are not conducted to describe consumer beliefs. Instead, they are intended to show how the trademark or content of the commercial influences respondents' perceptions or understanding of a product or commercial.

closed the tar content of the cigarettes being compared before questioning respondents about their cigarette preferences. Not surprisingly, respondents expressed a preference for the lower tar product. *Id.* 

<sup>85.</sup> This question was accepted by the court in U-Haul Int'l, Inc. v. Jartran, Inc., 522 F. Supp. 1238, 1249 (D. Ariz. 1981), *aff'd*, 681 F.2d 1159 (9th Cir. 1982).

<sup>86.</sup> Similarly, candidates in the first position on the ballot tend to attract extra votes when the candidates are not well known. Henry M. Bain & Donald S. Hecock, Ballot Position and Voter's Choice: The Arrangement of Names on the Ballot and its Effect on the Voter (1973).

<sup>87.</sup> Although rotation is desirable, many surveys are conducted with no attention to this potential bias. Since it is impossible to know in the abstract whether a question suffers much, little, or not at all from an order bias, lack of rotation should not preclude reliance on the answer to a particular question, but it should reduce the weight given that response.
Thus, the question is whether the commercial misleads the consumer into thinking that Product A is a superior pain reliever, not whether consumers hold inaccurate beliefs about the product. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that describes consumer impressions after they view the advertisement may reflect those preexisting beliefs rather than impressions produced by the commercial.

Surveys that record consumer impressions have a limited ability to answer questions about the origins of those impressions. The difficulty is that the consumer's response to any question on the survey may be the result of information or misinformation from sources other than the trademark the respondent is being shown or the commercial he or she has just watched. In a trademark survey attempting to show secondary meaning, for example, respondents were shown a picture of the stripes used on Mennen stick deodorant and asked, "Which brand would you say uses these stripes on their package?" <sup>88</sup> The court recognized that the high percentage of respondents who selected "Mennen" from an array of brand names may have represented "merely a playback of brand share." <sup>89</sup> That is, respondents asked to give a brand name may guess the one that is most familiar, generally the brand with the largest market share.

Some surveys attempt to reduce the impact of preexisting impressions on respondents' answers by instructing respondents to focus solely on the stimulus as a basis for their answers. Thus, the survey includes a preface ("based on the commercial you just saw") or directs the respondent's attention to the mark at issue (e.g., "these stripes on the package"). Such efforts are likely to be only partially successful. It is often difficult for respondents to identify accurately the source of their impressions.<sup>90</sup> The more routine the idea being tested (e.g., that the advertised pain reliever is more effective than others on the market; that the mark belongs to the brand with the largest market share), the more likely it is that the respondent's answer is influenced by preexisting impressions, by expectations about what commercials generally say, or by guessing, rather than by the actual content of the commercial message or trademark being evaluated.

It is possible to adjust many survey designs so that causal inferences about the effect of a trademark or an allegedly deceptive commercial become clear and unambiguous. By adding an appropriate control group, the survey expert can test directly the influence of the stimulus.<sup>91</sup> In the simplest version of a survey ex-

<sup>88.</sup> Mennen Co. v. Gillette Co., 565 F. Supp. 648, 652 (S.D.N.Y. 1983), *affd*, 742 F.2d 1437 (2d Cir. 1984). To demonstrate secondary meaning, "the [c]ourt must determine whether the mark has been so associated in the mind of consumers with the entity that it identifies that the goods sold by that entity are distinguished by the mark or symbol from goods sold by others." *Id*. at 652.

<sup>89.</sup> Id.

<sup>90.</sup> Sæ Richard E. Nisbett & Timothy D. Wilson, Telling More Than We Can Know: Verbal Reports on Mental Processes, 84 Psychol. Rev. 231 (1977).

<sup>91.</sup> See Shari S. Diamond, Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing, 13 Law & Hum. Behav. 239 (1989); Shari S. Diamond & Linda Dimitropoulos, Deception and Puffery in Advertising: Behavioral Science Implications for Regulation, in Advertising, Law, and the Social Sciences 21 (J. Lipton & B. D. Sales eds., 1994); Jacob Jacoby & Constance Small, Applied Marketing: The

periment, respondents are assigned randomly to one of two conditions.<sup>92</sup> Respondents assigned to the experimental condition view the allegedly deceptive commercial, and respondents assigned to the control condition either do not view any commercial or view a commercial that does not contain the allegedly deceptive material. The same questions are then asked of all respondents. If 40% of the viewers of the test commercial report that it conveys a deceptive message (e.g., the product has fewer calories than its competitor), that response rate is evaluated against the base background noise level obtained from the control group. If 40% of the viewers who watched the commercial without the allegedly deceptive message report getting the idea that the product has fewer calories than its competitor, then the 40% logically cannot be attributed to the content of the commercial being examined for deception. If the confusion level in the control group is significantly lower than 40%, then the increased confusion of respondents in the experimental group can be attributed only to the commercial they viewed. The difference cannot be the result of background noise, preexisting beliefs, or even a leading question, because all these explanations should have produced similar response levels in the experimental and control groups. Thus, the focus on the response level in a control group design is not on the absolute response level, but rather on the difference in response levels between the experimental and control groups.

Explicit attention to the value of control groups in trademark and deceptive advertising litigation is a relatively recent phenomenon. A LEXIS search using Lanham Act and control group revealed five district court cases since 1987, <sup>93</sup> and only one case before 1987, <sup>94</sup> in which surveys with control groups were discussed. Other cases, however, have described or considered surveys using control group designs without labeling the comparison group a control group.<sup>95</sup> Indeed, the relative absence of control groups in reported cases may reflect the

FDA Approach to Defining Misleading Advertising, 39 J. Mktg. 65 (1975). For a more general discussion of the role of control groups, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.C, in this manual.

<sup>92.</sup> Random assignment should not be confused with random selection. When respondents are assigned randomly to different treatment groups (e.g., respondents in each group watch a different commercial), the procedure ensures that within the limits of sampling error the two groups of respondents will be equivalent except for the different treatments they receive. Respondents selected for a mall intercept study, and not from a probability sample, may be assigned randomly to different treatment groups. Random selection, in contrast, describes the method of selecting a sample of respondents in a probability sample. See supra § III.C.

<sup>93.</sup> Conagra, Inc. v. George A. Hormel & Co., 784 F. Supp. 700 (D. Neb. 1992), *aff'd*, 990 F.2d 368 (8th Cir. 1993); Johnson & Johnson-Merck Consumer Pharmaceuticals Co. v. Smithkline Beecham Corp., No. 91-C0960, 1991 U.S. Dist. LEXIS 13689 (S.D.N.Y. Sept. 30, 1991), *aff'd*, 960 F.2d 294 (2d Cir. 1992); Goya Foods, Inc. v. Condal Distribs., Inc., 732 F. Supp. 453 (S.D.N.Y. 1990); Sturm, Ruger & Co. v. Arcadia Mach. & Tool, Inc., No. 85-8459, 1988 U.S. Dist. LEXIS 16451 (C.D. Cal. Nov. 7, 1988); Frisch's Restaurant, Inc. v. Elby's Big Boy, Inc., 661 F. Supp. 971 (S.D. Ohio 1987), *aff'd*, 849 F.2d 1012 (6th Cir. 1988).

<sup>94.</sup> American Basketball Ass'n v. AMF Voit, Inc., 358 F. Supp. 981 (S.D.N.Y.), aff'd, 487 F.2d 1393 (2d Cir. 1973).

<sup>95.</sup> See, e.g., Quality Inns Int'l, Inc. v. McDonald's Corp., 695 F. Supp. 198, 218 (D. Md. 1988) (survey revealed confusion between McDonald's and McSleep, but control survey revealed no confusion between McDonald's and McTavish).

fact that a survey with a control group produces less ambiguous findings, which leads to a resolution before a preliminary injunction hearing or trial occurs. <sup>96</sup>

Another more common use of control methodology is a control question. Rather than administering a control stimulus to a separate group of respondents, the survey asks all respondents one or more control questions along with the question about the product or service. In a trademark dispute, for example, a survey indicated that 7.2% of respondents believed that "The Mart" and "K-Mart" are owned by the same individuals. The court found no likelihood of confusion based on survey evidence that 5.7% of the respondents also thought that "The Mart" and "King's Department Store" were owned by the same source.<sup>97</sup>

Similarly, a standard technique used to evaluate whether a brand name is generic is to present survey respondents with a series of product or service names and ask them to indicate in each instance whether they believe the name is a brand name or a common name. By showing that 68% of respondents considered Teflon a brand name (a proportion similar to the 75% of respondents who recognized the acknowledged trademark Jell-O as a brand name, and markedly different from the 13% who thought aspirin was a brand name), the makers of Teflon retained their trademark.<sup>98</sup>

Every measure of opinion or belief in a survey reflects some degree of error. Control groups and control questions are the most reliable means for assessing response levels against the baseline level of error associated with a particular question.

# G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey?

Three primary methods are used to collect survey data: (1) in-person interviews, (2) telephone interviews, and (3) mail surveys.<sup>99</sup> The choice of a data collection method for a survey should be justified by its strengths and weaknesses.

#### 1. In-person interviews

Although costly, in-person interviews generally are the preferred method of data collection, especially when visual materials must be shown to the respondent under controlled conditions.<sup>100</sup> When the questions are complex and the in-

<sup>96.</sup> The paucity of control groups in surveys discussed in federal cases is not confined to Lanham Act litigation. A LEXIS search using *survey* and *control group* revealed fifty-five cases in which control group was used to refer to a methodological feature. The majority (thirty-five cases) referred to medical, physiological, or pharmacological experiments.

<sup>97.</sup> S.S. Kresge Co. v. United Factory Outlet, Inc., 598 F.2d 694, 697 (1st Cir. 1979).

<sup>98.</sup> E.I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502 (E.D.N.Y. 1975).

<sup>99.</sup> Methods also may be combined, as when the telephone is used to "screen" for eligible respondents who then are invited to participate in an in-person interview.

<sup>100.</sup> A mail survey also can include limited visual materials but cannot exercise control over when and how the respondent views them.

terviewers are skilled, in-person interviewing provides the opportunity to clarify or probe, the ability to implement complex skip sequences (in which the respondent's answer determines which question will be asked next), and the power to control the order in which the respondent answers the questions. As described in section V.A, appropriate training is necessary if these potential benefits are to be realized. Objections to the use of in-person interviews arise primarily from their high cost or, on occasion, from evidence of inept or biased interviewers.

#### 2. Telephone surveys

Telephone surveys offer a comparatively fast and low-cost alternative to in-person surveys and are particularly useful when the population is large and geographically dispersed. Telephone interviews (unless supplemented with mailed materials) can be used only when it is unnecessary to show the respondent any visual materials. Thus, an attorney may present the results of a telephone survey of jury-eligible citizens in a motion for a change of venue in order to provide evidence that community prejudice raises a reasonable suspicion of potential jury bias.<sup>101</sup> Similarly, potential confusion between a restaurant called McBagel's and the McDonald's fast-food chain was established in a telephone survey. Over objections from defendant McBagel's that the survey did not show respondents the defendant's print advertisements, the court found likelihood of confusion based on the survey, noting that "by soliciting audio responses [, the telephone survey] was closely related to the radio advertising involved in the case."<sup>102</sup> In contrast, when words are not sufficient because, for example, the survey is assessing reactions to the trade dress or packaging of a product that is alleged to promote confusion, a telephone survey alone does not offer a suitable vehicle for questioning respondents.<sup>103</sup>

In evaluating the sampling used in a telephone survey, the trier of fact should consider:

1. (when prospective respondents are not business personnel) whether some form of random digit dialing<sup>104</sup> was used instead of or to supplement telephone numbers obtained from telephone directories, be-

Survey Research

<sup>101.</sup> United States v. Partin, 320 F. Supp. 275 (E.D. La. 1970). For a discussion of surveys used in motions for change of venue, see Neal Miller, Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation, Part II, 40 Rutgers L. Rev. 467, 470–74 (1988); Nat'l Jury Project, Jurywork: Systematic Techniques (Elissa Krauss & Beth Bonora eds., 2d ed. 1983).

<sup>102.</sup> McDonald's Corp. v. McBagel's, Inc., 649 F. Supp. 1268, 1278 (S.D.N.Y. 1986).

<sup>103.</sup> Thompson Medical Co. v. Pfizer, Inc., 753 F.2d 208 (2d Cir. 1985); Incorporated Pub. Corp. v. Manhattan Magazine, Inc., 616 F. Supp. 370 (S.D.N.Y. 1985), aff'd without op., 788 F.2d 3 (2d Cir. 1986).

<sup>104.</sup> Random digit dialing provides coverage of households with both listed and unlisted telephone numbers by generating numbers at random from the frame of all possible telephone numbers. James M. Lepkowski, *Telephone Sampling Methods in the United States, in* Telephone Survey Methodology 81–91 (Robert M. Groves et al. eds., 1988).

cause up to 65% of all residential telephone numbers in some areas may be unlisted;  $^{\rm 105}$ 

- 2. whether the sampling procedures required the interviewer to sample within the household or business, instead of allowing the interviewer to administer the survey to any qualified individual who answered the telephone;<sup>106</sup> and
- 3. whether interviewers were required to call back at several different times of the day and on different days before dropping a potential respondent from the sample.

Telephone surveys that do not include these procedures may, like other nonprobability sampling approaches, be adequate for providing rough approximations. The vulnerability of the survey depends on the information being gathered. More elaborate procedures for achieving a representative sample of respondents are advisable if the survey instrument requests information that is likely to be different for individuals with listed and unlisted telephone numbers, different for individuals rarely at home and those usually at home, and so forth.

The report submitted by a survey expert who conducts a telephone survey should specify:

- 1. the procedures that were used to identify potential respondents;
- 2. the number of telephone numbers where no contact was made; and
- 3. the number of contacted potential respondents who refused to participate in the survey.

#### 3. Mail surveys

In general, mail surveys tend to be substantially less costly than both in-person and telephone surveys. <sup>107</sup> Although response rates for mail surveys are often low, researchers have obtained 70% response rates in some general public surveys and response rates of over 90% with certain specialized populations. <sup>108</sup> Procedures that encourage high response rates include multiple mailings, highly personalized communications, and prepaid return envelopes.

A mail survey will not produce a high response rate unless it begins with an accurate and up-to-date list of names and addresses for the target population. Even if the sampling frame is adequate, the sample may be unrepresentative if some individuals are more likely to respond than others. For example, if a survey targets a population that includes individuals with literacy problems, these indi-

<sup>105.</sup> In 1992, the percentage of households with unlisted numbers reached 65% in Las Vegas and 62% in Los Angeles. Survey Sampling, The Frame (March 1993). Studies comparing listed and unlisted household characteristics show some important differences. Lepkowski, *supra* note 104, at 76.

<sup>106.</sup> This is true only if the survey is sampling individuals. If the survey is seeking information on the household, more than one individual may be able to answer questions on behalf of the household.

<sup>107.</sup> Don A. Dillman, Mail and Other Self-Administered Questionnaires , in Handbook of Survey Research, supra note 1, at 359, 373.

<sup>108.</sup> Id. at 360.

viduals tend to be underrepresented among the respondents to a mail survey. Open-ended questions are generally of limited value on a mail survey because they depend entirely on the respondent to answer fully and do not provide the opportunity to probe or clarify unclear answers. Similarly, if eligibility to answer some questions depends on the answers to previous questions, such skip sequences may be difficult for some respondents to follow. Finally, because respondents complete mail surveys without supervision, survey personnel are unable to control the order in which respondents answer the questions. If it is crucial to have respondents answer questions in a particular order, a mail survey cannot be depended on to provide adequate data.<sup>109</sup>

109. Id. at 368-70.

Survey Research

### V. Surveys Involving Interviewers

#### A. Were the Interviewers Appropriately Selected and Trained?

A properly defined population or universe, a representative sample, and clear and precise questions can be depended on to produce trustworthy survey results only if "sound interview procedures were followed by competent interviewers."<sup>110</sup> Properly trained interviewers receive detailed instructions on everything they are to say to respondents, any stimulus materials they are to use in the survey, and how they are to complete the interview form. These instructions should be made available to the opposing party and to the trier of fact. Thus, interviewers should be told, and the interview form on which answers are recorded should indicate, which responses, if any, are to be read to the respondent. Interviewers also should be instructed to record verbatim the respondent's answers, to indicate explicitly whenever they repeat a question to the respondent, and to record any statements they make to or supplementary questions they ask the respondent.

Interviewers require training to ensure that they are able to follow directions in administering the survey questions. Some training in general interviewing techniques is required for most interviews (e.g., practice in pausing to give the respondent enough time to answer and in resisting invitations to express the interviewer's beliefs or opinions). Although practices vary, one treatise recommends at least five hours of training in general interviewing skills and techniques for new interviewers.<sup>111</sup>

The more complicated the survey instrument is, the more training and experience the interviewers require. Thus, if the interview includes a skip pattern (where, e.g., Questions 4–6 are asked only if the respondent says yes to Question 3, and Questions 8–10 are asked only if the respondent says no to Question 3), interviewers must be trained to follow the pattern. Similarly, if the questions require specific probes to follow up ambiguous responses, interviewers must receive instruction on when to use the probes and what to say. In some surveys, the interviewer is responsible for last-stage sampling (i.e., selecting the particular

<sup>110.</sup> Toys "R" Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983).

<sup>111.</sup> Eve Weinberg, Data Collection: Planning and Management, in Handbook of Survey Research, supra note 1, at 329, 332.

respondents to be interviewed), and training is especially crucial to avoid interviewer bias in selecting respondents who are easiest to approach or easiest to find.

Training and instruction of interviewers should include directions on the circumstances under which interviews are to take place (e.g., question only one respondent at a time out of the hearing of any other respondent). The trustworthiness of a survey is questionable if there is evidence that some interviews were conducted in a setting in which respondents were likely to have been distracted or in which others were present and could overhear. Such evidence of careless administration of the survey was one ground used by a court to reject as inadmissible a survey that purported to demonstrate consumer confusion.<sup>112</sup>

Some compromises may be accepted when surveys must be conducted swiftly. In the trademark and deceptive advertising area, the plaintiff's usual request is for a preliminary injunction, because a delay means irreparable harm. Nonetheless, careful instruction and training of interviewers who administer the survey and complete disclosure of the methods used for instruction and training are crucial elements that if compromised, seriously undermine the trustworthiness of any survey.

# B. What Did the Interviewers Know About the Survey and Its Sponsorship?

One way to protect the objectivity of survey administration is to avoid telling interviewers who is sponsoring the survey. Interviewers who know the identity of the survey's sponsor may affect results inadvertently by communicating to respondents their expectations or what they believe are the preferred responses of the survey's sponsor. To ensure objectivity in the administration of the survey, it is standard interview practice to conduct *double-blind research* whenever possible: both the interviewer and the respondent are blind to the sponsor of the survey and its purpose. Thus, the survey instrument should provide no explicit clues (e.g., a sponsor's letterhead appearing on the survey) and no implicit clues (e.g., reversing the usual order of the yes and no response boxes on the interviewer's form next to a crucial question, thereby potentially increasing the likelihood that *ro* will be checked<sup>113</sup>) about the sponsorship of the survey or the expected responses.

Nonetheless, in some cases (e.g., some government surveys), sponsorship disclosure to respondents (and thus to interviewers) is required. Such instances call for an evaluation of the likely biases introduced by interviewer or respondent

<sup>112.</sup> Toys "R" Us, 559 F. Supp. at 1204 (some interviews apparently were conducted in a bowling alley; some interviewees waiting to be interviewed overheard the substance of the interview while they were waiting). 113. Centaur Communications, Ltd. v. A/S/M Communications, Inc., 652 F. Supp. 1105, 1111 n.3 (S.D.N.Y.) (pointing out that reversing the usual order of response choices, yes or no, to no or yes may confuse interviewers as well as introduce bias), aff/d, 830 F.2d 1217 (2d Cir. 1987).

awareness. In evaluating the consequences of sponsorship awareness, it is important to consider (1) whether the sponsor has views and expectations that are apparent and (2) whether awareness is confined to the interviewers or involves the respondents. For example, if a survey concerning attitudes toward gun control is sponsored by the National Rifle Association, it is clear that responses opposing gun control are likely to be preferred. In contrast, if the survey on gun control attitudes is sponsored by the Department of Justice, the identity of the sponsor may not suggest the kind of responses the sponsor expects or would find acceptable. When interviewers are well trained, their awareness of sponsorship may be a less serious threat than respondent awareness. The empirical evidence for the effects of interviewers' prior expectations on respondents' answers generally reveals modest effects when the interviewers are well trained.<sup>114</sup>

# C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias?

Three methods are used to ensure that the survey instrument was implemented in an unbiased fashion and according to instructions. The first, monitoring the interviews as they occur, is done most easily when telephone surveys are used. A supervisor listens to a sample of interviews for each interviewer. Field settings make monitoring more difficult, but evidence that monitoring has occurred provides an additional indication that the survey has been reliably implemented.

Second, validation of interviews occurs when a sample of respondents is recontacted to ask whether the initial interviews took place and to determine whether the respondent was qualified to participate in the survey. The standard procedure for validation is to telephone a random sample of about 15% of the respondents.<sup>115</sup> Some attempts to reach the respondent will be unsuccessful, and occasionally a respondent will deny that the interview took place even though it did. Because the information checked is limited to whether the interview took place and whether the respondent was qualified, this validation procedure does not determine whether the initial interview as a whole was conducted properly. Nonetheless, this standard validation technique warns interviewers that their work is being checked and can detect gross failures in the administration of the survey.

A third way to verify that the interviews were conducted properly is to compare the work done by each individual interviewer. By reviewing the interviews and individual responses recorded by each interviewer, any response patterns or inconsistencies can be identified for further investigation.

<sup>114.</sup> See, e.g., Seymour Sudman et al., Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses, 6 Soc. Methods & Res. 171 (1977).

<sup>115.</sup> See, e.g., National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 515 (D.N.J. 1986).

## VI. Data Entry and Grouping of Responses

#### A. What Was Done to Ensure That the Data Were Recorded Accurately?

To analyze the results of a survey, the data obtained on each sampled element must be recorded, edited, and often coded before the results can be tabulated and processed. Procedures for data entry should include checks for completeness, checks for reliability and accuracy, and rules for resolving inconsistencies. Accurate data entry is maximized when responses are verified by duplicate entry and comparison, and when data entry personnel are unaware of the purposes of the survey.

# B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately?

Coding of answers to open-ended questions requires a detailed set of instructions so that decision standards are clear and responses can be scored consistently and accurately. Two trained coders should independently score the same subset of responses to check for the level of consistency in classifying responses. When the criteria used to categorize verbatim responses are controversial or allegedly inappropriate, those criteria should be sufficiently clear to reveal the source of disagreements. In such cases, the verbatim responses should be available so that they can be recoded using alternative criteria.<sup>116</sup>

<sup>116.</sup> See, e.g., Coca-Cola Co. v. Tropicana Prods., Inc., 538 F. Supp. 1091, 1094–96 (S.D.N.Y.) (plaintiff's expert stated that respondents' answers to the several open-ended questions revealed that 43% of respondents thought Tropicana was portrayed as fresh squeezed; the court's own tabulation found no more than 15% believed this was true), *rev'd on other grounds*, 690 F.2d 312 (2d Cir. 1982). S ee also McNeilab, Inc. v. American Home Prods. Corp., 501 F. Supp. 517 (S.D.N.Y. 1980).

#### VII. Disclosure and Reporting

## A. When Was Information About the Survey Methodology and Results Disclosed?

Objections to the definition of the relevant population, the method of selecting the sample, and the wording of questions generally are raised for the first time when the results of the survey are presented. By that time it is too late to correct methodological deficiencies that could have been addressed in the planning stages of the survey. The plaintiff in a trademark case<sup>117</sup> submitted a set of proposed survey questions to the trial judge who ruled that the survey results would be admissible at trial, while reserving the question of the weight the evidence would be given.<sup>118</sup> The court of appeals called this approach a commendable procedure and suggested that it would have been even more desirable if the parties had "attempt[ed] in good faith to agree upon the questions to be in such a survey."<sup>119</sup>

The Manual for Complex Litigation <sup>120</sup> recommends that parties be required, "before conducting any poll, to provide other parties with an outline of the proposed form and methodology, including the particular questions that will be asked, the introductory statements or instructions that will be given, and other controls to be used in the interrogation process." <sup>121</sup> The parties then are encouraged to attempt to resolve any methodological disagreements before the survey is conducted.<sup>122</sup> Although this passage in the previous edition of the Manual has been cited with apparent approval, <sup>123</sup> the prior agreement the Manual recommends has occurred rarely.

<sup>117.</sup> Union Carbide Corp. v. Ever-Ready, Inc., 392 F. Supp. 280 (N.D. Ill. 1975), rev'd, 531 F.2d 366 (7th Cir.), cert. denied, 429 U.S. 830 (1976).

<sup>118.</sup> Before trial, the presiding judge was appointed to the court of appeals, so the case was tried by another district court judge.

<sup>119.</sup> Union Carbide, 531 F.2d at 386. More recently, the Seventh Circuit recommended the filing of a motion in limine, asking the district court to determine the admissibility of a survey based on an examination of the survey questions and the results of a preliminary survey before the party undertakes the expense of conducting the actual survey. Piper Aircraft Corp. v. Wag-Aero, Inc., 741 F.2d 925, 929 (7th Cir. 1984).

<sup>120.</sup> MCL 3d, supra note 14, § 21.493.

<sup>121.</sup> Id.

<sup>122.</sup> Id.

<sup>123.</sup> E.g., National Football League Properties, Inc., v. New Jersey Giants, Inc., 637 F. Supp. 507, 514 n.3 (D.N.J. 1986).

Recent amendments to Rule 26 of the Federal Rules of Civil Procedure require extensive disclosure of the basis of opinions offered by testifying experts. These provisions, however, may not produce disclosure of survey materials, because parties are not obligated to disclose information about nontestifying experts. Parties considering whether to commission or use a survey for litigation are not obligated to present a survey that produces unfavorable results. Prior disclosure of a proposed survey instrument places the party that ultimately would prefer not to present the survey in the position of presenting damaging results or leaving the impression that the results are not being presented because they were unfavorable. Anticipating such a situation, parties do not decide whether an expert will testify until after the results of the survey are available.

Nonetheless, courts are in a position to encourage early disclosure and discussion even if they do not lead to agreement between the parties. Judge William C. Conner encouraged the parties to submit their survey plans for court approval to ensure their evidentiary value. The plaintiff McNeil did so and altered its research plan based on Judge Conner's recommendations.<sup>124</sup> Parties can anticipate that changes consistent with a judicial suggestion are likely to increase the weight given to, or at least the prospects of admissibility of, the survey.<sup>125</sup>

## B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics?

The completeness of the survey report is one indicator of the trustworthiness of the survey and the professionalism of the expert who is presenting the results of the survey. A survey report generally should describe in detail:

- 1. the purpose of the survey;
- 2. a definition of the target population and a description of the population that was actually sampled;
- 3. a description of the sample design, including the method of selecting respondents, the method of interview, the number of callbacks, respondent eligibility or screening criteria, and other pertinent information;
- a description of the results of sample implementation, including (a) the number of potential respondents contacted, (b) the number not reached, (c) the number of refusals, (d) the number of incomplete interviews or terminations, (e) the number of noneligibles, and (f) the number of completed interviews;
- 5. the exact wording of the questions used, including the actual questionnaire, interviewer directions, and visual exhibits;

<sup>124.</sup> McNeilab, Inc. v. American Home Prods. Corp., 848 F.2d 34, 36 (2d Cir. 1988) (discussing with approval the actions of the district court).

<sup>125.</sup> Larry C. Jones, Developing and Using Survey Evidence in Trademark Litigation, 19 Mem. St. U. L. Rev. 471, 481 (1989).

- 6. a description of any special scoring (e.g., grouping of verbatim responses into broader categories);
- 7. estimates of the sampling error where appropriate (i.e., in probability samples);
- 8. statistical tables clearly labeled and identified as to source of data, including the number of raw cases forming the base for each table, row, or column; and
- 9. copies of interviewer instructions, validation results, and codebooks.<sup>126</sup>

A description of the procedures and results of pilot testing is not included on this list. Survey professionals generally do not describe pilot testing in their reports. The Federal Rules of Civil Procedure, however, may require that a testifying expert disclose pilot work that serves as a basis for the expert's opinion. The situation is more complicated when a nontestifying expert conducts the pilot work and the testifying expert learns about the pilot testing only indirectly through the attorney's advice about the relevant issues in the case. Some commentators suggest that attorneys are obligated to disclose such pilot work.<sup>127</sup>

# C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents?

The respondents questioned in a survey generally do not testify in legal proceedings and are unavailable for cross-examination. Indeed, one of the advantages of a survey is that it avoids a repetitious and unrepresentative parade of witnesses. To verify that interviews occurred with qualified respondents, standard survey practice includes validation procedures,<sup>128</sup> the results of which should be included in the survey report.

Conflicts may arise when an opposing party asks for respondents' names and addresses in order to reinterview some survey respondents. The party introducing the survey or the survey organization that conducted the research generally resists supplying such information.<sup>129</sup> Professional surveyors as a rule guarantee confidentiality in an effort to increase participation rates and to encourage candid responses. Because failure to extend confidentiality may bias both the willingness of potential respondents to participate in a survey and their responses, the professional standards for survey researchers generally prohibit disclosure of respondent identity. "The use of survey results in a legal proceeding does not relieve the survey research organization of its ethical obligation to maintain in confidence all respondent-identifiable information or lessen the im-

<sup>126.</sup> These criteria were adapted from the Council of Am. Survey Res. Orgs., supra note 38, § III.B.

<sup>127.</sup> Yvonne C. Schroeder, Pretesting Survey Questions, 11 Am. J. Trial Advoc. 195 (1987).

<sup>128.</sup> See supra §V.C.

<sup>129.</sup> See, e.g. , Alpo Petfoods, Inc. v. Ralston Purina Co., 720 F. Supp. 194 (D.D.C. 1989), aff'd in part & vacated in part, 913 F.2d 958 (D.C. Cir. 1990).

portance of respondent anonymity."<sup>130</sup> Although no surveyor-respondent privilege currently is recognized, the need for surveys and the availability of other means to examine and ensure their trustworthiness argues for deference to legitimate claims for confidentiality in order to avoid seriously compromising the ability of surveys to produce accurate information.<sup>131</sup>

Copies of all questionnaires should be made available upon request so that the opposing party may have an opportunity to evaluate the raw data. All identifying information, such as the respondent's name, address, and telephone number, should be removed to ensure respondent confidentiality.

<sup>130.</sup> See, e.g., Council of Am. Survey Res. Orgs., *supra* note 38, § I.A.3.f. Similar provisions are contained in the By-Laws of the American Association for Public Opinion Research. 131. *In re* Litton Industries, Inc., No. 9123, 1979 FTC LEXIS 311, at \*13 & n.12 (June 19, 1979) (Order

<sup>131.</sup> In re Litton Industries, Inc., No. 9123, 1979 FTC LEXIS 311, at \*13 & n.12 (June 19, 1979) (Order Concerning the Identification of Individual Survey-Respondents with Their Questionnaires) (citing Frederick H. Boness & John F. Cordes, Note, *The Researcher-Subject Relationship: The Need for Protection and a Model Statute*, 62 Geo. LJ. 243, 253 (1973)).

#### **Glossary of Terms**

The following terms and definitions were adapted from a variety of sources, including: Handbook of Survey Research (Peter H. Rossi et al. eds., 1983); 1 Environmental Protection Agency, Survey Management Handbook (1983); Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991); Seymour Sudman, Applied Sampling (1976).

- Closed-Ended Question. A question that provides the respondent with a list of choices and asks the respondent to choose from among them.
- *Cluster Sampling.* Technique allowing for the selection of sample cases in groups or clusters, rather than on an individual basis; may significantly reduce field costs and increase sampling error.
- *Confidence Interval.* An indication of the probable range of error associated with a sample value obtained from a probability sample. Also, margin of error.
- Convenience Sample. A sample of elements selected because they were readily available.
- Double-Blind Research. Research in which the respondent and the interviewer are not given information that will alert them to the anticipated or preferred pat tern of response.
- *Error Score.* The degree of measurement error in an observed score (see true score).
- *Full-Filter Question.* A question asked of respondents to screen out those who do not have an opinion on the issue under investigation before asking them the question proper.
- Mall Intercept Survey. A survey conducted in a mall or shopping center in which potential respondents are approached by a recruiter (intercepted) and invited to participate in the survey.
- Multistage Sampling Design. Design in which sampling takes place in several stages, beginning with larger units (e.g., cities) and then proceeding with, for example, households or individuals within these units.

Nonprobability Sample. Any sample that does not qualify as a probability sample.

- Open-Ended Question. A question that requires the respondent to formulate his or her own response.
- *Order Effect.* A tendency of respondents to choose an item based in part on the order in which it appears in the question, questionnaire, or interview (see Primacy Effect, Recency Effect).
- *Parameter.* A characteristic of a population. For example, statistics are estimates of parameters.
- *Pilot Test.* A small field test replicating the field procedures planned for the fullscale survey; although the terms pilot test and pretest are sometimes used interchangeably, a pretest tests the questionnaire, while a pilot test generally tests proposed collection procedures as well.
- Population. The totality of elements (individuals, objects, or measurements) that have some common property of interest; the target population is the collection of elements that the researcher would like to study; the survey population is the population that is actually sampled and for which data may be obtained. Also, universe.
- Population Value, Population Parameter. The actual value of some characteristic in the population (e.g., the average age); a sample from the population is measured to estimate the population value.
- *Primacy Effect.* A tendency of respondents to choose early items from a list of choices; the opposite of a recency effect.
- Probability Sample. Sample selected so that every element in the population has a known nonzero probability of being included in the sample; a simple random sample is a probability sample.
- Probe. A follow-up question that an interviewer asks to obtain a more complete answer from a respondent (e.g., "Anything else?" "What kind of x do you mean?").
- Quasi-Filter Question. A question that offers a "no opinion" option to respondents as part of a set of response alternatives; used to screen out respondents who may not have an opinion on the issue under investigation.
- Random Sample. See Simple Random Sample.
- *Recency Effect.* A tendency of respondents to choose later items from a list of choices; the opposite of a primacy effect.
- Sample. A subset of a population or universe selected so as to yield information about the population as a whole.

- Sampling Error. The difference between the result obtained from a sample study and the result that would be obtained by attempting a complete study of all units in the sampling frame from which the sample was selected in the same manner and with the same care.
- Sampling Frame. The list of all objects, individuals, or other social organisms in the population of interest from which the sample is drawn.
- Simple Random Sample. The most basic type of probability sampling; each unit in the population has an equal probability of being in the sample, and all possible samples of a given size are equally likely to be selected.
- Skip Pattern, Skip Sequence. A sequence of items within a questionnaire that is to be answered only if the respondent gives a particular previous answer.
- Stratified Sampling. Permits the researcher to subdivide the population into mutually exclusive and exhaustive subpopulations, or strata; within these strata, separate samples are selected; results can be combined to form overall population estimates or used to report separate within-stratum estimates.
- Survey Population. See Population.
- Systematic Sampling. Consists of a random start and the selection of every *n*th member of the population; generally produces the same results as simple random sampling.
- Target Population. See Population.
- *True Score.* The underlying true score which is unobservable because there is always some error in measurement; the observed score = true score + error score.
- Universe. See Population.

#### **References on Survey Research**

William G. Cochran, Sampling Techniques (3d ed. 1977).

- Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire (1986).
- Thomas D. Cook & Donald T. Campbell, Quasi-Experimentation: Design and Analysis Issues for Field Settings (1979).
- Shari S. Diamond, *Methods for the Empirical Study of Law, in* Law and the Social Sciences (Leon Lipson & Stanton Wheeler eds., 1986).
- 1 Floyd J. Fowler, Survey Research Methods (2d ed. 1984).
- Robert M. Groves & Robert L. Kahn, Surveys by Telephone: A National Comparison with Personal Interviews (1979).
- Handbook of Survey Research (Peter H. Rossi et al. eds., 1983).
- Leslie Kish, Survey Sampling (1965).
- Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991).
- Questions About Questions: Inquiries into the Cognitive Bases of Surveys (Judith M. Tanur ed., 1992).
- Howard Schuman & Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context (1981).
- Seymour Sudman, Applied Sampling (1976).
- Seymour Sudman & Norman M. Bradburn, Response Effects in Surveys: A Review and Synthesis (1974).
- Telephone Survey Methodology (Robert M. Groves et al. eds., 1988).

## Reference Guide on Forensic DNA Evidence

Judith A. McKenna Joe S. Cecil Pamela Coukos

Judith A. McKenna, J.D., Ph.D., and Joe S. Cecil, J.D., Ph.D., are members of the staff in the Research Division of the Federal Judicial Center, Washington, D.C. Pamela Coukos, J.D., was a legal assistant in the Research Division of the Federal Judicial Center and is currently a law clerk in the United States District Court for the Eastern District of New York.

This page left blank intentionally for proper pagination when printing two-sided

### Contents

- I. Introduction 277
- II. Overview of RFLP Analysis 281
- III. Theory and Technique of RFLP Analysis 285
  - A. Is the Scientific Theory Underlying RFLP Valid? 285
  - B. Are the Laboratory Techniques Used in RFLP Analysis Valid and Reliable? 285
  - C. With Respect to Disputed Issues, What Are the Relevant Scientific Communities? 286
- IV. Sample Quantity and Quality 287
  - A. Did the Crime Sample Contain Enough DNA to Permit Accurate Analysis? 287
  - B. Was the Crime Sample of Sufficient Quality to Permit Accurate Analysis? 288
  - C. How Many Sources of DNA Are Thought to Be Represented in the Crime Sample? 288
- V. Laboratory Performance of the RFLP Analysis Offered as Evidence 291
  - A. Has the Testing Laboratory Demonstrated a Record of Proficiency and Quality Control Sufficient to Permit Confidence That the Tests Were Conducted Properly? 291
    - 1. Does the laboratory maintain appropriate documentation? 291
    - 2. Has the laboratory's procedure been validated? 292
    - 3. Has the laboratory been subjected to appropriate proficiency testing? With what results? 292
  - B. Was the Crime Sample in This Case Handled Properly? 293
    - 1. What opportunities for laboratory mislabeling were present? 293
    - 2. What sources of possible laboratory contamination were present? 293
  - C. Was the RFLP Analysis in This Case Conducted Properly? 294
- VI. Comparison of DNA Profiles 295

- A. Does the DNA Profile from the Crime Sample Appear to Be Consistent with the DNA Profile of the Suspect? 295
- B. What Measurement Standard Was Used to Determine That the Bands Are Similar Enough to Declare a Match? 296
- VII. Estimation of the Probability That the DNA Profiles Match by Coincidence 297
  - A. What Procedure Was Used to Estimate the Probability That the Individual Alleles Match by Coincidence? 298
  - B. How Were the Probability Estimates for Coincidental Matches of Individual Alleles Combined for an Overall Estimate of a Coincidental Match of the Entire DNA Profile? 300
    - 1. Product rule technique 300
    - 2. Modified ceiling principle technique 301
  - C. What Is the Relevant Comparison Population for Estimating the Probability of a Coincidental Allelic Match? 303
    - 1. Is the comparison population consistent with the population of possible sources of the DNA? 303
    - 2. Does the comparison population conform to characteristics that allow the estimation of the joint occurrence of matching alleles by multiplication of the probabilities of the individual alleles? 304

Appendix A: Schematic of Single-Locus Probe RFLP Analysis311Appendix B: Example of Illustrated DNA Patterns on Autorad313

Appendix C: TWGDAM Guidelines 315

Glossary of Terms 323

References on Forensic DNA Evidence 329

## I. Introduction

This reference guide addresses technical issues that arise in considering the admissibility of and weight to be accorded analyses of *forensic samples* of *deoxyribonucleic acid* (DNA). We address only the best-established form of forensic DNA identification analysis, known as *restriction fragment length polymorphism* (RFLP) *analysis.*<sup>1</sup> Scientific knowledge in this area is evolving rapidly, and analytical techniques based on different principles are being developed. Technologies based on the *polymerase chain reaction* (PCR) technique of DNA amplification may become the dominant modes of analysis in the relatively near future. PCR-based analysis techniques are discussed briefly in section IV.A.

In this guide we set forth an analytical framework judges can use when DNA identification evidence is offered by expert witnesses. Therefore we describe only the basic principles of the technique; we do not address the general expert witness qualification matters raised by Federal Rule of Evidence 702 or procedural issues such as discovery.<sup>2</sup> Although we cite both state and federal cases for illustrative purposes, this reference guide is not intended as a guide to the status of DNA evidence in any particular jurisdiction.<sup>3</sup>

DNA analysis is based on well-established principles of the wide genetic variability among humans and the presumed uniqueness of an individual's genetic makeup (identical twins excepted). Laboratory techniques for isolating and observing the DNA of human *chromosomes* have long been used in nonforensic scientific settings. The forensic application of the technique involves comparing a known DNA sample obtained from a suspect with a DNA sample obtained from the crime scene, and often with one obtained from the victim. Such analyses typically are offered to support or refute the claim that a criminal suspect contributed a biological specimen (e.g., semen or blood) collected at a crime

<sup>1.</sup> Both RFLP and PCR may be supplanted by new techniques that rely on direct sequencing of genes, which may avoid some of the chief problems associated with current techniques. *See, e.g.,* Alec J. Jeffreys et al., *Minisatellite Repeat Coding as a Digital Approach to DNA Typing,* 354 Nature 204 (1991).

<sup>2.</sup> See, e.g., Paul C. Giannelli, *Criminal Discovery*, *Scientific Evidence, and DNA*, 44 Vand. L. Rev. 793 (1991). See also Margaret A. Berger, Evidentiary Framework § II, in this manual.

<sup>&</sup>lt;sup>3</sup>. Some states have adopted statutes that expressly provide that such analyses are admissible as evidence in criminal proceedings under certain circumstances. States that have adopted statutes addressing the admissibility of DNA identification evidence include: Indiana (Ind. Code § 35-37-4-13 (1993)); Louisiana (La. Rev. Stat. Ann. §15:441.1 (West 1992 & Supp. 1994)); Maryland (Md. Code Ann., Cts. & Jud. Proc. §10-915 (1989 & Supp. 1993)); Minnesota (Minn. Stat. Ann. §§ 634.25–634.26, 634.30 (West Supp. 1994)); Tennessee (Tenn. Code Ann. § 24-7-117 (Supp. 1993)); and Virginia (Va. Code Ann. §19.2-270.5 (Michie 1990)).

scene.<sup>4</sup> For example, an analyst may testify on the basis of a report that includes the following:

Deoxyribonucleic acid (DNA) profiles for [the specific sites tested] were developed from specimens obtained from the crime scene, from the victim, and from the suspect. Based on these results, the DNA profiles from the crime scene match those of the suspect. The probability of selecting at random from the population an unrelated individual having a DNA profile matching the suspect's is approximately 1 in 200,000 in Blacks, 1 in 200,000 in Whites, and 1 in 100,000 in Hispanics.

An objection to this sort of testimony usually comes before the court when the defense moves to exclude the testimony and the report. Such a motion can be made before or during trial, depending on circumstances and the court's rules regarding in limine motions. Before the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,<sup>5</sup> such disputes often were aired in hearings devoted to determining whether the theory and techniques of DNA identification were generally accepted by the relevant scientific community and so satisfied the *Frye* standard.<sup>6</sup> Since *Daubert*, general acceptance remains an issue but only one of several a court may consider when DNA evidence is offered. Frequently disputed issues include the validity of applying the standard RFLP technique to crime samples, the proper interpretation of test results as showing a match, and the appropriate statistical determination of the probability of a coincidental match.

To address concerns that had been raised regarding the use of DNA identification evidence in forensic contexts, the National Research Council (NRC) of the National Academy of Sciences convened a committee of scientists, forensic scientists working in law enforcement, and legal scholars. The committee's report<sup>7</sup> offers conclusions about the scientific validity and reliability of DNA analysis and makes recommendations concerning its use as evidence in court. The report, which recommends that courts take judicial notice of both the theory underlying DNA analysis and the ability of the most commonly used method of testing to distinguish reliably between different sources of DNA, confirmed the conclusion several courts had reached. However, the committee also recognized actual and potential problems with the use of current DNA identification methods and recommended steps by which the scientific and forensic communities could address them. Throughout this guide, we cite, where appropriate, the

<sup>&</sup>lt;sup>4</sup>. Although the technique has other uses, this is the usual context in which such evidence comes before a federal court. Other considerations may arise where DNA analysis is used to narrow the field of suspects by comparing a crime sample with samples from a blood bank, to establish the commission of a crime where no body is found, or to establish parentage.

<sup>&</sup>lt;sup>5</sup>. 113 S. Ct. 2786 (1993).

<sup>&</sup>lt;sup>6</sup>. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

<sup>&</sup>lt;sup>7</sup>. Committee on DNA Technology in Forensic Science, National Research Council, DNA Technology in Forensic Science (1992) [hereinafter NRC Report].

NRC committee's position on issues still in dispute, but such citation should not be taken as endorsement of the committee's conclusions.

This page left blank intentionally for proper pagination when printing two-sided

## II. Overview of RFLP Analysis<sup>8</sup>

RFLP analysis is based on the observable variability of human genetic characteristics. Human genetic information (one's *genome*) is encoded primarily in chromosomal DNA, which is present in most body cells.<sup>9</sup> Except for sperm and egg cells, each DNA-carrying cell contains 46 chromosomes. Forty-four of these are arranged in homologous pairs of one *autosome* (nonsex chromosome) inherited from the mother and one autosome from the father. The cells also carry two sex chromosomes (an X from the mother and either an X or a Y from the father). Chromosomal DNA sequences vary in length and are made up of four organic bases (*adenine* (A), *cytosine* (C), *thymine* (T), and *guanine* (G)). A pairs only with T; C pairs only with G. These sequences of *base pairs* are arranged in long chains that form the twisted *double helix*, or ladder structure, of DNA. Thus, if the bases on one side of the helix or ladder are represented as CATAGAT, the complementary side would be GTATCTA.

Most DNA-carrying cells in a human contain the same information encoded in the approximately 3.3 billion base pairs per set of chromosomes in each cell.<sup>10</sup> More than 99% of the base pairs in human cells are the same for all individuals, which accounts for the many common traits that make humans an identifiable species. The remaining base pairs (about 3 million) are particular to an individual (identical twins excepted), which accounts for most of the wide variation that makes each person unique.

A *gene* (characteristic DNA sequence) is found at a particular site, or *locus*, on a particular chromosome. For instance, a gene for eye color is found at the same place or locus on the same chromosome in every individual.<sup>11</sup> Normal individuals have two copies of each gene at a given locus—one from the father and one from the mother. A locus on the DNA molecule where all humans have the

<sup>&</sup>lt;sup>8</sup>. More detailed descriptions of these principles may be found in the NRC Report, *supra* note 7, at 27–40; United States v. Jakobetz, 955 F.2d 786, 791–93 (2d Cir.), *cert. denied*, 113 S. Ct. 104 (1992); and in United States v. Yee, 134 F.R.D. 161, 169 (N.D. Ohio 1991), *aff'd sub nom*. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993).

<sup>&</sup>lt;sup>9</sup>. Mature red blood cells do not carry chromosomal DNA. However, white blood cells do carry chromosomal DNA and can be used for DNA analysis.

<sup>&</sup>lt;sup>10</sup>. The only significant universal exception to this fact is that a sex cell (i.e., an egg or sperm) contains only twenty-three chromosomes, which vary from cell to cell (e.g., from sperm to sperm). This accounts for the genetic individuality of offspring.

<sup>&</sup>lt;sup>11</sup>. This consistency of genetic localization has made it possible for geneticists to begin mapping the human genome.

same genetic code is called *monomorphic*. However, genes vary. An individual may receive the genetic code for blue eyes from his or her mother and the genetic code for brown eyes from his or her father. An alternative form of a gene is known as an *allele*.<sup>12</sup> A locus where the allele differs among individuals is called *polymorphic*, and the difference is known as a *polymorphism*.

Although some polymorphisms have been found to govern what makes individuals observably distinct from one another (e.g., eye color), others serve no known function. Among these noncoding DNA regions are some in which certain base pair sequences repeat in tandem many times (e.g., CATCATCAT... paired with GTAGTAGTA...). This is known as a *Variable Number of Tandem Repeats*, or a VNTR. The number of base pairs and the sequence of pairs vary from locus to locus on one chromosome and from chromosome to chromosome. RFLP analysis allows scientists to determine the size of a repetitive sequence. Because the length of these sequences (sometimes called *band size*) of base pairs is highly polymorphic, although not necessarily unique to an individual, comparison of several corresponding sequences of DNA from known (suspect) and unknown (forensic) sources gives information about whether the two samples are from the same source.

Appendix A is a detailed schematic of the steps in RFLP analysis. First, DNA is extracted from an evidence sample collected at the crime scene. Second, it is digested by a *restriction enzyme* that recognizes a particular known sequence called a *restriction site* and cuts the DNA there. The result is many DNA fragments of varying sizes. Third, digested DNA from the crime sample is placed in a well at the end of a lane in an agarose gel, which is a gelatin-like material so-lidified in a slab about five inches thick. Digested DNA from the suspect is placed in another well on the same gel. Typically, control specimens of DNA fragments of known size, and, where appropriate, DNA specimens obtained from a victim, are run on the same gel. Mild electric current applied to the gel slowly separates the fragments in each lane by length, as shorter fragments travel farther than longer, heavier fragments. This procedure is known as *gel electrophoresis*.

Fourth, the resulting array of fragments is transferred for manageability to a sheet of nylon by a process known as *Southern blotting*. Either during or after this transfer, the DNA is *denatured* ("unzipped") by heating, separating the double helix into single strands. The weak bonds that connect the two strands are susceptible to heat and salinity. The double helix can be unzipped or denatured without disrupting the chain on either side. Fifth, a *probe*—usually with a radioactive tag—is applied to the membrane. The probe is a single strand of DNA that *hybridizes* with (binds to) its complementary sequence when it is applied to

<sup>&</sup>lt;sup>12</sup>. When an individual's two alleles for a locus are the same (e.g., two blue-eye alleles), the individual is said to be *homozygous* for that locus. When the alleles are not the same, the individual is said to be *heterozygous* for that locus. See Office of Technology Assessment, U.S. Congress, Genetic Witness: Forensic Uses of DNA Tests 42 (1990) [hereinafter OTA Report].

the samples of denatured DNA. The DNA locus identified by a given probe is found by experimentation, and individual probes often are patented by their developers. Different laboratories may use different probes (i.e., they may test for alleles at different loci). Where different probes are used, test results are not comparable.

Finally, excess *hybridization* solution is washed off, and the nylon membrane is placed between at least two sheets of photographic film. Over time, the radioactive probe material exposes the film where the biological probe has hybridized with the DNA fragments. The result is an *autoradiograph*, or an *autorad*, a visual pattern of bands representing specific DNA fragments. An autorad that shows two bands in a single lane indicates that the source is heterozygous for that locus (i.e., he or she inherited a different allele from each parent). If the autorad shows only one band, the person may be homozygous for that allele (i.e., each parent contributed the same variant of the gene).<sup>13</sup> Together, the two alleles make up the person's *genotype* (genetic code) for the specific locus associated with the probe.

Once an appropriately exposed autorad is obtained, the probe is washed from the membrane, and the process is repeated with different probes that bind to different sequences of DNA. Three to five probes are typically used, the number depending in part on the amount of testable DNA recovered from the crime sample.<sup>14</sup> The result is a set of autorads, each of which shows the results of one probe.<sup>15</sup> In Appendix B there is a copy of an autorad with illustrative DNA patterns. Illustrations of the results of all probes depict an overlay of multiple films or the results from a multilocus probe.

If the two DNA samples are from the same source, and if the laboratory procedures are conducted properly, hybridized DNA fragments of approximately the same length should appear at the same point in the suspect and evidence specimen lanes. If, on visual inspection, the DNA band patterns for the suspect and the evidence sample appear to be aligned on the autorad, this impression is verified by a computerized measurement. If the two bands fall within a specified length, or *match window* (e.g.,  $\pm 2.5\%$  of band length), a match is declared for that probe or allele.<sup>16</sup> For forensic purposes, a match means that the patterns are

<sup>&</sup>lt;sup>13</sup>. However, the appearance of a single band may also indicate problems with the analysis, and the analyst should explore alternative explanations when homozygosity cannot be established (as by testing the DNA of the parents or performing other checks). *See* NRC Report, *supra* note 7, at 58.

<sup>&</sup>lt;sup>14</sup>. Another type of DNA analysis uses a *multilocus probe*, which hybridizes with multiple locations simultaneously, yielding a more complex DNA pattern. In general, this technique is most appropriate in paternity cases. In the United States it is not preferred for analyses in criminal cases. See NRC Report, *supra* note 7, at 40.

<sup>&</sup>lt;sup>15</sup>. As the membrane is compressed between two films for each probe, two autorads per probe actually result. However, one film per probe is checked during the process to see whether the process is complete. As a result, the image on the film removed first may be weaker than the final print and generally is not used.

<sup>&</sup>lt;sup>16</sup>. The match window is intended to accommodate measurement error or differences that may result when the tests are run repeatedly on the same sample.

consistent with the conclusion that the two DNA samples came from the same source.<sup>17</sup> Taken together, the results of the probes form the DNA profile.<sup>18</sup>

Reference Manual on Scientific Evidence

<sup>&</sup>lt;sup>17</sup>. The interpretation of autorads is discussed in more detail *infra* § VI. <sup>18</sup>. Throughout this reference guide, we use the term DNA profile to refer to the autorad pattern yielded by the test probes. If it were feasible to map the entire genome, it is believed that a unique profile would be obtained for each individual.

## III. Theory and Technique of RFLP Analysis

#### A. Is the Scientific Theory Underlying RFLP Valid?

The basic genetic theory underlying DNA profiling (e.g., the fundamental structure of DNA and the observable polymorphism of some areas of DNA) essentially is undisputed, as several courts have noted, and courts may want to consider whether the theory is now a proper subject for judicial notice.<sup>19</sup> The ability to discriminate between human genetic profiles, using enough test sites, is well accepted.<sup>20</sup>

## B. Are the Laboratory Techniques Used in RFLP Analysis Valid and Reliable?

The basic laboratory procedures used in RFLP analysis have been used in nonforensic settings for many years and are generally accepted by molecular biologists.<sup>21</sup> Most courts addressing the issue have concluded that there also appears to be a broad consensus among molecular biologists that properly conducted RFLP analysis can produce reliable information relevant to the identification of a forensic sample of DNA. Courts reaching this conclusion have applied the *Frye* test,<sup>22</sup> the reliability test,<sup>23</sup> a combination of the two,<sup>24</sup> and the *Daubert* 

<sup>&</sup>lt;sup>19</sup>. See, e.g., State v. Davis, 814 S.W.2d 593, 602 (Mo. 1991), cert. denied sub nom. Davis v. Missouri, 112 S. Ct. 911 (1992); People v. Castro, 545 N.Y.S.2d 985, 987 (1989).

<sup>&</sup>lt;sup>20</sup>. The NRC committee has recommended that courts take judicial notice that "[t]he study of DNA polymorphisms can, in principle, provide a reliable method for comparing samples," and that "[e]ach person's DNA is unique (with the exception of identical twins), although the actual discriminatory power of any particular DNA test will depend on the sites of DNA variation examined." NRC Report, *supra* note 7, at 149.

 <sup>&</sup>lt;sup>21</sup>. The Office of Technology Assessment has concluded that "no scientific doubt exists that technologies available today accurately detect genetic differences." OTA Report, *supra* note 12, at 59.
<sup>22</sup>. Smith v. Deppish, 807 P.2d 144, 157–59 (Kan. 1991); State v. Davis, 814 S.W.2d 593, 600–03 (Mo.

 <sup>&</sup>lt;sup>22</sup>. Smith v. Deppish, 807 P.2d 144, 157–59 (Kan. 1991); State v. Davis, 814 S.W.2d 593, 600–03 (Mo. 1991), *cert. denied sub nom.* Davis v. Missouri, 112 S. Ct. 911 (1992); State v. Ford, 392 S.E.2d 781, 783–84 (S.C. 1990); State v. Schwartz, 447 N.W.2d 422, 424–25 (Minn. 1989); State v. Woodall, 385 S.E.2d 253, 259 (W. Va. 1989); Glover v. State, 787 S.W.2d 544, 547–48 (Tex. Ct. App. 1990), *aff d*, 825 S.W.2d 127 (Tex. Crim. App. 1992); People v. Wesley, 633 N.E.2d 451 (N.Y. 1994).
<sup>23</sup>. State v. Brown, 470 N.W.2d 30, 32–33 (Iowa 1991); Caldwell v. State, 393 S.E.2d 436, 441 (Ga. 1990);

<sup>&</sup>lt;sup>23</sup>. State v. Brown, 470 N.W.2d 30, 32–33 (Iowa 1991); Caldwell v. State, 393 S.E.2d 436, 441 (Ga. 1990); Spencer v. Commonwealth, 384 S.E.2d 775, 783 & n.10 (Va. 1989), *cert. denied*, 493 U.S. 1036 (1990); Andrews v. State, 533 So. 2d 841, 846–47 (Fla. Dist. Ct. App. 1988) (applying a "relevancy/reliability" test).

<sup>&</sup>lt;sup>24</sup>. United States v. Jakobetz, 995 F.2d 786, 792-97 (2d Cir.), cert. denied, 113 S. Ct. 104 (1992).

analysis.<sup>25</sup> The forensic applicability of RFLP analysis also has been recommended as a subject for judicial notice.<sup>26</sup>

# C. With Respect to Disputed Issues, What Are the Relevant Scientific Communities?

Most courts have accepted molecular biologists as the relevant scientific community with respect to the laboratory techniques of isolating and probing the DNA. The statistical interpretation of the results has been more properly the province of population geneticists.<sup>27</sup> Individuals from other fields (e.g., genetic epidemiologists and biostatisticians) also may have the requisite background to testify about the analysis. Forensic scientists or laboratory technicians involved in the analysis often do not have a strong background in the relevant scientific discipline but may be knowledgeable about techniques of sample collection and preservation, forensic laboratory standards and procedures, and proficiency tests.

NRC Report, supra note 7, at 149.

<sup>27</sup>. United States v. Yee, 134 F.R.D. 161, 164 (N.D. Ohio 1991), *aff d sub nom*. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993); United States v. Jakobetz, 747 F. Supp. 250, 259–60 (D. Vt. 1990), *aff d*, 955 F.2d 786 (2d Cir.), *cert. denied*, 113 S. Ct. 104 (1992).

<sup>&</sup>lt;sup>25</sup>. United States v. Bonds, 12 F.3d 540, 554–68 (6th Cir. 1993) (assessing DNA identification evidence in light of Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993)).

<sup>&</sup>lt;sup>16</sup>. The NRC committee recommended that courts take judicial notice that

<sup>[</sup>t]he current laboratory procedure for detecting DNA variation (specifically, single-locus probes analyzed on Southern blots without evidence of band shifting) is fundamentally sound, although the validity of any particular implementation of the basic procedure will depend on proper characterization of the reproducibility of the system (e.g., measurement variation) and the inclusion of all necessary scientific controls.
### IV. Sample Quantity and Quality

RFLP analysis requires a suitable sample of DNA. Several factors may affect a sample's suitability for analysis. For each factor claimed to affect a particular analysis, the court may want to have the experts address whether its influence is likely to cause a false positive result (incorrect identification of the suspect as a potential source of the forensic DNA) or merely an inconclusive or uninterpretable result.<sup>28</sup>

### A. Did the Crime Sample Contain Enough DNA to Permit Accurate Analysis?

To be interpretable, the crime sample must contain enough DNA of sufficiently high molecular weight to allow isolation of longer DNA fragments, which are the most susceptible to degradation. Samples of blood, semen, or other DNA sources may be too small to permit analysis. We are aware of no evidence to suggest that small sample size increases the likelihood that an interpretable test will yield a false match for a given probe or allele. However, to the extent that small sample size precludes a full series of tests (three to five probes), it can significantly diminish the power of RFLP analysis to distinguish between DNA samples obtained from different individuals. In addition, the unavailability of additional DNA precludes repeated testing that might verify or refute the initial test.

The amount of testable DNA may be increased by a technique known as the polymerase chain reaction, or PCR amplification. PCR mimics DNA's self-replicating properties to make up to millions of copies of the original DNA sample in only a few hours.<sup>29</sup> Although the term PCR often is used loosely to refer to the

<sup>&</sup>lt;sup>28</sup>. Some factors, including contamination, may cause a false negative result, but courts seldom encounter the problem of false negatives in criminal cases.

<sup>&</sup>lt;sup>29</sup>. For descriptions of the PCR process, see OTA Report, *supra* note 12, at 47–50; George F. Sensabaugh & Cecilia Von Beroldingen, *The Polymerase Chain Reaction: Application to the Analysis of Biological Evidence, in* Forensic DNA Technology 63, 64–67 (Mark A. Farley & James J. Harrington eds., 1991); Lorne T. Kirby, DNA Fingerprinting: An Introduction 75–76 (1990); Russell Higuchi & Edward Blake, *Applications of the Polymerase Chain Reaction in Forensic Science, in* DNA Technology and Forensic Science 265, 266–68 (Jack Ballantyne et al. eds., 1989); Henry A. Erlich et al., *Recent Advances in the Polymerase Chain Reaction,* 252 Science 1643, 1649 (1991); NRC Report, *supra* note 7, at 40–42. PCR requires special attention to issues of quality control and contamination. Courts presented with evidence derived from PCR-based methods may want to consult the NRC Report, *supra* note 7, at 63–70, and the OTA Report, *supra* note 12, at 69–71.

entire process of replicating DNA and testing for the presence of matching alleles, the term properly refers only to the *replication* portion of that process. After amplifying a DNA sample with PCR, technicians must use other methods to determine whether a known and unknown sample match. Standard RFLP analysis can be used in many circumstances, but other techniques often are used, including a process using *sequence-specific oligonucleotide* (SSO) *probes*. Currently, one locus, called *HLA DQ*  $\alpha$ ,<sup>30</sup> is available for this process. The technique is faster than RFLP but less discriminating and therefore somewhat less powerful as an identification tool. A second technique, *amplified fragment length polymorphism* (AMP-FLP), is under development and may be used soon in criminal investigations. It relies on amplification of VNTR loci and usually uses gel electrophoresis. Other detection systems are either too novel or too unreliable to be addressed here.<sup>31</sup>

# B. Was the Crime Sample of Sufficient Quality to Permit Accurate Analysis?

Exposure to heat, moisture, and ultraviolet radiation can degrade the DNA sample. Samples also may have been contaminated by exposure to chemical or bacterial agents that alter DNA, interfere with the enzymes used in the testing process, or otherwise make DNA difficult to analyze. Such exposure is known as *environmental insult*.<sup>32</sup>

Although old samples of DNA may be analyzed successfully, attention must be given to possible sample degradation due to age. Again, we are aware of no evidence to suggest that age-degraded samples are likely to produce false positive results. Courts may want to ask the experts whether the research has addressed the effects of specific types of environmental insult likely in a particular case.

# C. How Many Sources of DNA Are Thought to Be Represented in the Crime Sample?

Often, the expected composition of a crime sample of DNA can be narrowed to a single perpetrator, a single victim, or both. However, a crime sample may be thought to include DNA from multiple sources, as where more than one person

<sup>&</sup>lt;sup>30</sup>. Human Leukocyte Antigens (HLAs) are antigens (foreign bodies that stimulate an immune system response) located on the surface of most cells (excluding red blood cells and sperm cells). HLAs differ among individuals and are associated closely with transplant rejection. HLA DQ  $\alpha$  is a particular class of HLA whose locus has been completely sequenced and thus can be used for forensic typing.

<sup>&</sup>lt;sup>31</sup>. Some of these other techniques include direct sequencing of DNA samples, which promises theoretically absolute identification (and raises all the issues and concerns associated with capturing an individual's genetic code), but the process is not yet sufficiently developed for use in forensic applications. *See* NRC Report, *supra* note 7, at 43–44.

<sup>&</sup>lt;sup>32</sup>. See, e.g., Dwight E. Adams et al., Deoxyribonucleic Acid (DNA) Analysis by Restriction Fragment Length Polymorphisms of Blood and Other Body Fluid Stains Subjected to Contamination and Environmental Insults, 36 J. Forensic Sci. 1284 (1991).

is thought to have contributed to the crime sample of blood or semen. Male and female DNA extracted from such a sample may be distinguished, as can samesex DNA where the alternative source is known and available for testing (e.g., a rape victim's husband). The presence of multiple, same-sex samples from unknown sources raises additional complications. Mixed samples can be difficult to interpret, although the intensity of different bands can offer clues. Courts are most likely to encounter the possibility of mixed samples when analysts report extra or anomalous bands; they should be alert to the possible explanations for such bands and inquire into the laboratory's procedures for testing those explanations.<sup>33</sup>

<sup>33</sup>. See NRC Report, supra note 7, at 58–59.

Forensic DNA Evidence

This page left blank intentionally for proper pagination when printing two-sided

## V. Laboratory Performance of the RFLP Analysis Offered as Evidence

Adherence to proper test procedures is critical to accurate comparison of DNA samples. Before the NRC issued its report, guidelines for laboratory procedures, quality control, and proficiency testing were promulgated jointly by the Technical Working Group on DNA Analysis Methods (TWGDAM) and the California Association of Criminalists Ad Hoc Committee on DNA Quality Assurance.<sup>34</sup> These guidelines (TWGDAM Guidelines) have been used by courts to assess the quality of individual test results and testing laboratories<sup>35</sup> and enjoy wide, albeit not universal, acceptance as appropriate laboratory procedures for RFLP analysis.<sup>36</sup> This section incorporates some of the TWGDAM Guidelines and can be used to identify disputed issues regarding the testing laboratory and the individual RFLP analysis. The report of the NRC also contains a detailed discussion of quality-assurance issues and strongly favors adoption of quality control and proficiency-testing standards.<sup>37</sup>

### A. Has the Testing Laboratory Demonstrated a Record of Proficiency and Quality Control Sufficient to Permit Confidence That the Tests Were Conducted Properly?

1. Does the laboratory maintain appropriate documentation?

The TWGDAM Guidelines state that the DNA laboratory "must maintain documentation on all significant aspects of the DNA analysis procedure, as well as any related documents or laboratory records that are pertinent to the analysis or

<sup>&</sup>lt;sup>34</sup>. TWGDAM is a practitioners' group, largely sponsored by the Federal Bureau of Investigation (FBI) and comprising representatives of the FBI, the National Institute of Standards and Technology, the Royal Canadian Mounted Police, and state forensic laboratories.

<sup>&</sup>lt;sup>35</sup>. State v. Schwartz, 447 N.W.2d 422, 426 (Minn. 1989). In particular, the protocol used by the FBI has been held to be sufficiently reliable. United States v. Jakobetz, 955 F.2d 786, 799–800 (2d Cir.), *cert. denied*, 113 S. Ct. 104 (1992); Prater v. State, 820 S.W.2d 429, 436 (Ark. 1991).

<sup>&</sup>lt;sup>36</sup>. United States v. Yee, 134 F.R.D. 161, 202–06 (N.D. Ohio 1991), *aff d sub nom*. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993). The NRC committee described the TWGDAM Guidelines as an "excellent starting point for a quality-assurance program," but recommended that certain technical recommendations be added. NRC Report, *supra* note 7, at 16, 99.

<sup>&</sup>lt;sup>37</sup>. See NRC Report, supra note 7, at 97–109.

interpretation of results, so as to create a traceable audit trail."38 The guidelines list seventeen areas that must be covered by the documentation.<sup>39</sup>

#### 2. Has the laboratory's procedure been validated?

The TWGDAM Guidelines stress the importance of validation "to acquire the necessary information to assess the ability of a procedure to reliably obtain a desired result, determine the conditions under which such results can be obtained and determine the limitations of the procedure."<sup>40</sup> Each locus or probe to be used in the analysis should be validated separately.<sup>41</sup> The court may want to ask whether the laboratory's procedures have been validated in accordance with the minimum standards set forth by section 4 of the TWGDAM Guidelines.<sup>42</sup>

3. Has the laboratory been subjected to appropriate proficiency testing? With what results?

Studies of the quality and accuracy of laboratory analysis in many forensic areas have raised profound concerns about the accuracy of evidence presented in criminal cases.43 Section 9 of the TWGDAM Guidelines recommends that DNA laboratories participate in appropriately designed proficiency-testing programs, preferably programs conducted by outside institutions.<sup>44</sup> The court may want to inquire about the relevant laboratory's participation in external or internal proficiency-testing programs, the results of such testing, and the significance of the results. Some commentators have suggested that, at least when an adequate history of repeated proficiency testing becomes available, expert testimony should report not only the likelihood of a coincidental match according to applicable *population genetics* but also the applicable laboratory error rates.<sup>4</sup>

Blind Proficiency Testing. Ideally, blind proficiency test specimens should be presented to the testing laboratory through a second agency. These samples should appear to the examiner/analyst as routine evidence. The blind proficiency test serves to evaluate all aspects of the laboratory examination procedure, including evidence handling, examination/testing and reporting. It is highly desirable that the DNA laboratory participate in a blind proficiency test program, and every effort should be made to implement such a program.

<sup>45</sup>. Arguments over what information juries should be given concerning laboratory error focus in part on considerations of how to give jurors the most diagnostic information possible in a particular case and in part on policy considerations of how to increase the diagnosticity of DNA identification evidence overall by increasing incentives for laboratories to improve their procedures and submit to large-scale proficiency testing. See, e.g., Michael J. Saks & Jonathan J. Koehler, What DNA "Fingerprinting" Can Teach the Law About the Rest of

<sup>8.</sup> Technical Working Group on DNA Analysis Methods (TWGDAM) & California Ass'n of Criminalists Ad Hoc Comm. on DNA Quality Assurance, Guidelines for a Quality Assurance Program for DNA Analysis, 18 Crime Laboratory Dig. 44, 52 (1991) [hereinafter TWGDAM Guidelines] (see infra Appendix C).

 <sup>&</sup>lt;sup>39</sup>. Id. §§ 3.1–3.17, at 52 (see infra Appendix C).
<sup>40</sup>. Id. § 4.1.1, at 53 (see infra Appendix C).

<sup>&</sup>lt;sup>41</sup>. *Id.* §§ 4.1.1–4.1.3, at 53 (*see infra* Appendix C).

<sup>&</sup>lt;sup>42</sup>. Id. § 4, at 53–56 (see infra Appendix C).

<sup>&</sup>lt;sup>43</sup>. See Edward J. Imwinkelried, The Debate in the DNA Cases Over the Foundation for the Admission of Scientific Evidence: The Importance of Human Error as a Cause of Forensic Misanalysis, 69 Wash. U.L.Q. 19, 25-27; and studies cited therein reporting error rates in forensic testing.

<sup>44.</sup> TWGDAM Guidelines, supra note 38, § 9.2, states:

Some also argue that the laboratory error rate should be combined with the probability of a coincidental match derived from the population analysis to yield an overall probability that the analyst would report a match when there was no true match.<sup>46</sup> This approach would incorporate the *Daubert* factor of laboratory error rate in such a way that it would go to weight, not admissibility. There is no reported federal decision in which a court has imposed this constraint on the expert's testimony.

In evaluating proficiency-testing results, a court may want to attach more weight to blind tests than to open tests and may inquire as to the type of samples that were actually tested (e.g., pristine samples or samples subjected to environmental insults typical of forensic casework). Anonymous testing may give some insights into whether the procedures are generally reliable, but nonanonymous testing provides stronger evidence of the quality of a laboratory's work. Several earlier studies of forensic laboratory work have found different error rates; differences were accounted for in part by whether false negatives (incorrect exclusions) and inconclusive results were counted as errors.<sup>47</sup> Because one purpose of proficiency testing is to help laboratories identify problem areas and correct them, it is important to focus on tests of a laboratory's current procedures where possible.

### B. Was the Crime Sample in This Case Handled Properly?

When the normal chain-of-custody evidence is adduced, the court should allow inquiry into potential sources of error in sample handling by laboratory personnel.

1. What opportunities for laboratory mislabeling were present?

It is especially important that the laboratory follow procedures for ensuring that the known sample is not mislabeled as, or mixed with, the crime sample. These errors can lead to false positive results that incriminate the wrong suspect.

2. What sources of possible laboratory contamination were present?

Potential contamination of the sample by reagents and other substances found in the laboratory is also of concern, although such contamination is less likely than mislabeling to create a false positive result. Contamination is of particular concern when PCR amplification is used, because even a small fragment of DNA from a foreign source may be amplified many times.

Forensic DNA Evidence

Forensic Science, 13 Cardozo L. Rev. 361, 368–70 (1991); and Jonathan J. Koehler, DNA Matches and Statistics: Important Questions, Surprising Answers, 76 Judicature 222, 228–29 (1993). For a discussion of the competing considerations, see Richard Lempert, Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes, 13 Cardozo L. Rev. 303, 323–28 (1991).

<sup>&</sup>lt;sup>46</sup>. For examples of ways such a combination might be derived, see Saks & Koehler, *supra* note 45, at 368\_70.

<sup>&</sup>lt;sup>47</sup>. See OTA Report, *supra* note 12, at 79.

### C. Was the RFLP Analysis in This Case Conducted Properly?

Minor departures from the TWGDAM Guidelines or other accepted standards are not necessarily fatal to the validity of the DNA evidence, and some courts have treated such departures as bearing on the weight of the evidence, not its admissibility.<sup>48</sup> Few courts have excluded DNA evidence because of a laboratory's failure to meet accepted standards, but a persuasive argument can be made that significant departures from accepted standards render the evidence unreliable and, hence, inadmissible. In any event, deviations from the laboratory's written procedures should be explored, and the court may want to have experts from both sides address the nature of errors that could be caused by deviations from the standards.<sup>49</sup>

Reference Manual on Scientific Evidence

<sup>&</sup>lt;sup>48</sup>. United States v. Yee, 134 F.R.D. 161, 206–07 (N.D. Ohio 1991), affd sub nom. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993).

<sup>&</sup>lt;sup>49</sup>. See Thomas M. Fleming, Annotation, Admissibility of DNA Identification Evidence, 84 A.L.R.4th 313, 342–46 (1991 & Supp. 1993); State v. Schwartz, 447 N.W.2d 422, 426 (Minn. 1989). See also State v. Jobe, 486 N.W.2d 407, 420 (Minn. 1992) (Frye hearing is still required on whether the testing laboratory was in compliance with appropriate standards and controls, but not on the basic RFLP testing procedures themselves). The split of authority over whether compliance with proper test procedures is a foundational matter that should govern the admissibility of laboratory evidence, and over which party should bear the burden of showing compliance or noncompliance, is discussed in Inwinkelried, *supra* note 43.

### VI. Comparison of DNA Profiles

The first step in interpreting the evidence is to determine if the DNA obtained from the crime sample is consistent with the DNA profile of the suspect or other known source. The RFLP procedure compares the length of the DNA fragments in the known sample with the length of the DNA fragments in the crime sample. Techniques used to declare profile matches vary in their details but typically use a two-step process of visual inspection and objective measurement of DNA bands that appear to present the same profile.

### A. Does the DNA Profile from the Crime Sample Appear to Be Consistent with the DNA Profile of the Suspect?

A pair of lanes represents the results of one genetic probe on the suspect and crime samples, and each pair should be evaluated independently of every other pair. Distinct bands appearing, on visual inspection, to be at the same position for the two samples suggest that both sources of DNA have fragments of the same length (i.e., the same alleles). Only such a pattern will implicate the suspect as the source of the crime sample. If the bands do not appear to match, the analyst must determine if the test excludes the suspect as the source of the crime sample or if the test is inconclusive because of some shortcoming of the samples or the laboratory procedure.<sup>50</sup>

When bands of different specimens from the same source do not appear, on visual inspection, to be in the same position in the lane, an exclusion usually is declared. Occasionally, the two profiles obtained will appear to be highly similar, even though the bands do not appear in the same lane position. When the bands are consistently offset or shifted relative to one another, some analysts may declare a match, attributing the differences in position to a phenomenon known

<sup>&</sup>lt;sup>50</sup>. See, e.g., State v. Woodall, 385 S.E.2d 253, 260 (W. Va. 1989) (DNA test inconclusive due to poor quality of forensic sample). The distinction is most likely to become relevant to federal courts in the context of post-conviction efforts to reopen cases in which DNA evidence was not offered by the prosecution at the original trial. The FBI has reported that approximately 35% of the interpretable samples it has tested yielded conclusive "no match" decisions that excluded primary suspects. NRC Report, *supra* note 7, at 156. See also State v. Hammond, 604 A.2d 793, 800 (Conn. 1992) (FBI forensic analyst testified that DNA testing, which confirmed the results of blood type testing, showed defendant accused of rape could not have been the source of the semen stain; the court rejected the state's argument that these tests were unreliable due to "inherent flaws in the testing procedures themselves, or from contamination of the sample").

as *band-shift*.<sup>51</sup> Band shifting may occur with different samples from the same source because of differences in DNA concentration or other conditions within a specimen (e.g., salt concentration). Arguably, a match could be declared with offset but similar patterns when the DNA band sizes are consistent relative to internal lane markers of known size, which are in the lanes with the suspect and crime samples.<sup>52</sup> Declaring a match on the assumption that patterns would have matched in the absence of band-shift is controversial; the NRC committee recommends that samples that show apparent band shifting be declared "inconclusive" until the accuracy and reliability of proposed corrections (e.g., *monomorphic probes*) have been demonstrated through research.<sup>53</sup> Often, but not always, RFLP analyses can be repeated to avoid band shifting.

# B. What Measurement Standard Was Used to Determine That the Bands Are Similar Enough to Declare a Match?

If the bands appear to be identical, the forensic analyst must measure the length of the bands by comparing their positions to standardized size markers included on the autorad. Repeated measurements of the same sample may differ slightly due to variation in laboratory materials and procedures, even if the suspect is the source of the crime sample. The standardized size markers are of known length (i.e., a known number of base pairs in a sequence), and the size of the sample fragment can be estimated in relation to these markers. A computer digitizes the images to be compared and determines whether they fall within a specified window of measurement variation. A valid analysis must apply an objective standard for declaring matching band patterns based on the variability observed when known crime samples are tested repeatedly.<sup>54</sup>

<sup>&</sup>lt;sup>51</sup>. Band-shift is discussed in the NRC Report, *supra* note 7, at 60–61, 140–41.

<sup>&</sup>lt;sup>52</sup>. See, e.g., Caldwell v. State, 393 S.E.2d 436, 442–43 (Ga. 1990) (allowing testimony that DNA from a forensic sample matched suspect after band shifting was adjusted by measuring the position of the bands relative to internal lane markers).

<sup>&</sup>lt;sup>37</sup>. NRC Report, *supra* note 7, at 61. See also People v. Keene, 591 N.Y.S.2d 733, 740 (N.Y. Sup. Ct. 1992) ("While the DNA principle and RFLP analysis are generally accepted in the scientific community, this court cannot find that the practice of using monomorphic probes to correct for band shift is a generally accepted test among molecular geneticists.").

<sup>&</sup>lt;sup>54</sup>. For example, using a match window of  $\pm 2.5\%$  of the observed length of the fragment to estimate the true length of the fragment, if the measurement instrument reports that the length of a fragment is 2000 base pairs (bp), the true length is considered to fall within the window defined by 1950–2050 bp. *See also* NRC Report, *supra* note 7, at 72 (recommending a "precise and objective matching rule for declaring whether two samples match").

# VII. Estimation of the Probability That the DNA Profiles Match by Coincidence

If a profile match is declared, it means only that the DNA profile of the suspect is consistent with that of the source of the crime sample. The crime sample may be from the suspect or from someone else whose profile, using the particular probes involved, happens to match that of the suspect. Expert testimony concerning the frequency with which the observed alleles are found in the appropriate comparison population is necessary for the finder of fact to make an informed assessment of the incriminating value of this match.<sup>55</sup>

The frequency with which an individual allele occurs in the comparison population is taken to be the probability of a coincidental match on that allele. These individual probabilities of a coincidental match are combined into an estimate of the probability of a coincidental match on the entire profile. This estimate is interpreted as the probability that a person selected at random from a comparison population would have a DNA profile that matches that of the crime sample. The probability estimate typically provided by a forensic expert cannot be interpreted strictly as the probability that an examiner will declare a match when the samples are actually from different sources. That probability is affected by other factors, the most important of which is the chance of laboratory error.

<sup>&</sup>lt;sup>55</sup>. Statistical testimony concerning the likelihood of a DNA profile matching by coincidence is necessary to assess the probative value of the matching profile. NRC Report, supra note 7, at 74 ("To say that two patterns match, without providing any scientifically valid estimate (or, at least, an upper bound) of the frequency with which such matches might occur by chance is meaningless."). See also People v. Barney, 8 Cal. App. 4th 798, 817 (Cal. Ct. App. 1992) ("The statistical calculation step is the pivotal element of DNA analysis, for the evidence means nothing without a determination of the statistical significance of a match of DNA patterns."); D. H. Kaye, The Forensic Debut of the National Research Council's DNA Report: Population Structure, Ceiling Frequencies and the Need for Numbers, 34 Jurimetrics J. 369, 381 (1994) ("As a legal matter, a completely unexplained statement of a 'match' should be inadmissible because it is too cryptic to be weighed fairly by the jury, ... [H]ow to present to a jury valid scientific evidence of a match is a legal rather than a scientific issue falling far outside the domain of the general acceptance test and the fields of statistics and population genetics. Thus, it would not be 'meaningless' to inform the jury that two samples match and that this match makes it more probable, in an amount that is not precisely known, that the DNA in the samples comes from the same person."). But see United States v. Martinez, 3 F.3d 1191, 1199 (8th Cir. 1993) (ruling that admitting evidence of a DNA profile match without evidence concerning the statistical probability of a coincidental match was not reversible error where defendant stipulated that statistical evidence was not required), cert. denied, 114 S. Ct. 734 (1994).

Differences in scientific opinion arise with respect to two main issues: (1) the appropriate method for computing the estimated probability of a coincidental match of a DNA profile; and (2) the selection of an appropriate comparison population. These issues are addressed below.

# A. What Procedure Was Used to Estimate the Probability That the Individual Alleles Match by Coincidence?

Ascertaining an allele's frequency in a given population is essentially an empirical exercise. A sample of individuals is drawn from the designated population, their DNA is examined with genetic probes used in forensic analysis, and a table of frequencies is developed. For example, the FBI has constructed frequency tables using a *fixed-bin method*, in which standardized size markers are used to define boundaries of bins into which are sorted the fragment sizes observed in a sample population. Table 1, for example, shows a distribution of 858 alleles observed in a Caucasian sample, using the probe known as TBQ7, for the locus D10S28. The probability of a band of a specific length occurring by chance is computed by dividing the number of bands falling in the assigned bin by the total number of alleles observed in the sample of persons tested for that probe.<sup>56</sup> Thus, in Table 1, fragment lengths falling in Bin 7 are more common (75/858, or 0.087) than fragment lengths falling in Bin 20 (12/858, or 0.014). Similar tables have been and continue to be constructed using different probes in different populations.<sup>57</sup> The size of a matching DNA band is cross-referenced in the table of frequencies generated by the examination of the appropriate comparison sample.

<sup>&</sup>lt;sup>56</sup>. If a bin contains fewer than five bands, the FBI merges that bin into an adjacent bin of higher frequency. This merging of bins is believed to yield a more conservative estimate of the probability of a random match. United States v. Yee, 134 F.R.D. 161, 172 (N.D. Ohio 1991), *aff'd sub nom*. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993). Ranajit Chakraborty & Kenneth K. Kidd, *The Utility of DNA Typing in Forensic Work*, 254 Science 1735, 1738 (1991). An alternative approach used by some testing laboratories is the *floating bin method*, in which the frequency of an allele in a data set is calculated by counting the number of alleles falling into a bin centered on the allele of interest with a width specified by the matching rule.

<sup>&</sup>lt;sup>57</sup>. Note, however, that courts have excluded DNA evidence when a laboratory's standards for declaring a match in individual samples have differed from the standards used in creating the allelic frequency tables. Under this analysis, if a  $\pm 2.5\%$  match window is used when the suspect's DNA is measured, the same standard should have been used when the alleles observed in the reference sample were sorted into bins. State v. Pennell, 584 A.2d 513, 521–22 (Del. Super. Ct. 1989) (court excluded evidence of population frequency because of different matching criteria for comparing prints with the database and for determining the population frequencies); People v. Castro, 545 N.Y.S.2d 985, 998 (1989) (DNA identification evidence declared inadmissible because, among other reasons, different criteria were used in comparing the sample prints with the population database).

Bin	Range (bp)	Count	Fraction
1	0–963	13	0.015
2	964–1077	44	0.051
3	1078–1196	38	0.044
4	1197–1352	15	0.017
5	1353-1507	34	0.040
6	1508–1637	67	0.078
7	1638–1788	75	0.087
8	1789–1924	71	0.083
9	1925–2088	40	0.047
10	2089–2351	51	0.059
11	2352-2522	16	0.019
12	2523–2692	14	0.016
13	2693–2862	36	0.042
14	2863-3033	42	0.049
15	3034–3329	41	0.048
16	3330-3674	56	0.065
17	3675-3979	39	0.045
18	3980-4323	62	0.072
19	4324–4821	58	0.068
20	4822-5219	12	0.014
21	5220-5685	6	0.007
22	5686-6368	23	0.027
23	6369–	5	0.006
Total		858	0.999

Table 1 Rebinned Caucasian Population Data for TBQ7 for the Locus D10S28

Source: Bruce Budowle et al., A Preliminary Report on Binned General Population Data on Six VNTR Loci in Caucasians, Blacks and Hispanics from the United States, 18 Crime Laboratory Dig. 9 tbl. 5 at 14 (1991).

*Note:* This table is presented for illustrative purposes; the FBI does not currently use the D10S28 locus in its analyses.

Assumptions of classical population genetics are used to estimate the probability that a person chosen at random from the specified population would exhibit the same genotype as the suspect's.<sup>58</sup> The greater the probability of such a coincidental match, the lower the incriminating value of the evidence.

### B. How Were the Probability Estimates for Coincidental Matches of Individual Alleles Combined for an Overall Estimate of a Coincidental Match of the Entire DNA Profile?

One of the most difficult and contentious issues in forensic use of DNA evidence is how to estimate the probability that two DNA profiles match by chance. This issue has become especially difficult in federal courts since the Supreme Court's decision in *Daubert*.<sup>59</sup> Because legal standards and scientific opinion are changing, judges are likely to benefit from information obtained from counsel on recent developments concerning estimation of the probability of a coincidental match.

What follows is a description of two techniques for estimating the probability of a coincidental match: the *product rule technique* and the *modified ceiling principle technique*, which was recommended by the NRC committee. The probability estimates can vary widely depending on the technique used.

### 1. Product rule technique

The product rule technique offers the most straightforward method of computing the probability of a matching DNA profile. To compute the probability of a random occurrence of a specific pattern of alleles in a DNA profile, the analyst multiplies the separate estimated probabilities of a random occurrence of each allele in the comparison population. When these individual probabilities are multiplied, the estimated probability of a distinctive pattern occurring at random may be less than one in several hundred thousand.<sup>60</sup>

The probability estimate resulting from the multiplication assumes that the individual alleles identified by genetic probes are *independent* of each other.<sup>61</sup> If the probabilities of the individual alleles are not independent (i.e., if certain alleles are likely to occur together in a person), multiplying the individual allele frequencies may underestimate or overestimate the true probability of matching alleles in the chosen population and thereby misstate the incriminating value of the evidence. Critics of the product rule technique contend that in some ethnic

<sup>&</sup>lt;sup>8</sup>. The probability of a genotype for two distinct alleles at a single locus is  $2p_1p_2$ , where  $p_1$  and  $p_2$  are the relative frequencies of the alleles at a locus. This computation assumes that  $p_1$  and  $p_2$  are independent, an issue discussed *infra* note 61. Using the example above, the probability of a coincidental match of a genotype represented by alleles falling in Bin 20 and Bin 7 would be 2(0.014)(0.087) or 0.0024, or 24 out of 10,000, or 1 out of 417. When only a single band appears in a lane, most forensic laboratories will estimate the probability of the genotype as  $2p_1$  (rather than  $p_1^-$ ) to leave open the possibility of a second undetected allele. *See* NRC Report, *supra* note 7, at 78.

<sup>&</sup>lt;sup>1</sup><sup>59</sup>. 113 S. Ct. 2786 (1993).

<sup>&</sup>lt;sup>60</sup>. For an example of such an analysis, see OTA Report, *supra* note 12, at 67.

<sup>&</sup>lt;sup>61</sup>. Two alleles will be independent if the occurrence of one allele is unrelated to the occurrence of another, much as tossing a pair of honest dice will yield two numbers that are independent.

subpopulations the alleles identified by commonly used genetic probes are not independent and that using a broad-based comparison population is therefore inappropriate. They prefer estimation techniques that do not require analysts to assume the *independence* of individual alleles in large comparison populations.<sup>62</sup>

#### 2. Modified ceiling principle technique The method recommended by the

NRC committee involves conservative interpretations of existing population data.<sup>63</sup> As a preliminary test, the laboratory should examine its population database to determine if it contains a sample that matches the profile of the multiple alleles of the crime sample. If no match is found across multiple genetic probes, the expert reports that "the DNA pattern was compared to a database containing N individuals from the population and no match was observed."<sup>64</sup> This is an extremely conservative approach that yields probative information but does not give the fact finder any information concerning the probability of finding a matching profile by chance in the larger population.

The NRC committee has proposed a modified ceiling principle technique, which takes advantage of the computation techniques of population genetics but which includes adjustments that the committee believes make it an "appropriately conservative" approach. Although it noted that recent empirical studies have detected no evidence of a departure from independence within or across commonly used genetic probes,<sup>65</sup> the NRC committee chose "to assume for the sake of discussion that *population substructure* may exist and provide a method for estimating population frequencies in a manner that adequately accounts for it."<sup>66</sup>

<sup>&</sup>lt;sup>62</sup>. This issue is discussed in greater detail *infra* § VII.C.2.

<sup>&</sup>lt;sup>63</sup>. The procedure described herein is the NRC committee's recommended interim solution for the computation of conservative estimates of a coincidental match of a specific DNA profile. The NRC report also calls for research that will assess the existence of *population substructures* by examining the frequency of alleles used for forensic identification in 100-person samples from fifteen to twenty genetically homogeneous ethnic populations in various regions of the world. This information would be combined in a reported frequency that represents the maximum possible likelihood of a coincidental match for any possible heritage, making it unnecessary to specify a comparison group based on an ethnic subpopulation. This technique is discussed in detail in the NRC Report, *supra* note 7, at 80–85. We present the modified ceiling principle interim technique because the more refined technique must await the findings of the proposed research on ethnic subpopulations.

<sup>&</sup>lt;sup>64</sup>. NRC Report, *supra* note 7, at 91. The simplicity of this counting technique is offset by its strong conservative bias; the estimate of the likelihood of a match is bounded by the number of persons in the data set, which rarely exceeds a few hundred. Furthermore, the counting technique fails to take advantage of information from population genetics regarding the rarity of a match across multiple alleles. *See* Neil J. Risch & B. Devlin, *On the Probability of Matching DNA Fingerprints*, 255 Science 717, 720 (1992) (concluding that the counting rule is "unnecessarily conservative"). For these reasons, the NRC committee found the counting technique, when used to the exclusion of other methods, to be too conservative for use in estimating the frequency of particular DNA patterns. NRC Report, *supra* note 7, at 76.

<sup>&</sup>lt;sup>65</sup>. See also NRC Report, *supra* note 7, at 80 ("Recent empirical studies concerning VNTR loci detected no deviation from independence within or across loci." (footnotes & citation omitted)); Richard Lempert, DNA, Science and the Law: Two Cheers for the Ceiling Principle, 34 Jurimetrics J. 41, 45–46 ("most empirical studies fail to find any forensically significant departures from Hardy-Weinberg or linkage equilibria within general populations" (footnote citing studies omitted)).

<sup>&</sup>lt;sup>66</sup>. NRC Report, *supra* note 7, at 80. The NRC committee noted four considerations supporting this policy choice:

The NRC committee recommends that two adjustments be made to population frequencies derived from existing data to permit conservative estimates of the likelihood of a coincidental matching profile. First, the 95% upper confidence limit for the estimated allele frequency is computed for each of the existing population samples (Black, White, Native American, etc.). This upper bound of the confidence interval is intended to accommodate the uncertainties in current population sampling.<sup>67</sup> Second, the largest of these upper confidence limit estimates, or 10%, whichever is greater, is used to compute the joint probability of a coincidental match on the DNA profile of the crime sample. A lower bound of 10% is intended to address concerns that current population data sets may be substructured in unknown ways that would yield misleading estimates of a coincidental matching profile for members of subpopulations. The resulting probabilities then are multiplied as in the product rule computations.<sup>68</sup> The upper confidence limit from among the existing samples, or a 10% lower bound, is used to provide a probability estimate that errs, if at all, in a conservative direction (i.e., that is more favorable to a suspect).<sup>69</sup>

<sup>(1)</sup> It is possible to provide conservative estimates of population frequency, without giving up the inherent power of DNA typing.

<sup>(2)</sup> It is appropriate to prefer somewhat conservative numbers for forensic DNA typing, especially because the statistical power lost in this way can often be recovered through typing of additional loci, where required.

<sup>(3)</sup> It is important to have a general approach that is applicable to any loci used for forensic typing. Recent empirical studies pertain only to the population genetics of the VNTR loci in current use. However, we expect forensic DNA typing to undergo much change over the next decade—including the introduction of different types of DNA polymorphisms, some of which might have different properties from the standpoint of population genetics.

<sup>(4)</sup> It is desirable to provide a method for calculating population frequencies that is independent of the ethnic group of the subject. Id.

The propriety of such policy considerations in shaping the recommendations of the NRC committee has been questioned, as have been some of their scientific underpinnings. *See* Lempert, *supra* note 65 (acknowledging validity of some criticisms of the NRC recommendations but concluding that using the ceiling principle is preferable to the way random match probabilities were presented before the NRC report); William C. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the "DNA War,*" 84 J. Crim. L. & Criminology 22, 80 (1993) ("A major theme of critics from all perspectives is that the 'ceiling principle' is not a principle of science. It is an arbitrary policy statement, and can be accepted or rejected only as such." (footnote omitted)); David H. Kaye, *DNA Evidence: Probability, Population Genetics, and the Courts,* 7 Harv. J.L. & Tech. 101, 147 n.189 (1993) ("Plainly, the NRC panel's desire for a single method of calculating an upper bound on genotype frequencies in any likely population or subpopulation is not a pronouncement about science, but a mere preference for one jurisprudential policy over another.").

<sup>&</sup>lt;sup>67</sup>. NRC Report, *supra* note 7, at 92.

<sup>&</sup>lt;sup>68</sup>. This approach is discussed in greater detail in the NRC Report, *supra* note 7, at 91–92.

<sup>&</sup>lt;sup>69</sup>. NRC Report, *supra* note 7, at 82–85.

C. What Is the Relevant Comparison Population for Estimating the Probability of a Coincidental Allelic Match?

Disputes over the appropriate comparison population have focused on cases in which the product rule technique has been used.<sup>70</sup> The extent to which the product rule technique may underestimate the probability of a coincidental match has been hotly disputed by population geneticists and other scholars.<sup>71</sup> The dispute centers on disagreements over the adequacy of commonly used comparison populations and the role of racial and ethnic subpopulations in probability estimation. Specifying the appropriate comparison population may be of considerable importance, as the estimates of a coincidental match can vary greatly depending on the population selected. For example, the prevalence of certain alleles may vary greatly across races-some alleles are common in Black populations and infrequent in White populations, and vice versa. If a DNA profile for a Black suspect is compared with frequency estimates based on a White population, the estimated likelihood of a chance match may be in error by some unknown amount. Similarly, if a DNA profile of a member of a subpopulation with a distinct frequency distribution of alleles is compared with frequency estimates based on an inappropriate larger population, an error of unknown magnitude may result. Concern over accuracy of estimates of a coincidental match has focused attention on the assumptions used in selecting a comparison population and the scientific validity of the methods used to estimate the probability of a coincidental match.

1. Is the comparison population consistent with the population of possible sources of the DNA?

If the modified ceiling principle technique is not used, an appropriate comparison population must be designated. Such a designation is guided by the characteristics of the population of individuals who might have been the source of the sample.<sup>72</sup> For example, if a rape victim saw her assailant and described him as White, and there is no more information to implicate a member of a specific subpopulation or group, the comparison population should be those who appear

<sup>&</sup>lt;sup>70</sup>. Although the modified ceiling principle technique was intended to eliminate the necessity of specifying a comparison group, a dispute has emerged over which databases are to be considered when estimating the ceiling frequency for alleles (i.e., all databases or only those for major population groups). Thompson, *supra* note 66, at 80–81. This issue involves differing interpretations of the NRC committee's intended definition of the modified ceiling principle rather than a dispute over issues of science.

<sup>&</sup>lt;sup>71</sup>. See, e.g., R. C. Lewontin & Daniel L. Hartl, *Population Genetics in Forensic DNA Typing*, 254 Science 1745 (1991); Chakraborty & Kidd, *supra* note 56; Eric S. Lander, *Invited Editorial: Research on DNA Typing Catching Up with Courtroom Application*, 48 Am. J. Hum. Genet. 819, 821 (1991); Risch & Devlin, *supra* note 64. This debate is summarized in the NRC Report, *supra* note 7, at 79–80; Thompson, *supra* note 66.

<sup>&</sup>lt;sup>72</sup>. NRC Report, *supra* note 7, at 85 ("[F]requencies should properly be based on the population of possible perpetrators, rather than on the population to which a particular suspect belongs." (footnotes omitted)). See *also* Lempert, *supra* note 45, at 310; and Kaye, *supra* note 66, at 137–38 & n.155.

to be White and who were in a position to commit the assault.<sup>73</sup> Comparisons based on members of populations or subpopulations who appear to be non-White therefore would be inappropriate. Similarly, where there is no information indicating the race or ethnicity of the perpetrator, the comparison population should be designated by the characteristics of those in a position to commit the assault. The race or ethnicity of the suspect is irrelevant.<sup>74</sup>

In other cases, however, the pool of alternative suspects may be limited to members of a distinct isolated community or a specific ethnic subgroup. This circumstance has arisen in federal courts where the defendant was a member of a Native American tribe and the crime occurred on the tribal reservation.<sup>75</sup> The typical databases of allele frequencies used for probability estimation address broader population groups, such as Blacks, Whites, Native Americans, Hispanics from the southeastern United States, and Hispanics from the southwestern United States. The extent to which broad racial and cultural comparison populations must correspond to the characteristics of the suspect population turns on the extent to which the distribution of alleles tested for in the forensic analysis differs for the suspect population and the comparison database. This remains a disputed issue among scientists. Difficulty in resolving this issue was responsible, in part, for the NRC committee's recommending a technique that does not require the designation of a suspect population.

2. Does the comparison population conform to characteristics that allow the estimation of the joint occurrence of matching alleles by multiplication of the probabilities of the individual alleles?

The comparison population must conform to assumptions that underlie the technique used to compute the estimate of a coincidental match, or at least conform sufficiently that minor deviations are of little consequence in computing the probability estimates. The computation of the probability of a random match by the product rule technique is based on the assumption that the individual alleles of the DNA profile are independent of one another. According to the principles of population genetics, the independence of alleles may be assumed only where the comparison population mixes freely and mates randomly (with respect to the alleles) such that the distribution of alleles within the comparison population is homogeneous.<sup>76</sup> If the comparison population

<sup>&</sup>lt;sup>73</sup>. See, e.g., United States v. Jakobetz, 955 F.2d 786 (2d Cir.) (rape committed at a rest stop along an interstate highway by person identified only as White), *cert. denied*, 113 S. Ct. 104 (1992).

<sup>&</sup>lt;sup>74</sup>. See Kaye, *supra* note 66, at 137–38 & n.155.

<sup>&</sup>lt;sup>75</sup>. E.g., United States v. Two Bulls, 918 F.2d 56 (8th Cir. 1990), and vacated, 925 F.2d 1127 (8th Cir. 1991); United States v. Martinez, 3 F.3d 1191 (8th Cir. 1993), cert. denied, 114 S. Ct. 734 (1994).

<sup>&</sup>lt;sup>76</sup> A comparison population whose members freely mix and randomly mate with respect to the relevant genes, resulting in a homogeneous distribution of alleles within the population, is said to be in *Hardy-Weinberg equilibrium*. Fleming, *supra* note 49, at 322. If the distribution of alleles within the population is not homogeneous, such a state may not exist, and the probe-detected alleles may not be independent, thereby violating an assumption of the product rule technique. *Id.* The assumptions of the product rule technique may also be violated if the detected alleles are in such proximity that they tend to be inherited together, resulting in

does not conform to these assumptions, the alleles may not be independent, and the computation of probability estimates may be incorrect.

Opponents of the product rule technique argue that it is inappropriate to use broad racial and cultural characteristics to specify the comparison population, because in reality such groups do not mix freely with respect to relevant genes. Broad racial groups, they argue, disguise subpopulations that would be more appropriate for comparison. More specifically, opponents charge that use of broad racial groups violates the assumption of independence justifying the multiplication of the separate probabilities assigned to each probe. For example, a White comparison group that includes diverse ethnic groups (e.g., persons of Polish, Italian, or Irish descent) may mask differences in the distribution of alleles among subgroups. Opponents charge that no meaningful estimate of the probability of a particular DNA profile can be developed without specifying a suitable subpopulation that meets the demanding assumptions of the product rule technique.<sup>77</sup> When alleles are not independent, as when a comparison group contains a substructure, the product rule technique may underestimate the probability that the forensic and suspect DNA patterns match by coincidence.

Proponents of the product rule technique acknowledge that some substructuring may exist in the comparison populations typically used; but they argue that it is inappropriate to apply the assumption strictly and that the probability estimates generally are accurate in spite of violations of the strict assumption regarding the absence of substructuring.<sup>78</sup> Furthermore, proponents claim that the conservative features of the fixed-bin method more than compensate for any underestimation.<sup>79</sup> The NRC committee notes that what little empirical evidence existed at the time of its report appeared to support this contention.<sup>80</sup>

At the time of this writing, two federal courts of appeals have approved the admission of probability estimates based on the product rule technique, and one court of appeals has approved a variation of the product rule technique with a

<sup>80</sup>. NRC Report, *supra* note 7, at 80:

Recent empirical studies concerning VNTR loci detected no deviation from independence within or across loci. Moreover, pairwise comparisons of all five-locus DNA profiles in the FBI database showed no exact matches; the closest match was a single threelocus match among 7.6 million pairwise comparisons. These studies are interpreted as indicating that multiplication of gene frequencies across loci does not lead to major inaccuracies in the calculation of genotype frequency—at least not for the specific polymorphic loci examined. (citations omitted)

See also Lempert, supra note 65, at 45-46.

*linkage disequilibrium. Id.* The latter problem often is avoided by probing for alleles of genes known to be on different chromosomes. *Id.* 

<sup>&</sup>lt;sup>77</sup>. Eric S. Lander, DNA *Fingerprinting on Trial*, 339 Nature 501 (1989); Lewontin & Hartl, *supra* note 71, at 1746.

<sup>&</sup>lt;sup>78</sup>. Chakraborty & Kidd, *supra* note 56, at 1736; OTA Report, *supra* note 12, at 68 ("Consensus exists that genetic departures as extreme as those for rare disease alleles do not exist for alleles detected by forensic DNA probes.").

<sup>&</sup>lt;sup>79</sup>. Chakraborty & Kidd, supra note 56, at 1738. For a discussion of the fixed-bin method, see supra VII.A.

conservative adjustment for substructuring. Before the Daubert decision, the U.S. Court of Appeals for the Second Circuit upheld the admission of probability estimates under Rule 702 of the Federal Rules of Evidence.<sup>81</sup> Soon after the Daubert decision, the U.S. Court of Appeals for the Sixth Circuit held that disputes over probability estimation techniques go to weight, not admissibility.<sup>82</sup> The district court had found probability estimates based on the product rule technique admissible under Frye's test of general acceptance within the scientific community. The Sixth Circuit reinterpreted these findings under the Daubert standards and concluded, "[I]t is clear from this record that the DNA evidence and testimony would have met the more liberal Rule 702 test adopted by the Supreme Court."83 The court noted that the theory and methods that support the FBI estimates can and have been tested and that they have received at least some degree of peer review and evaluation.<sup>84</sup> The court also found that the theory and methods were generally accepted by the relevant scientific community, or at least not disfavored by a substantial part of that community.<sup>85</sup> The court noted that the dispute over ethnic substructure goes to the accuracy of the probability estimates and thus to the weight of the evidence, not its admissibility.<sup>86</sup> Because the NRC report was not part of the trial record, the Sixth Circuit did not consider it or the merits of the modified ceiling principle technique. However, the court noted that the substance of the criticisms presented in the NRC report, including the possibility of ethnic substructure, was before the court in the form of expert testimony.<sup>87</sup>

<sup>82</sup>. United States v. Bonds, 12 F.3d 540, 564 (6th Cir. 1993), *aff* g United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991) ("This substructure argument involves a dispute over the accuracy of the probability results, and thus this criticism goes to the weight of the evidence, not its admissibility."). One additional federal court of appeals has considered but has not ruled on the admissibility of probability estimates under the product rule. An earlier decision by the U.S. Court of Appeals for the Eighth Circuit was vacated and set for *en banc* review, but the appeal was dismissed following the death of the appellant. United States v. Two Bulls, 918 F.2d 56 (8th Cir. 1990), *and vacated*, 925 F.2d 1127 (8th Cir. 1991). In a more recent decision, the Eighth Circuit was not required to decide whether it was error to admit evidence of a profile match without an accompanying probability estimate because the defendant had asked the court to exclude all statistical evidence of the probability of a match. United States v. Martinez, 3 F.3d 1191, 1198–99 (8th Cir. 1993), *cert. denied*, 114 S. Ct. 734 (1994).

<sup>83</sup>. Bonds, 12 F.3d at 557.

<sup>84</sup>. *Id.* at 558–60. The court acknowledged concern over the failure to conduct blind proficiency tests and the lack of specific information concerning error rates, but it noted that error rate is but one of the factors that bear on admissibility of scientific evidence. *Id.* at 560.

<sup>85</sup>. *Id.* at 562 ("[É]ven substantial criticism as to one theory or procedure will not be enough to find that the theory/procedure is not generally accepted. Only when a theory or procedure does not have the acceptance of most of the pertinent scientific community, and in fact a substantial part of the scientific community disfavors the principle or procedure, will it not be generally accepted.").

<sup>56</sup>. *Id.* at 564 (The potential of ethnic substructure does not mean that the theory and procedures used by the FBI are not generally accepted; it means only that there is a dispute over whether the results are as accurate as they might be and what, if any, weight the jury should give those results.).

<sup>87</sup>. Id. at 552–53.

<sup>&</sup>lt;sup>81</sup>. United States v. Jakobetz, 955 F.2d 786, 798–800 (2d Cir.) (no abuse of discretion where evidence was admitted under relevancy standard prior to *Daubert* and the trial judge concluded that conservative estimate arising from FBI's fixed-bin method compensates for any departure from assumptions of the analysis arising from a possible substructure in the comparison population), *cert. denied*, 113 S. Ct. 104 (1992).

The U.S. Court of Appeals for the Ninth Circuit upheld the admission of probability estimates computed in a way that takes into account some of the concerns expressed in the NRC report.<sup>88</sup> The defendant, a member of the Navajo tribe, challenged his conviction for rape and murder, contending, among other things, that the existing comparison databases underrepresent the members of his tribe.<sup>89</sup> Because members of his tribe form an isolated subgroup that may have a distinct genetic substructure, he contended that the probability estimates derived from existing databases would likely understate the probability of a random match between the forensic sample and members of his tribe. In the absence of a database with appropriate representation of members of the Navajo tribe, the prosecution used a technique similar to that used in the modified ceiling principle technique—allele frequencies were examined for several Native American tribal populations and the estimates were based on the highest frequency.<sup>90</sup>

The court found the probability estimates generated by a variation of the modified ceiling principle technique satisfied the standard of Rule 702 as interpreted in *Daubert*, and it went on to assess whether the probative value of such evidence was outweighed by the potential for prejudice under Rule 403.<sup>91</sup> Of particular concern was the possibility "that the jury will accept the DNA evidence as a statement of source probability (i.e., the likelihood that the defendant is the source of the evidentiary sample)"<sup>92</sup> rather than as an estimate of the rarity

<sup>89</sup>. *Id.* at 1155. The crime occurred in an isolated area of the Navajo tribal reservation, and people with an opportunity to commit the crime were members of the tribe.

<sup>30</sup>. The details of this technique are not reported in the opinion. A footnote quotes one of the government's witnesses as follows: "[T]hey looked at the allele frequencies in several American Indian segment tribal populations and picked up the one containing largest frequency (sic)." The court noted:

While not calculated pursuant to the NRC Report's controversial recommendation to adopt the ceiling principle, the one in 2563 probability that was introduced at trial was nonetheless arguably calculated on the basis of somewhat conservative statistical assumptions, was premised on the favorable assumption that the source of the sperm was a Native American and was emphasized at the expense of a much smaller probability of a random match that Government witnesses testified would be statistically defensible. (footnotes omitted).

Id. at 1158.

### <sup>91</sup>. The court noted:

Notwithstanding *Daubert's* express preference for exposing novel scientific theories and methodologies to the glare of the adversarial process, *Daubert* enjoins watchful assessment of the risk that a jury would assign undue weight to DNA profiling statistics even after hearing appellant's opposing evidence, the testimony of Government witnesses under vigorous cross-examination and the careful instructions of the district court on burdens of proof. Of particular concern is where the Government seeks to present probability testimony derived from statistical analysis, the third main phase of DNA profiling. Numerous hazards attend the courtroom presentation of statistical evidence of any sort. Accordingly, Rule 403 requires judicial vigilance against the risk that such evidence will inordinately distract the jury from or skew its perception of other, potentially exculpatory evidence lacking not so much probative force as scientific gloss. (footnotes omitted).

Id. at 1156.

<sup>92</sup>. *Id.* at 1156. In support of its concern, the court cited Lempert, *supra* note 45, at 306 ("Unfortunately, the careless presentation of evidence . . . may make it look as if the question of the rareness of the evidence

<sup>&</sup>lt;sup>88</sup>. United States v. Chischilly, 30 F.3d 1144 (9th Cir. 1994).

of the DNA profile. Estimating the probability that the defendant was the source of the crime sample requires not only an estimate of a coincidental match based on a comparison population but consideration of other evidence bearing on guilt or innocence. In addition, the court indicated concern that even when the source probability is correctly assessed, jurors might equate such probability estimates with guilt, thereby ignoring the possibility that there may be a noncriminal explanation of how the suspect could have been the source of the evidentiary sample. The solution to such potential problems, the court concluded, is careful oversight by the district court to limit the opportunity for misrepresentation of statistical evidence. The court of appeals approved the admission of the estimates, holding that the district court exercised such control by presenting statistical estimates as the probability of a random match, not the probability of the defendant's innocence, and by admitting probability estimates based on conservative statistical assumptions.<sup>93</sup>

As a result, federal courts are at odds with a number of state courts that have considered explicitly the NRC report, cited the modified ceiling principle technique with favor, and expressed the hope that it will gain general acceptance in the scientific community.<sup>94</sup> Unlike the federal courts that have addressed the

<sup>93</sup>. United States v. Chischilly, 30 F.3d at 1158.

DNA profile and the probability that the defendant's matching DNA is the source of the evidence profile are identical."); and Janet C. Hoeffel, Note, The Dark Side of DNA Profiling: Unreliable Scientific Evidence Meets the Criminal Defendant, 42 Stan. L. Rev. 465, 515 (1990) (asserting that "juries often erroneously equate the frequency of the accused's blood type in the population with the probability of innocence, discounting the other evidence in the case pointing to guilt or innocence, such as the fact that a close relative is also suspected"). See also William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials, The Prosecutor's Fallacy and the Defense Attorney's Fallacy, 11 Law & Hum. Behav. 167, 170–71 (1987); Jonathan J. Koehler, Error and Exaggeration in the Presentation of DNA Evidence at Trial, 34 Jurimetrics J. 21 (1993).

<sup>&</sup>lt;sup>94</sup>. Commonwealth v. Lanigan, 596 N.E.2d 311, 316 (Mass. 1992) ("This [ceiling] principle is entirely in keeping with the hope that we expressed in Commonwealth v. Curnin, that the scientific community would 'generally agree on a means of arriving at a conservative estimate of the probability of another person having the same alleles and thus resolve all uncertainties and variables in favor of the defense."" (citation omitted)); People v. Barney, 8 Cal. App. 4th 798, 821 (Cal. Ct. App. 1992) ("There must be some common ground, some sufficiently conservative method of determining statistical significance, as to which there is general scientific agreement.... The NRC report on DNA analysis appears to point the way to such common ground." (citations omitted)); State v. Vandebogart, 616 A.2d 483, 494 (N.H. 1992) ("The NRC asserts that the ceiling principle can account for any error caused by possible population substructure. Therefore, the admissibility of population frequency estimates do not necessarily await resolution of the population substructure issue, as long as the relevant scientific community generally accepts a method for calculating statistical probabilities."). Vargas v. State, Nos. 92-556, 92-557, 92-558 (consolidated), 1994 WL 231360, \*13 (Fla. App. 1 Dist. June 1, 1994) ("The discussion of the modified ceiling principle in the cases appears to confirm that a more conservative calculation may be possible, which would be generally accepted in the relevant scientific commu-nity, ..."); People v. Watson, 629 N.E.2d 634, 647–48 (III. App. 1994) ("Because the match of DNA patterns is a matter of substantial significance and because this case has the potential for becoming a significant precedent in this jurisdiction, we believe the trial court should be given the opportunity to determine whether the recently promulgated ceiling principle is appropriate under *Frye* for calculating the probability estimate to be applied to a match declaration in the present case. Accordingly, we remand this cause to the trial court for such a determination. At least in our view, the NRC Report, which was not previously available to the trial court, suggests that the DNA evidence should be admitted on the basis of this more conservative probability calculation for which the requisite consensus may now exist."); State v. Bloom, 516 N.W.2d 159, 160 (Minn. 1994) (testimony regarding statistical probability of a matching DNA profile derived by using the "interim ceil-

issue, many of the state courts continue to follow variations of the *Frye* standard and look to the NRC report as expressing a consensus of scientific opinion. Yet, while some state courts have embraced the modified ceiling principle technique, a portion of the scientific community has questioned its scientific validity.<sup>95</sup> The possibility of a convergence of scientific opinion around the ceiling principle technique seems remote, and the NRC plans to impanel another committee to consider scientific criticisms of the technique and more recent research.<sup>96</sup>

ing method" admissible as an exception to Minnesota's rule against admitting statistical probability evidence on the issue of identity in criminal prosecutions).

<sup>95</sup>. B. Devlin et al., Statistical Evaluation of DNA Fingerprinting: A Critique of the NRC's Report, 259 Science 748, 749 (1993); Peter Aldhous, Geneticists Attack NRC Report As Scientifically Flawed, 259 Science 755 (1993); Joel E. Cohen, The Ceiling Principle Is Not Always Conservative in Assigning Genotype Frequencies for Forensic DNA Testing, 51 Am. J. Hum. Genet. 1165, 1166 (1992) (letter to the editor); B. Devlin et al., Comments on the Statistical Aspects of the NRC's Report on DNA Typing, 39 J. Forensic Sci. 28 (1994). See also People v. Wallace, 14 Cal. App. 4th 651, 660 (Cal. Ct. App. 1993):

[R]ecent developments have shown that general acceptance may not be easily achieved. It appears that some proponents of DNA analysis, rather than attempting to come to terms with the NRC report or some other compromise on statistical calculation, have taken the offensive and attacked the report's proposed new methods of statistical calculation as unsound.

Forensic DNA Evidence

<sup>&</sup>lt;sup>96</sup>. Rorie Sherman, New Scrutiny for DNA Testing, Nat'l L.J., Oct. 18, 1993, at 3.

This page left blank intentionally for proper pagination when printing two-sided

Appendix A: Schematic of Single-Locus Probe RFLP Analysis This page left blank intentionally for proper pagination when printing two-sided

# Appendix B: Example of Illustrated DNA Patterns on Autorad

Autorad courtesy of Cellmark Diagnostics.

The autorad on the preceding page depicts DNA evidence from a criminal case in which blood stains on the pants and shirt of the defendant yielded DNA that was compared with the DNA samples known to be from the victim and the defendant. The autorad depicts DNA patterns in ten parallel lanes. The images in the lanes represent the following sources:

The lanes labeled " $\lambda$ ," "1kb," and "TS" show control samples of DNA. These serve as quality-control checks.

The lane labeled "D" shows the pattern obtained from a known DNA sample obtained from the defendant.

The lane labeled "jeans" shows the print of DNA from blood stains on the defendant's pants.

The two lanes labeled "shirt" show the prints of DNA from blood stains on the defendant's shirt.

The lane labeled "V" is a print of the DNA from a blood sample known to be from the victim.

The DNA from the blood stains on the defendant's clothing (lanes labeled "jeans" and "shirt") do not match his own blood sample ("D") but do match that of the victim ("V").

### Appendix C: TWGDAM Guidelines

The following excerpts were taken from Technical Working Group on DNA Analysis Methods & California Ass'n of Criminalists Ad Hoc Comm. on DNA Quality Assurance, *Guidelines for a Quality Assurance Program for DNA Analysis*, 18 Crime Laboratory Dig. 44, 52–56, 60–62 (1991).

#### 3. Documentation

The DNA laboratory must maintain documentation on all significant aspects of the DNA analysis procedure, as well as any related documents or laboratory records that are pertinent to the analysis or interpretation of results, so as to create a traceable audit trail. This documentation will serve as an archive for retrospective scientific inspection, reevaluation of the data, and reconstruction of the DNA procedure. Documentation must exist for the following topic areas:

3.1 Test Methods and Procedures for DNA Typing

This document must describe in detail the protocol currently used for the analytical testing of DNA. This protocol must identify the standards and controls required, the date the procedure was adopted and the authorization for its use. Revisions must be clearly documented and appropriately authorized.

3.2 Population Data Base to include number, source and ethnic and/or racial classification of samples.

3.3 Quality control of critical reagents (such as commercial supplies and kits which have expiration dates) to include lot and batch numbers, manufacturer's specifications and internal evaluations.

3.4 Case files/case notes - Must provide foundation for results and conclusions contained in formal report.

- 3.5 Data analysis and reporting
- 3.6 Evidence handling protocols
- 3.7 Equipment calibration and maintenance logs
- 3.8 Proficiency testing
- 3.9 Personnel training and qualification records
- 3.10 Method validation records
- 3.11 Quality assurance and audit records
- 3.12 Quality assurance manual

- 3.13 Equipment inventory
- 3.14 Safety manuals
- 3.15 Material safety data sheets
- 3.16 Historical or archival records
- 3.17 Licenses and certificates
- 4. Validation

4.1 General Considerations for Developmental Validation of the DNA Analysis Procedure

4.1.1 Validation is the process used by the scientific community to acquire the necessary information to assess the ability of a procedure to reliably obtain a desired result, determine the conditions under which such results can be obtained and determine the limitations of the procedure. The validation process identifies the critical aspects of a procedure which must be carefully controlled and monitored.

4.1.2 Validation studies must have been conducted by the DNA laboratory or scientific community prior to the adoption of a procedure by the DNA laboratory.

4.1.3 Each locus to be used must go through the necessary validation.

4.1.4 The DNA primers, probe(s) or oligonucleotides selected for use in the forensic DNA analysis must be readily available to the scientific community.

4.1.5 The validation process should include the following studies (Report of a Symposium on the Practice of Forensic Serology 1987, and Budowle *et al.* 1988):

4.1.5.1 Standard Specimens - The typing procedure should have been evaluated using fresh body tissues, and fluids obtained and stored in a controlled manner. DNA isolated from different tissues from the same individual should yield the same type.

4.1.5.2 Consistency - Using specimens obtained from donors of known type, evaluate the reproducibility of the technique both within the laboratory and among different laboratories.

4.1.5.3 Population Studies - Establish population distribution data in different racial and/or ethnic groups.

4.1.5.4 Reproducibility - Prepare dried stains using body fluids from donors of known types and analyze to ensure that the strain specimens exhibit accurate, interpretable and reproducible DNA types of profiles that match those obtained on liquid specimens.

4.1.5.5 Mixed Specimen Studies - Investigate the ability of the system to detect the components of mixed specimens and define the limitations of the system.

4.1.5.6 Environmental Studies - Evaluate the method using known or previously characterized samples exposed to a variety of environmental conditions. The samples should be selected to represent the types of specimens to be routinely analyzed by the method. They should resemble actual evidence materials as closely as possible so that the effects of factors such as matrix, age and degradative environment (temperature, humidity, UV) of a sample are considered.

4.1.5.7 Matrix Studies - Examine prepared body fluids mixed with a variety of commonly encountered substances (e.g. dyes, soil) and deposited on commonly encountered substrates (e.g. leather, denim).

4.1.5.8 Nonprobative Evidence - Examine DNA profiles in nonprobative evidentiary stain materials. Compare the DNA profiles obtained for the known liquid blood versus questioned blood deposited on typical crime scene evidence.

4.1.5.9 Nonhuman Studies - Determine if DNA typing methods designed for use with human specimens detect DNA profiles in nonhuman source stains.

4.1.5.10 Minimum Sample - Establish quantity of DNA needed to obtain a reliable typing result.

4.1.5.11 On-site Evaluation - Set up newly developed typing methods in the case working laboratory for on-site evaluation of the procedure.

4.1.5.12 It is essential that the results of the developmental validation studies be shared as soon as possible with the scientific community through presentations at scientific/professional meetings. It is imperative that details of these studies be available for peer review through timely publications in scientific journals.

4.2 Characterization of Loci

During the development of a DNA analysis system, basic characteristics of the loci must be determined and documented. (Baird 1989; AABB Standards Committee 1990.)

4.2.1 Inheritance - DNA loci used in forensic testing shall have been validated by family studies to demonstrate the mode of inheritance. Those DNA loci used in parentage testing should have a low frequency of mutation and/or recombination.

4.2.2 Gene Mapping - The chromosomal location of the polymorphic loci used for forensic testing shall be submitted to or recorded in the Yale Gene Library or the International Human Gene Mapping Workshop.

4.2.3 Detection - The molecular basis for detecting the polymorphic loci shall be documented in the scientific or technical literature.

4.2.3.1 For RFLP this includes the restriction enzyme and the probes used.

4.2.3.2 For PCR this includes the primers and probes if used.

4.2.4 Polymorphism - The type of polymorphism detected shall be known.

4.3 Specific Developmental Validation of RFLP Procedures

4.3.1 Restriction - The conditions and control(s) needed to ensure complete and specific restriction must be demonstrated.

4.3.2 Separation - Parameters for the reproducible separation of DNA fragments must be established.

4.3.3 Transfer - Parameters for the reproducible transfer of DNA fragments must be established.

4.3.4 Detection - The hybridization and stringency wash conditions necessary to provide the desired degree of specificity must be determined.

4.3.5 Sizing - The precision of the sizing procedure must be established.

4.4 Specific Developmental Validation of PCR Based DNA Procedures

4.4.1 Amplification

4.4.1.1 The PCR primers must be of known sequence.

4.4.1.2 Conditions and measures necessary to protect pre-amplification samples from contamination by post PCR materials should be determined (see Section 7.5).

4.4.1.3 The reaction conditions such as thermocycling parameters and critical reagent concentrations (primers, polymerase and salts) needed to provide the required degree of specificity must be determined.

4.4.1.4 The number(s) of cycles necessary to produce reliable results must be determined.

4.4.1.5 Potential for differential amplification must be assessed and addressed.

4.4.1.6 Where more than one locus is amplified in one sample mixture, the effects of such amplification on each system (alleles) must be addressed and documented.

4.4.2 Detection of PCR Product

The validation process will identify the panel of positive and negative controls needed for each assay described below.

4.4.2.1 Characterization without hybridization

When a PCR product is characterized directly, appropriate standards for assessing the alleles shall be established (e.g., size markers). 4.4.2.2 Characterization with hybridization (a) Hybridization and stringency wash conditions necessary to provide the desired degree of specificity must be determined.

(b) For assays in which the amplified target DNA is to be bound directly to a membrane, some mechanism should be employed to ensure that the DNA has been applied to the membrane.

(c) For assays in which the probe is bound to the membrane, some mechanism should be employed to show that adequate amplified DNA is present in the sample (e.g., a probe which reacts with any amplified allele or a product yield gel).

4.5 Internal Validation of Established Procedures (ASCLD 1986) Prior to implementing a new DNA analysis procedure, or an existing DNA procedure developed by another laboratory that meets the developmental criteria described under Section 4.1, the forensic laboratory must first demonstrate the reliability of the procedure in-house. This internal validation must include the following:

4.5.1 The method must be tested using known samples.

4.5.2 If a modification which materially effects the results of an analysis has been made to an analytical procedure, the modified procedure must be compared to the original using identical samples.

4.5.3 Precision (e.g., measurement of fragment lengths) must be determined by repetitive analyses to establish criteria for matching.

4.5.4 The laboratory must demonstrate that its procedures do not introduce contamination which would lead to errors in typing.

4.5.5 The method must be tested using proficiency test samples. The proficiency test may be administered internally, externally or collaboratively.

- 7. Analytical Procedures
  - 7.1 Sample Evaluation and Preparation

7.1.1 General characterization of the biological material should be performed prior to DNA analysis. Evidence samples submitted should be evaluated to determine the appropriateness for DNA analysis.

7.1.2 When semen is identified, a method of differential extraction should be employed, and when appropriate, each of the DNA fractions typed (see Section 4.1.5.10).

7.1.3 Testing of evidence and evidence samples should be conducted to provide the maximum information with the least consumption of the sample. Whenever possible, a portion of the original sample should be retained or returned to the submitting agency as established by laboratory policy.

7.2 DNA Isolation

7.2.1 The DNA isolation procedure should protect against sample contamination.

7.2.2 The effectiveness of the DNA isolation procedure should be evaluated by regular use of an appropriate cellular source of human DNA.

7.3 Procedures for Estimating DNA Recovery:

A procedure should be used for estimating the quality (extent of DNA degradation) and quantity of DNA recovered from the specimens. One or more of the following procedures may be employed to evaluate the effectiveness of the DNA recovery.

7.3.1 Yield Gel - Yield gels must include a set of high molecular weight DNA calibration standards for quantitative estimate of yield.

7.3.2 UV Absorbance - Absorbance and wavelength standards or a high molecular weight DNA calibration standard may be used.

7.3.3 Fluorescence - Approximate quantification of extracted DNA can be accomplished by comparison with known concentrations of high molecular weight DNA.

7.3.4 Hybridization - Quantitation with human/primate specific probes requires an appropriate set of human DNA standards.

7.4 Analytical Procedures for RFLP Analysis

7.4.1 Restriction Enzymes

7.4.1.1 Prior to its initial use, each lot of restriction enzyme should be tested against an appropriate viral, human or other DNA standard which produces an expected DNA fragment pattern under standard digestion conditions. The restriction enzyme should also be tested under conditions that will reveal contaminating nuclease activity.

7.4.1.2 Demonstration of Restriction Enzyme Digestion - Digestion of extracted DNA by the restriction enzyme should be demonstrated using a test gel which includes:

(a) Size Marker - Determines approximate size range of digested DNA.

(b) Human DNA Control - Measures the effectiveness of

restriction enzyme digestion of genomic human DNA.

7.4.2 Analytical Gel - The analytical gel used to separate restriction fragments must include the following:

7.4.2.1 Visual Marker - Visual or fluorescent markers which are used to determine the end point of electrophoresis.

7.4.2.2 Molecular Weight Size Markers - Markers which span the RFLP size range and are used to determine the size of unknown restriction fragments. Case samples must be bracketed by molecular weight size marker lanes. 7.4.2.3 Human DNA Control - A documented positive human DNA control of known type which produces a known fragment pattern with each probe and serves as a systems check for the following functions:

- (a) electrophoresis quality and resolution
- (b) sizing process
- (c) probe identity
- (d) hybridization efficiency
- (e) stripping efficiency

7.4.2.4 A procedure should be available to interpret altered migration of DNA fragments.

7.4.3 Southern Blots/Hybridization - The efficiency of blotting, hybridizations and stringency washes are monitored by the human DNA control and size markers.

7.4.4 Autoradiography - The exposure intensity is monitored by the use of multiple X-ray films or by successive exposures in order to obtain films of the proper intensity for image analysis.

7.4.5 Image and Data Processing - The functioning of image and data processing is monitored by the human DNA control allelic values.

7.5 Analytical Procedures for PCR Based Techniques

7.5.1 Internal Controls and Standards

The laboratory's QC guidelines should contain specific protocols to assess critical parameters in normal operations which include the following:

7.5.1.1 Negative controls to be included with each sample set are: (a) a reagent blank and (b) an amplification blank.

7.5.1.2 A human DNA known type must be introduced at the amplification step as a positive control and carried through the remainder of the typing.

7.5.1.3 Where appropriate, substrate controls should be collected from the evidence (e.g., unstained areas adjacent to stained areas, hair shafts adjacent to hair roots) and should be processed at the same time as the evidence samples.

7.5.1.4 Where feasible, the sample should be split for duplicate analysis as early as possible prior to amplification.

7.5.1.5 To characterize amplified fragment length polymorphisms, markers which span the allele size range must be used. Case samples must be bracketed by marker lanes.

This page left blank intentionally for proper pagination when printing two-sided
# **Glossary of Terms**

Many of the following terms are defined specifically within the context of forensic typing of human DNA. Some of these terms have broader or slightly different meanings in other molecular biology applications. These terms and definitions were adapted in part from the following: Committee on DNA Technology in Forensic Science, National Research Council, DNA Technology in Forensic Science (1992); Office of Technology Assessment, U.S. Congress, Genetic Witness: Forensic Uses of DNA Tests (1990); Lorne T. Kirby, DNA Fingerprinting: An Introduction (1990).

- Adenine (A). One of the four bases, or nucleotides, that make up the DNA double helix. Adenine only binds to thymine. See Nucleotide.
- Allele. An alternative form of a gene or VNTR at a specific locus. Some genes have two variants (e.g., an allele for eye color may be blue or brown); others have more. Alleles are inherited separately from each parent. At the same loci on any two homologous chromosomes, an individual may have two different alleles (heterozygous) or the same allele (homozygous).
- Amplified Fragment Length Polymorphism (AMP-FLP). A DNA identification technique that uses PCR-amplified DNA fragments of varying lengths based on VNTRs.
- Autoradiograph (Autorad). In RFLP analysis, the x-ray film (or print thereof) showing the positions of radioactively marked lengths (bands) of DNA.
- Autosome. Any chromosome other than the sex chromosomes X or Y. See Chromosome.
- *Band-Shift.* Movement of DNA fragments in one lane of a gel at a different rate than fragments of identical length in another lane, resulting in the same pattern "shifted" up or down relative to the comparison lane. Band-shift does not necessarily occur at the same rate in all portions of the gel.
- Band Size. Length of DNA fragment measured in base pairs. See Base Pair.
- Base Pair (bp). Two complementary nucleotides bonded together at the matching bases (A and T or C and G) to form one segment of the DNA double helix. The length of a DNA fragment often is measured in numbers of base

pairs (1 kilobase (kb) =1000 bp); base pair numbers also are used to describe the location of an allele on the DNA strand.

- *Chromosome*. A rod-like structure composed of DNA which carries part of the genome. Most normal human cells contain forty-six chromosomes, twenty-two autosomes and one sex chromosome (X) inherited from the mother, and twenty-two autosomes and one sex chromosome (either X or Y) inherited from the father.
- *Confidence Interval.* Interval set around an allele frequency that accounts for uncertainty in measurement of the allele. The upper bound of this interval (the upper confidence limit) may be used as a conservative estimate of the frequency of the allele.
- *Cytosine* (C). One of the four bases, or nucleotides, that make up the DNA double helix. Cytosine only binds to guanine. See Nucleotide.
- *Denature*, *Denaturation*. The process of splitting, as by heating, two complementary strands of the DNA double helix into single strands in preparation for hybridization with biological probes.
- Deoxyribonucleic Acid (DNA). Basic molecule of heredity. DNA is composed of nucleotide building blocks, each containing a base (A, C, G, or T), a phosphate, and a sugar. These nucleotides are linked together in a double helix—two strands of DNA molecules paired up at complementary bases (A with T, C with G). See Adenine, Cytosine, Guanine, Thymine.
- DNA *Profile*. The pattern of band lengths on an autorad representing the combined results of multiple probes.
- Double Helix. Ladder structure of DNA.
- *Environmental Insult.* Exposure of DNA to external agents such as heat, moisture, and ultraviolet radiation, or chemical or bacterial agents. Such exposure can interfere with the enzymes used in the testing process, or otherwise make DNA difficult to analyze.
- *Fixed-Bin Method.* In a fixed-bin method, preestablished bins are designated by certain absolute base pair ranges, and band sizes are sorted into these existing bins.
- *Floating Bin Method.* In a floating bin method, the bin is centered on the base pair length of the allele in question, and the width of the bin is defined by the laboratory's matching rule (e.g.,  $\pm 2.5\%$  of band size).
- *Forensic Sample.* A sample of DNA associated with the commission of a crime, such as a DNA sample obtained from a blood stain at a crime scene. See Suspect Sample.

- *Gel Electrophoresis.* In RFLP analysis, the process of sorting DNA fragments by size by applying an electric current to an agarose gel. The different-sized fragments move at different rates through the gel.
- *Gene*. A distinctive, ordered sequence of nucleotide base pairs on a chromosome. The gene is the fundamental unit of heredity; each gene provides a "code" for a specific biological characteristic.
- *Genome*. The complete genetic makeup of an organism, comprising 50,000–100,000 genes in humans. See Gene.
- *Genotype.* The genetic code of an organism (as distinguished from phenotype, which refers to how the genetic code expresses itself, as in physical appearance).
- *Guanine* (G). One of the four bases, or nucleotides, that make up the DNA double helix. Guanine only binds to cytosine. See Nucleotide.
- *Hardy-Weinberg Equilibrium.* A condition in which the allele frequencies within a large, random, intrabreeding population are unrelated to patterns of mating. In this condition, the occurrence of alleles from each parent will be independent and have a joint frequency estimated by the product rule. See Independence, Linkage Disequilibrium.
- *Heterozygous*. Having a different allele at a given locus on each of a pair of homologous chromosomes, one inherited from each parent. See Allele.
- HLA DQ  $\alpha$ . A particular class of Human Leukocyte Antigen (HLA) whose locus has been sequenced completely and thus can be used for forensic typing. See Human Leukocyte Antigen.
- Homologous Chromosomes. The forty-four autosomes in the normal human genome are in homologous pairs that share an identical sequence of genes, but may have different alleles at the same loci.
- *Homozygous*. Having the same allele at a given locus on each of a pair of homologous chromosomes, one inherited from each parent. See Allele.
- Human Leukocyte Antigen (HLA). Antigen (foreign body that stimulates an immune system response) located on the surface of most cells (excluding red blood cells and sperm cells). HLAs differ among individuals and are associated closely with transplant rejection. See HLA DQ  $\alpha$ .
- *Hybridize*, *Hybridization*. Pairing up of complementary strands of DNA from different sources at the matching base pair sites. For example, a primer with the sequence AGGTCT would bond with the complementary sequence TCCAGA on a DNA fragment.
- *Independent, Independence.* A condition in which the occurrence of alleles is uncorrelated, permitting accurate estimation of frequency of the joint occur-

rence by the product rule. In large populations, the presence of distinct subpopulations may violate this condition. See Hardy-Weinberg Equilibrium.

- *Linkage Disequilibrium.* The nonrandom association of one allele at one locus and another allele at a different locus (i.e., they appear together with greater frequency than expected by chance). See Hardy-Weinberg Equilibrium.
- Locus, Loci. A specific location or locations on a chromosome.
- *Match Window.* Size of a bin for DNA fragment length measurement and comparison purposes.
- Modified Ceiling Principle Technique. An adjustment to account for possible population substructure in statistical estimates of a coincidental match on a DNA profile. In multiplying allele frequencies, an expert would use the upper bound of the confidence interval for the subpopulation in which the allele frequency is most common. This ceiling frequency is chosen so that the allele frequencies used in the calculation exceed the allele frequency in any of the subgroups, providing theoretically more conservative (i.e., favoring the suspect) estimates of the likelihood of a matching DNA profile.
- Monomorphic, Monomorphism. A single form of a genetic trait.
- Monomorphic Probe. A probe that detects the same allele and hence the same pattern in everyone.
- *Multilocus Probe.* A probe that marks multiple sites (loci). RFLP analysis using a multilocus probe will yield an autorad showing a striped pattern of thirty or more bands. Rarely used now in forensic applications in the United States.
- *Nucleotide*. A unit of DNA consisting of a base (A, C, G, or T) and attached to a phosphate and a sugar group. See Deoxyribonucleic Acid.
- *Oligonucleotide*. A synthetic polymer made up of fewer than 100 nucleotides; used as a primer or a probe in PCR. See Primer.
- *Polymerase Chain Reaction (PCR).* A process that mimics DNA's own replication processes to make up to millions of copies of the original genetic material in a few hours.
- Polymorphic, Polymorphism. Multiple forms of a genetic trait.
- Population Genetics. The study of genetic composition of groups of individuals.
- *Population Substructure.* In population genetics, the theory that allele frequencies are not spread randomly within large heterogeneous racial groups, but vary greatly between smaller ethnic subpopulations which do not mix freely.
- *Primer*. An oligonucleotide that attaches to one end of a DNA fragment and provides a point for more complementary nucleotides to attach and replicate the DNA strand. See Oligonucleotide.

- *Probe.* In forensics, a short segment of DNA used to detect certain alleles. The probe hybridizes, or matches up, to a specific complementary sequence. Probes allow visualization of the hybridized DNA, either by radioactive tag (usually used for RFLP analysis) or biochemical tag (usually used for HLA DQ  $\alpha$ ).
- *Product Rule Technique*. Technique for calculating genotype frequencies by multiplying allele frequencies observed in a population database. Assumes that alleles are inherited independently.
- *Replication.* The synthesis of new DNA from existing DNA. See Polymerase Chain Reaction.
- *Restriction Enzyme.* Protein that cuts double-stranded DNA at specific base pair sequences (different enzymes recognize different sequences). See Restriction Site.
- Restriction Fragment Length Polymorphism (RFLP) Analysis. Analysis of individual variations in the lengths of DNA fragments produced by digesting sample DNA with a restriction enzyme.
- *Restriction Site.* A sequence marking the location at which a restriction enzyme cuts DNA into fragments. See Restriction Enzyme.
- Sequence-Specific Oligonucleotide (SSO) Probes. Also, Allele-Specific Oligonucleotide (ASO) Probes. Oligonucleotide probes used in a PCR-associated detection technique to identify the presence or absence of certain base pair sequences identifying different alleles. The probes are visualized by an array of dots rather than by the "bar codes" associated with RFLP analysis.
- *Single-Locus Probe.* A probe that only marks a specific site (locus). RFLP analysis using a single-locus probe will yield an autorad showing one band if the individual is homozygous, two bands if heterozygous.
- Southern Blotting. Named for its inventor, a technique by which processed DNA fragments, separated by gel electrophoresis, are transferred onto a nylon membrane in preparation for the application of biological probes.
- *Thymine* (*T*). One of the four bases, or nucleotides, that make up the DNA double helix. Thymine only binds to adenine. See Nucleotide.
- Variable Number of Tandem Repeats (VNTR). Multiple copies of virtually identical base pair sequences, arranged in succession at a specific locus on a chromosome. The number of repeats varies from individual to individual, thus providing a basis for individual recognition.

This page left blank intentionally for proper pagination when printing two-sided

# **References on Forensic DNA Evidence**

- Committee on DNA Technology in Forensic Science, National Research Council, DNA Technology in Forensic Science (1992).
- Forensic DNA Technology (Mark A. Farley & James J. Harrington eds., 1991).
- Lorne T. Kirby, DNA Fingerprinting: An Introduction (1990).
- Office of Technology Assessment, U.S. Congress, Genetic Witness: Forensic Uses of DNA Tests (1990).
- David H. Kaye, DNA Evidence: Probability, Population Genetics, and the Courts, 7 Harv. J.L. & Tech. 101 (1993).
- David H. Kaye, The Forensic Debut of the National Research Council's DNA Report: Population Structure, Ceiling Frequencies and the Need for Numbers, 34 Jurimetrics J. 369 (1994).

# **Reference Guide on Statistics**

David H. Kaye David A. Freedman

**David H. Kaye**, A.M., J.D., is Professor of Law, Arizona State University College of Law, Tempe, Arizona. **David A. Freedman**, Ph.D., is Professor of Statistics, University of California, Berkeley, California.

This page left blank intentionally for proper pagination when printing two-sided

# Contents

- I. Introduction 335
  - A. Varieties and Limits of Statistical Expertise 336
  - B. Procedures That Enhance Statistical Testimony 337
    - 1. Maintaining professional autonomy 337
    - 2. Disclosing other analyses 338
    - 3. Disclosing data and analytical methods before trial 338
    - 4. Presenting expert statistical testimony 339
- II. How Have the Data Been Collected? 341
  - A. Individual Measurements 341
    - 1. Is the measurement process reliable? 341
    - 2. Is the measurement process valid? 342
    - 3. Are the measurements recorded correctly? 342
  - B. Descriptive Surveys and Censuses 343
    - 1. What method is used to select the units to be measured? 343
    - 2. Of the units selected, which are measured? 345
  - C. Experiments 346
    - 1. What are the independent and dependent variables? 348
    - 2. What are the confounding variables? 348
    - 3. Can the results be generalized? 349
  - D. Observational Studies of Causation 350
- III. How Have the Data Been Presented? 353
  - A. Is the Data Display Sufficiently Complete? 353
  - B. Are Rates or Percentages Properly Interpreted? 355
  - C. Does a Graph Portray Data Fairly? 356
    - 1. Displaying distributions: histograms 356
    - 2. Displaying trends 357
    - 3. Displaying association: scatter diagrams 359
  - D. Is an Appropriate Measure Used for the Center of a Distribution? 360
  - E. Is an Appropriate Measure of Variability Used? 361

- F. Is an Appropriate Measure of Association Used? 362
  - 1. Percentage-related statistics 362
  - 2. Correlation coefficients 364
    - a. Is the association linear? 366
    - b. Do outliers influence the coefficient? 366
    - c. Does a third variable influence the coefficient? 367
  - 3. Regression lines 367
    - a. What are the slope and intercept? 368
    - b. What does the slope ignore? 369
    - c. What is the unit of analysis? 370
- IV. What Inferences Can Be Drawn from the Data? 373
  - A. Estimation 374
    - 1. What estimator should be used? 374
    - 2. What is the standard error? 375
    - 3. What do standard errors and confidence intervals mean? 376
  - B. P-Values and Hypothesis Tests 378
    - 1. What is the *p*-value? 378
    - 2. Is a difference statistically significant? 380
    - 3. Questions about hypothesis tests 381
      - a. What is the power of the test? 381
      - b. One-tailed versus two-tailed tests 382
      - c. How many tests have been performed? 383
      - d. What are the interval estimates? 384
      - e. What are the other explanations for the findings? 385
  - C. Posterior Probabilities 386

Appendix: Technical Details on the Standard Error, the Normal Curve, and the *p*-Value 389

Glossary of Terms 395

References on Statistics 413

# I. Introduction

*Statistics*, broadly defined, is the science and art of gaining information from data. For statistical purposes, *data* mean observations or measurements, expressed as numbers. A *statistic* may refer to a particular numerical value, derived from the data. Baseball statistics, for example, is the study of data about the game; a player's batting average is a statistic.

The field of statistics includes methods for (1) collecting data, (2) analyzing data, and (3) drawing inferences from data. This reference guide describes the underlying ideas of statistics as they relate to legal proceedings. Statistical assessments figure prominently in antitrust, discrimination, fraud, homicide, sexual assault, trademark, toxic tort, and many other kinds of cases.<sup>1</sup> Typically, the most difficult arguments about such studies concern their probative value.<sup>2</sup>

This reference guide focuses on the nature of statistical thinking rather than on the rules of evidence or substantive legal doctrine. We hope that the explanations provided, although summary and nonmathematical in form, will permit judges who are confronted with statistical testimony to understand more of the

<sup>1.</sup> See generally David C. Baldus & James W. L. Cole, Statistical Proof of Discrimination (1980); Statistics and the Law (Morris H. DeGroot et al. eds., 1986); The Evolving Role of Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989); Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990); 1 & 2 Joseph L. Gastwirth, Statistical Reasoning in Law and Public Policy (1988).

<sup>2.</sup> Statistical studies suitably designed to address a material issue generally will be admissible under the Federal Rules of Evidence. The hearsay rule rarely is a serious barrier to the presentation of statistical studies. Sæ Linda A. Bailey et al., Reference Guide on Epidemiology n.1, in this manual. Likewise, since most statisti cal methods relied on in court are described in textbooks and journal articles and are capable of producing useful results when carefully and appropriately applied, the methodology generally satisfies the "scientific knowledge" requirement articulated in Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2795 (1993). For a discussion of the implications and scope of Daubert generally, see Margaret A. Berger, Evidentiary Framework §§ I, III, in this manual: Bert Black et al., Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, 72 Tex. L. Rev. 715 (1994); Richard D. Friedman, The Death and Transfiguration of Frye, 34 Jurimetrics J. 133 (1994); Susan R. Poulter, Daubert and Scientific Evidence: Assessing Evidentiary Reliability in Toxic Tort Cases, 1993 Utah L. Rev. 1307; Symposium, Scientific Evidence After the Death of Frye, 15 Cardozo L. Rev. 1745 (1994). Of course, a particular study may use a method that is entirely appropriate for some problems, but that is so poorly executed that it should be inadmissible under Fed. R. Evid. 403 and 702. Or, the method may be inappropriate for the problem at hand and thus lack the "fit' spoken of in Daubert. 113 S. Ct. at 2796. Or, the study may rest on data of the type not reasonably relied on by statisticians or substantive experts and hence run afoul of Fed. R. Evid. 703. Šee, e.g., Faust F. Rossi, Expert Witnesses 43–98 (1991); Ronald L. Carlson, Policing the Bases of Modern Expert Testimony, 39 Vand. L. Rev. 577 (1986); Paul R. Rice, Inadmissible Evidence as a Basis for Expert Opinion Testimony: A Response to Professor Carlson , 40 Vand. L. Rev. 583 (1987); Michael C. McCarthy, Note, "Helpful" or "Reasonably Reliable"?: Analyzing the Expert Witness's Methodology Under Federal Rules of Evidence 702 and 703 Cornell L. Rev. 350 (1992). More often, however, the battle over statistical evidence concerns weight or sufficiency rather than admissibility.

terminology, to place the evidence in context, to appreciate its strengths and weaknesses, and to develop and apply legal doctrine governing the use of statistical evidence.

The reference guide is organized as follows:

- Section I provides an overview of the field and offers some suggestions about procedures to encourage the best use of statistical expertise in litigation.
- Section II addresses data collection. The design of a study is the most important determinant of its quality. Section II describes the design of surveys, controlled experiments, and observational studies. It indicates when these procedures are likely to produce useful data for various purposes.
- Section III discusses methods for extracting and summarizing the most important features of data. *Descriptive statistics* is the art of describing and summarizing data, and section III considers the meaning, useful ness, and limitations of such descriptive statistics as the mean, median, standard deviation, correlation coefficient, and slope of a regression line. These are the basic descriptive statistics, and most statistical analyses seen in court use them as building blocks.
- Section IV describes the logic of statistical inference, emphasizing its foundations and limitations. In particular, it explains statistical estimation, standard errors, confidence intervals, *p*-values, and hypothesis tests.

# A. Varieties and Limits of Statistical Expertise

For convenience, the field of statistics may be divided into three subfields: probability, theoretical statistics, and applied statistics. Theoretical statistics is the study of the mathematical properties of statistical procedures; probability theory plays a key role in this endeavor. Results may be used by applied statisticians who specialize in particular types of data collection, such as survey research, or in particular types of analysis, such as multivariate methods.

Statistical expertise is not confined to those with degrees in statistics. Because statistical reasoning underlies all empirical research, researchers in many fields are exposed to statistical ideas. Experts with advanced degrees in the physical, medical, and social sciences and some of the humanities may receive formal training in statistics. Such specializations as biostatistics, epidemiology, econometrics, and psychometrics are primarily statistical, with an emphasis on methods and problems most important to the related substantive discipline.

Experience with applied statistics is the best indication of the type of statistical expertise needed in court. By and large, individuals who think of themselves as specialists in using statistical methods—and whose professional careers demonstrate this orientation—are most likely to apply appropriate procedures and cor-

rectly interpret the results. <sup>3</sup> At the same time, the choice of which data to examine or how best to model a particular process may require subject matter expertise that a statistician may lack. Statisticians typically advise experts in substantive fields on the procedures for collecting data and usually analyze data collected by others. As a result, cases involving statistical evidence often are (or should be) "two-expert" cases of interlocking testimony.<sup>4</sup> A labor economist, for example, may supply a definition of the relevant labor market from which an employer draws its employees, and the statistical expert may contrast the racial makeup of those hired to the racial composition of the labor market. Naturally, the value of the statistical analysis depends on the substantive economic knowledge that informs it. <sup>5</sup>

# B. Procedures That Enhance Statistical Testimony

#### 1. Maintaining professional autonomy

Ideally, experts who conduct research for litigants should proceed with the same objectivity that they would apply in other contexts. Thus, if experts testify or if their results are used in testimony by others, they should be free to do whatever analysis and have access to whatever data are required to address the problems

<sup>3.</sup> Forensic scientists and technicians often testify to probabilities or statistics derived from studies or databases compiled by others, even though some of these experts lack the training or knowledge required to understand and apply the information. *Sæ* Andre A. Moenssens, *Foreword: Novel Scientific Evidence in Criminal Cases: Some Words of Caution*, 84 J. Crim. L. & Criminology 1, 19 (1993) ("Most forensic experts who use . . . [probability and] statistics have no idea of how the calculations were made, and are not statisticians themselves."). We believe that courts should be more discerning in assessing the qualifications of these experts to opine on matters that they cannot explain adequately. *See* Paul C. Giannelli, *Expert Testimony and the Confrontation Clause*, 22 Cap. U. L. Rev. 45 (1993). State v. Garrison, 585 P.2d 563 (Ariz. 1978), illus trates the problem. In a murder prosecution involving bite mark evidence, a dentist was allowed to testify that "the probability factor of two sets of teeth being identical in a case similar to this is, approximately, eight in one million," even though "he was unaware of the formula utilized to arrive at that figure other than that it was 'computerized.'" *Id*. at 566, 568.

<sup>4.</sup> Sometimes a single witness presents both the substantive underpinnings and the statistical analysis. Ideally, such a witness has extensive expertise in both fields, although less may suffice to qualify the witness under Fed. R. Evid. 702. In deciding whether a witness who clearly is qualified in one field may testify in a related area, courts should recognize that qualifications in one field do not necessarily imply qualifications in the other. See, e.g., United States *ex rel*. DiGiacomo v. Franzen, 680 F.2d 515, 516 (7th Cir. 1982) (state criminal-ist testified not only to her finding matching hairs but also to a study that she vaguely recalled on the probability of coincidental matches); Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224, 286 (N.D. Tex. 1980) (plaintiffs' expert "is an impressive expert on statistics, but not on compensation or other personnel practices"), modified in part , 521 F. Supp. 656 (N.D. Tex. 1981), vacated, 723 F.2d 1195 (5th Cir.), *cert. denied* , 469 U.S. 1073 (1984).

<sup>5.</sup> In Vuyanich, 505 F. Supp. at 319, defendant's statistical expert criticized the plaintiffs' statistical model for an implicit, but restrictive, assumption about male and female salaries. The district court accepted the model because the plaintiffs' expert had a "very strong guess" about the assumption, and her expertise included labor economics as well as statistics. *ld*. It is doubtful, however, that economic knowledge sheds much light on the assumption, and it would have been simple to perform a less restrictive analysis. In this case, the court may have been overly impressed with a single expert who combined substantive and statistical expertise. Once the issue is defined by legal and substantive knowledge, some aspects of the statistical analysis will turn on statistical considerations alone, and expertise in another subject will not be pertinent.

the litigation poses in a professionally responsible fashion.<sup>6</sup> Questions about the freedom of inquiry accorded to testifying experts and the scope and depth of experts' investigations may reveal the experts' approach to acquiring and extracting relevant information.

#### 2. Disclosing other analyses

Statisticians may analyze data using a variety of statistical models and methods. There is nothing underhanded in, and much to be said for, looking at the data in a variety of ways. To permit a fair evaluation of the analysis that the statistician may settle on, however, the testifying expert should explain the history behind the development of the final statistical approach.<sup>7</sup>

#### 3. Disclosing data and analytical methods before trial

The collection of data often is expensive, and data sets typically contain at least some minor errors or omissions. Careful exploration of alternative modes of analysis also can be expensive and time-consuming. To minimize the occurrence of distracting debates at trial over the accuracy of data and the choice of analytical techniques, and to permit informed expert discussions of method, pre-trial procedures should be used, particularly with respect to the accuracy and scope of the data, and to discover the methods of analysis. <sup>8</sup> Suggested procedures along these lines are available elsewhere.<sup>9</sup>

6. See The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 164 (recommending that the expert be free to consult with colleagues who have not been retained by any party to the litigation and that the expert receive a letter of engagement providing for these and other safeguards).

7. See, e.g., Mikel Aickin, Issues and Methods in Discrimination Statistics, in Statistical Methods in Discrimination Litigation 159 (David H. Kaye & Mikel Aickin eds., 1986). Some commentators have urged that counsel who know of other data samples or analyses that do not support the client's position should reveal this fact to the court rather than attempt to mislead the court by presenting only favorable results. The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 167; *d*. William W Schwarzer, *h* Defense of "Automatic Disclosure in Discovery," 27 Ga. L. Rev. 655, 658–59 (1993) ("[T]he lawyer owes a duty to the court to make disclosure of core information."). The Panel on Statistical Assessments as Evidence in the Courts also recommends that "if a party gives statistical data to different experts for competing analyses, that fact be disclosed to the testifying expert, if any." The Evolving Role of Statistical Assessments as Evidence in the courts, *supra* note 1, at 167. Whether and under what circumstances a particular statistical analysis might be so imbued with counsel's thoughts and theories of the case that it should receive protection as the attorney's work product is an issue beyond the scope of this reference guide.

8. See Fed. R. Civ. P. 16(c), 26(a) (2)(B) (Supp. 1993); Black et al., *supra* note 2, at 791. We also think that a pretrial procedure used in England deserves consideration. In most cases, Order 38, Rule 37, like Fed. R. Civ. P. 26(a)(2)(B), demands that an expert produce a written report before trial. Evidence Rules, S.I. 1989, No. 2427, *reprinted in* The Supreme Court Practice 83 (6th Cum. Supp. 1988). But Order 38, Rule 38 goes beyond the Federal Rules in explicitly authorizing the judge to require the experts to participate in the pretrial identification of disputed issues. Evidence Rules, S.I. 1987, No. 1423, *reprinted in* The Supreme Court Practice, *supra* at 83–84. This rule allows the court to

direct that there be a meeting "without prejudice" of such experts . . . for the purpose of identifying those parts of their evidence which are in issue. Where such a meeting takes place the experts may prepare a joint statement indicating those parts of their evidence on which they are, and those on which they are not, in agreement. *Id.* 

9. See The Special Comm. on Empirical Data in Legal Decision Making, Recommendations on Pretrial Proceedings in Cases with Voluminous Data, *reprinted in* The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, app. F. When the parties are alerted before trial to the criticisms of their

## 4. Presenting expert statistical testimony

The most common format for the presentation of evidence at trial is sequential. The plaintiff's witnesses are called first, one by one, without interruption except for cross-examination, and testimony is in response to specific questions rather than by an extended narration. Although traditional, this structure is not compelled by the Federal Rules of Evidence.<sup>10</sup> Some alternatives have been proposed that might be more effective in cases involving substantial statistical testimony. For example, when the reports of witnesses go together, the judge might allow their presentations to be combined and the witnesses to be questioned as a panel rather than sequentially. More narrative testimony might be allowed, and the expert might be permitted to give a brief tutorial on statistics as a preliminary to some testimony. Instead of allowing the parties to present their experts in the midst of all the other evidence, the judge might call for the experts for opposing sides to testify at about the same time. Some courts, particularly in bench trials, may have both experts placed under oath and, in effect, permit them to engage in a dialogue. In such a format, experts are able to say whether they agree or disagree on specific issues. The judge and counsel can interject questions. Such practices may improve the judge's understanding and reduce the tensions associated with the experts' adversarial role.<sup>11</sup>

Statistics

data or analyses, it may be possible to determine whether the putative problems have much effect on the results. *E.g.*, David H. Kaye, *Improving Legal Statistics*, 24 Law & Soc'y Rev. 1255 (1990).

<sup>10.</sup> See Fed. R. Evid. 611.

<sup>11.</sup> The Evolving Role of Statistical Assessments as Evidence in the Courts, supra note 1, at 174.

This page left blank intentionally for proper pagination when printing two-sided

# **II.** How Have the Data Been Collected?

An analysis is only as good as the data on which it rests. Along with an examination of the statistical analysis, therefore, it is important to verify the quality of the data collection<sup>12</sup> and to identify its limitations.

# A. Individual Measurements

### 1. Is the measurement process reliable?

In science, *reliability* refers to reproducibility of results.<sup>13</sup> A reliable measuring instrument returns consistent measurements of the same quantity. A scale, for example, is reliable if it reports the same weight for the same object time and again. It may not be accurate—it may always report a weight that is too high or one that is too low—but the perfectly reliable scale always reports the same weight for the same object. Its errors, if any, are systematic; they always point in the same direction.

Reliability can be ascertained by repeatedly measuring the same quantity. The predominant method of DNA identification, for instance, requires laboratories to determine the molecular weight of fragments of DNA. By making duplicate measurements of the same fragments, laboratories can determine the likelihood that two measurements of the same fragment will differ by a specified amount. <sup>14</sup> Ascertaining the usual range of such random error is essential in deciding whether an observed discrepancy between a crime sample and a suspect's

<sup>12.</sup> For introductory treatments of data collection, see, e.g., Stephen K. Campbell, Flaws and Fallacies in Statistical Thinking (1974); David Freedman et al., Statistics (2d ed. 1991); Darrell Huff, How to Lie with Statistics (1954); Jeffrey Katzer et al., Evaluating Information: A Guide for Users of Social Science Research (2d ed. 1982); David S. Moore, Statistics: Concepts and Controversies (2d ed. 1985); Robert S. Reichard, The Figure Finaglers (1974); Richard P. Runyon, Winning with Statistics: A Painless First Look at Numbers, Ratios, Percentages, Means, and Inference (1977); Hans Zeisel, Say It with Figures (6th ed. 1985).

<sup>13.</sup> Courts often use *reliable* to mean "that which can be relied on" for some purpose, such as establishing probable cause or crediting a hearsay statement when the declarant is not produced for confrontation. Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2795 n.9 (1993), for instance, distinguishes "evidentiary reliability," or "trustworthiness," from "scientific reliability," or "consistent results." Here, we use the term in the latter sense to clarify the many components of ultimate trustworthiness.

<sup>14.</sup> See, e.g., B. Budowle et al., Fixed-Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci, for Use in Forensic Comparisons, 48 Am. J. Hum. Genet. 841 (1991); B. S. Weir & B. S. Gaut, Matching and Binning DNA Fragments in Forensic Science , 34 Jurimetrics J. 9 (1993).

sample is sufficient to exclude the suspect as a possible source of the crime sample.  $^{\rm 15}$ 

In some social science studies, researchers examine recorded information and characterize it. For instance, in a study of death sentencing in Georgia, legally trained evaluators examined short summaries of cases and ranked them according to the defendant's culpability. <sup>16</sup> Two different aspects of reliability are worth considering. First, the "within-observer" variability of judgments should be small—the same evaluator should rate essentially identical cases the same way. Second, the "between-observer" variability should be small—different evaluators should rate the same cases the same way.

#### 2. Is the measurement process valid?

Reliability is necessary, but not sufficient, to ensure accuracy. In addition to reliability, *validity* is needed. A valid measuring instrument measures what it is supposed to. Thus, a polygraph measures certain physiological responses to stimuli. It may accomplish this task reliably. Nevertheless, it is not valid as a lie detector unless increases in pulse rate, blood pressure, and the like are well correlated with conscious deception.

When an independent and highly accurate way of measuring the *variable*<sup>17</sup> of interest is available, it may be used to validate the measuring system in question. Breathalyzer readings may be validated against alcohol levels found in blood samples. Employment test scores may be validated against job performance. A common measure of validity is the *correlation coefficient* between the *criterion* (job performance) and the predictor (the test score).<sup>18</sup>

# 3. Are the measurements recorded correctly?

Judging the adequacy of data collection may involve examining the process by which measurements are recorded and preserved. Are responses to interviews coded and logged correctly? Are all the responses to a survey included? If gaps or mistakes are present, do they appear randomly so they do not distort the results?

Once it is shown that measurements are reliable, valid, and properly recorded, inferences can be made. The purpose of collecting and analyzing the data may be to describe something, such as the prevalence of a blood type, or it

<sup>15.</sup> Committee on DNA Technology in Forensic Science, National Research Council, DNA Technology in Forensic Science 61–62 (1992).

<sup>16.</sup> David C. Baldus et al., Equal Justice and the Death Penalty: A Legal and Empirical Analysis 49-50 (1990).

<sup>17.</sup> For present purposes, a variable is a numerical characteristic of units in a study. For instance, in a survey of people, the unit of analysis is the person, and variables might include income (in dollars per year) and educational level (years of schooling completed). In a study of school districts, the unit of analysis is the district, and variables might include average family income of residents and average test scores of students.

<sup>18.</sup> E.g., Washington v. Davis, 426 U.S. 229, 252 (1976); Albemarle Paper Co. v. Moody, 422 U.S. 405, 430–32 (1975). Sæ discussion of the correlation coefficient *infra* § III.F.2. Various statistics are used to characterize the reliability of laboratory instruments, psychological tests, or human judgments. These include the standard deviation (SD) as well as the correlation coefficient. See infra § III.E.

may be to investigate a question of cause and effect, such as the deterrent effect of capital punishment.

# B. Descriptive Surveys and Censuses

A census measures some characteristic of every unit in a *population* of individuals or objects. A survey, alternatively, measures characteristics only in part of a population. The accuracy of the information collected in a census or survey depends on how the units are selected, which units are actually measured, and how the measurements are made.

# 1. What method is used to select the units to be measured?

By definition, a census seeks to measure every unit in the population. It may fall short of this goal, in which case the question must be asked whether the missing data are likely to differ in some systematic way from the data that are collected. The U.S. Bureau of the Census estimates that the past six censuses failed to count everyone, and there is evidence that the undercount is greater in certain subgroups of the population. Supplemental studies may enable statisticians to adjust for such omissions, but the adjustments may rest on uncertain assumptions.<sup>19</sup>

The methodological framework of a scientific survey is more complicated than that of a census. In surveys that use random sampling methods, a *sampling frame*, that is, an explicit list of units in the population, is created. Individual units then are selected by a kind of lottery procedure, and measurements are made on these sampled units. For example, a defendant charged with a notorious crime who seeks a change of venue may commission an opinion poll to show that popular opinion is so adverse and deep-rooted that it will be difficult to impanel an unbiased jury. The population consists of all persons in the jurisdiction who might be called for jury duty. A sampling frame here could be the list of these persons as maintained by appropriate officials.<sup>20</sup> In this case, the fit between the sampling frame and the population would be excellent.

20. If the jury list is not compiled properly from appropriate sources, it might be subject to challenge. See David Kairys et al., Jury Representativeness: A Mandate for Multiple Source Lists , 65 Cal. L. Rev. 776 (1977).

<sup>19.</sup> For conflicting views on proposed adjustments to the 1990 census, see Stephen E. Fienberg, *The New* York City Census Adjustment Trial: Witness for the Plaintiffs, 34 Jurimetrics J. 65 (1993); David A. Freedman, Adjusting the Census of 1990, 34 Jurimetrics J. 99 (1993); John E. Rolph, *The Census Adjustment Trial:* Reflections of a Witness for the Plaintiffs , 34 Jurimetrics J. 85 (1993); Kenneth W. Wachter, *The Census Adjustment Trial:* An Exchange, 34 Jurimetrics J. 107 (1993). See also Symposium , Undercount in the 1990 Census , 88 J. Am. Stat. Ass'n 1044 (1993). Similarly, the courts are divided over the legal standard governing claims that adjustment is statutorily or constitutionally compelled. *Compare* New York City v. United States Dep't of Commerce, 63 U.S.L.W. 2128 (2d Cir. 1994) (equal protection clause requires government to show compelling interest that could justify Secretary of Commerce's refusal to adjust 1990 census) with City of Detroit v. Franklin, 4 F.3d 1367 (6th Cir. 1993) (neither statutes nor constitution requires adjustment), *cert. denied*, 114 S. Ct. 1212 (1994); Tucker v. Dep't of Commerce, 958 F.2d 1411 (7th Cir.) (issue is not justicia-ble), *cert. denied*, 113 S. Ct. 407 (1992).

In other situations, the sampling frame may cover less of the population. In an obscenity case, for example, the defendant's poll of opinion about community standards<sup>21</sup> should identify all adults in the legally relevant community as the population, but obtaining the names of all such people may not be possible. If names from a telephone directory are used, people with unlisted numbers are excluded from the sampling frame. If these people, as a group, hold different opinions from those included in the sampling frame, the poll will not reflect this difference, no matter how many individuals are polled and no matter how well their opinions are elicited.<sup>22</sup> The poll's measurement of community opinion will be biased, although the magnitude of this bias may not be great.

Not all surveys use random selection. In some commercial disputes involving trademarks or advertising, the population of all potential purchasers of the products is difficult to identify. Some surveyors may resort to an easily accessible subgroup of the population, such as shoppers in a mall.<sup>23</sup> Such *convenience sam ples* may be biased by the interviewer's discretion in deciding whom to interview—a form of selection bias—and the refusal of some of those approached to participate—nonresponse bias.<sup>24</sup> Selection bias is acute when constituents write their representatives, listeners call into radio talk shows, or interest groups collect information from their members.<sup>25</sup> Selection bias also affects data from juryreporting services that gather information from readily available sources.<sup>26</sup>

22. A classic example of selection bias is the 1936 *Literary Digest* poll. After successfully predicting the winner of every U.S. presidential election since 1916, the *Digest* used the replies from 2.4 million respondents to predict that Alf Landon would win 57% to 43%. In fact, Franklin Roosevelt won by a landslide vote of 62% to 38%. See Freedman et al., *supra* note 12, at 306. The *Digest* was so far off, in part, because it chose names from telephone books, rosters of clubs and associations, city directories, lists of registered voters, and mail order listings. *Id.* at 306–08, A-13 n.6. In 1936, when only one household in four had a telephone, the people whose names appeared on such lists tended to be more affluent. Lists that overrepresented the affluent had worked well for sampling in earlier elections, when rich and poor voted along similar lines, but the bias in the sampling frame proved fatal when the Great Depression made economics a salient consideration for voters. *See* Judith M. Tanur, *Samples and Surveys*, *in* Perspectives on Contemporary Statistics 55, 57 (David C. Hoaglin & David S. Moore eds., 1992). Today, survey organizations conduct polls by telephone, but most voters have telephones, and these organizations select the numbers to call at random rather than sampling names from telephone books.

23. E.g., R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc., 511 F. Supp. 867, 876 (S.D.N.Y. 1980) (questioning the propriety of basing a "nationally projectable statistical percentage" on a suburban mall intercept study).

24. Nonresponse bias is discussed infra § II.B.2.

25. E.g., Pittsburgh Press Club v. United States, 579 F.2d 751, 759 (3d Cir. 1978) (tax-exempt club's mail survey of its members to show little sponsorship of income-producing uses of facilities was held to be inadmissible hearsay because it "was neither objective, scientific nor impartial"), *rev'd on other grounds*, 615 F.2d 600 (3d Cir. 1980). So, too, veterans groups collected instances of multiple myeloma (a form of cancer) among veterans of the Hiroshima and Nagasaki occupation forces. They claimed that the number of cases was un usual and called for government study and compensation. Such anecdotal evidence, based on a few cases without systematic comparison or data collection, may be an incentive for more careful investigation but may also reflect rumor and speculation rather than fact. In this instance, a committee of the National Research

<sup>21.</sup> On the admissibility of such polls, compare, *e.g.*, Saliba v. State, 475 N.E.2d 1181, 1187 (Ind. Ct. App. 1985) ("Although the poll did not . . . [ask] the interviewees . . . whether the particular film was obscene, the poll was relevant to an application of community standards") with United States v. Pryba, 900 F.2d 748, 757 (4th Cir.) ("Asking a person in a telephone interview as to whether one is offended by nudity, is a far cry from showing the materials . . . and then asking if they are offensive," so exclusion of the survey results was proper), *cert. denied*, 498 U.S. 924 (1990).

Various procedures are available to cope with selection bias. In quota sampling, the interviewer is instructed to interview so many women, so many older men, so many ethnic minorities, or the like. But quotas alone still leave vast discretion in selecting among the members of each category and therefore do not solve the problem of selection bias.

Probability sampling methods, in contrast, ideally are suited to avoid selection bias. Once the conceptual population is reduced to a tangible sampling frame, the units to be measured are selected by some kind of lottery that gives each unit in the sampling frame a known, nonzero probability of being chosen. Selection according to a table of random digits or the like<sup>27</sup> leaves no room for selection bias.<sup>28</sup>

#### 2. Of the units selected, which are measured?

Although probability sampling ensures that, within the limits of chance, the sample of units selected will be representative of the sampling frame, the question remains as to which units actually get measured. When objects like receipts (for an audit) or vegetation (for a study of the ecology of a region) are sampled, all can be examined. Human beings are more troublesome. Some may refuse to respond, and the survey should report the nonresponse rate. A large nonresponse rate warns of bias, <sup>29</sup> but it does not necessarily demonstrate bias. Supplemental

27. In simple random sampling, each unit has the same probability of being chosen. More complicated methods, such as stratified sampling and cluster sampling, have advantages in certain applications. In systematic sampling, every fifth, tenth, or hundredth (in mathematical jargon, every *n*th) unit in the sampling frame is selected. If the starting point is selected at random and the units are not in any special order, then this procedure is comparable to simple random sampling.

28. Before 1968, most federal districts used the "key man" system for compiling lists of eligible jurors. Individuals believed to have extensive contacts in the community would suggest names of prospective jurors, and the qualified jury wheel would be made up from those names. To reduce the risk of discrimination associated with this system, the Jury Selection and Service Act of 1968, 28 U.S.C. §§ 1861–1878 (1988), substituted the principle of "random selection of juror names from the voter lists of the district or division in which court is held." S. Rep. No. 891, 90th Cong., 1st Sess. 10 (1967), reprinted in 1968 U.S.C.C.A.N. 1792, 1793.

29. The 1936 Literary Digest election poll illustrates the danger. Only 24% of the 10 million people who received questionnaires returned them. Most of these respondents probably had strong views on the candidates, and most of them probably objected to President Roosevelt's innovative economic programs. This self-selection is likely to have biased the poll. Maurice C. Bryson, The Literary Digest Poll: Making of a Statistical Myth, 30 Am. Statistician 184 (1976); Freedman et al., supra note 12, at 307–08.

In United States v. Gometz, 730 F.2d 475, 478 (7th Cir.) (en banc), cert. denied, 469 U.S. 845 (1984), the Seventh Circuit recognized that "a low rate of response to juror questionnaires could lead to the underrepresentation of a group that is entitled to be represented on the qualified jury wheel." Nevertheless, the court held that under the Jury Selection and Service Act of 1968, 28 U.S.C. §§ 1861–1878 (1988), the clerk did not abuse his discretion by failing to take steps to increase a response rate of 30%. According to the court, "Congress wanted to make it possible for all qualified persons to serve on juries, which is different from forcing all qualified persons to be available for jury service." *Gometz*, 730 F.2d at 480. Although it might "be a good thing to

Council found no evidence that the rate of multiple myeloma for the "atomic veterans" was higher than that in similar populations. Moore, *supra* note 12, at 124.

<sup>26.</sup> For example, a study from the mid-1980s found that the average award in medical malpractice cases was \$962,258. The figure comes from a jury-reporting service that relies on newspaper accounts and other sources that are likely to report predominantly large awards. Kenneth Jost, *Still Warring Over Medical Malpractice*, A.B.A. J., May 1993, at 68, 71 (citing an interview with Neil Vidmar). On the limitations of jury verdict service and other reports of jury awards, see, *eg.*, Theodore Eisenberg & Thomas A. Henderson, Jr., *Inside the Quiet Revolution in Products Liability*, 39 UCLA L. Rev. 731, 765 n.100 (1992).

study may establish that the nonrespondents do not differ systematically from the respondents with respect to the characteristics of interest<sup>30</sup> or may permit the missing data to be imputed.<sup>31</sup>

In short, a convincing survey defines an appropriate population to study, uses an unbiased method for selecting units to measure, with a reliable and valid procedure for gathering information on the units selected for study, and succeeds in gathering information on all or a fair cross section of these units. When these goals are met, the sample tends to be representative of the population: The measurements within the sample describe fairly the characteristics in the population. It remains possible, however, that despite every precaution, the sample, being less than exhaustive, is not representative; proper statistical analysis helps address the magnitude of this risk, at least for probability samples.<sup>32</sup> Of course, surveys may be useful even if they fail to meet all of the criteria given above; but then, additional arguments are needed to justify the inferences.

# C. Experiments

In many cases, the court needs more than a description of a population. It seeks an answer to a question of causation. Would additional information in a securities prospectus disclosure have caused potential investors to behave any differently? Does the similarity in the names of two products lead consumers to buy one brand because of their familiarity with the other brand? Does capital punishment deter crime? Do food additives cause cancer?

follow up on persons who do not respond to a jury questionnaire," the court concluded that Congress merely "wanted to make it possible for all qualified persons to serve on juries" and "was not concerned with anything so esoteric as nonresponse bias." *Id.* at 479, 480, 482.

30. Even when demographic characteristics of the sample match those of the population, however, caution still is indicated. In the 1980s, a behavioral researcher sent out 100,000 questionnaires to explore how women viewed their relationships with men. Shere Hite, Women and Love: A Cultural Revolution in Progress (1987). She amassed a huge collection of anonymous letters from thousands of women disillusioned with love and marriage, and she wrote that these responses established that the "outcry" of some feminists "against the many injustices of marriage—exploitation of women financially, physically, sexually, and emotionally" is "just and accurate." *Id.* at 344. The outcry may indeed be justified, but this research does little to prove the point. About 95% of the 100,000 inquiries did not produce responses. The nonrespondents may have had less distressing experiences with men and therefore did not see the need to write autobiographical letters. Furthermore, this systematic difference would be expected within every demographic and occupational class. Therefore, the argument that the sample responses are representative because "those participating according to age, occupation, religion, and other variables known for the U.S. population at large in most cases quite closely mirrors that of the U.S. female population" is far from convincing. *Id.* at 777. In fact, the results of this non random sample differ dramatically from those of polls with better response rates. *See* Chamont Wang, Sense and Nonsense of Statistical Inference: Controversy, Misuse, and Subtlety 174–76 (1993). *See also* David Streitfeld, *Shere Hite and the Trouble with Numbers*, 1 Chance 26 (1988).

31. Methods for "imputing" missing data are sketched in, *e.g.*, Tanur, *supra* note 22, at 55. For more technical references, see, *e.g.*, Donald B. Rubin, Multiple Imputation for Nonresponse in Surveys (1987); Imputation and Editing of Faulty or Missing Survey Data (Faye Aziz & Fritz Scheuren eds., 1978). Efforts to fill in missing data can be problematic. The "easy case" is one in which the response rate is so high that even if all nonrespondents had responded in a way adverse to the proponent of the survey, the substantive conclusion would be unaltered.

32. See infra § IV.

Controlled experiments are, far and away, the best vehicle for establishing a causal relationship. Such experiments may exist before the commencement of the litigation. If so, it becomes the task of the lawyer and appropriate experts to explain this research to the court. Examples of such "off-the-shelf" research are experiments pinpointing conditions under which eyewitnesses tend to err in identifying criminals<sup>33</sup> or studies of how sex stereotyping affects perceptions of women in the workplace.<sup>34</sup> Even if no preexisting studies are available, a case-specific one may be devised:<sup>35</sup> A psychologist may simulate the conditions of a particular eyewitness's identification to see whether comparable identifications tend to be correct;<sup>36</sup> an organization investigating racial discrimination in the rental-housing market may send several "testers" (who, it is hoped, differ only in their race) to rent a property.<sup>37</sup>

A well-designed experiment shows how one variable responds to changes in variables under the control of the experimenter. Variables not directly controlled should be subject only to random fluctuations. For example, to verify that a fertilizer improves crop yields, it is insufficient only to report that the yield is high in a fertilized field. It may be that the yield would have been higher without the fertilizer. To compare the outcome with fertilizer to the outcome without fertilizer, two essentially identical fields can be planted, and fertilizer can be applied only to one field. If the conditions in the fields are nearly identical, any large difference in the yields must be the result of fertilizer. By definition, other possible causes have been eliminated.

To the extent that the two fields are not truly identical, but differ in a myriad of ways that are hard to specify but could affect the yield, the experiment may be replicated on many fields randomly assigned to be fertilized or not.<sup>38</sup> These strategies of control and randomization are the earmarks of good experiments.

33. E.g., State v. Chapple, 660 P.2d 1208, 1223–24 (Ariz. 1983) (reversing a conviction for excluding expert testimony about scientific research on eyewitness testimony). For citations to the case law and scientific literature, see, eg., 1 McCormick on Evidence § 206(A) (John William Strong ed., 4th ed. 1992).

35. For a review of the law on such pretrial experiments and a proposal that the parties be encouraged to cooperate in the design of such experiments, see 1 McCormick on Evidence, *supra* note 33, § 202.

36. Willem A. Wagenaar, The Proper Seat: A Bayesian Discussion of the Position of the Expert Witness, 12 Law & Hum. Behav. 499 (1988) (describing the difficulty of presenting the results of such an experiment to a court in the Netherlands).

<sup>34.</sup> The testimony of a social psychologist about stereotyping played a limited—and controversial—role in Price Waterhouse v. Hopkins, 490 U.S. 228, 235 (1989). Compare Gerald V. Barrett & Scott B. Morris, The American Psychological Association's Amicus Curiae Brief in Price Waterhouse v. Hopkins : The Values of Science Versus the Values of the Law , 17 Law & Hum. Behav. 201 (1993) with Susan T. Fiske et al., What Constitutes a Scientific Review?: A Majority Retort to Barrett and Morris, 17 Law & Hum. Behav. 217 (1993).

<sup>37.</sup> E.g., United States v. Youritan Constr. Co., 370 F. Supp. 643, 647 (N.D. Cal. 1973), aff'd in part, 509 F.2d 623 (9th Cir. 1975).

<sup>38.</sup> Statistically significant differences are those that are so large that they rarely would occur with an ineffectual fertilizer just because the fields randomly selected for the fertilizer treatment happen to be the best for growth. The techniques for establishing statistical significance are considered *infra* § IV.

### 1. What are the independent and dependent variables?

In investigating a possible cause-and-effect relationship, the variable that characterizes the effect is called the *dependent variable*, since it may depend on the causes.<sup>39</sup> In contrast, the variables that represent the causes are called *independent variables*.<sup>40</sup> In the fertilizer experiment, crop yield is the dependent variable. It depends on such independent variables as the density of planting, the level of irrigation or rainfall, the nature of the soil, and the extent of insect infestation. Listing such variables is a useful exercise because it focuses attention on which factors are under control (and can be excluded as causes of the observed differences) and which are not (and may mask a causal relation or give a false appearance of one).

### 2. What are the confounding variables?

A confounding variable is correlated with the independent variables and with the dependent variable. Since a confounding variable changes with one or more independent variables, it is generally not possible to determine whether changes in the independent variables caused changes in the dependent variable or whether changes in the confounding variable did—especially if the investigator did not collect data on the confounder. For example, many studies have been conducted to determine whether physical exercise increases life span. In one such study, the physical fitness of a large number of men was measured. Over the next sixteen years, about twice as many men in the lowest fitness quartile died as did men in the highest quartile.<sup>41</sup> One interpretation is that maintaining a high level of physical activity protects against death. However, both physical fitness and mortality are correlated with general health at the beginning of the study; thus, it is possible that the highly fit men lived longer, not because they exercised, but simply because they were healthier to begin with. <sup>42</sup> A disproportion ate number of healthier men in the high fitness group biases the study in favor of finding improved survival in that group.

Randomly assigning subjects to a treatment and a control group eliminates this problem.<sup>43</sup> In experiments on human beings, it is especially difficult to ensure that the treatment and control groups are identical, but with random selection the many factors not under the experimenter's control tend to balance out

<sup>39.</sup> Dependent variables also may be called response variables.

<sup>40.</sup> Independent variables also may be called factors or explanatory variables.

<sup>41.</sup> Leiv Sandvik et al., Physical Fitness as a Predictor of Mortality Among Healthy, Middle-Aged Norwegian Man, 328 New Eng. J. Med. 533 (1993). The lower mortality among the fittest men was attributable chiefly to a lower risk of dying of cardiovascular causes.

<sup>42.</sup> Gregory D. Curfman, The Health Benefits of Exercise: A Critical Reappraisal, 328 New Eng. J. Med. 574, 575 (1993).

<sup>43. &</sup>quot;A randomized, controlled trial of physical activity for the primary prevention of cardiovascular disease . . . has never been performed and is probably not feasible because of problems related to compliance and cost." *Id. But see* M. A. Fiatarone et al., *Exercise Training and Nutritional Supplementation for Physical Frailty in Very Elderly People*, 330 New Eng. J. Med. 1769 (1994).

in both groups.<sup>44</sup> A handful of well-designed studies is far more convincing than any number of biased ones.

Two insights are important. First, outcome figures from a treatment group without a control group reveal very little and can be misleading.<sup>45</sup> Comparisons are essential. Second, if the control group was obtained through random assignment before treatment, a difference in the outcomes between treatment and control groups may be accepted, within the limits of statistical error, as the true measure of the treatment effect.<sup>46</sup> However, if the control group was created in any other way, differences in the groups that existed before treatment may contribute to differences in the outcomes or mask differences that otherwise would be observed. Thus, observational studies succeed to the extent that their treatment and control groups are comparable, apart from the treatment.

### 3. Can the results be generalized?

All experiments are conducted with a sample of a certain population, at a certain place, at a certain time, and with a limited number of treatments. With respect to the sample studied, the experiment may be persuasive. It may have succeeded in controlling all confounding variables and in finding an unequivocally large difference between the treatment and control groups. If so, its "internal validity" will not be disputed; in the sample studied, the treatment has an effect.

But an issue of "external validity" remains. To extrapolate from the limiting conditions of an experiment always raises questions. If juries react differently to competing instructions on the law of insanity in cases of housebreaking and of incest, <sup>47</sup> would the difference persist if the charge were rape or murder? Would the failure of ex -convicts to react to transitory payments after release hold if conditions in the employment market were to change radically?<sup>48</sup>

Confidence in the appropriateness of an extrapolation cannot come from the experiment itself. It must come from knowledge about which outside factors

45. For an effort to identify circumstances in which such studies may be informative, see John C. Bailar III et al., Studies Without Internal Controls, in Medical Uses of Statistics, supra note 44, at 105.

46. The problem of statistical error is treated infra § IV.

<sup>44.</sup> Of course, the possibility that the two groups will not be comparable in some unrecognized way can never be eliminated. Random assignment, however, allows the researcher to compute the probability of seeing a large difference in the outcomes when the treatment actually has no effect. When this probability is small, the difference in the response is said to be "statistically significant." See infra § IV.B.2.

Randomization also ensures that the assignment of subjects to treatment and control groups is free from conscious or unconscious manipulation by investigators or subjects. Randomization may not be the only way to ensure such protection, but "it is the simplest and best understood way to certify that one has done so." Philip W. Lavori et al., Designs for Experiments—Parallel Comparisons of Treatment, in Medical Uses of Statistics 61, 66 (John C. Bailar III & Frederick Mosteller eds., 2d ed. 1992). To avoid ambiguity, the researcher should be explicit "about how the randomization was done (e.g., table of random numbers) and executed (e.g., by sealed envelopes prepared in advance)." Id.

<sup>47.</sup> See Rita James Simon, The Jury and the Defense of Insanity 58-59 (1967).

<sup>48.</sup> For an experiment on income support and recidivism, see Peter H. Rossi et al., Money, Work, and Crime: Experimental Evidence (1980). The interpretation of the data has proved controversial. See Hans Zeisel, Disagreement over the Evaluation of a Controlled Experiment, 88 Am. J. Soc. 378 (1982) (with commentary).

would or would not affect the outcome.<sup>49</sup> Sometimes, several experiments or other studies, each having different limitations, all point in the same direction.<sup>50</sup> Such convergent results strongly suggest the validity of the generalization.<sup>51</sup>

# D. Observational Studies of Causation

The bulk of the statistical studies seen in court are observational, not experimental. In an experiment the investigators select certain units for treatment. In an *observational study* the investigators have no control over who or what receives the treatment. Take the question of whether capital punishment deters murder. To do a randomized controlled experiment, people would have to be assigned randomly to a control group and a treatment group. The controls would know that they could not receive the death penalty for murder, while those in the treatment group would know they could be executed. The rate of subsequent murders by the subjects in these groups would be observed. Such an experiment is unacceptable—politically, ethically, and legally.<sup>52</sup>

Nevertheless, many studies of the deterrent effect of the death penalty have been conducted, all observational, and some have attracted judicial attention.<sup>53</sup> Researchers have catalogued differences in the incidence of murder in states with and without the death penalty, and they have analyzed changes in homicide rates and execution rates over the years. In such observational studies, investigators may speak of control groups (such as the states without capital punish-

50. This is the case, for example, with eight studies indicating that jurors who approve of the death penalty are more likely to convict in a capital case. Phoebe C. Ellsworth, *Some Steps Between Attitudes and Verdicts, in* Inside the Juror 42, 46 (Reid Hastie ed., 1993). Nevertheless, in Lockhart v. McCree, 476 U.S. 162 (1986), the Supreme Court held that the exclusion of opponents of the death penalty in the guilt phase of a capital trial does not violate the constitutional requirement of an impartial jury.

51. See Zeisel, supra note 12, at 252-62.

52. Cf. Experimentation in the Law: Report of the Federal Judicial Center Advisory Committee on Experimentation in the Law (Federal Judicial Center 1981) [hereinafter Experimentation in the Law] (study of ethical issues raised by controlled experimentation in the evaluation of innovations in the justice system).

53. See generally Hans Zeisel, The Deterrent Effect of the Death Penalty: Facts v. Faith, 1976 Sup. Ct. Rev. 317.

<sup>49.</sup> Such judgments are easiest in the natural sciences, but even here, there are problems. For example, it may be difficult to infer human reactions to substances that affect animals. First, there are inconsistencies across test species: A chemical may be carcinogenic in mice but not in rats. Extrapolation from rodents to humans is even more problematic. Second, to get measurable effects in animal experiments, chemicals are administered at very high doses. Results are extrapolated-using mathematical models-to the very low doses of concern in humans. However, there are many dose-response models to use and few grounds for choosing among them. Generally, different models produce radically different estimates of the "virtually safe dose" in humans. David A. Freedman & Hans Zeisel, From Mouse to Man: The Quantitative Assessment of Cancer Risks, 3 Stat. Sci. 3 (1988). For these reasons, many experts—and some courts in toxic tort cases—have concluded that evidence from animal experiments is generally insufficient to establish causation. Sæ Michael D. Green, Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation, 86 Nw. U. L. Rev. 643 (1992); Lois S. Gold et al., Rodent Carcinogens: Setting Priorities, 258 Science 261 (1992); D. Krewski et al., A Model-Free Approach to Low-Dose Extrapolation, 90 Envtl. Health Persp. 279 (1991); Susan R. Poulter, Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?, 7 High Tech. L.J. 189 (1993) (epidemiological evidence on humans is needed). See also Committee on Risk Assessment Methodology, National Research Council, Issues in Risk Assessment (1993).

ment) and of controlling for potentially confounding variables (e.g., worsening economic conditions).<sup>54</sup> However, association is not causation, and the causal inferences that can be drawn from such analyses rest on a less secure foundation than that provided by a controlled randomized experiment.<sup>55</sup>

Of course, observational studies can be very useful. The evidence that smoking causes lung cancer in humans, although largely observational, is compelling. In general, observational studies provide powerful evidence in the following circumstances.

- The association is seen in studies of different types among different groups. This reduces the chance that the observed association is due to a defect in one type of study or a peculiarity in one group of subjects.
- The association holds when the effects of plausible confounding variables are taken into account by appropriate statistical techniques, such as comparing smaller groups that are relatively homogeneous with respect to the factor. <sup>56</sup>

54. A procedure often used to control for confounding in observational studies is regression analysis. The underlying logic is described *infra* § III.F.3. The early enthusiasm for using multiple regression analysis to study the death penalty was not shared by reviewers. *Compare* Isaac Ehrlich, *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, 65 Am. Econ. Rev. 397 (1975) *with, e.g., Lawrence* R. Klein et al., *The Deterrent Effect of Capital Punishment: An Assessment of the Estimates, in Panel on Research on Deterrent and Incapacitative Effects, National Research Council, Deterrence and Incapacitation: Estimating the Effects of <i>Criminal Sanctions on Crime Rates* 336 (Alfred Blumstein et al. eds., 1978); Edward Leamer, *Let's Take the Con Out of Econometrics ,* 73 Am. Econ. Rev. 31 (1983); Richard O. Lempert, *Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment, 79 Mich. L. Rev.* 1177 (1981); Richard O. Lempert, *The Effect of Executions on Homicides: A New Look in an Old Light,* 29 Crime & Delinq. 88 (1983). 55. See, e.g., Experimentation in the Law, supra note 52, at 18:

[G]roups selected without randomization will [almost] always differ in some systematic way other than exposure to the experimental program. Statistical techniques can eliminate chance as a feasible explanation for the differences, ... [b]ut without randomization there are no certain methods for determining that observed differences between groups are not related to the preexisting, systematic difference ... [C]omparison between systematically different groups will yield ambiguous implications whenever the systematic difference affords a plausible explanation for apparent effects of the experimental program.

56. The idea is to control for the influence of a confounder by making comparisons separately within groups for which the confounding variable is nearly constant and therefore has little influence over the variables of primary interest. For example, smokers are more likely to get lung cancer than nonsmokers. Age, gender, social class, and region of residence are all confounders, but controlling for such variables does not really change the relationship between smoking and cancer rates. On the basis of observational studies, most experts believe that smoking does cause lung cancer (and many other diseases). For a recent review of the literature, see 38 International Agency for Research on Cancer (IARC), World Health Org., IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Tobacco Smoking (1986). However, the associations seen in observational studies, even good ones, can be misleading. For example, women with herpes are more likely to develop cervical cancer than women who have not been exposed to the virus. For a time, it was believed that herpes caused cancer. In other words, the association was thought to be causal. Later research suggests that herpes is only a marker of sexual activity. Women who have had multiple sexual partners are more likely to be exposed not only to herpes but also to human papilloma virus. Certain strains of papilloma virus seem to cause cervical cancer, while herpes does not. Apparently, the association between herpes and cervical cancer is not causal but is due to the effect of other variables. See Viral Etiology of Cervical Cancer (Richard Peto & Harald zur Hausen eds., 1986); The Epidemiology of Human Papillomavirus and Cervical Cancer (N. Muñoz et al. eds., 1992). For additional examples and discussion, see Freedman et al., supra note 12, at 11-25, 133-48.

There is a plausible explanation for the effect of the independent variables; thus, the causal link does not depend on the observed association alone. Other explanations linking the response to confounding variables should be less plausible.<sup>57</sup>

When these criteria are not fulfilled, observational studies may produce legitimate disagreement among experts, and there is no mechanical procedure for ascertaining who is correct. In the end, deciding whether associations are causal is not a matter of statistics, but a matter of good scientific judgment, and the questions that should be asked with respect to data offered on the question of causation can be summarized as follows:

- Was there a control group? If not, the study has little to say about causation.
- If there was a control group, how were subjects assigned to treatment or control: through a process under the control of the investigator (a controlled experiment) or a process outside the control of the investigator (an observational study)?
- If it was a controlled experiment, was the assignment made using a chance mechanism (randomization), or did it depend on the judgment of the investigator?
- If the data came from an observational study or a nonrandomized controlled experiment, how did the subjects come to be in treatment or in control groups? Are the groups comparable? What factors are confounded with treatment? What adjustments were made to take care of confounding? Were they sensible?<sup>58</sup>

<sup>57.</sup> David S. Moore & George P. McCabe, Introduction to the Practice of Statistics 202 (2d ed. 1993).

<sup>58.</sup> These questions are adapted from Freedman et al., *supra* note 12, at 25. As with controlled experiments, chance variation sometimes produces an apparent association between variables when none really exists (see infra § IV).

# **III.** How Have the Data Been Presented?

After data have been collected, they should be presented in a way that makes them intelligible and revealing. Huge quantities of data can be summarized with a few numbers or with graphical displays. However, the wrong summary or a distorted graph can mislead.<sup>59</sup>

# A. Is the Data Display Sufficiently Complete?

Selective presentation of numerical information is like quoting someone out of context. A television commercial for the Investment Company Institute (the mutual fund trade association) said that a \$10,000 investment made in 1950 in an average common stock mutual fund would have increased to \$113,500 by the end of 1972. The *Wall Street Journal* indicated that the same investment spread over all the stocks making up the New York Stock Exchange Composite Index would have grown to \$151,427. Mutual funds performed worse than the stock market as a whole.<sup>60</sup> In this example, and in many other situations, it is helpful to look beyond a single number to some comparison or benchmark that places the isolated figure into perspective.

Even complete and accurate data can mislead if changes in the process of collecting the data are not reported. For example, the number of petty larcenies reported in Chicago more than doubled between 1959 and 1960—not because of an abrupt crime wave—but because a new police commissioner introduced an improved reporting system. <sup>61</sup> For many years, researchers ignored New York City crime statistics because it was common practice for the precincts to underreport crime to protect the reputations of their neighborhoods. When New York City shifted to a centralized reporting system, burglary reports increased more than fourteenfold in three years. <sup>62</sup> During the 1970s, police officials in Washington, D.C., "demonstrated" the success of President Nixon's law-and-order campaign by valuing stolen goods at \$49, just below the \$50 threshold for inclusion in the Federal Bureau of Investigation's (FBI) Uniform Crime

<sup>59.</sup> See generally Campbell, supra note 12; Freedman et al., supra note 12; Huff, supra note 12; Katzer et al., supra note 12; Moore, supra note 12; Runyon, supra note 12; Zeisel, supra note 12.

<sup>60.</sup> Moore, supra note 12, at 128.

<sup>61.</sup> Id. at 129.

<sup>62.</sup> Mark H. Maier, The Data Game: Controversies in Social Science Statistics 80-81 (1991).

Reports. <sup>63</sup> Likewise, in the mid-1970s, the Indianapolis police department tripled the number of crime reports deemed "without merit," which hence went uncounted in the Uniform Crime Reports. <sup>64</sup>

Changes in the collection of data over the years are by no means limited to crime statistics. In 1971, President Nixon signed the National Cancer Act, calling for a war on cancer of the "same kind of concentrated effort that split the atom and took man to the moon."<sup>65</sup> Two decades and hundreds of billions of dollars later, advocates of the war on cancer recognize that no general cure is close at hand. They are encouraged, however, by the development of cures for some cancers, and they cite improved survival rates for other cancers. Some epidemiologists question the inference that the changes in survival rates reflect a successful assault on the disease. Because some kinds of cancers now are detected earlier, patients with these cancers merely appear to live longer.<sup>66</sup>

Almost all series of numbers that cover many years are affected by changes in definitions and collection methods. When considering time series data, it is worth looking for any sudden jumps, which may signal a change in definitions or data collection procedures.<sup>67</sup>

63. James P. Levine et al., Criminal Justice in America: Law in Action 99 (1986); Maier, *supra* note 62, at 81.

64. Maier, supra note 62, at 81; Harold E. Pepinsky & Paul Jesilow, Myths That Cause Crime 28 (1985).

65. As quoted in Ralph W. Moss, The Cancer Syndrome 16 (1980). See also Richard M. Nixon, Acting Against Cancer, Sat. Evening Post, July/Aug. 1986, at 67.

66. See Maier, supra note 62, at 55; James E. Enstrom & Donald F. Austin, Interpreting Cancer Survival Rates, 195 Science 847 (1977); Cancer: Illusory Progress?, Sci. Am., June 1987, at 29. For a more recent discussion of the difficulties in interpreting trends in incidence and death rates for cancers, see Tim Beardsley, A War Not Won, Sci. Am., Jan. 1994, at 130; National Cancer Inst., Evaluating the National Cancer Program: An Ongoing Process (1994) (transcript of the President's Cancer Panel meeting, Sept. 23, 1993, on file with the National Cancer Institute).

67. Moore, *supra* note 12, at 129. Another problem can arise from collapsing categories in a table. In Philip Morris, Inc. v. Loew's Theatres, Inc., 511 F. Supp. 855 (S.D.N.Y. 1980), and R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc., 511 F. Supp. 867 (S.D.N.Y. 1980), Philip Morris and R.J. Reynolds sought an injunction to stop the maker of Triumph low-tar cigarettes from running advertisements claiming that participants in a national taste test preferred Triumph to other brands. Plaintiffs alleged that claims that Triumph was a "national taste test winner" or Triumph "beats" other brands were false and misleading. An exhibit introduced by defendant contained the following data:

				Triumph some -	Triumph
	Triumph much	Triumph somewhat	Triumph about the	what worse than	much worse
	better than Merit	better than Merit	same as Merit	Merit	than Merit
Number	45	73	77	93	36
Percentage	14%	22%	24%	29%	11%

511 F. Supp. at 866. Only 14% + 22% = 36% of the sample preferred Triumph to Merit, while 29% + 11% = 40% preferred Merit to Triumph. *ld*. at 856. By selectively combining categories, however, defendant attempted to create a different impression. Since 24% found the brands about the same, and 36\% preferred Triumph, defendant claimed that a clear majority (36% + 24% = 60%) found Triumph "as good or better than Merit." *ld*. at 866. The court correctly resisted this chicanery, finding that defendant's test results did not support the advertising claims. *ld*. at 856–57. The statistical issues in these cases are discussed more fully in 2 Gastwirth, *supra* note 1, at 633–39. For a hypothetical, but strikingly similar example of selective collapsing of categories, see Richard P. Runyon, How Numbers Lie: A Consumer's Guide to the Fine Art of Numerical Deception 67–70 (1981).

Few summaries of data are intended to mislead; most try to bring out broad features of the data. All descriptive statistics, however, are simplifications, and there are times when the details they omit are important. The statistical analyst should be able to explain why the summary statistics used are sufficient to capture the relevant aspects of the data. For instance, the proportion of applicants who pass an entrance examination for a police academy is sufficient to indicate how significant a barrier the test is for that group of tested individuals. For this purpose, it is not necessary to know how each individual scored.<sup>68</sup> In other situations, a graph may reveal a pattern not evident from the summary statistic.<sup>69</sup>

# B. Are Rates or Percentages Properly Interpreted?

Rates and percentages effectively summarize data, but these statistics can be misinterpreted. A percentage is a summary that makes a comparison between two numbers. One number is the base, and the other number is compared with that base. When the base is small, actual numbers may be more revealing than percentages. For example, there were media accounts in 1982 of a crime wave by the elderly. The annual Uniform Crime Reports showed a near tripling of the crime rate by older people since 1964, while crimes by younger people only doubled. But people over 65 years of age account for less than 1% of all arrests. In 1980, for instance, there were only 151 arrests of the elderly for robbery out of 139,476 total robbery arrests.<sup>70</sup>

Usually, the small-base problem is obvious if the presentation is reasonably complete. An expert who says that 50% of the people interviewed had a certain opinion also should reveal how many individuals were contacted and how many expressed an opinion.<sup>71</sup> Then we know whether the 50% is 2 out of 4 or 500 out of 1,000.

Finally, there is the issue of which numbers to compare.<sup>72</sup> Researchers sometimes choose among alternative comparisons. It may be worthwhile to ask why they chose the one they did. Does it give a fair picture, or would another comparison give a different view? A government agency, for example, may want to compare the amount of service being given this year with that of earlier years—but what earlier year ought to be the baseline? If the first year of operation is used, a large percentage increase due to start-up problems for a new

<sup>68.</sup> If the analyst wants to examine the effect of the test on different subgroups, the proportions in each relevant subgroup must be considered. See, eg., Bouman v. Block, 940 F.2d 1211 (9th Cir.), cert. denied, 112 S. Ct. 640 (1991); 1 Gastwirth, supra note 1, at 254–55.

<sup>69.</sup> See infra § III.C.3.

<sup>70.</sup> Maier, supra note 62, at 83. See also Alfred Blumstein & Jacqueline Cohen, Characterizing Criminal Careers, 237 Science 985 (1987).

<sup>71.</sup> For a poll suggesting that male and female trial attorneys have different impressions of the behavior of male and female litigators but omitting the number of respondents by category, see Stephanie B. Goldberg, "Good Girl" Litigators, A.B.A. J., June 1993, at 33.

<sup>72.</sup> See, e.g., Runyon, supra note 67, at 75-79.

agency should be expected.<sup>73</sup> If last year is used as the baseline, was last year also part of an increasing service trend, or was it an unusually poor year? If the base year is not representative of the other years, the percentage may not portray the trend fairly.<sup>74</sup> No single question can be formulated to detect such distortions. The judge can ask for the numbers from which the percentages were obtained, and asking about the base can expose distortions. Ultimately, however, the judge must recognize which numbers relate to which issues—a species of clear thinking that is not reducible to a checklist.<sup>75</sup>

# C. Does a Graph Portray Data Fairly?

Graphs are useful for revealing key characteristics of a batch of numbers, trends over time, and the relationships among variables.<sup>76</sup>

#### 1. Displaying distributions: histograms

A graph commonly used to display the distribution of a batch of numbers is the *histogram*.<sup>77</sup> One axis shows the numbers, and the other indicates how often those fall within specified intervals (called a *bin* or a *class interval*). For example, we flipped a quarter ten times in a row and counted the number of heads in this "batch" of ten tosses. For 50 batches, we got the following data:

 7
 7
 5
 6
 8
 4
 2
 3
 6
 5
 4
 3
 4
 7
 4
 6
 8
 4
 7
 4
 5
 4
 3

 4
 4
 2
 5
 3
 5
 4
 2
 4
 4
 5
 7
 2
 3
 5
 4
 6
 8
 4
 7
 4
 5
 4
 3

 4
 4
 2
 5
 3
 5
 4
 2
 5
 3
 5
 6
 6
 4
 9
 10
 5
 5
 6
 6
 4
 9
 10
 5
 5
 6
 6
 4
 9
 10
 5
 5
 6
 6
 4
 9
 10
 5
 5
 6
 6
 4
 3
 4
 5
 7
 2
 3
 5
 4
 6
 4
 9
 10
 5
 5
 6
 6
 4
 3
 4
 5
 7
 2
 3
 5
 4
 6
 4

The data are shown in Figure 1 below (with a bin width of 1).

73. Cf. Michael J. Saks, Do We Really Know Anything About the Behavior of the Tort Litigation System — And Why Not?, 140 U. Pa. L. Rev. 1147, 1203 (1992) (using 1974 as the base year for computing the growth of federal product liability filings exaggerates growth because "1974 was the first year that product liability cases had their own separate listing on the cover sheets.... The count for 1974 is almost certainly an understatement....").

74. Katzer et al., supra note 12, at 106.

75. For some assistance in the task of coping with percentages, see Zeisel, supra note 12, at 1-24.

76. See generally William S. Cleveland, The Elements of Graphing Data (1985); Moore & McCabe, supra note 57, at 3–20.

77. For small batches of numbers, stem-and-leaf plots show all the values and how they are distributed. A stem-and-leaf plot for 11, 12, 23, 23, 23, 23, 33, 45, 69 is shown below:

1	12
2	3333
3	3
4	5
5	
6	9

The numbers to the left of the line are the first digits; those to the right are the second digits. Thus, the entry "2 | 3333" stands for "23, 23, 23, 23."

Reference Manual on Scientific Evidence

# Figure 1

Histogram showing how frequently various numbers of heads appeared in 50 batches of 10 tosses of a quarter. The bin width is 1.

## insert figure 1 here

Figure 1 shows how the number of heads per batch of ten tosses is distributed over the full range of possible values. The spread can be made to appear larger or smaller, however, by changing the scale of the horizontal axis. Likewise, the shape can be altered somewhat by changing the size of the bins.<sup>78</sup> It may be worth inquiring how the analyst chose the bin width.<sup>79</sup>

### 2. Displaying trends

Graphs that plot many values of a variable over time are useful for seeing trends. However, the scales on the axes matter. Figures 2 and 3 show how the scale of an axis can be changed to give a different appearance to the same data. <sup>80</sup> In Figure 2, the federal debt appears to skyrocket during the Reagan and Bush administrations, whereas in Figure 3, the federal debt grows steadily during the same years. The moral is simple: Pay attention to the markings on the axes to determine whether the scale is appropriate.

<sup>78.</sup> In Figure 1, all the bins have equal widths. The histogram is just like a bar graph. However, government agencies often publish economic and social data in tables with unequal intervals. The resulting histograms have unequal bin widths; bar heights are calculated so that the areas (height  $\times$  width) are proportional to the frequencies. In general, a histogram differs from a bar graph in that it represents frequencies by area, not height. See Freedman et al., supra note 12, at 29–40.

<sup>79.</sup> As the width of the bins decreases, the graph becomes more detailed. But the appearance becomes more ragged until finally the graph is effectively a plot of each datum. No general rule can be stated as to what bin width is optimal: "[T]he tolerable loss depends on the subject matter and the goal of the analysis." Cleveland, *supra* note 76, at 125.

<sup>80.</sup> The data are taken from figures in Howard Wainer, *Graphical Answers to Scientific Questions*, Chance, Fall 1993, at 48, 50. This flexibility in presentation applies to other types of graphs as well. See Runyon, *supra* note 67, at 37–39.









Reference Manual on Scientific Evidence

# 3. Displaying association: scatter diagrams

The relationship between two variables can be shown in a *scatter diagram* (also known as a scatterplot or scattergram). Data on income and education for a sample of 350 men aged 25 to 29 in Texas<sup>81</sup> provide an illustration. Each person in the sample corresponds to one dot in the diagram. As indicated in Figure 4, the horizontal axis shows this person's education, and the vertical axis shows his income. Person A completed 8 years of schooling (grade school) and had an income of \$19,000 dollars. Person B completed 16 years of schooling (college) and had an income of \$38,000.

## Figure 4

Plotting a scatter diagram. The horizontal axis shows educational level, and the vertical axis shows income.

#### insert figure 4 here

Figure 5 (next page) is the scatter diagram for all the Texas data. This scatter diagram confirms an obvious point. There is a positive association between income and education. In general, people with higher educational levels have higher incomes. However, there are many exceptions to this rule, and the association is not as strong as one might expect. The correlation coefficient is a numerical measure of the strength of the association.<sup>82</sup>

<sup>81.</sup> These data are from a public-use data tape, Bureau of the Census, U.S. Dep't of Commerce, for the Current Population Survey of March 1988. Income and education are self-reported. Income is truncated at \$100,000 and education (years of schooling completed) at 18 years.

<sup>82.</sup> For a discussion of correlation coefficients, see infra § III.F.2.
Figure 5 Scatter diagram for income and education; men aged 25 to 29 in Texas.<sup>83</sup>

insert figure 5 here

# D. Is an Appropriate Measure Used for the Center of a Distribution?

Perhaps the most familiar descriptive statistic is the arithmetic mean, or average. The *mean* of a batch of numbers lies somewhere in the middle of the data. The mean can be found by adding up all the numbers and dividing by how many there are. The *median* has a different definition. Half the numbers are bigger than the median, and half are smaller.<sup>84</sup> Yet a third statistic is the *mode* —the most common number in the data set. These measures have different properties.<sup>85</sup> The mean takes account of all the data—it involves the total of all the numbers—but, particularly with small data sets, a few unusually large or small

<sup>83.</sup> Education may be compulsory, but the Current Population Survey generally finds a small percentage of respondents who report very little schooling. Such respondents will be found at the lower left corner of the scatter diagram.

<sup>84.</sup> Technically, at least half the numbers are at least as large as the median, and at least half are as small as the median. When the distribution is symmetric, the mean equals the median. The values diverge, however, when the distribution is asymmetric, or skewed.

<sup>85.</sup> How big an error do you make in replacing every number by the "center" of the batch? (1) The mode minimizes the number of errors; for the mode, all "errors" count the same, no matter what their sizes are. Consequently, similar distributions can have very different modes, and the mode is rarely useful. (2) The median minimizes a different measure of error—the sum of all the differences (treating positive and negative differences the same) between the center and the data points. (3) The mean minimizes the sum of the squared differences.

observations can cause it to shift substantially. The median, in contrast, is more resistant to such *outliers*.

Which statistic is most useful depends on the purpose of the analysis. For example, what should be made of a report that the average award in malpractice cases skyrocketed from \$220,000 in 1975 to more than \$1 million in 1985?<sup>86</sup> It might be noted that the median award almost certainly was far less than \$1 million<sup>87</sup> and that the apparently explosive growth may be nothing more than the addition of a tiny fraction of very large awards. Still, if the issue is whether insurers were experiencing more costs from jury verdicts, then the mean is the more appropriate statistic. The total of the awards is related directly to the mean,<sup>88</sup> but this figure cannot be recovered from the median.<sup>89</sup>

# E. Is an Appropriate Measure of Variability Used?

The location of the center of a batch of numbers reveals nothing about the variations that these numbers exhibit.<sup>90</sup> Statistical measures of variability include the *range*, the interquartile range, the mean absolute deviation, and the *standard deviation*. The range is the difference between the high and the low. It seems natural, and it indicates the maximum spread in the numbers, but it is generally the most unstable because it depends entirely on the most extreme values. The interquartile range is the difference between the 25th and 75th percentiles.<sup>91</sup> It contains 50% of the numbers and is more resistant to changes in the extreme values. The mean absolute deviation depends on all the numbers. It is calculated by averaging the differences between each number and the mean. The

86. Jost, supra note 26, at 68, 70-71.

Another dispute over the choice of the mean or the median involves the Railroad Revitalization and Regulatory Reform Act, 49 U.S.C. § 11503, which forbids the taxation of railroad property at a higher rate than other commercial and industrial property. To compare the rates, tax authorities often use the mean, but railroads prefer the median. See David A. Freedman, *The Mean Versus the Median: A Case Study in 4-R Act Litigation*, 3 J. Bus. & Econ. Stat. 1 (1985).

88. To get the total, just multiply the mean by the number of awards. The more pertinent figure is not the total of jury awards, but actual claims experience, including settlements.

89. These and related statistical issues are pursued further in, e.g., Eisenberg & Henderson, supra note 26, at 731, 764–72; Scott Harrington & Robert E. Litan, Causes of the Liability Insurance Crisis, 239 Science 737, 740–41 (1988); Saks, supra note 73, at 1147, 1248–54.

90. The numbers 1, 2, 5, 8, 9 have 5 as their mean and median. So do the numbers 5, 5, 5, 5, 5. In the first batch, the numbers vary considerably about their mean; in the second, the numbers do not vary at all.

91. By definition, 25% of the data fall below the 25th percentile. The median is the 50th percentile.

<sup>87.</sup> A study of cases in North Carolina reported an "average" (mean) award of \$367,737 and a median award of only \$36,500. *Id.* at 71. In TXO Prod. Corp. v. Alliance Resources Corp., 113 S. Ct. 2711 (1993), briefs portraying punitive damages awards as being out of control reported mean punitive awards some ten times larger than the median awards described in briefs defending the current system of punitive damages. See Michael Rustad & Thomas Koenig, *The Supreme Court and Junk Social Science: Selective Distortion in Amicus Briefs*, 72 N.C. L. Rev. 91, 145–47 (1993). The two measures differ so dramatically because the mean allows a few huge awards to overwhelm the effects of many smaller ones.

standard deviation is like the mean absolute deviation except that the squared differences<sup>92</sup> from the mean are averaged, and the square root is extracted.<sup>93</sup>

There are no hard-and-fast rules as to which statistic is the best. In general, the bigger these measures of spread are, the more the numbers are dispersed. Particularly in small data sets, the standard deviation can be influenced heavily by a few outlying values. To remove this influence, the mean and the standard deviation can be recomputed with the outliers discarded.<sup>94</sup> Beyond this, any of the statistics can be supplemented with a figure that displays much of the data.<sup>95</sup>

# F. Is an Appropriate Measure of Association Used?

Many cases involve statistical association. Does an employer's requirement of passing a test for promotion have an exclusionary effect that depends on race? Does the salary of workers depend on gender? Does the incidence of murder vary with the rate of executions for convicted murderers? Do consumer purchases of a product depend on the presence or absence of a product warning?

Statistics, such as the mean and the standard deviation, describe each variable in isolation. They do not describe the extent to which two variables are associated. This section will discuss statistics—percentages, proportions, ratios, correlation coefficients, and slopes of regression lines—that can be used to describe the association between two variables.<sup>96</sup>

93. The square root of 100 is 10. Taking the square root corrects for the fact that the variance is on a different scale than the measurements themselves. If the measurements are of length in inches, the variance is in square inches. Taking the square root changes back to inches.

94. Alternatively, a five-number summary, which lists the smallest value, the 25th percentile, the median, the 75th percentile, and the largest value, may be given. The five-number summary may be presented as a boxplot. If the five numbers were 10, 25, 40, 65, and 90, the boxplot would look like the following:



There are many variations on this idea in which the boundaries of the box or the whiskers extending from it represent different points in the distribution.

95. The measures of variability discussed above depend on the units of measurement. To facilitate comparisons of the variability of different distributions, another statistic known as the *coefficient of variation* often is used. It is the standard deviation expressed as a percentage of the mean. Consider the batch of numbers 1, 4, 4, 7, 9. The mean is 25/5 = 5, the variance is (16 + 1 + 1 + 4 + 16)/5 = 7.6, and the standard deviation is  $\sqrt{7.6} = 2.8$ . The coefficient of variation is 2.8/5 = 56%.

<sup>96</sup> 96. Even if there is an association, however, there will often be a second issue: Is the association causal? For instance, women may be paid less than men because of gender discrimination; or, the difference may be due to the influence of other *covariates*, such as education or experience. On the question of causation, see *supra* §§ II.C–D, which explains why controlled experiments are the best way to eliminate other variables as possible causes of an observed association.

Reference Manual on Scientific Evidence

<sup>92.</sup> If a difference is 10, the squared difference is  $10 \times 10 = 100$ . The mean of the squared differences is known as the variance.

#### 1. Percentage-related statistics

Percentages often are used to describe the association between two variables. Suppose that a university consisting of only two colleges, engineering and business, admits 550 out of 1,400 students: 350 out of 800 male applicants are admitted, but only 200 out of 600 female applicants are admitted. Such data commonly are displayed in the form of a table:<sup>97</sup>

# Table 1 Admissions by Gender

Decision	Male	Female	Total
Admit	350	200	550
Deny	450	400	850
Total	800	600	1,400

The entries in Table 1 indicate that 350/800 = 44% of the men are admitted, compared with only 200/600 = 33% of the women. The resulting selection ratio (used by the Equal Employment Opportunity Commission (EEOC) in its "80% rule") <sup>98</sup> is 33/44 = 75%, meaning that, on average, women have 75% the chance of admission that men have. <sup>99</sup> Another way to express the disparity is to subtract the two percentages: 44 percentage points – 33 percentage points = 11 percentage points.

One difficulty with the simple difference, however, is that it is inevitably small when the two percentages are both close to zero. If the selection rate for men is 5% and that for women is 1%, the difference is only 4 percentage points; yet, on average, women have only 1/5 the chance of men to be selected—and that may be of real concern.

The ratio of the selection rates also has its problems. In the last example, if the selection rates are 5% and 1%, the exclusion rates are 95% and 99%, respectively. The corresponding ratio is 99/95 = 104%, meaning that women have, on average, 104% the chance of men to be rejected. The underlying facts are the same, of course, but this formulation sounds much less disturbing.<sup>100</sup>

<sup>97.</sup> A table of this sort also is called a cross-tabulation, or a contingency table. Table 1 is "two-by-two" because it has two rows and two columns, not counting rows or columns containing the totals.

<sup>98.</sup> The EEOC generally regards any procedure that selects candidates from the least successful group at a rate less than 80% of the rate for the most successful group as having an adverse impact. EEOC Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (1993).

<sup>99.</sup> The analogous statistic used in epidemiology is called the relative risk. A variation on this idea is the relative difference in the proportions, which expresses the proportion by which the probability of selection is reduced. Baldus & Cole, supra note 1, § 5.1; Kairys et al., supra note 20, at 776, 789–90.

<sup>100.</sup> The Illinois Department of Employment Security tried to exploit this feature of the ratio in Council 31, Am. Fed'n of State, County & Mun. Employees v. Ward, 978 F.2d 373 (7th Cir. 1992). In January 1985, the department laid off 8.6% of the blacks on its staff in comparison with 3.0% of the whites on its staff. *Id.* at 375. Recognizing that these layoffs ran afoul of the 80% rule if analyzed in terms of those selected to be laid

Another statistic, the *odds ratio*, avoids this asymmetry. If 5% of male applicants are admitted, the *odds* of a man being admitted are 5%/95% = 1/19; the odds of a woman being admitted are 1%/99% = 1/99. The ratio of these quantities is (1/99)/(1/19) = 19/99. The odds ratio for rejection instead of acceptance is the same, except that the order is reversed.<sup>101</sup> Likewise, when the odds of an admitted applicant being a man as opposed to the odds of a denied applicant being a man is considered, the odds ratio also becomes 99/19.

Although the odds ratio has desirable mathematical properties, <sup>102</sup> its meaning may be less clear than that of the selection ratio or the simple difference. To gauge the magnitude of the association implicit in a two-by-two table, any of the statistics presented here may be considered.

Finally, to illustrate the point that association does not necessarily imply causation, consider again the hypothetical admission data in Table 1. Applicants can be classified not only by gender and admission but also by the college to which they applied, as in Table 2:

#### Table 2

Admissions by Gender and College

	Engi	Engineering		Business	
Decision	Male	Female	Male	Female	
Admit	300	100	50	100	
Deny	300	100	150	300	

The entries in Table 2 add up to the entries in Table 1. Yet, there is no association between gender and admission in either college; men and women are admitted in identical percentages.

Combining two colleges with no association produces a university in which gender is associated strongly with admission. The explanation for this paradox: the business college, to which most of the women applied, is hard to get into; the engineering college, to which most of the men applied, is easier to get into. This example illustrates a common issue in discrimination cases: the effect of other, often unreported, variables on an observed association.<sup>103</sup> When a study is

103. The example is taken from Moore, *supra* note 12, at 205–06, and inspired by more complex data on graduate admissions in 1973 at the University of California at Berkeley analyzed in P. J. Bickel et al., *Sex Bias* 

off—since 3.0%/8.6% = 35%, which is far less than 80%—the department instead presented the selection ratio for retention. *Id.* at 375–76. Since black employees were retained at 91.4%/97.0% = 94% of the white rate, use of a retention rate analysis showed no adverse impact. *Id.* at 376. When a subsequent wave of layoffs was challenged as discriminatory, the department argued "that its retention rate analysis is the right approach to this case and that . . [it] shows conclusively that the layoffs did not have a disparate impact," because they comported with the 80% rule. *Id.* at 379. The Seventh Circuit disagreed and, in reversing an order granting summary judgment to defendants on other grounds, left it to the district court on remand "to decide what method of proof is most appropriate." *Id.* 

<sup>101.</sup> For women, the odds of rejection are 99 to 1; for men, 19 to 1. The ratio of these odds is 99/19.

<sup>102.</sup> See, e.g., Finkelstein & Levin, supra note 1, at 2–4; Joseph L. Fleiss, Statistical Methods for Rates and Proportions 56–99 (2d ed. 1981); Steve Selvin, Statistical Analysis of Epidemiologic Data app. C (1991).

said to have omitted important variables, some experts find it helpful to consider how large a value of the omitted variable would be needed to explain away the reported results.<sup>104</sup>

# 2. Correlation coefficients

Two variables are positively correlated when their values tend to go up or down together.<sup>105</sup> Consider the scatter diagram for income and education in Figure 5. As a rule, people with below-average educational levels also have below-average incomes, while people with higher educational levels generally have higher incomes. The association is positive. The correlation coefficient (usually denoted by r) is a single number that measures the strength of a linear association. Figure 6 shows the values of r for several scatter diagrams.

## Figure 6

The correlation coefficient measures the strength of linear association.

$$r = 0.0$$
  $r = 0.5$   $r = 0.9$ 

# insert figure 6

A correlation coefficient of 0 indicates no linear association between the variables, while a coefficient of +1 indicates a perfect linear relationship: All the dots in the scatter diagram fall on a straight line that slopes up. The maximum value for r is +1. Sometimes, there is a negative association between two variables. Large values of one variable tend to go with small values of the other. The age of a car and its fuel economy in miles per gallon provide an example. Negative association is indicated by negative values for r. The extreme case is an

in Graduate Admissions: Data from Berkeley, 187 Science 398 (1975). See also Freedman et al., supra note 12, at 16–19. Table 2 is an instance of Simpson's Paradox. See generally Myra L. Samuels, Simpson's Paradox and Related Phenomena, 88 J. Am. Stat. Ass'n 81 (1993).

<sup>104.</sup> See, e.g., Joseph L. Gastwirth, Methods for Assessing the Sensitivity of Statistical Comparisons Used in Title VII Cases to Omitted Variables, 33 Jurimetrics J. 19 (1992); Joseph L. Gastwirth, Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims, 11 Law & Ineq. J. 151 (1992).

<sup>105.</sup> Many statistics and displays are available to investigate correlation. The most common are the correlation coefficient and the scatter diagram.

r of -1, indicating that all the points in the scatter diagram lie on a straight line that slopes down.

Moderate associations are the general rule in the social sciences. Correlations larger than about 0.7 are unusual. For example, the correlation between college grades and first-year law school grades is under 0.3 at most law schools, while the correlation between LSAT scores and first-year law school grades is generally about 0.4. <sup>106</sup> The correlation between heights of fraternal twins is about 0.5, while the correlation between heights of identical twins is about 0.95. In Figure 5, the correlation between income and education is 0.43. The correlation coefficient cannot capture all the underlying information. Several questions may arise in this regard, and we consider them in turn.

a. Is the association linear?

The correlation coefficient is designed to measure linear association. Figure 7 shows a strong nonlinear pattern with a correlation close to 0.

#### Figure 7

The correlation coefficient only measures linear association. The scatter diagram shows a strong nonlinear association with a correlation coefficient of nearly 0.

insert figure 7 here

# b. Do outliers influence the coefficient?

The correlation coefficient can be distorted by outliers—a few points that are far removed from the bulk of the data. The left-hand panel in Figure 8 shows that one outlier (lower right-hand corner) can reduce a perfect correlation to nearly

<sup>106.</sup> Linda F. Wightman, Predictive Validity of the LSAT: A National Summary of the 1990–1992 Correlation Studies 10 (1993) (draft final report on data from 167 law schools); *cf.* Linda F. Wightman & David G. Muller, An Analysis of Differential Validity and Differential Prediction for Black, Mexican American, Hispanic, and White Law School Students 11–13 (1990). A combination of LSAT and undergraduate grade point average has a higher correlation with first-year law school grades than either item alone. The multiple correlation coefficient is typically about 0.5. Wightman, *supra* at 10.

nothing. Conversely, the right-hand panel shows that one outlier (upper right-hand corner) can raise a correlation from 0 to nearly 1.

# Figure 8

The correlation coefficient can be distorted by outliers. The left-hand panel shows an outlier (in the lower right-hand corner) that destroys a nearly perfect correlation. The right-hand panel shows an outlier (in the upper right-hand corner) that changes the correlation from 0 to nearly 1.

insert figure 8 here

c. Does a third variable influence the coefficient?

The correlation coefficient measures the association between two variables. Investigators—and the courts—may be more interested in causation. However, association is not necessarily the same as causation. Indeed, the association between two variables may be driven largely by a third variable that has been omitted from the analysis. For instance, among schoolchildren, there is an association between shoe size and vocabulary. However, learning more words does not cause feet to grow bigger, and swollen feet do not make children more articulate. In this case, the third variable is easy to spot—age. In more realistic examples, the driving variable may be more difficult to identify.

Of course, in many other examples the association really does reflect causation, but a large correlation coefficient is not enough to warrant this conclusion. Technically, third variables are called "confounders," or "confounding variables." The basic methods for dealing with a confounding variable involve controlled experiments<sup>107</sup> or the application, typically through a technique called *multiple regression*, <sup>108</sup> of *statistical controls*.<sup>109</sup>

<sup>107.</sup> See supra § II.C.2.

<sup>108.</sup> Multiple regression analysis is discussed in Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual.

<sup>109.</sup> For the reasons stated *supra* § II.D, efforts to control confounding in observational studies are generally less convincing than randomized controlled experiments.

#### 3. Regression lines

The *regression line* can be used to describe a linear trend in the data. The regression line for income on education is shown in Figure 9. The height of the line estimates the average income for a given educational level. For example, the average income for people with 8 years of education is estimated at \$9,600, indicated by the height of the line at 8 years; the average income for people with 16 years of education is estimated at about \$23,200.

# Figure 9

The regression line for income and education, and its estimates.

## insert figure 9

Figure 10 repeats the scatter diagram for income and education (see Figure 5); the regression line is plotted too. In a general way, the line shows the average trend of income as education increases. Thus, the regression line indicates the extent to which a change in one variable (income) is associated with a change in another variable (education).

#### a. What are the slope and intercept?

The regression line can be described in terms of its slope and intercept.<sup>110</sup> In Figure 10, the slope is \$1,700 per year. On average, each additional year of education is associated with an additional \$1,700 of income. Next, the intercept is -\$4,000. This is an estimate of the average income for people with 0 years of education. The estimate is not a good one, for such people are far from the center

<sup>110.</sup> The regression line, like any straight line, has an equation of the form y = nx + b. Here, *m* is the slope, that is, the change in *y* per unit change in *x*. The slope is the same anywhere along the line. Mathematically, that is what distinguishes straight lines from curves. The intercept *b* is the value of *y* when *x* is 0. The slope of a line is akin to the grade of a road; the intercept tells you the starting elevation. For example (Figure 9), the regression line estimates an average income of \$23,200 for people with 16 years of education. This may be computed from the slope and intercept as follows:

<sup>(</sup>  $1,700\,$  per year )  $\,\times\,16\,$  years  $\,-\,$   $\,54,000\,=\,$   $\,$27,200\,$   $-\,$   $\,$4,000\,$   $=\,$   $\,$23,200\,$ 

of the diagram. In general, estimates based on the regression line become less trustworthy as you move away from the bulk of the data.

Figure 10

Scatter diagram for income and education; the regression line indicates the trend.

insert figure 10

b. What does the slope ignore?

The slope has the same limitations as the correlation coefficient in measuring the degree of association.<sup>111</sup> It only measures linear relationships, it may be influenced by outliers, and it does not control for the effect of other variables. Although the slope of \$1,700 per year of education presents each additional year of education as having the same value, some years of schooling surely are worth more and others less. Likewise, the association between education and income graphed in Figure 10 is partly causal, but there are other factors to consider as well, including family backgrounds. People with college degrees probably come from more affluent and better educated families than people who drop out after grade school. They have other advantages besides extra education. Such factors

<sup>111.</sup> In fact, the correlation coefficient is the slope of a regression line with the variables in standardized form, that is, measured in terms of standard deviations away from the mean.

must have some effect on income. This is why statisticians use the guarded language of "on average" and "associated with."<sup>112</sup>

# c. What is the unit of analysis?

If the association between the characteristics of individuals is of interest, those characteristics should be measured on individuals. Sometimes, however, the individual data are not available, but rates or averages are. "Ecological" correlations are computed from such rates or averages; however, ecological correlations generally overstate the strength of an association. An example makes the point. The Bureau of the Census divides the United States into nine geographic areas. The average income and average education can be determined for the men living in each region. The correlation coefficient for these nine pairs of averages turns out to be 0.7.<sup>113</sup> However, geographic regions do not attend school and do not earn incomes. People do. The correlation for income and education for men in the United States is only about 0.4.<sup>114</sup> The correlation for regional averages overstates the correlation for individuals—a common tendency for such ecological correlations.<sup>115</sup>

Scatter diagrams and regression lines are used often in voting rights cases, where the unit of analysis is the voting precinct. Each point in Figure 11 shows data for a precinct in the 1982 Democratic primary election for auditor in Lee County, South Carolina. The horizontal axis shows the percentage of registrants who are white. The vertical axis shows the turnout rate for the white candidate.<sup>116</sup> The regression line is plotted too.

<sup>112.</sup> Many investigators would use multiple regression to isolate the effects of one variable on another—for instance, the independent effect of education on income. Such efforts, like all attempts to infer causation from observational data (see supra § II), may run into problems. See David A. Freedman, As Others See Us: A Case Study in Path Analysis, 12 J. Educ. Stat. 101 (1987).

<sup>113.</sup> See Freedman et al., supra note 12, at 140-41 (using 1988 Current Population Survey).

<sup>114</sup> Id. at 140 (using 1988 Current Population Survey).

<sup>115.</sup> The ecological correlation uses only the average figures, but within each region there is a lot of spread about the average. The ecological correlation overlooks this individual variation.

<sup>116.</sup> By definition, this turnout rate equals the number of votes for the candidate, divided by the number of registrants; the rate is computed separately for each precinct.

# Figure 11

Turnout rate for the white candidate plotted against the percentage of registrants who are white. Precinct-level data, 1982 Democratic primary for auditor, Lee County, South Carolina.

# insert figure 11

Source: Data are from James W. Loewen & Bernard Grofman, Recent Developments in Methods Used in Vote Dilution Litigation, 21 Urb. Law. 589, tbl. 1, at 591 (1989).

In this sort of diagram, the slope is often interpreted as the difference between the white turnout rate and the black turnout rate for the white candidate; the intercept would be interpreted as the black turnout rate for the white candidate. However, the validity of such estimates is contested in the statistical literature. The problem comes from the ecological nature of the regression, that is, making the voting precinct the unit of analysis rather than the individual voter. <sup>117</sup>

117. The secrecy of the ballot box prevents one from obtaining voting data on individuals, although exit polls may provide some information. For further discussion of the problem of ecological regression in this context, see Symposium, Statistical and Demographic Issues Underlying Voting Rights Cases, 15 Evaluation Rev. 659 (1991); James W. Loewen & Bernard Grofman, Recent Developments in Methods Used in Vote Dilution Litigation, 21 Urb. Law. 589, tbl. 1, at 591 (1989); Stephen P. Klein & David A. Freedman, Ecological Regression in Voting Rights Cases, Chance, Summer 1993, at 38; Stephen P. Klein et al., Ecological Regression Versus the Secret Ballot, 31 Jurimetrics J. 393 (1991); Arthur Lupia & Kenneth McCue, Why the 1980s Measures of Racially Polarized Voting Are Inadequate for the 1990s, 12 Law & Pol'y 353 (1990).

This page left blank intentionally for proper pagination when printing two-sided

# **IV.** What Inferences Can Be Drawn from the Data?

The inferences that reasonably may be drawn from a study depend on the quality of the data. As discussed in section II, the data may not address the issue of interest, or may be systematically in error, or may be difficult to interpret due to confounding. We turn now to an additional concern—*random error*. <sup>118</sup> Are patterns in the data the result of chance? Would a pattern wash out if more data were collected? If measurements on individual units are unreliable, <sup>119</sup> the errors may combine to produce a false pattern. Even if the measurements on individual units are free from error, the sample may not be representative of the population.

The laws of probability are central to analyzing random error. By applying these laws, the statistician can assess the likely impact of chance error, using standard errors, confidence intervals, significance probabilities, hypothesis tests, or posterior probability distributions. The following example illustrates the ideas. An employer plans to use a standardized examination to select trainees from a pool of 5,000 male and 5,000 female applicants. This total pool of 10,000 applicants is the statistical population. Under Title VII of the Civil Rights Act, if the proposed examination excludes a disproportionate number of women, the employer must show that the exam is job related.<sup>120</sup>

To see whether there is disparate impact, the employer administers the exam to a sample of 50 men and 50 women drawn at random from the population of job applicants. In the sample, 29 of the men but only 19 of the women pass; the sample pass rates are therefore 29/50 = 58% and 19/50 = 38%. The employer announces that it will use the exam anyway, and several applicants bring an action under Title VII.

Disparate impact seems clear. The difference in sample pass rates is 20 percentage points: 58% - 38% = 20%. The employer argues, however, that the disparity could just reflect random error. After all, only a small number of people

<sup>118.</sup> Random error is also called *sampling error*, *chance error*, or statistical error. Econometricians use the parallel concept of *random disturbance term*.

<sup>119.</sup> See supra § II.A.1.

<sup>120.</sup> The seminal case is Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971). The requirements and procedures for the validation of tests can go beyond a simple showing of job-relatedness. See, e.g., Richard R. Reilly, Validating Employee Selection Procedures, in Statistical Methods in Discrimination Litigation, supra note 7, at 133; Michael Rothschild & Gregory J. Werden, Title VII and the Use of Employment Tests: An Illustration of the Limits of the Judicial Process, 11 J. Legal Stud. 261 (1982).

took the test, and the sample just may have happened to include disproportionate numbers of high-scoring men and low-scoring women. Clearly, even if there was no overall difference in pass rates for male and female applicants, in some samples men will outscore women. A statistician then might be asked to address such topics as the following:

- Estimation. Plaintiffs use the difference of 20 percentage points between the sample men and women to estimate the disparity between all male and female applicants. How good is this estimate? Precision can be expressed using the standard error or a confidence interval.
- Statistical Significance. Suppose the defendant is right—in the population of all 5,000 male and 5,000 female applicants, the pass rates are equal; there is no disparate impact. How likely is it that a random sample of 50 men and 50 women will produce a disparity of 20 percentage points or more? This chance is known as a *p*-value. Statistical significance is determined by reference to the *p*-value, and hypothesis testing is the technique for computing *p*-values or determining statistical significance. <sup>121</sup>
- Posterior probability . Given the observed disparity of 20 percentage points in the sample, what is the probability that—in the population as a whole—men and women have equal pass rates? This question is of direct interest to the courts. However, within the framework of classical statistical theory, such a posterior probability has no meaning.<sup>122</sup> For a subjectivist statistical cian, posterior probabilities may be computed using Bayes' rule.

# A. Estimation

#### 1. What estimator should be used?

An *estimator* is a statistic computed from sample data and used to estimate a numerical characteristic of the population. For example, the difference in pass rates for a sample of men and women is used to estimate the corresponding disparity in the population of all applicants. In our sample, the pass rates were 58% and 38%; the difference in pass rates for the whole population is estimated to be 20 percentage points: 58% - 38% = 20%. In more complex problems, statisticians may have to choose among several estimators. Generally, estimators that tend to make smaller errors are preferred. However, this idea can be made precise in more than one way,<sup>123</sup> leaving room for judgment in selecting an estimator.

<sup>121.</sup> Hypothesis testing is also called significance testing.

<sup>122.</sup> This classical framework is also called "objective" or "frequentist." Contrast with the subjectivist approach; see infra § IV.C.

<sup>123.</sup> Furthermore, reducing error in one context may increase error in other contexts; there may also be a trade-off between accuracy and simplicity.

# 2. What is the standard error?

The estimate of 20 percentage points is likely to be off, at least by a little, due to random error. The standard error gives the likely magnitude of this random error. <sup>124</sup> Whenever possible, an estimate should be accompanied by its standard error. <sup>125</sup> In our example, the standard error is about 10 percentage points: The estimate of 20 percentage points is likely to be off by about 10 percentage points or so, in either direction. <sup>126</sup> Since the pass rates for all 5,000 men and 5,000 women are unknown, we cannot say exactly how far off the estimate is going to be, but 10 percentage points gauges the likely magnitude of the error.

Confidence intervals make the idea more precise. Statisticians who say the population difference falls within plus-or-minus 1 standard error of the sample difference would be correct about 68% of the time. To write this more compactly, we can abbreviate standard error as SE. A 68% confidence interval is the range

estimate - 1 SE to estimate + 1 SE

In our example, the 68% confidence interval goes from 10 to 30 percentage points. If a higher confidence level is wanted, the interval must be widened. The 95% confidence interval is about

estimate -2 SE to estimate +2 SE

This runs from 0 to 40 percentage points. Although 95% confidence intervals are used commonly, there is nothing special about 95%. For example, a 99.7% confidence interval is about

estimate -3 SE to estimate +3 SE

This stretches from -10 to 50 percentage points.<sup>127</sup>

124. Standard errors are also called standard deviations, and courts seem to prefer the latter term, as do many authors. See infra notes 145, 149.

125. The standard error can also be used to measure reproducibility of estimates from one random sample to another. See *infra* the Appendix.

<sup>126.</sup> The standard error depends on the pass rates of men and women in the sample and on the size of the sample. Chance error is smaller for larger samples, so the standard error goes down as sample size goes up. The Appendix gives the formula for computing the standard error of a difference in rates based on random samples. Generally, the formula for the standard error must take into account the method used to draw the sample and the nature of the estimator. Statistical expertise is needed to choose the right formula.

<sup>127.</sup> A negative value, such as -10%, indicates that for the whole population, more women than men are estimated to pass the test. The 68%, 95%, and 99.7% come from the normal curve. See infra the Appendix. When there are samples of reasonable size, an estimator like the pass rate difference will follow the normal curve fairly well. Statisticians call this the *central limit theorem*. The probability that our estimator will be within 2 standard errors of the true population figure is approximately equal to the area under the normal curve between -2 and +2. This area is about 95%. For a more complete description of the normal curve and its use in large samples, see, *e.g.*, Freedman et al., *supra* note 12, at 73–89, 282–302. Of course, many estimators do not follow the normal curve, and other procedures then must be used to obtain confidence intervals.

A confidence interval is based on the standard error. If the standard error is small, the estimate probably is close to the truth. If the standard error is large, the estimate may be seriously wrong.

#### 3. What do standard errors and confidence intervals mean?

An estimate based on a sample will differ from the exact population value due to random error; the standard error measures the likely size of the random error. Confidence intervals are a technical refinement, and confidence is a term of art.<sup>128</sup> For a given confidence level, a narrower interval indicates a more precise estimate. For a given sample size, increased confidence can be attained only by widening the interval. A high confidence level alone means very little, but a high confidence level resulting in a small interval is impressive.<sup>129</sup> It indicates that the random error in the sample estimate is low.

Both the standard error and the confidence interval are derived using a particular model of statistical error. A statistical model expresses the way random error works and generally contains parameters that characterize the population from which the samples were drawn. <sup>130</sup> The data in our example came from a random sample, and that guaranteed the validity of the statistical calculations. <sup>131</sup>

128. In the standard frequentist theory of statistics, one cannot make probability statements about population characteristics. *See, e.g.*, Freedman et al., *supra* note 12, at 351–53; *infra* § IV.B.1. Because of the limited technical meaning of confidence, it has been argued that the term is misleading and should be replaced by a more neutral one, such as frequency coefficient, in courtroom presentations. David H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 Wash. L. Rev. 1333, 1354 (1986).

129. Conversely, a broad interval signals that random error is substantial. In Cimino v. Raymark Indus., Inc., 751 F. Supp. 649 (E.D. Tex. 1990), the district court drew certain random samples from more than 6,000 pending asbestos cases, tried these cases, and used the results to estimate the total award to be given to all plaintiffs in the 6,000 cases. The court then held a hearing to determine whether the samples were large enough to provide accurate estimates. *Id.* at 664. The court's expert, an educational psychologist, testified that the estimates were accurate because the samples matched the population on such characteristics as race and the percentage of plaintiffs still alive. *Id.* However, the matches occurred only in the sense that population characteristics fell within very broad 99% confidence intervals computed from the samples. The court thought that matches within the 99% confidence intervals proved more than matches within 95% intervals. *Id.* Unfortunately, this is backwards. It is not very impressive to be correct in a few instances with a 99% confidence intervals are broad enough to ensure coverage 99% of the time. *Cf.* Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 Stan. L. Rev. 815 (1992).

130. In our example, one parameter is the pass rate of the 5,000 male applicants; another parameter is the pass rate of the 5,000 female applicants. These two parameters determine the probabilities of observing the various possible values for the sample difference, according to a set of mathematical equations. The statistical problem consists of working backwards from the sample data to the population parameters.

When the parameters are known, the analyst may use the model to find the probability of an observed outcome (or one like it). This approach is common in cases alleging discrimination in the selection of jurors. *E.g.*, Castaneda v. Partida, 430 U.S. 482, 496 (1977); David H. Kaye, *Statistical Evidence of Discrimination in Jury Selection, in* Statistical Methods in Discrimination Litigation, *supra* note 7, at 13. *Cf.* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (1977) (computing probabilities of selecting black teachers). Although such problems in applied probability theory are not explicitly treated in this reference guide, the Appendix presents some relevant calculations.

131. Partly because the Supreme Court used models giving rise to variables that are approximately normal in *Hazelwood* and *Castaneda*, courts and attorneys sometimes are skeptical of models and analyses that produce other types of random variables. See, e.g., EEOC v. Western Elec. Co., 713 F.2d 1011 (4th Cir. 1983), discussed in David H. Kaye, Ruminations on Jurimetrics: Hypergeometric Confusion in the Fourth Circuit, 26 Jurimetrics J. 215 (1986). But cf. Branion v. Gramly, 855 F.2d 1256 (7th Cir. 1988) (questioning an apparently

The choice of an appropriate model in other situations may be less obvious.<sup>132</sup> When a model does not describe well the process giving rise to the data, the estimate and its standard error are less probative.<sup>133</sup>

Furthermore, the standard error and the confidence interval generally ignore systematic errors, such as selection bias or nonresponse bias.<sup>134</sup> For example, one court—reviewing studies of whether a particular drug causes birth defects—observed that mothers of children with birth defects may be more likely to remember taking a drug during pregnancy than women with normal children.<sup>135</sup> This selective recall would bias comparisons between samples from the two groups of women. The standard error for the estimated difference in drug usage between the two groups ignores this bias, as does the confidence interval.<sup>136</sup>

132. For examples of legal interest, see, eg., Mary W. Gray, Can Statistics Tell Us What We Do Not Want to Hear?: The Case of Complex Salary Structures, 8 Stat. Sci. 144 (1993); Arthur P. Dempster, Employment Discrimination and Statistical Science, 3 Stat. Sci. 149 (1988). One statistician describes the issue as follows:

[A] given data set can be viewed from more than one perspective, can be represented by a model in more than one way. Quite commonly no unique model stands out as "true" or correct; justifying so strong a conclusion might require a depth of knowledge that is simply lacking. So it is not unusual for a given data set to be analyzed in several apparently reasonable ways. If conclusions are qualitatively concordant, that is regarded as grounds for placing additional trust in them. But more often, only a single model is applied, and the data are analyzed in accordance with it . . . .

Desirable features in a model include (i) tractability, (ii) parsimony, and (iii) realism. That there is some tension among these is not surprising.

*Tractability*. A model that is easy to understand and to explain is tractable in one sense. Computational tractability can also be an advantage, though with cheap computing available not too much weight can be given to it.

*Parsimony*. Simplicity, like tractability, has a direct appeal, not wisely ignored—but not wisely over-valued either. If several models are plausible and more than one of them fits adequately with the data, then in choosing among them, *one* criterion is to prefer a model that is simpler than the other models.

*Realism*.... First, does the model reflect well the actual .... [process that generated the data]? This question is really a host of questions, some about the distributions of the random errors, others about the mathematical relations among the [variables and] parameters. The second aspect of realism is sometimes called robustness: If the model is *false* in certain respects, how badly does that affect estimates, significance test results, etc., that are based on the flawed model?

Lincoln E. Moses, The Reasoning of Statistical Inference, in Perspectives on Contemporary Statistics, supra note 22, at 107, 117–18.

133. It still may be helpful to consider the standard error, perhaps as a minimal estimate for statistical uncertainty.

134. For a discussion of such systematic errors, see supra § II.B.

135. Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307, 311–12 (5th Cir.), modified, 884 F.2d 166 (5th Cir. 1989), cert. denied, 494 U.S. 1046 (1990).

136. In *Brock*, the court held that the confidence interval took account of bias (in the form of selective recall) as well as random error. 874 F.2d at 311–12. With respect, we disagree. Even if sampling error were nonexistent, which would be the case if one could interview every woman who had a child in the period that the drug was available, selective recall would produce a difference in the percentages of reported drug exposure between mothers of children with birth defects and those with normal children. In this hypothetical situa-

arbitrary assumption of normality), cert. denied , 490 U.S. 1008 (1989), discussed in David H. Kaye, Statistics for Lawyers and Law for Statistics, 89 Mich. L. Rev. 1520 (1991). Whether a given variable is normally distributed is an empirical or statistical question, not a matter of law. That a particular model has been used in a previous case may be of limited value in deciding whether it is appropriate in the case at bar. See generally Statistical Methods in Discrimination Litigation, supra note 7, at iii; Laurens Walker & John Monahan, Social Facts: Scientific Methodology as Legal Precedent, 76 Cal. L. Rev. 877 (1988).

Likewise, the standard error does not address problems inherent in using convenience samples rather than random samples.<sup>137</sup>

# B. *p*-values and Hypothesis Tests

1. What is the *p*-value?

In our example, 50 men and 50 women were drawn at random from 5,000 male and 5,000 female applicants. An exam was administered to this sample, and in the sample, the pass rates for the men and women were 58% and 38%, respectively; the sample difference in pass rates was 58 - 38 = 20 percentage points. The *p*-value answers the following question: If the pass rates among all 5,000 male applicants and 5,000 female applicants were identical, how probable would it be to find a discrepancy as large as or larger than the 20 percentage point difference observed in our sample? The question is delicate, because the pass rates in the population are unknown—that is why a sample was taken in the first place.

The assertion that the pass rates in the population are the same is called the *null hypothesis*. The null hypothesis asserts that there is no difference between men and women in the whole population—differences in the sample are due to the luck of the draw. The *p*-value is the probability of getting data as extreme as, or more extreme than, the actual data, given that the null hypothesis is true:

p = Pr (extreme data | null hypothesis in model)

If the null hypothesis is true, there is only a 5% chance of getting a difference in the pass rates of 20 percentage points or more.<sup>138</sup> The *p*-value for the observed discrepancy is 5%, or .05.

In such examples, small *p*-values are evidence of disparate impact, while large *p*-values are evidence against disparate impact. Regrettably, multiple negatives are involved here. The null hypothesis asserts no difference in the population—that is, no disparate impact. Small *p*-values argue against the null hypothesis; that is, small *p*-values argue there is disparate impact. Generally, by indicating that the magnitude of the observed difference is improbable if the null hypothesis is true, small *p*-values undermine the null hypothesis. The smaller the *p*-value for a given study, the more surprising it would be to see such differences under the null hypothesis. Conversely, large *p*-values indicate that the data are compatible with the null hypothesis.

However, since p is calculated by assuming the null hypothesis, the p-value cannot give the chance that this hypothesis is true. The p-value merely gives the

tion, the standard error would vanish. Therefore, the standard error can disclose nothing about the impact of selective recall.

<sup>137.</sup> See supra § II.B.1.

<sup>138.</sup> See infra the Appendix.

chance of getting evidence against the null hypothesis as strong or stronger than the evidence at hand—assuming the null hypothesis is correct. No matter how many samples are obtained, the null hypothesis is either always right or always wrong. Chance affects the data, not the hypothesis. With the frequency interpretation of chance, there is no meaningful way to assign a numerical probability to the null hypothesis, or to any alternative hypothesis, for that matter.<sup>139</sup>

Computing *p*-values requires statistical expertise. Many methods are available, but only some will fit the occasion.<sup>140</sup> Sometimes standard errors will be part of the analysis, while other times they will not be. Sometimes a difference of 2 standard errors will imply a *p*-value of about .05, other times it will not. In general, the *p*-value depends on the model and its parameters, the size of the sample, and the sample statistics.

Because the *p*-value is affected by sample size, it does not measure the extent or importance of a difference.<sup>141</sup> Suppose, for instance, that the 5,000 male and 5,000 female job applicants would differ in their pass rates, but only by 1 percentage point. This difference might not be enough to make a case of disparate impact, but by including enough men and women in the sample, the data could be made to have an impressively small *p*-value. This *p*-value would confirm that the 5,000 men and 5,000 women have different pass rates, but it would not show the difference is substantial.

Likewise, in considering whether two quantities are correlated <sup>142</sup> in a population from which a random sample has been drawn, the *p*-value depends on the correlation in the sample as well as on the number of data points. Statistical significance may result from a small correlation and a large number of points. In short, the *p*-value does not measure the strength or importance of an association. <sup>143</sup>

140. See, e.g., Thomas J. Sugrue & William B. Fairley, A Case of Unexamined Assumptions: The Use and Misuse of the Statistical Analysis of Castaneda/Hazelwood in Discrimination Litigation, 24 B.C. L. Rev. 925 (1983).

141. Some opinions seem to equate small *p*-values with gross or substantial disparities. *E.g.*, Craik v. Minnesota St. Univ. Bd., 731 F.2d 465, 479 (8th Cir. 1984). Other courts have emphasized the need to decide whether the underlying sample statistics reveal that a disparity is large. *E.g.*, McCleskey v. Kemp, 753 F.2d 877, 892–94 (11th Cir. 1985), *aff* d, 481 U.S. 279 (1987).

142. See supra § III.F.2.

143. The conventional procedures used to compute a *p*-value for a correlation depend on the normality of the underlying process for generating the data. The scatter diagram itself gives some useful clues as to whether this assumption is satisfied. Basically, the scatter diagram should be roughly circular or oval in shape. The diagrams in Figure 6 confirm the assumption of normality. The diagram in Figure 5 is incompatible with the assumption, because the cloud of points widens as one moves from left to right along the horizontal axis. In the

<sup>139.</sup> See, e.g., The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 196–98; David H. Kaye, *Statistical Significance and the Burden of Persuasion*, Law & Contemp. Probs., Autumn 1983, at 13. Some opinions suggest a contrary view. *E.g.*, Fudge v. Providence Fire Dep't, 766 F.2d 650, 658 (1st Cir. 1985) ("Widely accepted statistical techniques have been developed to determine the likelihood an observed disparity resulted from mere chance."); Capaci v. Katz & Besthoff, Inc., 711 F.2d 647, 652 (5th Cir. 1983) ("the highest probability of unbiased hiring was  $5.367 \times 10^{-20}$ "), *cert. denied*, 466 U.S. 927 (1984). Such statements appear to confuse the probability of the kind of outcome observed, which is computed under some model of chance, with the probability that chance is the explanation for the outcome. (In scientific notation,  $10^{20}$  is one followed by twenty zeros, and  $10^{-20}$  is the reciprocal of that number. The proverbial "one in a million" is more dryly expressed as  $1 \times 10^{-6}$ .)

#### 2. Is a difference statistically significant?

Statistical significance is determined by comparing a *p*-value to a preestablished value, the *significance level* .<sup>144</sup> If an observed difference is in the middle of the distribution that would be expected under the null hypothesis, there is no surprise. The sample data are of the type that often would be seen when the null hypothesis is true: The difference is not significant, and the null hypothesis cannot be rejected. Conversely, if the sample difference is far from the expected value—according to the null hypothesis—the sample is unusual: The difference is significant, and the null hypothesis is rejected. In our example, the 20 percentage point difference in pass rates for the men and women in the sample, whose *p*-value was about .05, would be significant at the .05 level. If the threshold were set lower, for instance at .01, the result would not be significant.

In practice, statistical analysts use certain preset significance levels—typically .05 or .01.<sup>145</sup> The .05 level is the most common in social science, and an analyst who speaks of "significant" results without specifying the threshold probably is using this level.<sup>146</sup> An unexplained reference to "highly significant" results probably means that  $\rho$  is less than .01.<sup>147</sup>

Since the term "significant" is merely a label for certain kinds of *p*-values, it is subject to the same limitations as are *p*-values themselves. Significant differences are evidence that something besides random error is at work, but they are not evidence that this "something" is legally or practically important. Statisticians distinguish between statistical and practical significance to make the point. When

jargon of the field, Figure 5 shows *heteroscedasticity*, while the diagrams in Figure 6 are *homoscedastic*. Figures 7 and 8 are also incompatible with the normality assumption; Figure 7 shows a strong nonlinear pattern, while Figure 8 has outliers. Under other circumstances, too, procedures may be available to test the adequacy of a model's assumptions. See, e.g., David C. Hoaglin, *Diagnostics*, in Perspectives on Contemporary Statistics, *supra* note 22, at 123; David A. Belsley et al., Regression Diagnostics: Identifying Influential Data and Sources of Collinearity (1980). Sometimes, however, assumptions are tested only by introducing other assumptions that are even more obscure.

<sup>144.</sup> Statisticians use the Greek letter  $\alpha$  (*alpha*) to denote the significance level;  $\alpha$  gives the chance of obtaining a significant result, assuming that the null hypothesis is true. Thus,  $\alpha$  represents the chance of what is variously termed a false rejection of the null hypothesis, a type I error, a false positive, or a false alarm. For example, suppose  $\alpha = 5\%$ . If investigators do many studies, and the null hypothesis happens to be true in each case, then about 5% of the time they would obtain significant results—and falsely reject the null hypothesis.

<sup>145.</sup> The Supreme Court implicitly referred to this practice in Castaneda v. Partida, 430 U.S. 482, 496 n.17 (1977), and Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 311 n.17 (1977). In these footnotes, the Court described the null hypothesis as "suspect to a social scientist" when a statistic from "large samples" falls more than "two or three standard deviations" from its expected value under the null hypothesis. Although the Court did not say so, these differences produce *p*-values of about .05 and .01 when the statistic is normally distributed. The Court's standard deviation is our standard error.

<sup>146.</sup> Some have intimated that data not significant at the .05 level should be disregarded. This view is challenged in, eg., Kaye, supra note 128, at 1344 & n.56, 1345. But see Paul Meier et al., What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule, 1984 Am. B. Found. Res. J. 139, 152.

<sup>147.</sup> Merely labeling results as significant or not significant without providing the underlying information that goes into this conclusion is of limited value. See, e.g., John C. Bailar, III & Frederick Mosteller, Guidelines for Statistical Reporting in Articles for Medical Journals: Amplifications and Explanations, in Medical Uses of Statistics, supra note 44, at 313, 316.

practical significance is lacking—when the size of a disparity or correlation is negligible—there is no reason to worry about statistical significance.<sup>148</sup>

As noted above, it is easy to mistake the *p*-value for the probability that there is no difference. Likewise, if results are significant at the .05 level, it is tempting to conclude that the null hypothesis has only a 5% chance of being correct.<sup>149</sup> This temptation should be resisted. From the frequentist perspective, statistical hypotheses are either true or false—probabilities govern the samples, not the models and hypotheses. The significance level tells us what is likely to happen when the null hypothesis is correct; it cannot tell us the probability that the hypothesis is true. Significance comes no closer to expressing the probability that the null hypothesis is true than does the underlying *p*-value.<sup>150</sup>

- 3. Questions about hypothesis tests
- a. What is the power of the test?

When a *p*-value is high, findings are not significant, and the null hypothesis is not rejected. There are at least two possible explanations:

- 1. There is no difference in the population—the null hypothesis is true; or
- 2. There is some difference in the population—the null hypothesis is false but, by chance, the data are of the kind expected under the null hypothesis.<sup>151</sup>

If the *power* of a statistical study is low, the second is a reasonable explanation for the data. Power is the chance that a statistical test will declare an effect when there is an effect to declare.<sup>152</sup> This chance depends on the size of the effect and

148. E.g., Waisome v. Port Auth., 948 F.2d 1370, 1376 (2d Cir. 1991) ("[T]hough the disparity was found to be statistically significant, it was of limited magnitude . . . .") (citations omitted).

149. E.g., *id.* at 1376 ("Social scientists consider a finding of two standard deviations sig nificant, meaning there is about one chance in 20 that the explanation for a deviation could be random ...."); Rivera v. City of Wichita Falls, 665 F.2d 531, 545 n.22 (5th Cir. 1982) ("A variation of two standard deviations would indicate that the probability of the observed outcome occurring purely by chance would be approximately five out of 100; that is, it could be said with a 95% certainty that the outcome was not merely a fluke."); Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224, 272 (N.D. Tex. 1980) ("[I]f a 5% level of significance is used, a sufficiently large t-statistic for the coefficient indicates that the chances are less than one in 20 that the true coefficient is actually zero."), *vacated*, 723 F.2d 1195 (5th Cir.), *cert. denied*, 469 U.S. 1073 (1984).

150. For more discussion, see Kaye, supra note 1 39.

151. Tests also may reject—or fail to reject—because the statistical model does not fit the situation. See in fra § IV.B.3.e.

152. More precisely, power is the probability of rejecting the null hypothesis when the alternative hypothesis is right. Typically, this depends on the values of unknown parameters, as well as on the preset significance level ( $\alpha$ ). See *supra* notes 130, 144. Therefore, no single number gives the power of the test. The expert can specify particular values for the parameters and significance level and compute the power of the test accordingly. See *infra* the Appendix for an example. Power may be denoted by the Greek letter  $\beta$  (*beta*).

Accepting the null hypothesis when the alternative is true is known as a false acceptance of the null hypothesis, a type II error, a false negative, or a missed signal. The chance of a false negative may be computed from the power, as  $1 - \beta$ . Frequentist hypothesis testing keeps the risk of a false positive to a specified level (such as  $\alpha = .05$ ) and then tries to minimize the chance of a false negative  $(1 - \beta)$  for that value of  $\alpha$ . Regrettably, the notation is in some degree of flux; many authors use  $\beta$  to denote the chance of a false negative; then, it is  $\beta$  that should be minimized.

the size of the sample. Discerning subtle differences in the population requires large samples.

When a study with low power fails to show a significant effect, one should not treat the negative result as strong proof that there is no effect. The study is described more fairly as inconclusive than as negative.<sup>153</sup> In contrast, when studies have a good chance of detecting a meaningful association, failure to obtain significant findings can be persuasive evidence that there is no effect to be found.<sup>154</sup>

#### b. One-tailed versus two-tailed tests

In many cases, a statistical test can be either *one-tailed* or *two-tailed*. The second method will generally produce a *p*-value twice as big as the first method. Since small *p*-values are evidence against the null hypothesis, a one-tailed test seems to produce stronger evidence than a two-tailed test. However, this difference is largely illusory.<sup>155</sup>

Some courts have expressed a preference for two-tailed tests, <sup>156</sup> but a rigid rule is not required if *p*-values and significance levels are used as clues rather than as

153. In our pass rate example, with  $\alpha = .05$ , power to detect a difference of 10 percentage points between the male and female job applicants is only about 1/6. See infra the Appendix. Not seeing a "significant" difference therefore provides only weak proof that the difference between men and women is smaller than 10 percentage points. We prefer estimates accompanied by standard errors to tests, because the former seem to make the state of the statistical evidence clearer: The estimated difference is  $20 \pm 10$  percentage points, indicating that a difference of 10 percentage points is quite compatible with the data.

154. Some formal procedures are available to aggregate results across studies. See In R Paoli R.R. Yard PCB Litig., 916 F.2d 829 (3d Cir. 1990), cert. denied sub nom. General Elec. Co. v. Knight, 499 U.S. 961 (1991). In principle, the power of the collective results will be greater than the power of each study. See, e.g., The Handbook of Research Synthesis 226–27 (Harris Cooper & Larry V. Hedges eds., 1994); Larry V. Hedges & Ingram Olkin, Statistical Methods for Meta-Analysis (1985); Jerome P. Kassirer, Clinical Trials and Meta-Analysis: What Do They Do for Us?, 327 New Eng. J. Med. 273, 274 (1992) ("[C]umulative meta-analysis rep resents one promising approach."); National Research Council, Combining Information: Statistical Issues and Opportunities for Research (1992). Unfortunately, these procedures have their own limitations. E.g., Diana Petitti, Meta-Analysis, Decision Analysis, Cost-Effectiveness Analysis in Medicine: Methods for Quantitative Synthesis of Information (1994); Michael Oakes, Statistical Inference: A Commentary for the Social and Behavioral Sciences 157 (1986) ("a retrograde development"); Charles Mann, Meta-Analysis in the Breech, 249 Science 476 (1990).

155. In our pass rate example, the *p*-value of the test is approximated by a certain area under the normal curve. The one-tailed procedure uses the tail area under the curve to the right of 2, giving p = .025. The two-tailed procedure contemplates the area to the left of -2, as well as the area to the right of 2. Now there are two tails, and p = .05. According to formal statistical theory, the choice between one tail and two sometimes can be made by considering the exact form of the alternative hypothesis. The null hypothesis held that pass rates were equal for men and women in the whole population of applicants. The alternative hypothesis may exclude a priori the possibility that women have a higher pass rate and hold that more men will pass than women. This symmetric alternative suggests a one-tailed test. Conversely, the alternative hypothesis may simply be that pass rates for men and women in the whole population are unequal. This symmetric alternative admits the possibility that women may score higher than men and points to a two-tailed test. See, e.g., Freedman et al., *supra* note 12, at 495–98.

156. See, e.g., Baldus & Cole, supra note 1, at 308 n.35a (1980 & Supp. 1987); The Evolving Role of Statistical Assessments as Evidence in the Courts, supra note 1, at 38–40 (citing EEOC v. Federal Reserve

Some commentators have claimed that the cutoff for significance should be chosen to equalize the chance of a false positive and a false negative, on the ground that this criterion corresponds to the "more-probable-than-not" burden of proof. Unfortunately, the argument is fallacious because  $\alpha$  and  $\beta$  apply to data, not hypotheses. See supra § IV.B.1.

mechanical devices for deferring to or dismissing statistical proofs. One-tailed tests make it easier to reach a threshold like .05, but if .05 is not used as a magic line, then the choice between one tail and two is less important—as long as the choice and its effect on the *p*-value are made explicit.<sup>157</sup>

#### c. How many tests have been performed?

Repeated applications of significance testing complicate the interpretation of a significance level. If enough studies are conducted, random error almost guarantees that some will yield significant findings, even when there is no real effect.<sup>158</sup> Consider the problem of deciding whether a coin is biased. The probability that a fair coin will produce ten heads when tossed ten times is  $(1/2)^{10} = 1/1,024$ . Observing ten heads in the first ten tosses, therefore, would be strong evidence that the coin is biased. Nevertheless, if a fair coin is tossed a few thousand times, it is likely that at least one string of ten consecutive heads will appear. The test—looking for a run of ten heads—has been repeated far too often.<sup>159</sup>

The problem of multiple testing can affect statistical models with many possible equations and parameters. Almost any large data set—even pages from a table of random digits—will contain some unusual pattern that can be uncovered by a diligent search. Having detected the pattern, the analyst can perform a statistical test for it, blandly ignoring the search effort. Statistical significance is bound to follow. Ten heads in the first ten tosses means one thing; a run of ten heads in a few thousand tosses of a coin means another.

There are statistical methods for coping with multiple looks at the data, which permit the calculation of meaningful *p*-values in certain cases.<sup>160</sup> However, no general solution is available, and the existing methods would be of little help in the typical case where analysts have run through a variety of regression models to arrive at the one considered the most satisfactory. In these situations, courts

Bank, 698 F.2d 633 (4th Cir. 1983), rev'd on other grounds sub nom. Cooper v. Federal Reserve Bank, 467 U.S. 867 (1984)); Kaye, supra note 128, at 1358 n.113; David H. Kaye, *The Numbers Game: Statistical Inference in Discrimination Cases*, 80 Mich. L. Rev. 833 (1982) (citing Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977)). An argument for one-tailed tests is made by Richard Goldstein, *Two Types of Statistical Errors in Employment Discrimination Cases*, 26 Jurimetrics J. 32 (1985).

<sup>157.</sup> One-tailed tests at the .05 level are viewed as weak evidence—no weaker standard is commonly used in the technical literature. But see Richard Lempert, Statistics in the Courtroom: Building on Rubinfeld, 85 Colum. L. Rev. 1098, 1099 (1985) ("[T]he values of social science are not the values of law.").

<sup>158.</sup> Since research that fails to uncover significance is not usually published, reviews of the literature may produce an unduly large number of studies finding statistical significance. E.g., Stuart J. Pocock et al., Statistical Problems in the Reporting of Clinical Trials: A Survey of Three Medical Journals , 317 New Eng. J. Med. 426 (1987).

<sup>159.</sup> For advice on spotting comparable abuses in biomedical studies, see James L. Mills, *Data Torturing*, 329 New Eng. J. Med. 1196 (1993).

<sup>160.</sup> See, e.g., Yosef Hochberg & Ajit C. Tamhane, Multiple Comparison Procedures (1987); Rupert G. Miller, Jr., Simultaneous Statistical Inference (2d ed. 1981); Peter H. Westfall & S. Stanley Young, Resampling-Based Multiple Testing: Examples and Methods for p Value Adjustment (1993); Joseph L. Gastwirth & Samuel W. Greenhouse, Estimating a Common Relative Risk: Application in Equal Employment, 82 J. Am. Stat. Ass'n 38 (1987); Robert Follett & Finis Welch, Testing for Discrimination in Employment Practices, Law & Contemp. Probs., Autumn 1983, at 171; Kaye, supra note 130, at 13.

should not be overly impressed with claims that estimates are significant. Instead, they should be asking how analysts developed their models.<sup>161</sup>

# d. What are the interval estimates?

Statistical significance depends on the *p*-value, and the *p*-value depends on sample size. Therefore, a significant effect may be small. Conversely, an effect that is not significant may be large.<sup>162</sup> By inquiring into the magnitude of an effect, courts can avoid being misled by *p*-values. To focus attention where it belongs—on the actual size of an effect and the reliability of the statistical analysis—the court may ask for an *interval estimate*.<sup>163</sup> Seeing a plausible range of values for the quantity of interest enables the court to decide whether this quantity is large or small and to consider the statistical uncertainty in the estimate.

In our example, the 95% confidence interval for the difference in the pass rates of men and women ranged from 0 to 40 percentage points. Our best estimate is that the pass rate for men is 20 percentage points higher than that for women; and the difference may plausibly be as little as 0 or as much as 40 percentage points. The *p*-value does not yield this information. The confidence interval contains the information provided by a significance test—and more.<sup>164</sup> For instance, significance at the .05 level can be read off the 95% confidence interval. In our example, 0 is at the extreme edge of the 95% confidence interval; thus, we have significant evidence that the true difference in pass rates between male and female applicants is not 0. But there are values very close to 0 inside the interval. This may help us consider whether the difference is practically significant.

In contrast, suppose a significance test fails to reject the null hypothesis. The confidence interval may prevent the mistake of thinking there is positive proof for the null hypothesis. To illustrate, let us change our example slightly: 29 men and 20 women passed the test. The 95% confidence interval goes from -2 to 38 percentage points. Because a difference of 0 falls within the 95% confidence interval, the null hypothesis—that the true difference is 0—cannot be rejected at the .05 level. However, the interval extends to 38 percentage points, indicating

163. An interval estimate may be composed of a *point estimate* —like the sample mean used to estimate the population mean—together with its standard error, or the two can be combined into a confidence interval. The first alternative may be more informative.

164. Accordingly, it has been argued that courts should demand confidence intervals (whenever they can be computed) to the exclusion of explicit significance tests and *p*-values. Kaye, *supra* note 128, at 1349 n.78; f. Bailar & Mosteller, *supra* note 147, at 317.

<sup>161.</sup> See, e.g., Persi Diaconis, Theories of Data Analysis: From Magical Thinking Through Classical Statistics, in Exploring Data Tables, Trends, and Shapes 1, 8–9 (David C. Hoaglin et al. eds., 1985); Frank T. Denton, Data Mining As an Industry, 67 Rev. Econ. & Stat. 124 (1985); David A. Freedman, A Note on Screening Regression Equations, 37 Am. Statistician 152 (1983). Intuition may suggest that the more variables included in the model, the better. However, this idea often seems to be wrong. Complex models may reflect only accidental features of the data. Standard statistical tests offer little protection against this possibility when the analyst has tried a variety of models before settling on the final specification.

<sup>162.</sup> See supra § IV.B.1.

that the population difference could be substantial. Lack of significance does not exclude this possibility.<sup>165</sup>

#### e. What are the other explanations for the findings?

The *p*-value of a statistical test is computed on the basis of a model for the data—the null hypothesis. Usually, the test is made in order to argue for the *al*-ternative hypothesis—another model. However, on closer examination, both models may prove to be unreasonable.<sup>166</sup> A small *p*-value indicates the occurrence of something besides random error; the alternative hypothesis should be viewed as one possible explanation out of many for the data.<sup>167</sup>

In Mapes Casino, Inc. v. Maryland Casualty Co.,<sup>168</sup> for example, the court recognized the importance of explanations that the proponent of the statistical evidence had failed to consider. In this action to collect on an insurance policy, Mapes Casino sought to quantify the amount of its loss due to employee defalcation. The casino argued that certain employees were using an intermediary to cash in chips at other casinos. It established that over an eighteen-month period the win percentage at its craps tables was 6%, compared with an expected value of 20%. The court recognized that the statistics were probative of the fact that something was wrong at the craps tables—the discrepancy was too large to explain as the mere product of random chance. However, the court was not convinced by the plaintiff's alternative hypothesis. The court pointed to other possible explanations (such Runyonesque activities as skimming, scamming, and crossroading) that might have accounted for the discrepancy without implicating the suspect employees.<sup>169</sup> In short, rejection of the null hypothesis does not leave the proffered alternative hypothesis as the only viable explanation for the data.<sup>170</sup>

165. We have used two-sided intervals corresponding to two-tailed tests. One-sided intervals corresponding to one-tailed tests also are available.

166. Often, the null and alternative hypotheses are statements about possible ranges of values for parameters in a common statistical model. Computations of standard errors, *p*-values, and power all take place within the confines of this basic model. The statistical analysis looks at the relative plausibility for competing values of the parameters but makes no global assessment of the reasonableness of the basic model. Inquiry by the court may be advisable.

167. See, e.g., Paul Meier & Sandy Zabell, Benjamin Peirce and the Howland Will, 75 J. Am. Stat. Ass'n 497 (1980) (competing explanations in a forgery case). Outside the legal realm there are many intriguing examples of the tendency to think that a small *p*-value is definitive proof of an alternative hypothesis, even though there are other plausible explanations for the data. See, e.g., Freedman et al., *supra* note 12, at 503–04; C.E.M. Hansel, ESP: A Scientific Evaluation (1966).

168. 290 F. Supp. 186 (D. Nev. 1968).

169. *Id.* at 193. Skimming consists of taking off the top before counting the drop; scamming is cheating by collusion between dealer and player; and crossroading involves professional cheaters among the players. *Id.* In plainer language, the court seems to have ruled that the casino itself might be cheating, or there could have been cheaters other than the particular employees identified in the case. At the least, plaintiff's statistical evidence did not rule such possibilities out of bounds.

170. Compare EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 312 & n.9, 313 (7th Cir. 1988) (EEOC's regression studies showing significant differences did not establish liability because surveys and testimony supported the rival hypothesis that women generally had less interest in commission sales positions) with EEOC v. General Tel. Co. of N.W., Inc., 885 F.2d 575 (9th Cir. 1989) (unsubstantiated rival hypothesis of lack of interest in nontraditional jobs insufficient to rebut prima facie case of gender discrimination), cert. denied , 498 U.S. 950 (1990); cf. supra § II.C (problem of confounding).

# C. Posterior Probabilities

Standard errors, *p*-values, and significance tests are often used to assess random error. These assessments rely on the sample data and are justified in terms of the operating characteristics of the statistical procedures.<sup>171</sup> However, this frequen tist approach does not permit the statistician to compute the probability that a particular hypothesis is correct, given the data.<sup>172</sup>

In the Bayesian approach, probabilities represent subjective degrees of belief rather than objective facts. This approach allows the calculation of posterior probabilities for various hypotheses given the data.<sup>173</sup> However, such probabilities must be "personal," for they reflect not just the data, but also the statistician's, or perhaps the fact finder's, <sup>174</sup> subjective *prior probabilities* —that is, degree of belief about the hypotheses, prior to obtaining the data.<sup>175</sup>

171. Operating characteristics are the expected value and standard error of estimators, probabilities of error for statistical tests, and so forth.

172. See supra § IV.B.1. Consequently, quantities such as *p*-values or confidence levels cannot be compared directly with numbers like .95 or .50 that might be thought to quantify the burden of persuasion in civil or criminal cases. See David H. Kaye, Hypothesis Testing in the Courtroom , in Contributions to the Theory and Application of Statistics 331 (Alan E. Gelfand ed., 1987); David H. Kaye, Apples and Oranges : Confidence Coefficients and the Burden of Persuasion , 73 Cornell L. Rev. 54 (1987).

173. See, e.g., Joseph B. Kadane, A Statistical Analysis of Adverse Impact of Employer Decisions, 85 J. Am. Stat. Ass'n 925 (1990) (analysis of data in an age discrimination case); David H. Kaye, Statistical Evidence of Discrimination, 77 J. Am. Stat. Ass'n 773, 780 (1982) (Bayesian analysis of the data in Swain v. Alabama, 380 U.S. 202 (1965)); Kaye, supra note 156, at 848–52 (analysis of data from Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977)).

174. E.g., Michael O. Finkelstein & William B. Fairley, A Bayesian Approach to Identification Evidence, 83 Harv. L. Rev. 489 (1970); Kadane, supra note 173. But see Laurence H. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 Harv. L. Rev. 1329 (1971) (arguing that efforts to describe the impact of evidence on a juror's subjective probabilities would unduly impress jurors and undermine the presumption of innocence and other legal values).

175. See generally David H. Kaye, Introduction: What is Bayesianism?, in Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism 1 (Peter Tillers & Eric D. Green eds., 1988); Brian Skyrms, Choice and Chance: An Introduction to Inductive Logic (3d ed. 1986).

Reference Manual on Scientific Evidence

To date, such analyses rarely have been used in court,<sup>176</sup> and the question of their forensic value has been aired primarily in academic literature.<sup>177</sup> Some statisticians favor Bayesian methods, <sup>178</sup> and some legal commentators have proposed their use in certain kinds of cases.<sup>179</sup>

176. See The Evolving Role of Statistical Assessments as Evidence in the Courts, *supra* note 1, at 193. The one area where Bayesian techniques are often used is parentage testing in civil cases. *Compare* State v. Spann, 617 A.2d 247, 257 (N.J. 1993) *with* Plemel v. Walter, 735 P.2d 1209, 1215 (Or. 1987).

177. See, e.g., Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism, supra note 175; Symposium, Decision and Inference in Litigation, 13 Cardozo L. Rev. 253 (1991).

178. Donald Å. Berry, Inferences Using DNA Profiling in Forensic Identification and Paternity Cases, 6 Stat. Sci. 175, 180 (1991); Stephen E. Fienberg & Joseph B. Kadane, The Presentation of Bayesian Statistical Analyses in Legal Proceedings, 32 Statistician 88 (1983); Stephen E. Fienberg & Mark J. Schervish, The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking , 66 B.U. L. Rev. 771 (1986); Kadane, supra note 173; Kathryn Roeder, DNA Fingerprinting: A Review of the Controversy, 9 Stat. Sci. 222 (1994); cf. I. W. Evett et al., An Illustration of the Advantages of Efficient Statistical Methods for RFLP Analysis in Forensic Science , 52 Am. J. Hum. Genet. 498, 499 (1993) (favoring presentation of the likelihood ratio for expressing the weight of DNA evidence). Nevertheless, many statisticians question the general applicability of Bayesian techniques: The results of the analysis may be substantially influenced by the prior probabilities, which in turn may be quite arbitrary.

179. E.g., Ira Mark Ellman & David Kaye, Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity?, 54 N.Y.U. L. Rev. 1131 (1979); David H. Kaye, DNA Evidence: Probability, Population Genetics, and the Courts, 7 Harv. J.L. & Tech. 101 (1993); Joseph C. Bright et al., Statistical Sampling in Tax Audits, 13 Law & Soc. Inquiry 305 (1988); authorities cited supra note 174.

Bayesian procedures are sometimes defended on the ground that the beliefs of any rational observer must conform to the Bayesian rules. However, the definition of "rational" is purely formal. See Peter C. Fishburn, The Axioms of Subjective Probability, 1 Stat. Sci. 335 (1986); David Kaye, The Laws of Probability and the Law of the Land, 47 U. Chi. L. Rev. 34 (1979).

This page left blank intentionally for proper pagination when printing two-sided

# Appendix: Technical Details on the Standard Error, the Normal Curve, and the *P*-Value

This appendix describes several calculations for our pass rate example. The population consisted of all 5,000 men and 5,000 women in the applicant pool. By way of illustration, suppose that the pass rates for these men and women were 60% and 35%, respectively; so the population difference is 60 - 35 = 25 percentage points. We chose 50 men and 50 women at random from the population. In our sample, the pass rate for the men was 58%, and the pass rate for the women was 38%; thus, the sample difference was 58 - 38 = 20 percentage points. Another sample might have pass rates of 62% and 36%, for a sample difference of 62 - 36 = 26 percentage points. And so forth.

In principle, we can consider the set of all possible samples from the population and make a list of the corresponding differences. This is a long list. Indeed, the number of distinct samples of 50 men and 50 women that can be formed is immense—nearly  $5 \times 10^{240}$ , or 5 followed by 240 zeros. Our sample difference was chosen at random from this list. Statistical theory enables us to make some precise statements about the list and hence about the chances in the sampling procedure.

- The average of the list—that is, the average of the differences over the  $5 \times 10^{240}$  possible samples—equals the difference between the pass rates of all 5,000 men and 5,000 women. In more technical language, the expected value of the sample difference equals the population difference. Even more tersely, the sample difference is an unbiased estimator of the population difference.
- The standard deviation (SD) of the list—that is, the standard deviation of the differences over the  $5 \times 10^{240}$  possible samples—is equal to: <sup>180</sup>

<sup>180.</sup> See, e.g., Freedman et al., supra note 12, at 337; Moore & McCabe, supra note 57, at 590–91. The standard error for the sample difference equals the standard deviation of the list of all possible sample differences, making the connection between standard error and standard deviation. If we drew two samples at random, the difference between them would be on the order of  $\sqrt{2} \approx 1.4$  times this standard deviation. The standard error can therefore be used to measure reproducibility of sample data. See supra notes 125–26. On the standard deviation, see supra § III.E; see also Freedman et al., supra note 12, at 67.

$$\sqrt{\frac{5,000 - 50}{5,000 - 1}} \times \sqrt{\frac{P_{\text{men}}(1 - P_{\text{men}})}{50}} + \frac{P_{\text{women}}(1 - P_{\text{women}})}{50}$$
(1)

In equation (1),  $P_{\text{men}}$  stands for the proportion of the 5,000 male applicants who would pass the exam, and  $P_{\text{women}}$  stands for the corresponding proportion of women. When  $P_{\text{men}} = 60\%$  and  $P_{\text{women}} = 35\%$ , the standard deviation of the sample differences would be 9.6 percentage points:

$$\sqrt{\frac{5,000-50}{5,000-1}} \times \sqrt{\frac{.60(1-.60)}{50}} + \frac{.35(1-.35)}{50} = .096$$
(2)

# Figure 12

The distribution of the sample difference in pass rates when  $P_{\text{men}} = 60\%$  and  $P_{\text{women}} = 35\%$ .

#### insert figure 12 here

Figure 12 shows the histogram for the sample differences.<sup>181</sup> The graph is drawn so the area between two values gives the relative frequency of sample dif-

Reference Manual on Scientific Evidence

<sup>181.</sup> The probability histogram in Figure 12 shows the distribution of the sample differences, indicating the relative likelihood of the various ranges of possible values; likelihood is represented by area. The lower horizontal scale shows standard units, that is, deviations from the expected value relative to the standard error. In our example, the expected value is 25 percentage points and the standard error is 9.6 percentage points. Thus, 35 percentage points would be expressed as (35 - 25)/9.6 = 1.04 standard units. The vertical scale shows probability per standard unit. See Freedman et al., supra note 12, at 75, 289.

ferences falling in that range, among all  $5 \times 10^{240}$  possible samples. For instance, take the range from 20 to 30 percentage points. About half the area under the histogram falls into this range. Therefore, given our assumptions, there is about a 50% chance that a sample chosen at random will have a male-female pass rate difference between 20 and 30 percentage points. The "central limit theorem" establishes that the histogram for the sample differences follows the normal curve, at least to a good approximation. Figure 12 shows this curve for comparison. The main point is that chances for the sample difference can be approximated by areas under the normal curve.

Generally, we do not know the pass rates  $P_{\text{men}}$  and  $P_{\text{women}}$  in the population. We chose 60% and 35% just by way of illustration. Statisticians would use the pass rates in the sample—58% and 38%—to estimate the pass rates in the population. Substituting the sample pass rates in equation (1) yields:

$$\sqrt{\frac{5,000 - 50}{5,000 - 1}} \times \sqrt{\frac{.58(1 - .58)}{50}} + \frac{.38(1 - .38)}{50} = .097$$
(3)

That is about 10 percentage points—the standard error reported in section IV.A.2.<sup>182</sup>

To sum up, the histogram for the sample differences follows the normal curve, centered at the population difference. The spread is given by the standard error. That is why confidence levels can be based on the standard error, with confidence levels read off the normal curve: 68% of the area under the curve is between -1 and 1, 95% is between -2 and 2, and 99.7% is between -3 and 3, approximately.

We turn to *p*-values.<sup>183</sup> Consider the null hypothesis that the men and women in the population have the same overall pass rates. In that case, the sample differences are centered at 0, because  $P_{\text{men}} - P_{\text{women}} = 0$ . Since the overall pass rate in the sample is 48%, we use this value to estimate both  $P_{\text{men}}$  and  $P_{\text{women}}$  in equation (1):

$$\sqrt{\frac{5,000-50}{5,000-1}} \times \sqrt{\frac{.48(1-.48)}{50}} + \frac{.48(1-.48)}{50} = .099$$
(4)

Again, the standard error (SE) is about 10 percentage points. The observed difference of 20 percentage points is 20/10 = 2.0 SEs. As shown in Figure 13, differences of that magnitude or larger have about a 5% chance of occurring:

<sup>182.</sup> There is little difference between equations (2) and (3)—the standard error does not depend strongly on the pass rates. 183. See supra § IV.B.1.

About 5% of the area under the normal curve lies beyond  $\pm 2$ . (In Figure 13, this tail area is shaded.) The *p*-value is about 5%.<sup>184</sup>

# Figure 13

*P*-value for observed difference of 20 percentage points, computed using the null hypothesis. The chance of getting a sample difference of 20 points in magnitude (or more) is about equal to the area under the normal curve beyond  $\pm 2$ . That shaded area is about 5%.

#### insert figure 13 here

Finally, we calculate power.<sup>185</sup> We are making a two-tailed test at the .05 level. Instead of the null hypothesis, we assume an alternative: In the applicant pool, 55% of the men would pass, and 45% of the women. So there is a difference of 10 percentage points between the pass rates. The distribution of sample differences would now be centered at 10 percentage points (see Figure 14). Again, the sample differences follow the normal curve. The true SE is about 10 percentage points by equation (1), and the SE estimated from the sample will be about the same. On that basis, only sample differences larger than 20 percentage

Reference Manual on Scientific Evidence

<sup>184.</sup> Technically, the *p*-value is the chance of getting data as extreme as, or more extreme than, the data at hand. See *supra* § IV.B.1. That is the chance of getting a difference of 20 percentage points or more on the right, together with the chance of getting -20 or less on the left. This chance equals the area under the histogram to the right of 19, together with the area to the left of -19. (The rectangle whose area represents the chance of getting a difference of 20 is included, and likewise for the rectangle above -20.) The area under the histogram in turn may be approximated by the area under the normal curve beyond ±1.9, which is 5.7%. See, e.g., Freedman et al., *supra* note 12, at 291. Keeping track of the edges of the rectangles is called the "continuity correction." As a technical matter, the histogram is computed assuming pass rates of 48% for the men and the women. Other values could be dealt with in a similar way. See *infra* note 187.

<sup>185.</sup> See supra note 152.

points or smaller than -20 points will be declared significant. <sup>186</sup> About 1/6 of the area under the normal curve in Figure 14 lies in this region. <sup>187</sup> Therefore, the power of the test against the specified alternative is only about 1/6. In the figure, it is the shaded area that corresponds to power.

Figures 12, 13, and 14 have the same shape: The central limit theorem is at work. However, the histograms are centered differently, because the values of  $P_{\text{men}}$  and  $P_{\text{women}}$  are different in all three figures. Figure 12 is centered at 25 percentage points, reflecting our illustrative values of 60% and 35% for the pass rates. Figure 13 is centered at 0, because it is drawn according to the requirements of the null hypothesis. Figure 14 is centered at 10 percentage points, because the alternative hypothesis is used to determine the center, rather than the null hypothesis.

# Figure 14

Power when  $P_{\text{men}} = 55\%$  and  $P_{\text{women}} = 45\%$ . The chance of getting a significant difference (at the 5% level, two-tailed) is about equal to the area under the normal curve, to the right of +1 or to the left of -2. That shaded area is about 1/6. Power is about 1/6, or 17%.

<sup>186.</sup> The null hypothesis asserts a difference of 0: In Figure 13, 20 percentage points is 2 SEs to the right of the value expected under the null hypothesis; likewise, -20 is 2 SEs to the left. However, Figure 14 takes the alternative hypothesis to be true; on that basis, the expected value is 10 instead of 0, so 20 is 1 SE to the right of the expected value, while -20 is 3 SEs to the left.

<sup>187.</sup> Let t = sample difference/SE, where the SE is estimated from the data, as in equation (4). One for mal version of our test rejects the null hypothesis if  $|t| \ge 2$ . To find the power, we replace the estimated SE by the true SE, computed as in equation (1), and we replace the probability histogram by the normal curve. These approximations are quite good. The size can be approximated in a similar way, given a common value for the two population pass rates. Of course, more exact calculations are possible. See *supra* note 184.

This page left blank intentionally for proper pagination when printing two-sided

# **Glossary of Terms**

The following terms and definitions are adapted from a variety of sources, including: Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990), and David Freedman et al., Statistics (2d ed. 1991).

- Alpha (α). Also, size. A symbol often used to denote the probability of a Type I error. See Type I Error.
- Alternative Hypothesis. A statistical hypothesis that is contrasted with the null hypothesis in a significance test. See Statistical Hypothesis; Significance Test.
- Area Sample. An area sample is a probability sample in which the sampling frame is a list of geographical areas (i.e., one makes a list of areas, chooses some at random, and interviews people in the selected areas). This is a cost effective way to draw a sample of people. See Probability Sample; Sampling Frame.
- Bayes' Rule. An investigator may start with a subjective probability (the "prior") that expresses degrees of belief about a parameter or a hypothesis. Then data are collected according to some statistical model. Bayes' rule gives a procedure for combining the prior with the data to compute the "posterior" probability, which expresses the investigator's beliefs about the parameter or hypothesis given the data.
- Beta ( $\beta$ ). A symbol used sometimes to denote power and sometimes to denote the probability of a type II error. See Type II Error; Power.
- *Bias.* A systematic tendency for an estimate to be too high or too low. An estimate is unbiased if the bias is 0. See Nonsampling Error.
- Bin. A class interval in a histogram. See Class Interval; Histogram.
- Binary Variable. A variable that has only two possible values (e.g., the gender of an employee).
- Binomial Distribution. A distribution for the number of occurrences in repeated, independent trials where the probabilities are fixed. For example, the number of heads out of 100 tosses of a coin follows a binomial distribution. (The probability of heads is 1/2 on each toss.) When the probability is not too
close to 0 or 1 and the number of trials is large, the binomial distribution has about the same shape as the normal distribution. See Normal Distribution; Poisson Distribution.

- Bootstrapping. A procedure for estimating sampling error by generating a simulated population from the sample, then repeatedly drawing samples from this population.
- Categorical Data; Categorical Variable. See Qualitative Variable.
- Central Limit Theorem. Shows that under suitable conditions, the probability histogram for a sum (or average, or rate) will follow the normal curve.
- Chance Error. See Random Error; Sampling Error.
- *Chi-Squared* ( $\chi^2$ ). A statistic that measures the distance between the data and expected values computed from a statistical model. If  $\chi^2$  is too large to explain by chance, the data contradict the model. The definition of large depends on the context. See Statistical Hypothesis; Significance Test.
- *Class Interval.* Also, bin. The base of a rectangle in a histogram; the area of the rectangle shows the percentage of observations in the class interval. See Histogram.
- *Cluster Sample.* A type of random sample. For example, a statistician might take households at random, then interview all the people in the selected households. This is a cluster sample of people: A cluster consists of all the people in a selected household. Generally, clustering reduces the cost of interviewing.
- Coefficient of Determination. A statistic (more commonly known as R<sup>2</sup>) that describes how well a regression equation fits the data. See R-Squared.
- Coefficient of Variation. A statistic that measures spread relative to the center of the distribution: SD/average, or SE/expected value.
- *Conditional Probability.* The probability that one event will occur given that another has occurred.
- Confidence Coefficient. See Confidence Interval.
- Confidence Interval. An estimate, expressed as a range, for a quantity in a population. If an estimate from a large sample is unbiased, a 95% confidence interval is the range from two standard errors below to two standard errors above the estimate. Intervals obtained this way cover the true value about 95% of the time, and 95% is the confidence level, or the confidence coefficient. See Unbiased Estimator; Standard Error.

Confidence Level. See Confidence Interval.

Confounding. See Confounding Variable; Observational Study.

- Confounding Variable; Confounder. A variable that is correlated with the independent variables and the dependent variable. When confounding is suspected, an association between the dependent and independent variable may not be causal. See Controlled Experiment; Observational Study.
- Consistency; Consistent. See Consistent Estimator.
- Consistent Estimator. An estimator that tends to become more and more accurate as the sample size grows. (Inconsistent estimators, which do not become more accurate as the sample size grows, are generally not used by statisticians.)
- Content Validity. The extent to which a skills test is appropriate to its intended purpose, as evidenced by a set of questions that adequately reflect the domain being tested.
- Continuous Variable. A variable that has arbitrarily fine gradations, such as a person's height.
- Control Group. See Controlled Experiment.
- Control for. Statisticians "control for" the effects of confounding variables in nonexperimental data by making comparisons for smaller and more homogeneous groups of subjects or by using regression models. See Regression Model.
- Controlled Experiment. An experiment where the investigators determine which subjects are put into the treatment group and which are put into the control group. Subjects in the treatment group are exposed by the investigators to some influence—the treatment; those in the control group are not so exposed. For instance, in an experiment to evaluate a new drug, subjects in the treatment group are given the drug, while subjects in the control group are given some other therapy. The outcomes in the two groups are compared to see whether the new drug works. Randomization—that is, randomly assigning subjects to each group—is usually the best way to assure that any observed difference between the two groups comes from the treatment rather than preexisting differences. Of course, in many situations, a randomized controlled experiment is impractical, and investigators must then rely on observational studies.
- Convenience Sample. Also, grab sample. A nonrandom sample of units; for in stance, for a "mall sample," the interviewer picks respondents from the crowd in a shopping mall.
- Correlation Coefficient. A number between –1 and 1 that indicates the extent of the linear association between two variables. Often, the correlation coefficient is abbreviated as *r*.

- *Covariance.* A quantity that describes the statistical interrelationship of two variables.
- *Covariate.* A variable that is related to other variables of primary interest in a study.
- *Criterion.* The variable against which a skills test or other selection procedure is validated. See Predictive Validity.
- Data. Observations or measurements, usually of units in a sample taken from a larger population.
- Dependent Variable. See Independent Variable; Regression Model.
- Descriptive Statistic. A statistic, such as the mean or the standard deviation, used to summarize data.
- Differential Validity. Differences in the relationship between skills test scores and outcome measures across different subgroups of test takers.
- Discrete Variable. A variable that has only a finite number of possible values, such as the number of automobiles owned by a household.
- Random Disturbance Term. See Error Term.
- Double -Blind Experiment . An experiment with human subjects in which neither the diagnosticians nor the subjects know who is in the treatment group or the control group. This is accomplished by giving a placebo treatment to subjects in the control group.
- Dummy Variable. Generally, a dummy variable takes only the values 0 or 1 and distinguishes one group of interest from another. For example, in a regression study of salary differences between men and women in a firm, the analyst may include a dummy variable for gender and statistical controls such as education and experience to adjust for productivity differences between men and women. The dummy variable would be defined as 1 for the men, 0 for the women. See Regression Model.
- Econometrics. The statistical study of economic issues.
- Epidemiology. Statistical study of disease or injury in human populations.
- *Error Term.* The part of a statistical model that describes random error (i.e., the impact of chance factors unrelated to variables in the model). In econometric models, the error term is called a random disturbance term.
- *Estimator.* A sample statistic used to estimate a population parameter. For instance, the sample mean commonly is used to estimate the population mean. The term "estimator" connotes a statistical procedure, while an "estimate" connotes a particular numerical result.

Expected Value. See Random Variable.

- *Fisher's Exact Test.* When comparing two sample proportions (e.g., the proportions of whites and blacks getting a promotion), an investigator may want to test the null hypothesis that promotion does not depend on race. Fisher's exact test is one way to arrive at a *p*-value. The calculation is based on the hypergeometric distribution. See Hypergeometric Distribution; Statistical Hypothesis; Significance Test; *p*-Value.
- *Fixed Significance Level.* Also, alpha, size. A preset level, such as 0.05 or 0.01. If the *p*-value of a test falls below this level, the result is deemed statistically significant. See Significance Test.
- Frequency Distribution. Shows how often specified values occur in a data set.
- Gaussian Distribution. See Normal Distribution.
- General Linear Model. Expresses the dependent variable as a linear combination of the independent variables plus an error term whose components may be dependent and have differing variances. See Error Term; Linear Combination; Regression Model; Variance.
- Grab Sample. See Convenience Sample.
- Heteroscedastic. See Scatter Diagram.
- Histogram. A plot showing how observed values fall within specified intervals, called bins or class intervals. Generally, matters are arranged so the area under the histogram, but over a class interval, gives the frequency or relative frequency of data in that interval. In a probability histogram, the area gives the chance of observing a value that falls in the corresponding interval.
- Homoscedastic. See Scatter Diagram.
- Hypergeometric Distribution. Suppose a sample is drawn at random without replacement from a finite population. The number of times that items of a certain type come into the sample is given by the hypergeometric distribution.
- Hypothesis Test. See Significance Test.
- *Independence.* Events are independent when the probability of one is unaffected by the occurrence or nonoccurrence of the other.
- Independent Variable. The independent variable is used in a regression model to predict values of the dependent variable. For instance, the unemployment rate has been used as the independent variable in a model for predicting the crime rate; the latter is the dependent variable in this application. See Regression Model.
- Indicator Variable. See Dummy Variable.
- Interval Estimate. A confidence interval; or, a point estimate coupled with a standard error. See Confidence Interval; Standard Error.

Least Squares. See Least-Squares Estimator; Regression Model.

- Least-Squares Estimator. An estimator that is computed by minimizing the sum of the squared residuals. See Residual.
- Linear Combination. To obtain a linear combination of two variables, the first variable is multiplied by some constant, the second variable is multiplied by another constant, and the two products are added (e.g., 2u + 3v is a linear combination of u and v).
- Loss Function. Statisticians may evaluate estimators according to a mathematical formula involving the errors (i.e., differences between actual values and estimated values). The loss may be the total of the squared errors or the total of the absolute errors, etc. Loss functions seldom quantify real losses but may be useful summary statistics and may prompt the construction of useful statistical procedures.
- Mean. The mean is one way to find the center of a batch of numbers: Add up the numbers, and divide by how many there are. Weights may be employed, too, as in weighted mean or weighted average. Also, the expected value of a random variable; average. See Random Variable.
- Median. The median is another way to find the center of a batch of numbers. The median is the fiftieth percentile. Half the numbers are larger, and half are smaller. (To be very precise, at least half the numbers are greater than or equal to the median; at least half the numbers are less than or equal to the median; for small data sets, the median may not be uniquely defined.)
- *Meta-Analysis.* Attempts to combine information from all studies in a certain collection.
- Mode. The most commonly observed value.
- Multicollinearity. Also, collinearity. The existence of correlations among the independent variables in a regression model. See Independent Variable; Regression Model.
- Multiple Comparison. An examination of more than one test statistic relating to the same data set. Multiple comparisons complicate the interpretation of a *p*-value. For example, if twenty divisions of a company are examined for disparities, and one division is found to have a disparity significant at the 0.05 level, the result is not surprising; indeed, it should be expected under the null hypothesis.
- Multiple Correlation Coefficient. A number that indicates the extent to which one variable can be predicted as a linear combination of other variables. Its magnitude is the square root of R<sup>2</sup>. See Linear Combination; R-Squared; Regression Model.

- Multiple Regression. A regression equation that includes two or more independent variables. See Regression Model.
- Multistage Cluster Sample. A probability sample drawn in stages, usually after stratification; the last stage will involve drawing a cluster. See Cluster Sample; Probability Sample; Stratified Random Sample.
- Natural Experiment. An observational study in which treatment and control groups have been formed by some natural development; however, the assignment of subjects to groups is judged akin to randomization. See Observational Study.
- Nonsampling Error. A catch-all term for sources of error in a survey, other than sampling error. Nonsampling errors cause bias. One example is selection bias: The sample is drawn in a way that tends to exclude certain subgroups in the population. A second example is nonresponse bias: People who do not respond to a survey are usually different from respondents. A final example: Response bias arises, for instance, if the interviewer uses a loaded question.
- Normal Distribution. Also, Gaussian distribution. The density for this distribution is the famous bell-shaped curve. Statistical terminology notwithstanding, there is nothing wrong with a distribution that differs from the normal.
- Null Hypothesis. A hypothesis that there is no difference between two groups from which samples are drawn. See Statistical Hypothesis.
- Observational Study. A study in which subjects select themselves into groups; investigators then compare the outcomes for the different groups. For example, studies of smoking are generally observational. Subjects decide whether or not to smoke; the investigators compare the death rate for smokers with the death rate for nonsmokers. In an observational study, the groups may differ in important ways that the investigators do not notice; controlled experiments minimize this problem. The critical distinction is that in a controlled experiment, the investigators intervene to manipulate the circumstances of the subjects; in an observational study, the investigators are passive observers. (Of course, running a good observational study is hard work and may be quite useful.)
- Observed Significance Level. See p-Value.
- Odds. The probability that an event will occur divided by the probability that it will not. For example, if the chance of rain tomorrow is 2/3, then the odds on rain are (2/3)/(1/3) = 2/1, or 2 to 1.
- Odds Ratio. A measure of association, often used in epidemiology. For instance, if 10% of all people exposed to a chemical develop a disease, compared with 5% of people who are not exposed, the odds of the disease in the exposed group are 10/90 = 1/9, compared with 5/95 = 1/19 in the unexposed group. The odds ratio is 19/9 = 2.1. An odds ratio of 1 indicates no association.

- One-Sided Hypothesis. Excludes the possibility that a parameter could be, for example, less than the value asserted in the null hypothesis. A one-sided hypothesis leads to a one-tailed test. See Statistical Hypothesis; Significance Test.
- One-Tailed Test. See Significance Test.
- *Outlier.* An observation that is far removed from the bulk of the data. Outliers may indicate a faulty measurement; they may exert undue influence on a summary statistic, such as the mean or the correlation coefficient.
- *p*-Value. The output of a statistical test. The probability of getting, just by chance, a test statistic as large as or larger than the observed value. Large *p*-values are consistent with the null hypothesis; small *p*-values undermine this hypothesis. However, *p* itself does not give the probability that the null hypothesis is true. If *p* is smaller than 5%, the result is said to be statistically significant. If *p* is smaller than 1%, the result is highly significant. The *p*-value is also called the observed significance level. See Statistical Hypothesis; Significance Test.
- Parameter. A numerical characteristic of a population or of a model. See Probability Model.
- Percentile. To get the 90th percentile, for instance, of a data set, the data are arrayed from the smallest value to the largest. Then 90% of the values fall below the 90th percentile, and 10% fall above. (To be very precise, at least 90% of the data are at the 90th percentile or below; at least 10% of the data are at the 90th percentile or above.) The 50th percentile is the median. When the LSAT first was scored on a 10–50 scale in 1982, a score of 32 placed a test taker at the 50th percentile; a score of 40 was at the 90th percentile (approximately).
- Point Estimate. An estimate of the value of a quantity expressed as a single number.
- Poisson Distribution. The Poisson distribution is a limiting case of the binomial distribution, when the number of trials is large and the common probability is small. The parameter of the approximating Poisson distribution is the number of trials times the common probability, which gives the "expected" number of events. When this number is large, the Poisson distribution may be approximated by a normal distribution.

Population. Also, universe. All the units of interest to the researcher.

Posterior Probability. See Bayes' Rule.

*Power.* The probability that a statistical test will reject the null hypothesis. To compute power, the analyst has to fix the size of the test and specify parameter values outside the range given in the null hypothesis. A powerful test has

a good chance of detecting an effect, when there is an effect to be detected. See Significance Test; Beta.

- *Practical Significance.* Substantive importance. Statistical significance does not necessarily establish practical significance. Small differences can be statistically significant in large samples.
- *Predictive Validity.* A psychological or skills test has predictive validity to the extent that test scores are well correlated with later performance or, more generally, with outcomes that the test is intended to predict.
- Prior Probability. See Bayes' Rule.
- Probability. Chance, on a scale from 0 to 1. Impossibility is represented by 0, certainty by 1. Equivalently, chances may be quoted in percentages; 100% corresponds to 1; 5% to .05; and so forth.
- Probability Density. Describes the probability distribution for a random variable. The chance that the random variable falls in an interval equals the area below the density and above the interval. See Probability Distribution; Variable.
- Probability Distribution. Gives probabilities for possible values of a random variable. Often, the distribution is described in terms of the density. See Probability Density.
- Probability Histogram. See Histogram.
- Probability Model. Relates probabilities of outcomes to parameters; also, Statistical Model. The latter connotes unknown parameters.
- Probability Sample. A sample drawn from a sampling frame by some objective chance mechanism; each unit has a known probability of being sampled. Such samples are expensive to draw but minimize selection bias.
- *Psychometrics.* The study of psychological measurement and testing.
- Qualitative Variable; Quantitative Variable. A qualitative or categorical variable describes qualitative features of subjects in a study (e.g., marital status— never married, married, widowed, divorced, separated). A quantitative variable describes numerical features of the subjects (e.g., height, weight, income). This is not a hard-and-fast distinction, because qualitative features may be given numerical codes, as in a dummy variable. Quantitative variables may be classified as discrete or continuous. Concepts like the mean and the standard deviation apply only to quantitative variables. See Discrete Variable; Continuous Variable.
- Quartile. The 25th or 75th percentile. See Percentile.

- R-Squared (R<sup>2</sup>). Measures how well a regression equation fits the data. R<sup>2</sup> varies between 0 (no association) and 1 (perfect fit). Generally, R<sup>2</sup> does not measure the validity of underlying assumptions. See Regression Model.
- Random Error. Sources of error that are haphazard in their effect. These are reflected in the error term of a statistical model. Some authors refer to random error as chance error or sampling error. See Regression Model.
- Random Variable. A variable whose possible values occur according to some probability mechanism. For example, if you throw a pair of dice, the total number of spots is a random variable. The chance of two spots is 1/36, the chance of three spots is 2/36, and so forth; the most likely number is seven, with a chance of 6/36. The expected value of a random variable is the weighted average of the possible values; the weights are the probabilities. In our example, the expected value is

$$\frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7$$
$$+ \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 = 7$$

In many problems, the weighted average is computed with respect to the density; then sums must be replaced by integrals. The expected value need not be a possible value for the random variable. Generally, a random variable will be somewhere around its expected value, but it will be off (in either direction) by something like 1 standard error or so. See Standard Error.

Randomization. See Controlled Experiment.

- Randomized Controlled Experiment. A controlled experiment in which subjects are placed into the treatment and control groups at random—as if by lot. See Controlled Experiment.
- Range. The difference between the biggest and smallest values in a batch of numbers.
- Regression Coefficient. A constant in a regression equation. See Regression Model.
- Regression Diagnostics. Procedures intended to check whether the assumptions of a regression model are appropriate.
- Regression Equation. See Regression Model.
- *Regression Line.* The graph of a regression equation with only one dependent variable and one independent variable.
- Regression Model. A regression model attempts to combine the values of certain variables (the independent variables) to obtain expected values for another

variable (the dependent variable). A hypothetical example illustrates the idea. An analyst might try to predict salaries of employees in a firm using education, experience—and a dummy variable for gender, taking the value 1 for men and 0 for women. Here, salary is the dependent variable (the variable being predicted), while education, experience, and the dummy are the independent variables (the variables entered into the equation to make the predictions).

Sometimes, "regression model" refers to a statistical model for the data; if no qualifications are made, the model will generally be linear, and errors will be assumed independent, with common variance. At other times, "regression model" refers to an equation estimated from data.

In our example, salary (dollars per year) is predicted from education (years of schooling completed) and experience (years with the company)—along with the dummy variable man, taking the value 1 for male employees and 0 for female employees. The model is

salary = 
$$a + b \times education + c \times experience + d \times man + u$$
 (1)

Equation (1) is a statistical model for the data, with unknown parameters a, b, c, d; these parameters are regression coefficients; a often is called the intercept, and u is an error term, with a component for each employee.

The parameters in equation (1) are estimated from the data using least squares. If the estimated coefficient *d* for the dummy variable turns out to be positive and statistically significant (by a *t*-test), that would be taken as evidence of disparate impact: Men earn more than women, even after adjusting for differences in background factors that might affect productivity. Education and experience would be entered into equation (1) as statistical controls, precisely in order to claim that adjustment had been made for differences in background.

Suppose the estimated equation turns out as follows:

predicted salary = 
$$\$7,100 + \$1,300 \times \text{education} + \$2,200 \times \text{experience} + \$700 \times \text{man}$$
 (2)

According to equation (2), every extra year of education is worth on average \$1,300; similarly, every extra year of experience is worth on average \$2,200; and most important, men receive a premium of \$700 over women with the same education and experience, on average.

Some numerical examples will illustrate equation (2). A male employee with 12 years of education (high school) and 10 years of experience would have a predicted salary of

$$\$7,100 + \$1,300 \times 12 + \$2,200 \times 10 + \$700 \times 1 =$$
  
 $\$7,100 + \$15,600 + \$22,000 + \$700 = \$45,400$  (3)

A similarly situated female employee has a predicted salary of only

$$\$7,100 + \$1,300 \times 12 + \$2,200 \times 10 + \$700 \times 0 =$$
  
 $\$7,100 + \$15,600 + \$22,000 + \$0 = \$44,700$  (4)

Notice the impact of the dummy variable: \$700 is added to equation (3) but not to equation (4).

A male employee with 16 years of education (college) and 6 years of experience would have a predicted salary of

$$\$7,100 + \$1,300 \times 16 + \$2,200 \times 6 + \$700 \times 1 =$$
  
 $\$7,100 + \$20,800 + \$13,200 + \$700 = \$41,800$  (5)

A similarly situated female employee has a predicted salary of only

$$\$7,100 + \$1,300 \times 16 + \$2,200 \times 6 + \$700 \times 0 =$$
  
 $\$7,100 + \$20,800 + \$13,200 + \$0 = \$41,100$  (6)

In equation (1), u is an error term, with one component for each employee; these components are random errors. Equation (2) has corresponding residuals. For each employee, there is a difference (or residual) between the salary predicted from the equation and the actual salary:

$$actual = predicted + residual$$
 (7)

The residuals are approximations to the random errors in equation (1).

A critical step in the argument is stablishing that the coefficient d of the dummy variable in equation (1) is "statistically significant." This step depends on the statistical assumptions built into the model. For instance, each extra year of education is assumed to be worth the same (on average) across all levels of experience, for both men and women; similarly, each extra year of experience is worth the same across all levels of education, for both men and women; furthermore, the premium paid to men does not depend systematically on education or experience. Ability, quality of education, and quality of experience are assumed not to make any systematic difference to the predictions of the model.

Moreover, there are technical assumptions that must be made about the error term *u*: for instance, that its components—the random errors—are inde-

Reference Manual on Scientific Evidence

pendent from person to person in the data set but have the same variance. Some assumptions of this general nature will be found to underlie typical applications of regression techniques; such assumptions should be identified and their reasonableness assessed.

The term "predicted" in equation (2) has a specialized meaning, since the analyst has available the data being predicted. For that reason, statisticians often refer to "fitted values" rather than to "predicted values." See Random Error; Independence; Least Squares; Regression Model; Multiple Regression; *t*-Test; Dummy Variable; Random Variable; Variance.

- *Relative Risk.* A measure of association used in epidemiology. For instance, if 10% of all people exposed to a chemical develop a disease, compared with 5% of people who are not exposed, the disease occurs twice as frequently among the exposed people: The relative risk is 10%/5% = 2. A relative risk of 1 indicates no association.
- *Reliability.* The extent to which a measuring instrument gives the same results on repeated measurement of the same thing.
- *Residual.* The difference between an actual and a predicted value. The predicted value comes typically from a regression equation and also is called the "fitted value." See Regression Model; Independent Variable.
- *Risk.* Expected loss. "Expected" means on average, over the various data sets that could be generated by the statistical model under examination. Usually, risk cannot be computed exactly but has to be estimated, because the parameters in the statistical model are unknown and must be estimated. See Loss Function; Random Variable.
- *Robust.* A statistic or procedure that does not change much when data or assumptions are slightly modified.
- Sample. A set of units collected for study.
- Sample Size. The number of units in a sample.
- Sampling Distribution. The distribution of the values of a statistic, over all possible samples from a population. For example, suppose a random sample is drawn. Some values of the sample mean are more likely, others are less likely. The sampling distribution specifies the chance that the sample mean will fall in one interval rather than another.
- Sampling Error. A sample is part of a population. When a sample is used to estimate a numerical characteristic of the population, the estimate is likely to differ from the population value, because the sample is not a perfect microcosm of the whole. If the estimate is unbiased, the difference between the estimate and the exact value is sampling error. More generally,

estimate = true value + bias + sampling error

Sampling error is also called chance error or random error. See Standard Error.

- Sampling Frame. A list of units designed to represent the entire population as completely as possible. The sample is drawn from the frame.
- Scatter Diagram. Also, scatterplot, scattergram. A graph showing the relationship between two variables in a study; each dot represents one subject. One variable is plotted along the horizontal axis, the other variable is plotted along the vertical axis. A scatter diagram is homoscedastic when the spread is more or less the same inside any vertical strip. If the spread changes from one strip to another, the diagram is heteroscedastic.
- Sensitivity. In clinical medicine, the probability that a test for a disease will give a positive result given that the patient has the disease. Sensitivity is analogous to the power of a statistical test.
- Sensitivity Analysis. Analyzing data in different ways to see how results depend on methods or assumptions.
- Significance Level. See Fixed Significance Level; p-Value.
- Significance Test. Also, statistical test, hypothesis test, test of significance; statistical hypothesis; *p*-value; *t*-test. A significance test involves formulating a statistical hypothesis and a test statistic, computing a *p*-value, and comparing *p* with some preestablished value to decide if the test statistic is significant. The idea is to see whether the data conform to the predictions of the null hypothesis. Generally, a large test statistic goes with a small *p*-value, and small *p*-values would undermine the null hypothesis.

For instance, suppose that a random sample of male and female employees was given a skills test, and the mean scores of the men and women were different in the sample. To judge whether the difference is due to sampling error, a statistician might consider the implications of competing hypotheses about the difference in the population. The null hypothesis would say that on average, in the population, men and women have the same scores: The difference observed in the data is then just due to sampling error. A one-sided alternative hypothesis would be that on average, in the population, men score higher than women. A one-tailed test would reject the null hypothesis if the sample of men score substantially higher than the women—so much so that the difference is hard to explain on the basis of sampling error.

In contrast, the null could be tested against the two-sided alternative hypothesis that on average, in the population, men score differently than women—higher or lower. The corresponding two-tailed test would reject the null hypothesis if the sample of men score substantially higher—or substantially lower—than the women.

The one-tailed and two-tailed tests would both be based on the same data and use the same *t*-statistic. However, if the men in the sample score higher than the women, the one-tailed test would give a *p*-value only half as large as the two-tailed test (i.e., the one-tailed test would appear to give stronger evidence against the null hypothesis).

Significant. See p-Value; Practical Significance; Significance Test.

- Simple Random Sample. A random sample in which each unit in the sampling frame has the same chance of being sampled. For example, the statistician takes a unit at random (as if by lottery), sets it aside, takes another at random from what is left, and so forth.
- Size. The size of a statistical test is a synonym for alpha ( $\alpha$ ). See Alpha.
- Specificity. In clinical medicine, the probability that a test for a disease will give a negative result given that the patient does not have the disease. Specificity is analogous to  $1 \alpha$ , where  $\alpha$  is the significance level of a statistical test.
- Spurious Correlation. When two variables are correlated, one is not necessarily the cause of the other. The vocabulary and shoe size of children in elementary school, for instance, are correlated—but learning more words does not make their feet grow. Such noncausal correlations are said to be spurious. (Originally, the term seems to have been applied to the correlation between two rates with the same denominator: Even if the numerators are unrelated, the common denominator will create some association.)
- Standard Deviation (SD). Indicates how far a typical element deviates from the average. For instance, in round numbers, the average height of women aged eighteen and over in the United States is 5 feet 4 inches, and the SD is 3 inches. Typical woman are about 5 feet 4 inches in height; they are off this something like 3 inches.

For distributions that follow the normal curve, about 95% of the elements are in the range "mean -2 SD" to "mean +2 SD". Deviations from the average that exceed 3 or 4 SDs are extremely unusual. Many authors use standard deviation also to mean standard error.

- Standard Error (SE). Indicates the likely size of the sampling error in an estimate. Many authors use the term "standard deviation" instead of standard error.
- Standard Error of Regression. Indicates how actual values differ (in some average sense) from the fitted values in a regression model. See Regression Model.
- Standardization. See Standardized Variable.
- Standardized Variable. Transformed to have a mean of 0 and a variance of 1. This involves two steps: (1) subtract the mean, and (2) divide by the standard deviation.

- Statistic. A number that characterizes or summarizes data. A statistic refers to a sample; a parameter or a true value refers to a population or a probability model.
- Statistical Control. See Control for.
- Statistical Hypothesis. Data may be governed by a probability model; parameters are numerical characteristics describing features of the model. Generally, a statistical hypothesis is a statement about the parameters in a probability model. The null hypothesis may assert that certain parameters have specified values or fall in specified ranges; the alternative hypothesis would specify other values or ranges. The null hypothesis is compared with the data using a test statistic; the null hypothesis may be rejected if there is a statistically significant difference between the data and the predictions of the null hypothesis.

Typically, the investigator seeks to demonstrate the alternative hypothesis; the null hypothesis would explain the findings as a result of mere chance, and the investigator uses a significance test to rule out this explanation. See Significance Test.

Statistical Model. See Probability Model.

Statistical Significance. See p-Value; Significance Test.

- Statistically Significant. See p-Value.
- Stratified Random Sample. A type of probability sample. The analyst divides the population up into relatively homogeneous groups called strata and draws a random sample separately from each stratum.
- Stratum; Strata. See Stratified Random Sample.
- *t-Statistic*. A test statistic used to make the *t*-test. The *t*-statistic tells you how far away an estimate is from its expected value, relative to the standard error. The expected value is computed using the null hypothesis. Some authors refer to the *t*-statistic, others to the *z*-statistic, especially when the sample is large. A *t*-statistic larger than 2 or 3 in absolute value makes the null hypothesis rather unlikely—the estimate is too many standard errors away from its expected value. See Statistical Hypothesis; Significance Test; *t*-Test.
- *t*-Test. A statistical test based on the *t*-statistic. Large *t*-statistics are beyond the usual range of sampling error. For example, if *t* is larger than 2 or smaller than -2, the estimate is statistically significant at the 5% level: Such values of *t* are hard to explain on the basis of sampling error. The scale for *t*-statistics is tied to areas under the normal curve. For instance, a *t*-statistic of 1.5 is not very striking, because 13%, or 13/100, of the area under the normal curve is outside the range from -1.5 to 1.5. Conversely, *t* = 3 is remarkable: Only 3/1,000 of the area lies outside the range from -3 to 3. This discussion is

based on having a reasonably large sample; in that context, many authors refer to the *z*-test rather than the *t*-test.

For small samples drawn at random from a population known to be normal, the *t*-statistic follows "Student's *t*-distribution" (when the null hypothesis holds) rather than the normal curve; larger values of *t* are required to achieve significance. See Statistical Hypothesis; Significance Test; *p*-Value.

- Test Statistic. A statistic used to judge whether data conform to the null hypothesis. The parameters of a probability model determine expected values for the data; differences between expected values and observed values are measured by a test statistic. Test statistics include the chi-squared statistic ( $\chi^2$ ) and the t-statistic. Generally, small values of the test statistic are consistent with the null hypothesis; large values lead to rejection. See Statistical Hypothesis; *p*-Value; *t*-Statistic; Chi-Squared.
- *Time Series.* A series of data collected over time—for instance, the Gross National Product of the United States from 1940 to 1990.
- *Two* -Sided Hypothesis. An alternative hypothesis asserting that the values of a pa rameter are different from—either greater than or less than—the value asserted in the null hypothesis. See Statistical Hypothesis; Significance Test.
- *Two-Tailed Test.* See Significance Test.
- *Type I Error.* A statistical test makes a type I error when (a) the null hypothesis is in fact true, and (b) the test rejects the null hypothesis (i.e., there is a false alarm). For instance, a study of two groups may show some difference between samples from each group, even when there is no difference in the population. When a statistical test deems the difference to be significant in this situation, it makes a type I error. See Statistical Hypothesis; Significance Test.
- *Type II Error.* A statistical test makes a type II error when (a) the null hypothesis is in fact not true, and (b) the test fails to reject the null hypothesis (i.e., there is a false negative). For instance, there may not be a significant difference between samples from two groups when, in fact, the groups are different. See Statistical Hypothesis; Significance Test.
- Unbiased Estimator. An estimator that is correct on average, over the possible data sets. The estimates have no systematic tendency to fall high or low.
- Uniform Distribution. For example, if an investigator picks a whole number at random from 1 to 100, it has the uniform distribution: All values are equally likely. Similarly, one gets a uniform distribution by picking a real number at random between 0.75 and 3.25: The chance of landing in an interval is proportional to the length of the interval. The uniform distribution, without further qualification, is presumably on the unit interval (which goes from 0 to 1).

- Validity. The extent to which a test instrument measures what it is supposed to, rather than something else. The validity of a standardized test often is indicated, in part, by the correlation coefficient between the test scores and some outcome measure.
- Variable. A property of units in a study, which varies from one unit to another (e.g., incomes of households) in a study of households; employment status of individuals (employed, unemployed, not in labor force) in a study of people.

Variance. The square of the standard deviation. See Standard Deviation.

z-Statistic. See t-Statistic.

z-Test. See t-Test.

## **References on Statistics**

**General Surveys** 

David Freedman et al., Statistics (2d ed. 1991).

- Darrell Huff, How to Lie with Statistics (1954).
- Gregory A. Kimble, How to Use (and Misuse) Statistics (1978).
- David S. Moore, Statistics: Concepts and Controversies (2d ed. 1985).
- David S. Moore & George P. McCabe, Introduction to the Practice of Statistics (2d ed. 1993).
- Michael Oakes, Statistical Inference: A Commentary for the Social and Behavioral Sciences (1986).
- Perspectives on Contemporary Statistics (David C. Hoaglin & David S. Moore eds., 1992).
- Statistics: A Guide to the Unknown (Judith M. Tanur et al. eds., 2d ed. 1978).

Hans Zeisel, Say It with Figures (6th ed. 1985).

#### **Reference Works for Lawyers and Judges**

David C. Baldus & James W. L. Cole, Statistical Proof of Discrimination (1980).

David W. Barnes & John M. Conley, Statistical Evidence in Litigation: Methodology, Procedure, and Practice (1986).

James Brook, A Lawyer's Guide to Probability and Statistics (1990).

- The Evolving Role of Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989).
- Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers (1990).
- Statistical Methods in Discrimination Litigation (David H. Kaye & Mikel Aickin eds., 1986).

### **General Reference**

International Encyclopedia of Statistics (William H. Kruskal & Judith M. Tanur eds., 1978).

## **Reference Guide on Multiple Regression**

Daniel L. Rubinfeld

**Daniel L. Rubinfeld**, J.D., Ph.D., is Professor of Law and Professor of Economics, University of California, Berkeley.

## Contents

- I Introduction 419
- II. Research Design: Model Specification 423
  - A. What Is the Specific Question That Is Under Investigation by the Expert? 423
  - B. What Model Should Be Used to Evaluate the Question at Issue? 423
    - 1. Choosing the dependent variable 424
    - 2. Choosing the explanatory variable that is relevant to the issues in the case 424
    - 3. Choosing the additional explanatory variables 424
    - 4. Choosing the functional form of the multiple regression model 426
    - 5. Choosing multiple regression as a method of analysis 427
- III. Interpreting Regression Results 429
  - A. What Is the Practical as Opposed to the Statistical Significance of Regression Results? 429
    - 1. When should statistical tests of significance be used? 430
    - 2. What is the appropriate level of statistical significance? 431
    - 3. Should statistical tests be one-tailed or two-tailed? 431
  - B. Are the Regression Results Robust—Sensitive to Changes in Assumptions and Procedures? 432
    - 1. What evidence exists that the explanatory variable causes changes in the dependent variable? 432
    - 2. To what extent are the explanatory variables correlated with each other? 434
    - 3. To what extent are individual errors in the regression model independent? 435
    - 4. To what extent are the regression results sensitive to individual data points? 436
    - 5. To what extent are the data subject to measurement error? 437
- IV. The Expert 439
- V. Presentation of Statistical Evidence 441

- A. What Disagreements Exist Regarding Data on Which the Analysis Is Based? 441
- B. What Database Information and Analytical Procedures Will Aid in Resolving Disputes over Statistical Studies? 442

Appendix: The Basics of Multiple Regression 445 Glossary of Terms 463

References on Multiple Regression 469

### I. Introduction

Multiple regression analysis is a statistical tool for understanding the relationship between two or more variables.<sup>1</sup> Multiple regression involves a variable to be explained—called the *dependent variable*—and additional *explanatory variables* that are thought to produce or be associated with changes in the dependent variable.<sup>2</sup> For example, a multiple regression analysis might estimate the effect of the number of years of work on salary. Salary would be the dependent variable to be explained; years of experience would be the explanatory variable.

Multiple regression analysis is sometimes well suited to the analysis of data about competing theories in which there are several possible explanations for the relationship among a number of explanatory variables.<sup>3</sup> Multiple regression typically uses a single dependent variable and several explanatory variables to assess the statistical data pertinent to these theories.

In a case alleging sex discrimination in salaries, for example, a multiple regression analysis would examine not only sex, but also other explanatory variables of interest, such as education and experience.<sup>4</sup> The employer–defendant might use multiple regression to argue that salary is a function of the employee's

In disparate treatment cases involving claims of gender discrimination, plaintiffs typically use multiple regression analysis to isolate the influence of gender on employment decisions relating to a particular job or job benefit, such as salary. The first step in such a regression analysis is to specify all of the possible "legitimate" (i.e., nondiscriminatory) factors that are likely to significantly affect the dependent variable and which could account for disparities in the treatment of male and female employees. By identifying those legitimate criteria that affect the decision making process, individual plaintiffs can make predictions about what job or job benefits similarly situated employees should ideally receive, and then can measure the difference between the predicted treatment and the actual treatment of those employees, plaintiffs in a disparate treatment case can argue that the net "residual" difference represents the unlawful effect of discriminatory animus on the allocation of jobs or job benefits. (citations omitted)

<sup>1.</sup> A variable is anything that can take on two or more values (e.g., the daily temperature in Chicago).

<sup>2.</sup> Explanatory variables in the context of a statistical study are also called independent variables. See David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.C.1, in this manual. Kaye and Freedman also offer a brief discussion of regression analysis. *Id.* § III.F.3.

<sup>3.</sup> Multiple regression is one type of statistical analysis involving several variables. Other types include matching analysis, stratification, analysis of variance, probit analysis, logit analysis, discriminant analysis, and factor analysis.

<sup>4.</sup> Thus, in Ottaviani v. State Univ. of N.Y., 875 F.2d 365, 367 (2d Cir. 1989), cert. denied, 493 U.S. 1021 (1990), the court stated:

education and experience, and the employee–plaintiff might argue that salary is also a function of the individual's sex.

Multiple regression also may be useful (1) in determining whether or not a particular effect is present; (2) in measuring the magnitude of a particular effect; and (3) in forecasting what a particular effect would be, but for an intervening event. In a patent infringement case, for example, a multiple regression analysis could be used to determine (1) whether the behavior of the alleged infringer affected the price of the patented product; (2) the size of the effect; and (3) what the price of the product would have been had the alleged infringement not occurred.

Over the past several decades the use of regression analysis in court has grown widely.<sup>5</sup> Although multiple regression analysis has been used most frequently in cases of sex and race discrimination<sup>6</sup> and antitrust violation,<sup>7</sup> other applications have ranged across a variety of cases, including those involving census undercounts,<sup>8</sup> voting rights,<sup>9</sup> the study of the deterrent effect of the death penalty,<sup>10</sup> and intellectual property.<sup>11</sup>

6. Recent discrimination cases using multiple regression analysis include King v. General Elec. Co., 960 F.2d 617 (7th Cir. 1992), and Tennes v. Massachusetts Dep't of Revenue, No. 88-C3304, 1989 WL 157477 (N.D. Ill. Dec. 20, 1989) (age discrimination); EEOC v. General Tel. Co. of N.W., 885 F.2d 575 (9th Cir. 1989), *cert. denied*, 498 U.S. 950 (1990), Churchill v. International Business Machs., Inc., 759 F. Supp. 1089 (D.N.J. 1991), and Denny v. Westfield State College, 880 F.2d 1465 (1st Cir. 1989) (sex discrimination); Black Law Enforcement Officers Ass'n v. City of Akron, 920 F.2d 932 (6th Cir. 1990), Bazemore v. Friday, 848 F.2d 476 (4th Cir. 1988), Bridgeport Guardians, Inc. v. City of Bridgeport, 735 F. Supp. 1126 (D. Conn. 1990), *affd*, 933 F.2d 1140 (2d Cir.), *cert. denied*, 112 S. Ct. 337 (1991), and Dicker v. Allstate Life Ins. Co., No. 89-C4982, 1193 WL 62385 (N.D. Ill. Mar. 5, 1993) (race discrimination).

7. Recent antitrust cases using multiple regression analysis include *In re* Chicken Antitrust Litig., 560 F. Supp. 963, 993 (N.D. Ga. 1980); and United States v. Brown Univ., 805 F. Supp. 288 (E.D. Pa. 1992), *rev'd*, 5 F.3d 658 (3d Cir. 1993) (price fixing of college scholarships).

8. See, e.g., Carey v. Klutznick, 508 F. Supp. 420, 432–33 (S.D.N.Y. 1980) (use of reasonable and scientifically valid statistical survey or sampling procedures to adjust census figures for the differential undercount is constitutionally permissible), stay granted, 449 U.S. 1068 (1980), rev'd on other grounds, 653 F.2d 732 (2d Cir. 1981), cert. denied, 455 U.S. 999 (1982); Young v. Klutznick, 497 F. Supp. 1318, 1331 (E.D. Mich. 1980), rev'd on other grounds, 652 F.2d 617 (6th Cir. 1981), cert. denied, 455 U.S. 939 (1982); Cuomo v. Baldrige, 674 F. Supp. 1089 (S.D.N.Y. 1987).

9. Multiple regression analysis was used in suits charging that at-large area-wide voting was instituted to neutralize black voting strength. Multiple regression demonstrated that the race of the candidates and that of the electorate was a determinant of voting. See Kirksey v. City of Jackson, 461 F. Supp. 1282, 1289 (S.D. Miss. 1978), aff'd, 663 F.2d 659 (5th Cir. 1981); Brown v. Moore, 428 F. Supp. 1123, 1128–29 (S.D. Ala. 1976), aff'd without op., 575 F.2d 298 (5th Cir. 1978), vacated sub nom. Williams v. Brown, 446 U.S. 236 (1980); Bolden v. City of Mobile, 423 F. Supp. 384, 388 (S.D. Ala. 1976), aff'd, 571 F.2d 238 (5th Cir. 1978), stay denied, 436 U.S. 902 (1978), and rev'd, 446 U.S. 55 (1980); Jeffers v. Clinton, 730 F. Supp. 196, 208–09 (E.D. Ark. 1989), aff'd, 498 U.S. 1019 (1991); and League of United Latin Am. Citizens, Council No. 4434 v. Clements, 986 F.2d 728, 774–87 (5th Cir.), reh'g en banc, 999 F.2d 831 (5th Cir. 1993), cert. denied, 114 S. Ct. 878 (1994). For a recent update on statistical issues involving voting rights cases, see Daniel L. Rubinfeld, Statistical and Demographic Issues Underlying Voting Rights Cases, 15 Evaluation Rev. 659 (1991), and the associated articles within Vol. 15 (a special symposium issue devoted to statistical and demographic issues underlying voting rights cases).

10. See, e.g., Gregg v. Georgia, 428 U.S. 153, 184–86 (1976). For a critique of the validity of the deterrence analysis, see National Research Council, Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates (Alfred Blumstein et al. eds., 1978). Multiple regression methods have been used to

<sup>5.</sup> There were only 2 WESTLAW references to multiple regression in federal cases between 1960 and 1969, 26 references between 1970 and 1979, 204 references between 1980 and 1989, and 73 references since 1990.

Multiple regression analysis can be a source of valuable scientific testimony in litigation. However, when inappropriately used, regression analysis can confuse important issues while having little, if any, probative value. In *EEOC v. Sears, Roebuck & Co.*, in which Sears was charged with discrimination against women in hiring practices, the Seventh Circuit acknowledged that "[m]ultiple regression analyses, designed to determine the effect of several independent variables on a dependent variable, which in this case is hiring, are an accepted and common method of proving disparate treatment claims."<sup>12</sup> However, the court affirmed the district court's findings that the "E.E.O.C.'s regression analyses did not 'accurately reflect Sears' complex, nondiscriminatory decision-making processes'" and that "'E.E.O.C.'s statistical analyses [were] so flawed that they lack[ed] any persuasive value.'" <sup>13</sup> Serious questions also have been raised about the use of multiple regression analysis in census undercount cases and in death penalty cases.<sup>14</sup>

Moreover, in interpreting the results of a multiple regression analysis, it is important to distinguish between *correlation* and *causality*. Two variables are correlated when the events associated with the variables occur more frequently together than one would expect by chance. For example, if higher salaries are associated with a greater number of years of work experience, and lower salaries are associated with fewer years of experience, there is a positive correlation between the two variables. However, if higher salaries are associated with less experience, and lower salaries are associated with more experience, there is a negative correlation between the two variables.

A correlation between two variables does not imply that one event causes the second to occur. Therefore, in making causal inferences, it is important to avoid *spurious correlation*.<sup>15</sup> Spurious correlation arises when two variables are closely related but bear no causal relationship because they are both caused by a third, unexamined variable.

For example, there might be a negative correlation between the age of certain skilled employees of a computer company and their salaries. One should not

12. 839 F.2d 302, 324 n.22 (7th Cir. 1988).

13. *Id.* at 348, 351 (quoting EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1342, 1352 (N.D. Ill. 1986)). The district court comments specifically on the "severe limits of regression analysis in evaluating complex decision-making processes." 628 F. Supp. at 1350.

14. With respect to the census, see Stephen E. Fienberg. The New York City Census Adjustment Trial: Witness for the Plaintiffs, 34 Jurimetrics J. 65 (1993); John E. Rolph, The Census Adjustment Trial: Reflections of a Witness for the Plaintiffs, 34 Jurimetrics J. 85 (1993); David A. Freedman, Adjusting the Census of 1990, 34 Jurimetrics J. 107 (1993). Concerning the death penalty, see Richard Lempert, Capital Punishment in the '80's: Reflections on the Symposium, 74 J. Crim. L. & Criminology 1101 (1983).

15. See Linda A. Bailey et al., Reference Guide on Epidemiology § IV.A (Confounding Variables), and David H. Kaye & David A. Freedman, Reference Guide on Statistics §§ II.C.2, III.F.2.c, in this manual.

evaluate whether the death penalty was applied discriminately on the basis of race. See McClesky v. Kemp, 481 U.S. 279, 292-94 (1987).

<sup>11.</sup> See Polaroid Corp. v. Eastman Kodak Co., No. 76-1634-MA, 1990 WL 324105, at \*29, 62–63 (D. Mass. Oct. 12, 1990) (damages due to patent infringement), *amended by* No. 76-1634-MA, 1991 WL 4087 (D. Mass. Jan. 11, 1991); and Estate of Vane v. The Fair, Inc., 849 F.2d 186, 188 (5th Cir. 1988), *cert. denied*, 488 U.S. 1008 (1989) (lost profits due to copyright infringement).

conclude from this correlation that the employer has necessarily discriminated against the employees on the basis of their age. A third, unexamined variable— the level of the employees' technological skills—could explain differences in productivity and, consequently, differences in salary. Or, consider a patent infringement damage case in which increased sales of an allegedly infringing product are associated with a lower price of the patented product. This correlation would be spurious if the two products have their own noncompetitive market niches and the lower price is due to a decline in the production costs of the patented product.

Causality cannot be inferred by data analysis alone—rather, one must infer that a causal relationship exists on the basis of an underlying causal theory that explains the relationship between the two variables. Even when an appropriate theory has been identified, causality can never be inferred directly—one must also look for empirical evidence that there is a causal relationship. Conversely, the presence of a non-zero correlation between two variables does not guarantee the existence of a relationship; it could be that the *model* does not reflect the correct interplay among the explanatory variables. In fact, the absence of correlation does not guarantee that a causal relationship does not exist. Rather, lack of correlation could occur if (1) there are insufficient data; (2) the data are measured inaccurately; (3) the data do not allow multiple causal relationships to be sorted out; or (4) the model is specified wrongly.

There is a tension between any attempt to reach conclusions with near certainty and the inherently probabilistic nature of multiple regression analysis. In general, statistical analysis involves the formal expression of uncertainty in terms of probabilities. The reality that statistical analysis generates probabilities that there are relationships should not be seen in itself as an argument against the use of statistical evidence. The only alternative might be to use less reliable anecdotal evidence.

This reference guide addresses a number of procedural and methodological issues that are relevant in considering the admissibility of, and weight to be accorded to, the findings of multiple regression analyses. It also suggests some standards of reporting and analysis that an expert presenting multiple regression analyses might be expected to meet. Section II discusses research design—how the multiple regression framework can be used to sort out alternative theories about a case. Section III concentrates on the interpretation of the multiple regression results, from both a statistical and a practical point of view. Section IV briefly discusses the qualifications of experts. In section V the emphasis turns to procedural aspects associated with the use of the data underlying regression analyses. Finally, the Appendix delves into the multiple regression framework in further detail; it also contains a number of specific examples that illustrate the application of the technique. A list of statistical references and a glossary are also included.

## II. Research Design: Model Specification

Multiple regression allows the expert to choose among alternative theories or hypotheses and assists the expert in sorting out correlations between variables that are plainly spurious from those that reflect valid relationships.

## A. What Is the Specific Question That Is Under Investigation by the Expert?

Research begins with a clear formulation of a research question. The data to be collected and analyzed must relate directly to the immediate issue; otherwise, appropriate inferences cannot be drawn from the statistical analysis. For example, if the question at issue in a patent damage case is what price the plaintiff's product would have been but for the sale of the defendant's infringing product, sufficient data must be available to allow the expert to account statistically for the important factors that determine the price of the product.

#### B. What Model Should Be Used to Evaluate the Question at Issue?

Model specification involves several steps, each of which is fundamental to the success of the research effort. Ideally, a multiple regression analysis builds on a theory that describes the variables to be included in the study. For example, the theory of labor markets might lead one to expect salaries in an industry to be related to workers' experience and the productivity of workers' jobs. A belief in discrimination would lead one to add a variable or variables reflecting discrimination to the model.

Models are often characterized in terms of *parameters* —numerical characteristics of the model. In the labor market example, one parameter might reflect the increase in salary associated with each additional year of job experience. Multiple regression uses a *sample*, or a selection of data, from the *population*, or all the units of interest, to obtain *estimates* of the values of the parameters of the model—an estimate associated with a particular explanatory variable is a *regres sion coefficient*.

Failure to develop the proper theory, failure to choose the appropriate variables, and failure to choose the correct form of the model can *bias* substantially the statistical results, that is, create a systematic tendency for an estimate of a model parameter to be too high or too low.

#### 1. Choosing the dependent variable

The variable to be explained should be the appropriate variable for analyzing the question at issue. <sup>16</sup> Suppose, for example, that pay discrimination among hourly workers is a concern. One choice for the dependent variable is the hourly wage rate of the employees; another choice is the annual salary. The distinction is important, because annual salary differences may be due in part to differences in hours worked. If the number of hours worked is the product of worker preferences and not discrimination, the hourly wage is a good choice. If the number of hours is related to the alleged discrimination, annual salary is the more appropriate dependent variable to choose.<sup>17</sup>

#### 2. Choosing the explanatory variable that is relevant to the issues in the case

The explanatory variable that allows the evaluation of alternative hypotheses must be chosen appropriately. Thus, in a discrimination case, the *variable of in*-terest may be the race or sex of the individual. In an antitrust case, it may be a variable that takes on the value 1 to reflect the presence of the alleged anticompetitive behavior and a value 0 otherwise.<sup>18</sup>

#### Choosing the additional explanatory variables

An attempt should be made to identify the additional known or hypothesized explanatory variables, some of which are measurable and may support alternative substantive hypotheses that can be accounted for by the regression analysis. Thus, in a discrimination case, a measure of the skill level of the work may provide an alternative explanation—lower salaries were the result of inadequate skills.<sup>19</sup>

However, in Trout v. Garrett, 780 F. Supp. 1396, 1414 (D.D.C. 1991), the court ruled (in the damage phase of the trial) that the extent of civilian employees' prehire work experience was not an appropriate variable

<sup>16.</sup> In multiple regression analysis, the dependent variable is usually a continuous variable that takes on a range of numerical values. When the dependent variable is categorical, taking only two or three values, modified forms of multiple regression, such as probit or logit analysis, are appropriate. For an example of the use of the latter, see EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 325 (7th Cir. 1988) (EEOC used weighted logit analysis to measure the impact of variables, such as age, education, job type experience, and product line experience, on the female percentage of commission hires). *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.C.1, in this manual.

<sup>17.</sup> In job systems in which annual salaries are tied to grade or step levels, the annual salary corresponding to the job position could be more appropriate.

<sup>18.</sup> Explanatory variables may vary by type, which will affect the interpretation of the regression results. Thus, some variables may be continuous, taking on a wide range of values, while others may be categorical, taking on only two or three values.

<sup>19.</sup> In Ottaviani v. State Univ. of N.Y., 679 F. Supp. 288, 306–08 (S.D.N.Y. 1988), affd, 875 F.2d 365 (2d Cir. 1989), cert. denied, 493 U.S. 1021 (1990), the court ruled (in the liability phase of the trial) that the university showed there was no discrimination in either placement into initial rank or promotions between ranks, so rank was a proper variable in multiple regression analysis to determine whether women faculty members were treated differently from men.

Not all possible variables that may influence the dependent variable can be included if the analysis is to be successful—some cannot be measured, and others may make little difference.<sup>20</sup> If a preliminary analysis shows the unexplained portion of the multiple regression to be unacceptably high, the expert may seek to discover whether some previously undetected variable is missing from the analysis.<sup>21</sup>

Failure to include a major explanatory variable that is correlated with the variable of interest in a regression model may cause an included variable to be credited with an effect that actually is caused by the excluded variable.<sup>22</sup> In general, omitted variables that are correlated with the dependent variable reduce the probative value of the regression analysis.<sup>23</sup> This may lead to inferences made from regression analyses that do not assist the trier of fact.<sup>24</sup>

Omitting variables that are not correlated with the variable of interest is, in general, less of a concern, since the parameter that measures the effect of the variable of interest on the dependent variable is estimated without bias. Suppose, for example, that the effect of a policy introduced by the courts to encourage child support has been tested by randomly choosing some cases to be handled according to current court policies and other cases to be handled according to a new, more stringent policy. The effect of the new policy might be measured by a multiple regression using payment success as the dependent variable and a 0 or 1 explanatory variable (1 if the new program was applied; 0 if it was not). Failure

20. The summary effect of the excluded variables shows up as a random error term in the regression model, as does any modeling error. See *infra* the Appendix for details.

21. A very low R-square ( $R^2$ ) is one indication of an unexplained portion of the multiple regression model that is unacceptably high. For reasons discussed in the Appendix, however, a low  $R^2$  does not necessarily imply a poor model (and vice versa).

22. Technically, the omission of explanatory variables which are correlated with the variable of interest can cause biased estimates of regression parameters.

23. The effect tends to be important, the stronger the relationship between the omitted variable and the dependent variable, and the stronger the correlation between the omitted variable and the explanatory variables of interest.

24. See Bazemore v. Friday, 751 F.2d 662, 671–72 (4th Cir. 1984) (upholding the district court's refusal to accept a multiple regression analysis as proof of discrimination by a preponderance of the evidence, the court of appeals stated that, although the regression used four variable factors, consisting of race, education, tenure, and job title, the failure to use other factors, including pay increases which varied by county, precluded their introduction into evidence), *aff'd in part*, *vacated in part*, 478 U.S. 385 (1986).

Note, however, that in Sobel v. Yeshiva Univ., 839 F.2d 18, 33, 34 (2d Cir. 1988), *cert. denied*, 490 U.S. 1105 (1989), the court made clear that "a [Title VII] defendant challenging the validity of a multiple regres sion analysis [has] to make a showing that the factors it contends ought to have been included would weaken the showing of salary disparity made by the analysis," by making a specific attack and "a showing of relevance for each particular variable it contends . . . ought to [be] includ[ed]" in the analysis, rather than by simply attacking the results of the plaintiffs' proof as inadequate for lack of a given variable.

Also, in *Bazemore v. Friday*, the Court, declaring that the Fourth Circuit's view of the evidentiary value of the regression analyses was plainly incorrect, stated that "[n]ormally, failure to include variables will affect the analysis' probativeness, not its admissibility. Importantly, it is clear that a regression analysis that includes less than 'all measurable variables' may serve to prove a plaintiff's case." 478 U.S. 385, 400 (1986) (footnote omitted).

in a regression analysis to compute back pay in employment discrimination. According to the court, in cluding the prehire level would have resulted in a finding of no sex discrimination, despite a contrary conclusion in the liability phase of the action. *Id. See also* Stuart v. Roache, 951 F.2d 446 (1st Cir. 1991) (allowing only three years of seniority to be considered due to prior discrimination), *cert. denied*, 112 S. Ct. 1948 (1992).

to include an explanatory variable that reflected the age of the husbands involved in the program would not affect the court's evaluation of the new policy, since men of any given age are as likely to be affected by the old as the new policy. Choosing the court's policy by chance has ensured that the omitted age variable is not correlated with the policy variable.

Bias caused by the omission of important variables that are related to the included variables of interest can be a serious problem.<sup>25</sup> Nevertheless, it is possible to account for bias qualitatively if the expert has knowledge (even if not quantifiable) about the relationship between the omitted variable and the explanatory variable. Suppose, for example, that the plaintiff's expert in a sex discrimination pay case is unable to obtain quantifiable data that reflect the skills necessary for a job, and that, on average, women are more skillful than men. Suppose also that a regression of the wage rate of employees (the dependent variable) on years of experience and a variable reflecting the sex of each employee (the explanatory variable) suggests that men are paid substantially more than women with the same experience. Because differences in skill levels have not been taken into account, the expert may conclude reasonably that the wage difference measured by the regression is a conservative estimate of the true discriminatory wage difference.

The precision of the measure of the effect of a variable of interest on the dependent variable is also important.<sup>26</sup> In general, the more complete the explained relationship between the included explanatory variables and the dependent variable, the more precise the results. Note, however, that the inclusion of explanatory variables that are irrelevant (i.e., that are not correlated with the dependent variable) reduces the precision of the regression results. This can be a source of concern when the sample size is small, but it is not likely to be of great consequence when the sample size is large.

#### 4. Choosing the functional form of the multiple regression model

Choosing the proper set of variables to be included in the multiple regression model does not complete the modeling exercise. The expert must also choose the proper form of the regression model. The most frequently selected form is the *linear regression* model (described in the Appendix). In this model the magnitude of the change in the dependent variable associated with the change in any of the explanatory variables is the same no matter what the level of that explanatory variable. For example, one additional year of experience might add \$5,000 to salary, irrespective of the previous experience of the employee.

<sup>25.</sup> See also Linda A. Bailey et al., Reference Guide on Epidemiology § IV.A, and David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.C.2, in this manual.

<sup>26.</sup> A more precise estimate of a parameter is an estimate with a smaller standard error. See infra the Appendix for details.

In some instances, however, there may be reason to believe that changes in explanatory variables will have differential effects on the dependent variable as the values of the explanatory variables change. In this case, the expert should consider the use of a *nonlinear model*. Failure to account for nonlinearities can lead to either overstatement or understatement of the effect of a change in the value of an explanatory variable on the dependent variable.

One particular type of nonlinearity involves the interaction among several variables. An *interaction variable* is the product of two other variables that are included in the multiple regression model. The interaction variable allows the expert to take into account the possibility that the effect of a change in one variable on the dependent variable may change as the level of another explanatory variable changes. For example, in a salary discrimination case, the inclusion of a term that interacts a variable measuring experience with a variable representing the sex of the employee (1 if a female employee, 0 if a male employee) allows the expert to test whether the sex differential varies with the level of experience. A significant negative estimate of the parameter associated with the sex variable suggests that inexperienced women are discriminated against, while a significant negative estimate of the interaction parameter suggests that the extent of discrimination increases with experience.<sup>27</sup>

Note that insignificant *coefficients* in a model with interactions may suggest a lack of discrimination, while a model without interactions may suggest the contrary. It is especially important to account for the interactive nature of the discrimination; failure to do so may lead to false conclusions concerning discrimination.

#### 5. Choosing multiple regression as a method of analysis

There are many multivariate statistical techniques other than multiple regression that are useful in legal proceedings. Some statistical methods are appropriate when nonlinearities are important.<sup>28</sup> Others apply to models in which the dependent variable is discrete, rather than continuous.<sup>29</sup> Still others have been applied predominantly to respond to methodological concerns arising in the context of discrimination litigation.<sup>30</sup>

<sup>27.</sup> For further details, see infra the Appendix.

<sup>28.</sup> These techniques include, but are not limited to, piecewise linear regression, polynomial regression, maximum likelihood estimation of models with nonlinear functional relationships, and autoregressive and moving average time-series models. *See*, *e.g.*, Robert S. Pindyck & Daniel L. Rubinfeld, Econometric Models & Economic Forecasts 101–04, 117–20, 238–44, 472–560 (3d ed. 1991).

<sup>29.</sup> For a discussion of probit and logit analysis, techniques that are useful in the analysis of qualitative choice, see *id.* at 248-81.

<sup>30.</sup> The correct model for use in salary discrimination suits is a subject of debate among labor economists. As a result, some have begun to evaluate alternatives approaches. These include urn models (Bruce Levin & Herbert Robbins, Urn Models for Regression Analysis, with Applications to Employment Discrimination Studies, Law & Contemp. Probs., Autumn 1983, at 247); and reverse regression (Delores A. Conway & Harry V. Roberts, Reverse Regression, Fairness, and Employment Discrimination, 1 J. Bus. & Econ. Stat. 75 (1983)). But see Arthur S. Goldberger, Redirecting Reverse Regressions, 2 J. Bus. & Econ. Stat. 114 (1984), and Arlene S.

It is essential that a valid statistical method be applied to assist with the analysis in each legal proceeding. Therefore, the expert should be prepared to explain why any chosen method, including regression, was more suitable than the alternatives.

Ash, The Perverse Logic of Reverse Regression, in Statistical Methods in Discrimination Litigation 85 (David H. Kaye & Mikel Aickin eds., 1986).

Reference Manual on Scientific Evidence

## **III.** Interpreting Regression Results

Regression results can be interpreted in purely statistical terms, through the use of significance tests, or they can be interpreted in a more practical, nonstatistical manner. While an evaluation of the *practical significance* of regression results is almost always relevant in the courtroom, tests of *statistical significance* are appropriate only in particular circumstances.

# A. What Is the Practical as Opposed to the Statistical Significance of Regression Results?

Practical significance means that the magnitude of the effect being studied is not de minimis—it is sufficiently important substantively for the court to be concerned. For example, if the average wage rate is \$10.00 per hour, a wage differential between men and women of \$0.10 per hour is likely to be deemed practically insignificant because the differential represents only 1% (\$0.10/\$10.00) of the average wage rate. <sup>31</sup> That same difference could be statistically significant, however, if a sufficiently large sample of men and women was studied. <sup>32</sup> The reason is that statistical significance is determined, in part, by the number of observations in the data set.

Other things being equal, the statistical significance of a regression coefficient increases as the sample size increases. Often, results that are practically significant are also statistically significant.<sup>33</sup> It is possible with a large data set to find a number of statistically significant coefficients that are practically insignificant. Similarly, it is also possible (especially when the sample size is small) to obtain

<sup>31.</sup> There is no specific percentage threshold above which a result is practically significant. Practical significance must be evaluated in the context of a particular legal issue. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.2, in this manual.

<sup>32.</sup> Practical significance also can apply to the overall credibility of the regression results. Thus, in McClesky v. Kemp, 481 U.S. 279 (1987), coefficients on race variables were statistically significant, but the Court declined to find them legally or constitutionally significant.

<sup>33.</sup> In Melani v. Board of Higher Educ., 561 F. Supp. 769, 774 (S.D.N.Y. 1983), a Title VII suit was brought against the City University of New York (CUNY) for allegedly discriminating against female instructional staff in the payment of salaries. One approach of the plaintiff's expert in the case was to use multiple regression analysis. The coefficient on the variable that reflected the sex of the employee was approximately equal to \$1,800 when all years of data were included. Practically (in terms of average wages at the time) and statistically (in terms of a 5% significance test) this result was significant. Thus, the court stated that "[p]laintiffs have produced statistically significant evidence that women hired as CUNY instructional staff since 1972 received substantially lower salaries than similarly qualified men." (emphasis added). *Id.* at 781.

results that are practically significant but statistically insignificant. Suppose, for example, that an expert undertakes a damage study in a patent infringement case and predicts but-for sales—what sales would have been had the infringement not occurred—using data that predate the period of alleged infringement. If data limitations are such that only three or four years of pre-infringement sales are known, the difference between but-for sales and actual sales during the period of alleged infringement could be practically significant but statistically insignificant.

#### 1. When should statistical tests of significance be used?

A test of a specific contention—a *hypothesis test* —often assists the court in determining whether a violation of the law has occurred in areas where direct evidence is inaccessible or inconclusive. For example, an expert might use hypothesis tests in race and sex discrimination cases to determine the presence of discriminatory effect.

Statistical evidence alone never can prove with absolute certainty the worth of any substantive theory. However, by providing evidence contrary to the view that a particular form of discrimination has not occurred, for example, the multiple regression approach can aid the trier of fact in assessing the likelihood that discrimination has occurred.<sup>34</sup>

Tests of hypotheses are appropriate in a *cross-section analysis*, when the data underlying the regression study have been chosen as a sample of a population at a particular point in time, and in a *time-series analysis*, when the data being evaluated cover a number of time periods. In either case, the expert may want to evaluate a specific hypothesis, usually relating to a question of liability or to the determination of whether there is measurable impact of an alleged violation. Thus, in a sex discrimination case, an expert may want to evaluate a *null hypothesis* of no discrimination against the *alternative hypothesis* that discrimination takes a particular form. <sup>35</sup> Alternatively, in an antitrust damage proceeding, the expert may want to test a null hypothesis of no impact against the alternative hypothesis that there was legal impact. In either type of case, it is important to realize that rejection of the null hypothesis and believe that an alternative explanation other than one involving legal liability accounts for the results.

Often, the null hypothesis is stated in terms of a particular regression parameter being equal to 0. For example, in a wage discrimination case, the null hypothesis would be that there is no wage difference between sexes. If a negative difference is observed (meaning that women earn less than men after the expert

<sup>34.</sup> See International Bhd. of Teamsters v. United States, 431 U.S. 324 (1977) (the Court inferred discrimination from overwhelming statistical evidence by a preponderance of the evidence).

<sup>35.</sup> Tests are also appropriate when comparing the outcomes of a set of employer decisions with those that would have been obtained had the employer chosen differently from among the available options.

has controlled statistically for legitimate alternative explanations), the difference is evaluated as to its statistical significance using the *t*-test.<sup>36</sup> The *t*-test uses the *t*-statistic to evaluate the hypothesis that a model parameter takes on a particular value, usually 0.

#### 2. What is the appropriate level of statistical significance?

In most scientific work, the level of statistical significance required to reject the null hypothesis (i.e., to obtain a statistically significant result) is set conventionally at .05, or 5%.<sup>37</sup> The significance level measures the probability that the null hypothesis will be rejected incorrectly, assuming that the null hypothesis is true. In general, the lower the percentage required for statistical significance, the more difficult it is to reject the null hypothesis; therefore, the lower the probability that one will err in doing so. While the 5% criterion is typical, reporting of more stringent 1% significance tests or less stringent 10% tests can also provide useful information.

In doing a statistical test, it is useful to compute an observed significance level, or *p*-value. The *p*-value associated with the null hypothesis that a regression coefficient is 0 is the probability that a coefficient of this magnitude or larger could have occurred by chance if the null hypothesis were true. If the *p*-value were less than or equal to 5%, the expert would reject the null hypothesis in favor of the alternative hypothesis; if the *p*-value were greater than 5%, the expert would fail to reject the null hypothesis.<sup>38</sup>

#### 3. Should statistical tests be one-tailed or two-tailed?

When the expert evaluates the null hypothesis that a variable of interest has no *association* with a dependent variable against the alternative hypothesis that there is an association, a *two-tailed test* that allows for the effect to be either positive or negative is usually appropriate. A *one-tailed test* would usually be applied when the expert believes, perhaps on the basis of other direct evidence presented at trial, that the alternative hypothesis is either positive or negative, but not both. For example, an expert might use a one-tailed test in a patent infringement case

<sup>36.</sup> The *t*-test is strictly valid only if a number of important assumptions hold. However, for many regression models, the test is approximately valid if the sample size is sufficiently large. See *infra* the Appendix for a more complete discussion of the assumptions underlying multiple regression.

<sup>37.</sup> See, e.g., Palmer v. Shultz, 815 F.2d 84, 92 (D.C. Cir. 1987) ("the .05 level of significance . . . [is] certainly sufficient to support an inference of discrimination") (quoting Segar v. Smith, 738 F.2d 1249, 1283 (D.C. Cir. 1984), cert. denied, 471 U.S. 1115 (1985)). See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.2, in this manual.

<sup>38.</sup> The use of 1%, 5%, and, sometimes, 10% rules for determining statistical significance remains a subject of debate. One might argue, for example, that when regression analysis is used in a price-fixing antitrust case to test a relatively specific alternative to the null hypothesis (e.g., price fixing), a somewhat lower level of confidence (a higher level of significance, such as 10%) might be appropriate. Otherwise, when the alternative to the null hypothesis is less specific, such as ther ratue alternative of "effect" (e.g., the price increase is caused by the increased cost of production, increased demand, a sharp increase in advertising, or price fixing), a high level of confidence (associated with a low significance level, such as 1%) may be appropriate.
if he or she strongly believed that the effect of the alleged infringement on the price of the infringed product was either 0 or negative. (The sales of the infringing product competed with the sales of the infringed product, thereby lowering the price.)

Because one-tailed tests produce *p*-values that are one-half the size of the *p*-value using a two-tailed test, the choice of a one-tailed test makes it easier for the expert to reject a null hypothesis. Correspondingly, the choice of a two-tailed test makes null hypothesis rejection less likely. Since there is some arbitrariness involved in the choice of an alternative hypothesis, courts should avoid relying solely on sharply defined statistical tests.<sup>39</sup> Reporting the *p*-value should be encouraged, since it conveys useful information to the court, whether or not a null hypothesis is rejected.

# B. Are the Regression Results Robust—Sensitive to Changes in Assumptions and Procedures?

The issue of *robustness* —whether regression results are sensitive to slight modifications in assumptions (e.g., that the data are measured accurately)—is of vital importance for the courts. If the assumptions of the regression model are valid, standard statistical tests can be applied. However, when the assumptions of the model are imprecise, standard tests can overstate or understate the significance of the results.

The violation of an assumption does not necessarily invalidate a regression analysis, however. In some cases in which the assumptions of multiple regression analysis fail, there are more advanced statistical methods that are appropriate. Consequently, experts should be encouraged to provide additional information that goes to the issue of whether regression assumptions are valid, and if they are not valid, the extent to which the regression results are robust. The following questions highlight some of the more important assumptions of regression analysis.

1. What evidence exists that the explanatory variable causes changes in the dependent variable?

In the multiple regression framework, the expert often assumes that changes in explanatory variables affect the dependent variable, but changes in the dependent variable do not affect the explanatory variables—that is, there is no *feedback*.<sup>40</sup> In making this assumption, the expert draws the conclusion that a correlation between an explanatory variable and the dependent variable is due to

<sup>39.</sup> Courts have shown a preference for two-tailed tests. *See* Palmer v. Shultz, 815 F.2d 84, 95–96 (D.C. Cir. 1987) (rejecting the use of one-tailed tests, the court found that because some appellants were claiming overselection for certain jobs, a two-tailed test was more appropriate in Title VII cases). *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.B.3.b, in this manual.

<sup>40.</sup> When both effects occur at the same time, this is described as simultaneity.

the effect of the former on the latter and not vice versa. Were the assumption not valid, spurious correlation might cause the expert and the trier of fact to reach the wrong conclusion.<sup>41</sup>

Figure 1 illustrates this point. In Figure 1(a), the dependent variable, Price, is explained through a multiple regression framework by three explanatory variables, Demand, Cost, and Advertising, with no feedback. In Figure 1(b), however, there is feedback since Price affects Demand, and Demand, Cost, and Advertising affect Price. Cost and Advertising, however, are not affected by Price. As a general rule, there is no direct statistical test for determining the direction of causality. Rather the expert, when asked, should be prepared to defend his or her assumption based on an understanding of the underlying behavior of the firms or individuals involved.

Although there is no single approach that is entirely suitable for estimating models when the dependent variable affects one or more explanatory variables, one possibility is for the expert to drop the questionable variable from the regression to determine whether the variable's exclusion makes a difference. If it does not, the issue becomes moot. Second, the expert can expand the multiple regression model by adding one or more equations that explain the relationship between the explanatory variable in question and the dependent variable.

<sup>41.</sup> This is especially important in litigation, because it is possible for the defendant (if responsible, for example, for price fixing or discrimination) to affect the values of the explanatory variables and thus to bias the usual statistical tests that are used in multiple regression.





Suppose, for example, that in a salary-based sex discrimination suit the defendant's expert considers employer-evaluated test scores to be an appropriate explanatory variable for the dependent variable, salary. If the plaintiff were to provide information that the employer adjusted the test scores in a manner that penalized women, the assumption that salaries were determined by test scores and not that test scores were affected by salaries might be invalid. If it is clearly inappropriate, the test-score variable should be removed from consideration. Alternatively, the information about the employer's use of the test scores could be translated into a second equation in which a new dependent variable, test score, is related to workers' salary and sex. A test of the hypothesis that salary and sex affect test scores would provide a suitable test of the absence of feedback.

2. To what extent are the explanatory variables correlated with each other?

It is essential in multiple regression analysis that the explanatory variable of interest not be correlated perfectly with one or more of the other explanatory variables. If there were perfect correlation between two variables, the expert could not separate out the effect of the variable of interest on the dependent variable from the effect of the other variable. Suppose, for example, that in a sex discrimination suit a particular form of job experience is determined to be a valid source of high wages. If all men had the requisite job experience and all women did not, it would be impossible to tell whether wage differentials between men and women were due to sex discrimination or differences in experience.

When two or more explanatory variables are correlated perfectly—that is, when there is *perfect collinearity* —one cannot estimate the regression parameters. When two or more variables are highly, but not perfectly, correlated—that is, when there is *multicollinearity* —the regression can be estimated, but some concerns remain. The greater the multicollinearity between two variables, the less precise are the estimates of individual regression parameters (even though there is no problem in estimating the joint influence of the two variables and all other regression parameters).

Fortunately, the reported regression statistics take into account any multicollinearity that might be present.<sup>42</sup> It is important to note as a corollary, however, that a failure to find a strong relationship between a variable of interest and a dependent variable need not imply that there is no relationship.<sup>43</sup> A relatively small sample, or even a large sample with substantial multicollinearity, may not provide sufficient information for the expert to determine whether there is a relationship.

#### 3. To what extent are individual errors in the regression model independent?

If the parameters of a multiple regression model were calculated using the entire universe of data (the population), the estimates might still measure the model's population parameters with error. Errors can arise for a number of reasons, including (a) the failure of the model to include the appropriate explanatory variables; (b) the failure of the model to reflect any nonlinearities that might be present; and (c) the inclusion of inappropriate variables in the model. (Of course, fur ther sources of error will arise if a sample of the population is used to estimate the regression parameters.)

It is useful to view the cumulative effect of all of these sources of modeling error as being represented by an additional variable, the *error term*, in the multiple regression model. An important assumption in multiple regression analysis is that the error term and each of the explanatory variables are *independent* of each other. (If the error term and the explanatory variable are independent, they are not correlated with each other.) To the extent this is the case, the expert can estimate the parameters of the model without bias; the magnitude of the error term will affect the precision with which a model parameter is estimated, but will not cause that estimate to be consistently too high or too low.

Multiple Regression

<sup>42.</sup> See Denny v. Westfield State College, 669 F. Supp. 1146, 1149 (D. Mass. 1987) (the court accepted the testimony of one expert that "the presence of multicollinearity would merely tend to *overestimate* the amount of error associated with the estimate . . . . In other words, *P*-values will be artificially higher than they would be if there were no multicollinearity present.") (emphasis added).

<sup>43.</sup> If a variable of interest and another explanatory variable are highly correlated, dropping the second variable from the regression can be instructive. If the coefficient on the variable of interest becomes significant, a relationship between the dependent variable and the variable of interest is suggested.

The assumption of *independence* may be inappropriate in a number of circumstances. In some cases, failure of the assumption makes multiple regression analysis an unsuitable statistical technique; in other cases, modifications or adjustments within the regression framework can be made to accommodate the failure.

The independence assumption may fail, for example, in a study of individual behavior over time, in which an unusually high error value in one time period is likely to lead to an unusually high value in the next time period. For example, if an economic forecaster underpredicted this year's Gross National Product (GNP), he or she is likely to underpredict next year's as well; the factor that caused the prediction error (e.g., an incorrect assumption about Federal Reserve policy) is likely to be a source of error in the future.

Alternatively, the assumption of independence may fail in a study of a group of firms at a particular point in time, in which error terms for large firms are systematically higher than error terms for small firms. For example, an analysis of the profitability of firms may not accurately account for the importance of advertising as a source of increased sales and profits. To the extent that large firms advertise more than small firms, the regression errors would be large for the large firms and small for the small firms.

In some cases, there are statistical tests that are appropriate for evaluating the independence assumption.<sup>44</sup> If the assumption has failed, the expert should ask first whether the source of the lack of independence is the omission of an important explanatory variable from the regression. If so, that variable should be included when possible, or the potential effect of its omission should be estimated when inclusion is not possible. If there is no important missing explanatory variable, the expert should apply one or more procedures that modify the standard multiple regression technique to allow for more accurate estimates of the regression parameters.<sup>45</sup>

#### 4. To what extent are the regression results sensitive to individual data points?

Estimated regression coefficients can be highly sensitive to particular data points. Suppose, for example, that one data point deviates greatly from its expected value, as indicated by the regression equation, while the remaining data points show little deviation. It would not be unusual in this situation for the co-

<sup>44.</sup> In a time-series analysis, the correlation of error values over time, the *serial correlation*, can be tested (in most cases) using a Durbin-Watson test. The possibility that some disturbance terms are consistently high in magnitude while others are systematically low, *heteroscedasticity*, can also be tested in a number of ways. *See*, *e.g.*, Pindyck & Rubinfeld, *supra* note 28, at 126–56.

<sup>45.</sup> When serial correlation is present, a number of closely related statistical methods are appropriate, including generalized differencing (a type of generalized least-squares) and maximum-likelihood estimation. When heteroscedasticity is the problem, weighted least-squares and maximum-likelihood estimation are appropriate. *See, e.g.*, Pindyck & Rubinfeld, *supra* note 28, at 126–56. All these techniques are readily available in a number of statistical computer packages. They also allow one to perform the appropriate statistical tests of the significance of the regression coefficients.

efficients in a multiple regression to change substantially if the data point were removed from the sample.

Evaluating the robustness of multiple regression results is a complex endeavor. Consequently, there is no agreed on set of tests for robustness which analysts should apply. In general, it is important to explore the reasons for unusual data points. If the source is an error in recording data, the appropriate corrections can be made. If all the unusual data points have certain characteristics in common (e.g., they all are associated with a supervisor who consistently gives high ratings in an equal pay case), the regression model should be modified appropriately.

One generally useful diagnostic technique is to see to what extent the estimated parameter changes as each data point (or points) in the regression analysis is dropped from the sample. An *influential data point*—a point that causes the estimated parameter to change substantially—should be studied further to see whether mistakes were made in the use of the data or whether important explanatory variables were omitted.<sup>46</sup>

#### 5. To what extent are the data subject to measurement error?

In multiple regression analysis it is assumed that variables are measured accurately.<sup>47</sup> If there are measurement errors in the dependent variable, estimates of regression parameters will be less accurate, though they will not necessarily be biased. However, if one or more *independent variables* are measured with error, the corresponding parameter estimates are likely to be biased, typically toward 0.<sup>48</sup>

To understand why, suppose that the dependent variable, salary, is measured without error, and the explanatory variable, experience, is subject to measurement error. (Seniority or years of experience should be accurate, but the type of experience is subject to error, since applicants may overstate previous job responsibilities.) As the measurement error increases, the estimated parameter associated with the experience variable will tend toward 0—eventually, there will be no relationship between salary and experience.

It is important for any source of measurement error to be carefully evaluated. In some circumstances, little can be done to correct the measurement error problem; the regression results must be interpreted in that light. In other cases, however, measurement errors can be corrected by finding a new, more reliable data source. Finally, alternative estimation techniques (using related variables

<sup>46.</sup> A more complete and formal treatment of the robustness issue appears in David A. Belsley et al., Regression Diagnostics: Identifying Influential Data and Sources of Collinearity (1980).

<sup>47.</sup> Inaccuracy can occur not only in the precision by which a particular variable is measured, but also in the precision with which the variable to be measured corresponds to the appropriate theoretical construct specified by the regression model.

<sup>48.</sup> Other coefficient estimates are likely to be biased as well.

that are measured without error) can be applied to remedy the measurement error problem in some situations.  $^{\rm 49}$ 

49. See, e.g., Pindyck & Rubinfeld, supra note 28, at 157–79 (discussion of instrumental variables estimation).

# **IV.** The Expert

Multiple regression analysis is taught to students in an extremely diverse set of fields, including statistics, economics, political science, sociology, psychology, anthropology, public health, and history. Consequently, any individual with substantial training in and experience with multiple regression and other statistical methods may be qualified as an expert.<sup>50</sup> A doctoral degree in a discipline that teaches theoretical or applied statistics, such as economics, history, and psychology, usually signifies to other scientists that the proposed expert meets this preliminary test of the qualification process.

The decision to qualify an expert in regression analysis rests with the court. Clearly, the proposed expert should be able to demonstrate an understanding of the discipline. Publications relating to regression analysis in peer-reviewed journals, active memberships in related professional organizations, courses taught on regression methods, and practical experience with regression can indicate a professional's expertise. However, the expert's background and experience with the specific issues and tools that are applicable to a particular case should also be considered during the qualification process.

50. A proposed expert whose only statistical tool is regression analysis may not be able to judge when a statistical analysis should be based on an approach other than regression.

# V. Presentation of Statistical Evidence

The costs of evaluating statistical evidence can be reduced and the precision of that evidence increased if the discovery process is used effectively. The following questions should be considered in evaluating the admissibility of statistical evidence.<sup>51</sup> These considerations are motivated by two concerns: (1) Has the expert provided sufficient information to replicate the multiple regression analysis? (2) Are the methodological choices that the expert made reasonable, or are they arbitrary and unjustified?

# A. What Disagreements Exist Regarding Data on Which the Analysis Is Based?

In general, a clear and comprehensive statement of the underlying research methodology is a requisite part of the discovery process. The expert should be encouraged to reveal both the nature of the experimentation carried out and the sensitivity of the results to the data and to the methodology. The following are suggestions of a number of useful requirements that can substantially improve the discovery process.

- 1. To the extent possible, the parties should be encouraged to agree to use a common database. Early agreement on a common database, even if disagreement about the significance of the data remains, can help focus the discovery process on the important issues in the case.
- 2. A party that offers data to be used in statistical work, including multiple regression analysis, should be encouraged to provide the following to the other parties: (a) a hard copy of the data when available and manageable in size, along with the underlying sources; (b) computer disks or tapes on which the data are recorded; (c) complete documentation of the disks or tapes; (d) computer programs that were used to generate the data (in hard copy, on a computer disk or tape, or both); and (e) documentation of such computer programs.

<sup>51.</sup> See also David H. Kaye & David A. Freedman, Reference Guide on Statistics § I.B, in this manual.

- 3. A party offering data should make available the personnel involved in the compilation of such data to answer the other parties' technical questions concerning the data and the methods of collection or compilation.
- 4. A party proposing to offer an expert's regression analysis at trial should ask the expert to fully disclose: (a) the database and its sources; <sup>52</sup> (b) the method of collecting the data; and (c) the methods of analysis. When possible, this disclosure should be made sufficiently in advance of trial so that the opposing party can consult its experts and prepare cross-examination. The court must decide on a case-by-case basis where to draw the disclosure line.
- 5. An opposing party should be given the opportunity to object to a database or to a proposed method of analysis of the database to be offered at trial. Objections may be to simple clerical errors or to more complex issues relating to the selection of data, the construction of variables, and, on occasion, the particular form of statistical analysis to be used. Whenever possible, these objections should be resolved before trial.
- 6. The parties should be encouraged to resolve differences as to the appropriateness and precision of the data to the extent possible by informal conference. The court should make an effort to resolve differences before trial.
- B. What Database Information and Analytical Procedures Will Aid in Resolving Disputes over Statistical Studies?<sup>33</sup>
  - 1. The expert should state clearly the objectives of the study, as well as the time frame to which it applies and the statistical population to which the results are being projected.
  - 2. The expert should report the units of observation (e.g., consumers, businesses, or employees).
  - 3. The expert should clearly define each variable.
  - 4. The expert should clearly identify the sample of data being studied, <sup>54</sup> as well as the method by which the sample was obtained.

<sup>52.</sup> These sources would include all variables used in the statistical analyses conducted by the expert, not simply those variables used in a final analysis on which the expert expects to rely.

<sup>53.</sup> For a more complete discussion of these requirements, see The Evolving Role of Statistical Assessments as Evidence in the Courts app. F at 256 (Stephen E. Fienberg ed., 1989) (Recommended Standards on Disclosure of Procedures Used for Statistical Studies to Collect Data Submitted in Evidence in Legal Cases).

<sup>54.</sup> The sample information is important because it allows the expert to make inferences about the underlying population.

- 5. The expert should reveal if there are missing data, whether caused by a lack of availability (e.g., in business data) or nonresponse (e.g., in survey data), and the method used to handle the missing data (e.g., deletion of observations).
- 6. The expert should report investigations that were made into errors associated with the choice of variables and assumptions underlying the regression model.
- 7. If samples have been chosen randomly from a population (i.e., *probabil-ity sampling* procedures have been used),<sup>55</sup> the expert should make a good faith effort to provide an estimate of a *sampling error*, the measure of the difference between the sample estimate of a parameter (such as the *mean* of a dependent variable under study) and the (unknown) population parameter (the population mean of the variable).<sup>56</sup>
- 8. If probability sampling procedures have not been used, the expert should report the set of procedures that were used to minimize sampling errors.

<sup>55.</sup> In probability sampling, each representative of the population has a known probability of being in the sample. Probability sampling is ideal because it is highly structured, and in principle, it can be replicated by others. Nonprobability sampling is less desirable because it is often subjective, relying to a large extent on the judgment of the expert.

<sup>56.</sup> Sampling error is often reported in terms of standard errors or confidence intervals. See infra the Appendix for details.

# **Appendix: The Basics of Multiple Regression**

## I. Introduction

This appendix illustrates, through examples, the basics of multiple regression analysis in legal proceedings.

Often, visual displays are used to describe the relationship between variables that are used in multiple regression analysis. Figure 2 is a *scatterplot* that relates scores on a job aptitude test (shown on the x-axis) and job performance ratings (shown on the y-axis). Each point on the scatterplot shows where a particular individual scored on the job aptitude test and how his or her job performance was rated. For example, the individual represented by Point A in Figure 2 scored 49 on the job aptitude test and had a job performance rating of 62.



The relationship between two variables can be summarized by a *correlation coefficient*, which ranges in value from -1 (a perfect negative relationship) to +1 (a perfect positive relationship). Figure 3 depicts three possible relationships between the job aptitude variable and the job performance variable. In Figure 3(a) there is a positive correlation: In general, higher job performance ratings are associated with higher aptitude test scores, and lower job performance ratings are associated with lower aptitude test scores. In Figure 3(b) the correlation is negative: Higher job performance ratings are associated with lower aptitude test scores, and lower aptitude test scores, and lower aptitude test scores. In Figure 3(b) the correlation is negative: Higher job performance ratings are associated with higher aptitude test scores. Note: Higher aptitude test scores are associated with higher aptitude test scores are associated with higher aptitude test scores. In Figure 3(b) the correlation is negative: Higher job performance ratings are associated with lower aptitude test scores. In Figure 3(c).



Job Aptitude Test Score

Multiple regression analysis goes beyond the calculation of correlations; it is a method in which a regression line is used to relate the average of one variable—the dependent variable—to the values of other explanatory variables. As a result, regression analysis can be used to predict the values of one variable using the values of others. For example, if average job performance ratings depend on aptitude test scores, regression analysis can use information about test scores to predict job performance.

A regression line is the best-fitting straight line through a set of points in a scatterplot. If there is only one explanatory variable, the straight line is defined by the equation:

$$Y = a + bX$$

In the equation above, *a* is the *intercept* of the line with the *y*-axis when X equals 0, and *b* is the *slope*—the amount of vertical change in the line for each unit of change in the horizontal direction. In Figure 4, for example, when the aptitude test score is 0, the predicted (average) value of the job performance rating is the intercept, 18.4. Also, for each additional point on the test score, the job performance rating increases .73 units, which is given by the slope .73. Thus, the estimated regression line is:

$$Y = 18.4 + 73 X$$

The regression line typically is estimated using the standard method of *least-squares*, where the values of *a* and *b* are calculated so that the sum of the squared deviations of the points from the line are minimized. In this way, positive deviations and negative deviations of equal size are counted equally, and large deviations are counted more then small deviations. In Figure 4 the deviation lines are vertical because the equation is predicting job performance ratings from aptitude tests scores, not aptitude test scores from job performance ratings.



The important variables that systematically might influence the dependent variable, and for which data can be obtained, typically should be included explicitly in a statistical model. All remaining influences, which should be small individually, but can be substantial in the aggregate, are included in an additional random error term.<sup>57</sup> Multiple regression is a procedure that separates the systematic effects (associated with the explanatory variables) from the random effects (associated with the error term) and also offers a method of assessing the success of the process.

## II. Linear Regression Model

When there is an arbitrary number of explanatory variables, the *linear regression* model takes the following form:

$$\mathcal{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \kappa + \beta_k X_k + \varepsilon \tag{1}$$

where Y represents the dependent variable, such as the salary of an employee, and  $X_1 \dots X_k$  represent the explanatory variables (e.g., the experience of each employee and his or her sex, coded as a 1 or 0, respectively). The error term  $\varepsilon$ represents the collective unobservable influence of any omitted variables. In a

<sup>57.</sup> It is clearly advantageous for the random component of the regression relationship to be small relative to the variation in the dependent variable.

linear regression each of the terms being added involves unknown parameters,  $\beta_0$ ,  $\beta_1$ , ...,  $\beta_k$ ,<sup>58</sup> which are estimated by "fitting" the equation to the data using least-squares.

Most statisticians use the least-squares regression technique because of its simplicity and its desirable statistical properties. As a result, it also is used frequently in legal proceedings.

#### A. An Example

Suppose an expert wants to analyze the salaries of women and men at a large publishing house to discover whether a difference in salaries between employees with similar years of work experience provides evidence of discrimination. <sup>59</sup> To begin with the simplest case, Y, the salary in dollars per year, represents the dependent variable to be explained, and  $X_1$  represents the explanatory variable—the number of years of experience of the employee. The regression model would be written:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \tag{2}$$

In equation (2),  $\beta_0$  and  $\beta_1$  are the parameters to be estimated from the data, and  $\varepsilon$  is the random error term. The parameter  $\beta_0$  is the average salary of all employees with no experience. The parameter  $\beta_1$  measures the average effect of an additional year of experience on the average salary of employees.

#### B. Regression Line

Once the parameters in a regression equation, such as equation (1), have been estimated, the *fitted values* for the dependent variable can be calculated. If we denote the estimated regression parameters, or regression coefficients, for the model in equation (1) by  $b_0, b_1, \ldots, b_{k'}$  the fitted values for Y, denoted  $\hat{Y}$ , are given by:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + K b_k X_k$$
(3)

Figure 5 illustrates this for the example involving a single explanatory variable. The data are shown as a scatter of points; salary is on the vertical axis and years of experience is on the horizontal axis. The estimated regression line is drawn through the data points. It is given by:

<sup>58.</sup> The variables themselves can appear in many different forms. For example, Y might represent the logarithm of an employee's salary, and  $X_1$  might represent the logarithm of the employee's years of experience. The logarithmic representation is appropriate when Y increases exponentially as X increases—for each unit in crease in X, the corresponding increase in Y becomes larger and larger. For example, if an expert were to graph growth of U.S. population (Y) over time (t), an equation of the form  $\log(Y) = \beta_0 + \beta_1 \log(t)$  might be appropriate.

<sup>59.</sup> The regression results used in this example are based on data for 1,715 men and women, which were used by the defense in a sex discrimination case against the New York Times that was settled in 1978. Professor Orley Ashenfelter, of the Department of Economics, Princeton University, provided the data.

$$\ddot{Y} = \$15,000 + \$2,000 X_1$$
 (4)

Thus, the fitted value for the salary associated with an individual's years of experience  $X_{1i}$  is given by:

$$Y_i = b_0 + b_1 X_{1i}$$
 (at Point B)

The intercept of the straight line is the average value of the dependent variable when the explanatory variable (or variables) is equal to 0; the intercept  $b_0$  is shown on the vertical axis in Figure 5. Similarly, the slope of the line measures the (average) change in the dependent variable associated with a unit increase in an explanatory variable; the slope  $b_1$  also is shown. In equation (4), the intercept \$15,000 indicates that employees with no experience earn \$15,000 per year. The slope parameter implies that each year of experience adds \$2,000 to an "average" employee's salary.





Now, suppose that the salary variable is related simply to the sex of the employee. The relevant indicator variable, often called a *dummy variable*, is  $X_2$ , which is equal to 1 if the employee is male, and 0 if the employee is female. Suppose the regression of salary Y on  $X_2$  yields the following result:

$$Y = $30, 449 + $10,979 X_2$$

The coefficient \$10,979 measures the difference between the average salary of men and the average salary of women.<sup>60</sup>

#### 1. Regression Residuals

For each data point, the regression *residual* is the difference between the actual and fitted values of the dependent variable. Suppose, for example, that we are studying an individual with three years of experience and a salary of \$27,000. According to the regression line in Figure 5, the average salary of an individual with three years of experience is \$21,000. Since the individual's salary is \$6,000 higher than the average salary, the residual (the individual's salary minus the average salary) is \$6,000.

In general, the residual *e* associated with a data point, such as Point A in Figure 5, is given by:

$$e = Y_i - \ddot{Y}_i$$

Each data point in the figure has a residual, which is the error made by the least-squares regression method for that individual.

#### 2. Nonlinearities

Nonlinear models account for the possibility that the effect of an explanatory variable on the dependent variable may vary in magnitude as the level of the explanatory variable changes. One useful nonlinear model uses interactions among variables to produce this effect. For example, suppose that

$$S = \beta_1 + \beta_2 SEX + \beta_3 EXP + \beta_4 (EXP \times SEX) + \varepsilon$$
(5)

where S is annual salary, SEX is equal to 1 for women and 0 for men, EXP represents years of job experience, and  $\varepsilon$  is a random error term. The coefficient  $\beta_2$  measures the difference in average salary (across all experience levels) between men and women for employees with no experience. The coefficient  $\beta_3$  measures the effect of experience on salary for men (when SEX = 0), and the coefficient  $\beta_4$  measures the difference in the effect of experience on salary between men and women. It follows, for example, that the effect of one year of experience on salary for men is  $\beta_3$ , while the comparable effect for women is  $\beta_3 + \beta_4$ .<sup>61</sup>

The difference, \$41,428 - \$30,449, is \$10,979.

<sup>60.</sup> To understand why, note that when  $X_2$  equals 0, the average salary for women is \$30,449 + \$10,979 × 0 = \$30,449

Correspondingly, when X, equals 1, the average salary for men is

 $<sup>\$30,449 + \$10,979 \</sup>times 1 = \$41,428$ 

<sup>61.</sup> Estimating a regression in which there are interaction terms for all explanatory variables, as in equation (5), is essentially the same as estimating two separate regressions, one for men and one for women.

#### III. Interpreting Regression Results

To understand how regression results are interpreted, the earlier example associated with Figure 5 can be expanded to consider the possibility of an additional explanatory variable—the square of the number of years of experience,  $X_3$ . The  $X_3$  variable is designed to capture the fact that for most individuals, salaries increase with experience, but eventually salaries tend to level off. The estimated regression line using the third additional explanatory variable, as well as the first explanatory variable for years of experience ( $X_1$ ) and the dummy variable for sex ( $X_2$ ), is

$$\ddot{Y} = \$14,085 + \$2,323 X_1 + \$1,675 X_2 - \$36 X_3$$
 (6)

The importance of including relevant explanatory variables in a regression model is illustrated by the change in the regression results after the  $X_3$  and  $X_2$  variables are added. The coefficient on the variable  $X_2$  measures the difference in the salaries of men and women while holding the effect of experience constant. The differential of \$1,675 is substantially lower than the previously measured differential of \$10,979. Clearly, failure to control for job experience in this example leads to an overstatement of the difference in salaries between men and women.

Now consider the interpretation of the explanatory variables for experience,  $X_1$  and  $X_3$ . The positive sign on the  $X_1$  coefficient shows that salary increases with experience. The negative sign on the  $X_3$  coefficient indicates that the rate of salary increase decreases with experience. To see the combined effect of the variables  $X_1$  and  $X_3$ , some simple calculations can be made. For example, consider how the average salary of women ( $X_2 = 0$ ) changes with the level of experience. As experience increases from 0 to 1 year, the average salary increases by \$2,251, from \$14,085 to \$16,336. However, women with 2 years of experience earn only \$2,179 more than women with 1 year of experience, and women with 3 years of experience earn only \$2,127 more than women with 2 years. Further, women with 7 years of experience earn \$28,582 per year, which is only \$1,855 more than the \$26,727 earned by women with 6 years of experience.<sup>62</sup> Figure 6 illustrates the results; the regression line shown is for women's salaries; the corresponding line for men's salaries would be parallel and \$1,675 higher.

62. These numbers can be calculated by substituting different values of  $X_1$  and  $X_3$  in equation (6).



### IV. Determining the Precision of the Regression Results

Least-squares regression provides not only parameter estimates that indicate the direction and magnitude of the effect of a change in the explanatory variable on the dependent variable, but also an estimate of the reliability of the parameter estimates and a measure of the overall goodness-of-fit of the regression model. Each of these factors is considered in turn.

#### A. Standard Errors of the Coefficients and *t*-Statistics

Estimates of the true but unknown parameters of a regression model are numbers that depend on the particular sample of observations under study. If a different sample were used, a different estimate would be calculated.<sup>63</sup> If the expert continued to collect more and more samples and generated additional estimates, as might happen when new data became available over time, the estimates of each parameter would follow a *probability distribution* (i.e., the expert could determine the percentage or frequency of the time that each estimate occurs). This probability distribution can be summarized by a mean and a measure of

Multiple Regression

<sup>63.</sup> The least-squares formula that generates the estimates is called the least-squares estimator, and its values vary from sample to sample.

dispersion around the mean, a standard deviation, that usually is referred to as the standard error of the coefficient , or the standard error .<sup>64</sup>

Suppose, for example, that an expert is interested in estimating the average price paid for a gallon of unleaded gasoline by consumers in a particular geographic area of the United States at a particular point in time. The mean price for a sample of ten gas stations might be \$1.25, while the mean for another sample might be \$1.29, and the mean for a third, \$1.21. On this basis, the expert also could calculate the overall mean price of gasoline to be \$1.25 and the standard deviation to be \$0.04.

Least-squares regression generalizes this result, by calculating means whose values depend on one or more explanatory variables. The standard error of a regression coefficient tells the expert how much the parameter estimate is likely to vary from sample to sample. The greater the variation in parameter estimates from sample to sample, the larger the standard error and consequently the less reliable the regression results. Small standard errors imply results that are likely to be similar from sample to sample, while results with large standard errors show more variability.

Under appropriate assumptions, the least-squares *estimators* provide "best" determinations of the true underlying parameters. <sup>65</sup> In fact, least-squares has several desirable properties. First, least-squares estimators are unbiased. Intuitively, this means that if the regression were calculated over and over again with different samples, the average of the many estimates obtained for each coefficient would be the true parameter. Second, least-squares estimators are *consistent*; if the sample were very large, the estimates obtained would come close to the true parameters. Third, least-squares is *efficient*, in that its estimators have the smallest variance among all (linear) unbiased estimators.

If the further assumption is made that the probability distribution of each of the error terms is known, statistical statements can be made about the precision of the coefficient estimates. For relatively large samples (often, thirty or more data points will be sufficient for regressions with a small number of explanatory variables), the probability that the estimate of a parameter lies within an interval of 2 standard errors around the true parameter is approximately .95, or 95%. A frequent, although not always appropriate, assumption in statistical work is that the error term follows a *normal distribution*, from which it follows that the estimated parameters are normally distributed. The normal distribution has the property that the area within 1.96 standard errors of the mean is equal to 95% of the total area. Note that the normality assumption is not necessary for least-

<sup>64.</sup> See David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

<sup>65.</sup> The necessary assumptions of the regression model include (a) the model is specified correctly; (b) errors associated with each observation are drawn randomly from the same probability distribution and are independent of each other; (c) errors associated with each observation are independent of the corresponding observations for each of the explanatory variables in the model; and (d) no explanatory variable is correlated perfectly with a combination of other variables.

squares to be used, since most of the properties of least-squares apply regardless of normality.

In general, for any parameter estimate *b*, the expert can construct an interval around *b* such that there is a 95% probability that the interval covers the true parameter. This 95% confidence interval <sup>66</sup> is given by:

$$b \pm 1.96 \times (\text{standard error of } b)^{67}$$
 (7)

The expert can test the hypothesis that a parameter is actually equal to 0—often stated as testing the null hypothesis—by looking at its t-statistic, which is defined as:

$$t = \frac{b}{\text{standard error of } b}$$
(8)

If the *t*-statistic is less than 1.96 in magnitude, the 95% confidence interval around *b* must include 0.<sup>68</sup> Because this means that the expert cannot reject the hypothesis that  $\beta$  equals 0, the estimate, whatever it may be, is said to be not statistically significant. Conversely, if the *t*-statistic is greater than 1.96 in absolute value, the expert concludes that the true value of  $\beta$  is unlikely to be 0 (intuitively, *b* is "too far" from 0 to be consistent with the true value of  $\beta$  being 0). In this case, the expert rejects the hypothesis that  $\beta$  equals 0 and calls the estimate statistically significant. If the null hypothesis  $\beta$  equals 0 is true, using a 95% confidence level will cause the expert to falsely reject the null hypothesis 5% of the time. Consequently, results often are said to be significant at the 5% level. <sup>69</sup>

As an example, consider a more complete set of regression results associated with the salary regression described in equation (6):

$$Y = \$14,085 + \$2,323 X_1 + \$1,675 X_2 - \$36 X_3 (1,577) (140) (1,435) (3.4) (9) t = \$.9 16.5 1.2 -10.8$$

The standard error of each estimated parameter is given in parentheses directly below the parameter, and the corresponding *t*-statistics appear below the standard error values.

Consider the coefficient on the dummy variable  $X_2$ . It indicates that \$1,675 is the best estimate of the mean salary difference between men and women.

<sup>66.</sup> Confidence intervals are used commonly in statistical analyses because the expert can never be certain that a parameter estimate is equal to the true population parameter.

<sup>67.</sup> If the number of data points in the sample is small, the standard error must be multiplied by a number larger than 1.96.

<sup>68.</sup> The *t*-statistic applies to any sample size. As the sample gets large, the underlying distribution, which is the source of the *t*-statistic (the student's *t* distribution), approximates the normal distribution.

<sup>69.</sup> A t-statistic of 2.57 in magnitude or greater is associated with a 99% confidence level, or a 1% level of significance, that includes a band of 2.57 standard deviations on either side of the estimated coefficient.

However, the standard error of \$1,435 is large in relation to its coefficient \$1,675. Because the standard error is relatively large, the range of possible values for measuring the true salary difference, the true parameter, is great. In fact, a 95% confidence interval is given by:

 $1,675 \pm 1,435 \times 1.96 = 1,675 \pm 2,813$ 

In other words, the expert can have 95% confidence that the true value of the coefficient lies between -\$1,138 and \$4,488. Because this range includes 0, the effect of sex on salary is said to be insignificantly different from 0 at the 5% level. The *t* value of 1.2 is equal to \$1,675 divided by \$1,435. Because this *t*-statistic is less than 1.96 in magnitude (a condition equivalent to the inclusion of a 0 in the above confidence interval), the sex variable again is said to be an insignificant determinant of salary at the 5% level of significance.

Note also that experience is a highly significant determinant of salary, since both the  $X_1$  and the  $X_3$  variables have *t*-statistics substantially greater than 1.96 in magnitude. More experience has a significant positive effect on salary, but the size of this effect diminishes significantly with experience.

#### B. Goodness-of-Fit

Reported regression results usually contain not only the point estimates of the parameters and their standard errors or *t*-statistics, but also other information that tells how closely the regression line fits the data. One statistic, the *standard error* of the regression (SER), is an estimate of the overall size of the regression residuals.<sup>70</sup> An SER of 0 would occur only when all data points lie exactly on the regression line—an extremely unlikely possibility. Other things being equal, the larger the SER, the poorer the fit of the data to the model.

For a normally distributed error term, the expert would expect approximately 95% of the data points to lie within 2 SERs of the estimated regression line, as shown in Figure 7 (in Figure 7 the SER is approximately \$5,000).

*R-square* ( $\mathbb{R}^2$ ) is a statistic that measures the percentage of variation in the dependent variable that is accounted for by all the explanatory variables.<sup>71</sup> Thus,  $\mathbb{R}^2$  provides a measure of the overall goodness-of-fit of the multiple regression equation.<sup>72</sup> Its value ranges from 0 to 1. An  $\mathbb{R}^2$  of 0 means that the explanatory variables explain none of the variation of the dependent variable; an  $\mathbb{R}^2$  of 1 means that the explanatory variables explain the variables explain the variation in the dependent variable perfectly. The  $\mathbb{R}^2$  associated with equation (9) is .56. This implies that the three explanatory variables explain 56% of the variation in salaries.

<sup>70.</sup> More specifically, it is a measure of the standard deviation of the regression error *e*. It sometimes is called the root mean square error of the regression line.

<sup>71.</sup> The variation is the square of the difference between each Y value and the average Y value, summed over all the Y values.

<sup>72.</sup>  $R^2$  and SER provide similar information, because  $R^2$  is approximately equal to  $1 - SER^2$ /Variance of Y.

Figure 7 Standard Error of the Regression



Experience (X<sub>1</sub>)

What level of  $\mathbb{R}^2$ , if any, should lead to a conclusion that the model is satisfactory? Unfortunately, there is no clear-cut answer to this question, since the magnitude of  $\mathbb{R}^2$  depends on the characteristics of the data series being studied and, in particular, whether the data vary over time or over individuals. Typically, an  $\mathbb{R}^2$  is low in cross-sectional studies in which differences in individual behavior are explained. It is likely that these individual differences are caused by many factors that cannot be measured. As a result, the expert cannot hope to explain most of the variation. In time-series studies, in contrast, the expert is explaining the movement of aggregates over time. Since most aggregate time series have substantial growth, or trend, in common, it will not be difficult to "explain" one time series using another time series, simply because both are moving together. It follows as a corollary that a high  $\mathbb{R}^2$  does not by itself mean that the variables included in the model are the appropriate ones.

As a general rule, courts should be reluctant to rely solely on a statistic such as  $R^2$  to choose one model over another. Alternative procedures and tests are available.<sup>73</sup>

Multiple Regression

<sup>73.</sup> These include *F*-tests and specification error tests. See Pindyck & Rubinfeld, *supra* note 28, at 107–13, 149–55, 224–28.

#### C. Sensitivity of Least-Squares Regression Results

The least-squares regression line can be sensitive to extreme data points. This sensitivity can be seen most easily in Figure 8. Assume initially that there are only three data points, A, B, and C, relating information about  $X_1$  to the variable Y. The least-squares line describing the best-fitting relationship between Points A, B, and C is represented by Line 1. Point D is called an *outlier* because it lies far from the regression line that fits the remaining points. When a new, best-fitting least-squares line is reestimated to include Point D, Line 2 is obtained. Figure 8 shows that the outlier Point D is an influential data point, since it has a dominant effect on the slope and intercept of the least-squares line. Because least squares attempts to minimize the sum of squared deviations, the sensitivity of the line to individual points sometimes can be substantial.<sup>74</sup>

#### Figure 8

Least-Squares Regression



What makes the influential data problem even more difficult is that the effect of an outlier may not be seen readily if deviations are measured from the final regression line. The reason is that the influence of Point D on Line 2 is so substantial that its deviation from the regression line is not necessarily larger than the deviation of any of the remaining points from the regression line.<sup>75</sup> Although

<sup>74.</sup> This sensitivity is not always undesirable. In some cases it may be much more important to predict Point D when a big change occurs than to measure the effects of small changes accurately.

<sup>75.</sup> The importance of an outlier also depends on its location in the data set. Outliers associated with relatively extreme values of explanatory variables are likely to be especially influential.

they are not as popular as least-squares, alternative estimation techniques that are less sensitive to outliers, such as *robust estimation*, are available.

## V. Reading Multiple Regression Computer Output

Statistical computer packages that report multiple regression analyses vary to some extent in the information they provide and the form that the information takes. The following table contains a sample of the basic computer output that is associated with equation (6).

#### Table 1 Regression Output

Dependent Variable: Y		SSE 62346266124		F-Test	174.71	
		DFE 561		Prob > F	0.0001	
		MSE 111134164		$\mathbb{R}^2$	0.5560	
			Standard			
Variable	DF	Parameter Esti	mate Error	t-stat	$\mathbf{Prob} >  t $	
Intercept	1	14084.89	1577.484	8.9287	.0001	
X <sub>1</sub>	1	1675.11	1435.422	1.1670	.2437	
X <sub>2</sub>	1	2323.17	140.70	16.5115	.0001	
X <sub>3</sub>	1	-36.71	3.41	-10.7573	.0001	

Note: SSE = sum of squared errors; DFE = degrees of freedom associated with the error term; MSE = mean square error; DF = degrees of freedom; t-stat = t-statistic; Prob = probability.

Beginning with the lower portion of Table 1, note that the parameter estimates, the standard errors, and the *t*-statistics match the values given in equation (9). <sup>76</sup> The variable "Intercept" refers to the constant term  $\beta_0$  in the regression. The column DF represents *degrees of freedom*. The "1" signifies that when the computer calculates the parameter estimates, each variable that is added to the linear regression adds an additional constraint that must be satisfied. The column labeled "Prob > |t|" lists the two-tailed *p*-values associated with each estimated parameter; the *p*-value measures the observed significance level—the probability of getting a test statistic as extreme or more extreme than the observed number if the model parameter is in fact 0. The very low *p*-values on the variables  $X_2$  and  $X_3$  imply that each variable is statistically significant at less than the 1% level—both highly significant results. On the contrary, the  $X_1$  coefficient is only significant at the 24% level, implying that it is insignificant at the traditional 5% level. Thus, the expert cannot reject with confidence the null hypoth-

<sup>76.</sup> Computer programs give results to more decimal places than are meaningful. This added detail should not be seen as evidence that the regression results are exact.

esis that salaries do not differ by sex after the expert has accounted for the effect of experience.

The top portion of Table 1 provides data that relate to the goodness-of-fit of the regression equation. The *sum of squared errors* (SSE) measures the sum of the squares of the regression residuals—the sum that is minimized by the least-squares procedure. The degrees of freedom associated with the error term (DFE) is given by the number of observations minus the number of parameters that were estimated. The *mean square error* (MSE) measures the variance of the error term (the square of the standard error of the regression). MSE is equal to SSE divided by DFE.

The R<sup>2</sup> of .556 indicates that 55.6% of the variation in salaries is explained by the regression variables,  $X_1$ ,  $X_2$ , and  $X_3$ . Finally, the *F*-test is a test of the null hypothesis that all regression coefficients (except the intercept) are jointly equal to 0—that there is no association between the dependent variable and any of the explanatory variables. This is equivalent to the null hypothesis that R<sup>2</sup> is equal to 0. In this case, the *F*-ratio of 174.71 is sufficiently high that the expert can reject the null hypothesis with a very high degree of confidence (i.e., with a 1% level of significance).

### VI. Forecasting

In general, a forecast is a prediction made about the values of the dependent variable using information about the explanatory variables. Often, *ex ante fore*-*casts* are performed; in this situation, values of the dependent variable are predicted beyond the sample (e.g., beyond the time period in which the model has been estimated). However, *ex post forecasts* are frequently used in damage analyses.<sup>77</sup> An ex post forecast has a forecast period such that all values of the dependent and explanatory variables are known; ex post forecasts can be checked against existing data and provide a direct means of evaluation.

For example, to calculate the forecast for the salary regression discussed above, the expert uses the estimated salary equation:

$$Y = \$14,085 + \$2,323 X_1 + \$1,675 X_2 - \$36 X_3$$
(10)

To predict the salary of a man with two years experience, the expert calculates:

$$Y(2) = \$14,085 + \$2,323 \times 2 + \$1,675 - \$36 \times 2^2 = \$20,262$$
(11)

<sup>77.</sup> Frequently, in cases involving damages, the question arises as to what the world would have been like had a certain event not taken place. For example, in a price-fixing antitrust case, the expert can ask what the price of a product would have been had a certain event associated with the price-fixing agreement not occurred. If prices would have been lower, the evidence suggests impact. If the expert can predict how much lower they would have been, the data can help the expert develop a numerical estimate of the amount of damages.

The degree of accuracy of both ex ante and ex post forecasts can be calculated provided that the model specification is correct and the errors are normally distributed and independent. The statistic is known as the *standard error of forecast* (SEF). The SEF measures the standard deviation of the forecast error that is made within a sample in which the explanatory variables are known with certainty.<sup>78</sup> The SEF can be used to determine how accurate a given forecast is. In equation (11), the SEF associated with the forecast of \$20,262 is approximately \$5,000. If a large sample size is used, the probability is roughly 95% that the predicted salary will be within 1.96 standard errors of the forecasted value. In this case, the appropriate 95% interval for the prediction is \$20,262 ± \$5,000 × 1.96 (\$10,822 to \$30,422). Because the estimated model does not explain salaries effectively, the SEF is large, as is the 95% interval. A more complete model with additional explanatory variables would result in a lower SEF and a smaller 95% interval for the prediction.

There is a danger when using the SEF, which applies to the standard errors of the estimated coefficients as well. The SEF is calculated on the assumption that the model includes the correct set of explanatory variables and the correct functional form. If the choice of variables or the functional form is wrong, the estimated forecast error may be misleading; in some cases, it may be smaller, perhaps substantially smaller, than the true SEF; in other cases, it may be larger, for example, if the wrong variables happen to capture the effects of the correct variables.

<sup>78.</sup> There are actually two sources of error implicit in the SEF. The first source arises because the estimated parameters of the regression model may not be exactly equal to the true regression parameters. The second source is the error term itself; when forecasting, the expert typically sets the error equal to 0 when a turn of events not taken into account in the regression model may make it appropriate to make the error positive or negative.



Experience (X<sub>1</sub>)

The difference between the SEF and the SER is shown in Figure 9. The SER measures deviations within the sample. The SEF is more general, since it calculates deviations within or without the sample period. In general, the difference between the SEF and the SER increases as the values of the explanatory variables increase in distance from the mean values. Figure 9 shows the 95% prediction interval created by the measurement of 2 SEFs about the regression line.

## **Glossary of Terms**

The following terms and definitions are adapted from a variety of sources, including A Dictionary of Epidemiology (John M. Last ed., 1983) and Robert S. Pindyck & Daniel L. Rubinfeld, Econometric Models & Economic Forecasts (3d ed. 1991).

Alternative Hypothesis. See Hypothesis Test.

- Association. The degree of statistical dependence between two or more events or variables. Events are said to be associated when they occur more frequently together than one would expect by chance.
- *Bias.* Any effect at any stage of investigation or inference tending to produce results that depart systematically (either too high or too low) from the true values. A biased estimator of a parameter differs on average from the true parameter.
- Coefficient. An estimated regression parameter.
- Confidence Interval. An interval that contains a true regression parameter with a given degree of confidence.
- Consistent Estimator. An estimator that tends to become more and more accurate as the sample size grows.
- *Correlation.* A statistical means of measuring the association between variables. Two variables are correlated positively if, on average, they move in the same direction; two variables are correlated negatively if, on average, they move in opposite directions.
- *Cross-Section Analysis.* A type of multiple regression analysis in which each data point is associated with a different unit of observation (e.g., an individual or a firm) measured at a particular point in time.
- Degrees of Freedom. The number of observations in a sample minus the number of estimated parameters in a regression model. A useful statistic in hypothesis testing.
- Dependent Variable. The variable to be explained or predicted in a multiple regression model.

- Dummy Variable. A variable that takes on only two values, usually 0 and 1, with one value indicating the presence of a characteristic, attribute, or effect and the other value indicating absence.
- *Efficient Estimator.* An estimator of a parameter that produces the greatest precision possible.
- *Error Term.* A variable in a multiple regression model that represents the cumulative effect of a number of sources of modeling error.
- Estimate. The calculated value of a parameter based on the use of a particular sample.
- *Estimator.* The sample statistic that estimates the value of a population parameter (e.g., a regression parameter); its values vary from sample to sample.
- *Ex Ante Forecast.* A prediction about the values of the dependent variable that go beyond the sample; consequently, the forecast must be based on predictions for the values of the explanatory variables in the regression model.
- *Explanatory Variable.* A variable that partially explains or predicts the movement of a dependent variable.
- *Ex Post Forecast.* A prediction about the values of the dependent variable made during a period in which all the values of the explanatory and dependent variables are known. Ex post forecasts provide a useful means of evaluating the fit of a regression model.
- *F-test.* A statistical test (based on an *F*-ratio) of the null hypothesis that a group of explanatory variables are jointly equal to 0. When applied to all the explanatory variables in a multiple regression model, the *F*-test becomes a test of the null hypothesis that R<sup>2</sup> equals 0.
- *Feedback.* When changes in an explanatory variable affect the values of the dependent variable, and changes in the dependent variable also affect the explanatory variable. When both effects occur at the same time, the two variables are described as being determined simultaneously.
- *Fitted Value.* The estimated value for the dependent variable; in a linear regression this value is calculated as the intercept plus a weighted average of the values of the explanatory variables, with the estimated parameters used as weights.
- Heteroscedasticity. When the disturbance or error associated with a multiple regression model has a nonconstant variance; that is, the error values associated with some observations are typically high, whereas the values associated with other observations are typically low.
- Hypothesis Test. A statement about the parameters in a multiple regression model. The null hypothesis may assert that certain parameters have speci-

fied values or ranges; the alternative hypothesis would specify other values or ranges.

- *Independence.* When two variables are not correlated with each other (in the population).
- Independent Variable. An explanatory variable that affects the dependent variable but is not affected by the dependent variable.
- Influential Data Point. A data point whose addition to a regression sample causes one or more estimated regression parameters to change substantially.
- Interaction Variable. The product of two explanatory variables in a regression model. Used in a particular form of nonlinear model.
- *Intercept.* The value of the dependent variable when each of the explanatory variables takes on the value of 0.
- Least-Squares. A common method for estimating regression parameters. Leastsquares minimizes the sum of the squared differences between the actual values of the dependent variable and the values predicted by the regression equation.
- Linear Model. A model having the property that the magnitude of the change in the dependent variable associated with the change in any of the explanatory variables is the same no matter what the level of that variable.
- Linear Regression. A regression model in which the effect of a change in each of the explanatory variables on the dependent variable is the same, no matter what the values of those explanatory variables.
- Mean (Sample). An average of the outcomes associated with a probability distribution, where the outcomes are weighted by the probability that each will occur.
- Mean Square Error (MSE). The estimated variance of the regression error, calculated as the average of the sum of the squares of the regression residuals.
- Model. A representation of an actual situation.
- Multicollinearity. Arises in multiple regression analysis when two or more variables are highly correlated. Substantial multicollinearity can cause regression parameters to be estimated imprecisely, as reflected in relatively high standard errors.
- Multiple Regression Analysis. A statistical tool for understanding the relationship between two or more variables.
- Multivariate Analysis. A set of techniques used to study the variation in several variables simultaneously.

- Nonlinear Model. A model having the property that changes in explanatory variables will have differential effects on the dependent variable as the values of the explanatory variables change.
- Normal Distribution. A bell-shaped probability distribution having the property that about 95% of the distribution lies within two standard deviations of the mean.
- Null Hypothesis. In regression analysis the null hypothesis states that the results observed in a study with respect to a particular variable are no different from what might have occurred by chance, independent of the effect of that variable. See Hypothesis Test.
- One-Tailed Test. A hypothesis test in which the alternative to the null hypothesis that a parameter is equal to 0 is for the parameter to be either positive or negative, but not both.
- *Outlier.* A data point that is more than some appropriate distance from a regression line that is estimated using all the other data points in the sample.
- *p*-Value. The probability of getting a test statistic as extreme or more extreme than the observed value. The larger the *p*-value, the more likely the null hypothesis is true.
- Parameter. A numerical characteristic of a population or a model.
- Perfect Collinearity. When two (or more) variables are explanatory variables are correlated perfectly.
- Population. All the units of interest to the researcher; also, universe.
- *Practical Significance.* Substantive importance. Statistical significance does not ensure practical significance, since, with large samples, small differences can be statistically significant.
- Probability Distribution. The process that generates the values of a random vari able. A probability distribution lists all possible outcomes and the probability that each will occur.
- Probability Sampling. A process by which a sample of a population is chosen so that each observation has a known probability of being selected.
- Random. Governed by chance; not completely determined by other factors.
- Random Error. Random error (sampling error) is due to chance when the result obtained in the sample differs from the result that would be obtained if the entire population were studied.
- Regression Coefficient. The estimate of a population parameter obtained from a regression equation that is based on a particular sample; also, regression parameter.

- *Residual.* The difference between the actual value of a dependent variable and the value predicted by the regression equation.
- *Robust.* A statistic or procedure that does not change much when data or assumptions are slightly modified.
- Robust Estimation. An alternative to least-squares estimation that is less sensitive to outliers.
- *R-Square* (R<sup>2</sup>). A statistic that measures the percentage of the variation in the dependent variable that is accounted for by all of the explanatory variables in a regression model. R-square is the most commonly used measure of goodness-of-fit of a regression model.
- Sample. A set of units selected for a study; a subset of a population.
- Sampling Error. A measure of the difference between the sample estimate of a parameter and the population parameter.
- Scatterplot. A graph showing the relationship between two variables in a study; each dot represents one subject. One variable is plotted along the horizontal axis; the other variable is plotted along the vertical axis.
- Serial Correlation. The correlation of the values of regression errors over time.
- *Slope.* The change in the dependent variable associated with a 1-unit change in an explanatory variable.
- Spurious Correlation. When two variables are correlated, but one is not the cause of the other.
- Standard Deviation. The square root of the variance of a random variable. The variance is a measure of the spread of a probability distribution about its mean; it is calculated as a weighted average of the squares of the deviations of the outcomes of a random variable from its mean.
- Standard Error of the Coefficient; Standard Error. A measure of the variation of a parameter estimate or coefficient about the true parameter. The standard error is a standard deviation that is calculated from the probability distribution of estimated parameters.
- Standard Error of Forecast (SEF). An estimate of the standard deviation of the forecast error; it is based on forecasts made within a sample in which the values of the explanatory variables are known with certainty.
- Standard Error of the Regression (SER). An estimate of the standard deviation of the regression error; it is calculated as an average of the squares of the residuals associated with a particular multiple regression analysis.
- Statistical Significance. Used to evaluate the degree of association between a dependent variable and one or more explanatory variables. If the calculated *p*value is smaller than 5%, the result is said to be statistically significant (at the

Multiple Regression
5% level). If p is less than 5%, the result is statistically insignificant (at the 5% level).

- *t*-Statistic. A test statistic that describes how far an estimate of a parameter is from its hypothesized value (i.e., given a null hypothesis). If a *t*-statistic is sufficiently large (in absolute magnitude), an expert can reject the null hypothesis.
- *t*-Test. A test of the null hypothesis that a regression parameter takes on a particular value, usually 0. The test is based on the *t*-statistic.
- *Time-Series Analysis.* A type of multiple regression analysis in which each data point is associated with a particular unit of observation (e.g., an individual or a firm) measured at different points in time.
- *Two-Tailed Test.* A hypothesis test in which the alternative to the null hypothesis that a parameter is equal to 0 is for the parameter to be either positive or negative, or both.
- *Variable.* Any attribute, phenomenon, condition, or event that can have different values.
- Variable of Interest. The explanatory variable that is the focal point of a particular study or legal issue.
- Weighting. Weighting is used when statistics such as the mean and standard deviation are calculated. If ten observations are equally likely to occur, then each is weighted 0.1 when the appropriate statistics are calculated. However, if the first five observations are three times as likely to occur as the second five, the first five receive weights of 0.15, and the second five receive weights of 0.05. (In each case the sum of the weights is 1.0.)

# **References on Multiple Regression**

- Thomas J. Campbell, Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet, 36 Stan. L. Rev. 1299 (1984).
- Arthur P. Dempster, Employment Discrimination and Statistical Science, 3 Stat. Sci. 149 (1988).
- The Evolving Role of Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989).
- Michael O. Finkelstein, The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases, 80 Colum. L. Rev. 737 (1980).
- Michael O. Finkelstein & Hans Levenbach, Regression Estimates of Damages in Price-Fixing Cases, Law & Contemp. Probs., Autumn 1983, at 145.
- Franklin M. Fisher, Statisticians, Econometricians, and Adversary Proceedings, 81 J. Am. Stat. Ass'n 277 (1986).
- Franklin M. Fisher, Multiple Regression in Legal Proceedings, 80 Colum. L. Rev. 702 (1980).
- Joseph L. Gastwirth, Methods for Assessing the Sensitivity of Statistical Comparisons Used in Title VII Cases to Omitted Variables, 33 Jurimetrics J. 19 (1992).
- Note, Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal, 89 Harv. L. Rev. 387 (1975).
- Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 Colum. L. Rev. 1048 (1985).
- Daniel L. Rubinfeld & Peter O. Steiner, *Quantitative Methods in Antitrust Litigation*, Law & Contemp. Probs., Autumn 1983, at 69.
- Symposium, Statistical and Demographic Issues Underlying Voting Rights Cases, 15 Evaluation Rev. 659 (1991).

# **Reference Guide on Estimation of Economic Losses in Damages Awards**

Robert E. Hall Victoria A. Lazear

**Robert E. Hall**, Ph.D., is Professor, Department of Economics and Senior Fellow, Hoover Institution, Stanford University, Stanford, California. **Victoria A. Lazear**, M.S., is Partner, Applied Economics Partners, Menlo Park, California.

# Contents

- I. Introduction 477
- II. Expert's Qualifications 479
- III. Issues Common to Most Damages Studies 481
  - A. Characterization of the Harmful Event 481
    - 1. How was the plaintiff harmed and what legal principles govern compensation for the harm? 481
    - 2. Are the parties disputing differences in the plaintiff's economic environment absent the harmful event? 483
    - 3. Is there disagreement about the causal effect of the injury? 485
    - 4. Is there disagreement about alternative nonharmful conduct of the defendant in projecting the plaintiff's earnings but for the harmful event? 486
    - 5. Are losses measured before or after the plaintiff's income taxes? 487
    - 6. Is there disagreement about the costs that the plaintiff would have incurred but for the harmful event? 488
    - 7. Is there a dispute about the costs of stock options? 489
  - B. Mitigation and Earnings Before Trial 490
    - 1. Is there a dispute about mitigation? 490
  - C. Prejudgment Interest 491
    - 1. Do the parties agree about how to calculate prejudgment interest? 493
  - D. Projections of Future Earnings 493
    - 1. Is there disagreement about the projection of profitability but for the harmful event? 493
    - 2. Is there disagreement about the plaintiff's actual earnings after the harmful event? 494
    - 3. Do the parties use constant dollars for future losses, or is there escalation for inflation? 494
  - E. Discounting Future Losses 495

- 1. Are the parties using a discount rate properly matched to the projection in constant dollars or escalated terms? 495
- 2. Is one of the parties assuming that discounting and earnings growth offset one another? 496
- 3. Is there disagreement about the interest rate used to discount future lost value? 497
- 4. Is one of the parties using a capitalization factor? 497
- 5. Is one party using the appraisal approach to valuation and the other the discounted-income approach? 499
- F. Other Issues Arising in General in Damages Measurement 499
  - 1. Is there disagreement about the role of subsequent unexpected events? 499
  - 2. How should damages be apportioned among the various stakeholders? 500
- IV. Subject Areas of Economic Loss Measurement 503
  - A. Personal Lost Earnings 503
    - 1. Is there a dispute about projected earnings but for the harmful event? 503
    - 2. What benefits are part of damages? 503
    - 3. Is there a dispute about mitigation? 504
    - 4. Is there disagreement about how the plaintiff's career path should be projected? 505
    - 5. Is there disagreement about how earnings should be discounted to present value? 506
    - 6. Is there disagreement about subsequent unexpected events? 507
    - 7. Is there disagreement about retirement and mortality? 507
  - B. Intellectual Property Damages 508
    - 1. Is there disagreement about what fraction of the defendant's sales would have gone to the plaintiff? 508
    - Is there disagreement about the effect of infringement or misappropriation on prices as well as quantities (price erosion)? 509
    - 3. Is there a dispute about whether the lost-profit calculation includes contributions from noninfringing features of the work or product (apportionment)? 509
    - 4. Do the parties disagree about whether the defendant could have designed around the plaintiff's patent? 510
    - 5. Is there disagreement about how much of the defendant's advantage actually came from infringement (apportionment)? 511

- 6. Is there disagreement about how to combine the plaintiff's loss and the defendant's gain in a way that avoids double counting? 511
- C. Antitrust Damages 511
  - 1. Is there disagreement about the scope of the damages? 512
  - 2. Is there a dispute about the causal link between the misconduct and the measured damages? 512
  - 3. Is there a dispute about how conditions would differ absent the challenged misconduct? 513
- D. Securities Damages 515
  - 1. Is there disagreement about when the adverse information affected the market? 515
  - 2. Is there disagreement about how to take proper account of turnover of the securities? 515
- E. Liquidated Damages 516
  - 1. Is there a dispute about the proper application of a provision for liquidated damages? 516

Appendix: Example of a Damages Study 517 Glossary of Terms 521 References on Damages Awards 523

# I. Introduction

This reference guide identifies areas of dispute that arise frequently when economic losses are at stake. Even though such evidence differs from other topics presented in this manual, it is included because expert testimony is commonly offered on these issues. This reference guide discusses the application of economic analysis within the established legal framework for damages. It is not a commentary on the legal framework. It does not lay out a comprehensive theory of damages measurement, nor does it describe the applicable law. We give only a few legal citations where courts have introduced economic principles into damages.

This reference guide has three major sections. Section II discusses the qualifications required of experts who quantify damages. Section III considers issues common to most studies of economic damages (the harmful event, pretrial earnings and mitigation, prejudgment interest, future earnings and losses, subsequent events, and apportionment). Section IV considers the major subject areas of economic loss measurement (personal lost earnings, intellectual property losses, antitrust losses, securities losses, and liquidated damages).

Our discussion follows the structure of the standard damages study, as shown in Figure 1. We assume that the defendant has been found liable for damages for a harmful event he or she committed sometime in the past. The plaintiff is entitled to recover monetary damages for losses occurring before and possibly after the time of the trial. The top line of Figure 1 measures the losses before trial; the bottom line measures the losses after trial.<sup>1</sup>

The defendant's harmful act has reduced the plaintiff's *earnings*, or stream of economic value. Earnings are the stream of economic value received in the form of compensation by a worker, the profit earned by a business, or one-time receipts, such as the proceeds from the sale of property. They are measured net of any associated costs.

The essential features of a study of losses are the quantification of the reduction in earnings, the calculation of interest on past losses, and the application of financial *discounting* to future losses. The losses are measured as the difference

<sup>1.</sup> Our scope here is limited to losses of actual dollar income. However, economists have a growing role in the measurement of non-dollar damages, including pain and suffering and the hedonic value of life. See generally W. Kip Viscusi, Reforming Products Liability (1991).

between the earnings the plaintiff would have received if the harmful event had not occurred and the earnings the plaintiff has or will receive, given the harmful event. The plaintiff may be entitled to interest for losses occurring before the trial. Losses occurring after trial will normally be discounted. The majority of damages studies fit this format, so we have used it as the basic organization of this reference guide.<sup>2</sup>

We use numerous brief examples to explain the disputes that can arise. These examples are not full case descriptions; they are deliberately stylized. They attempt to capture the types of disagreements about damages that arise in practical experience, though they are purely hypothetical. In many examples, the dispute involves factual as well as legal issues. We do not try to resolve the disputes in these examples. We hope that the examples will help clarify the legal and factual disputes that need to be resolved before or at trial.

Each area of potential dispute is introduced with a question. It is our hope that the majority of disputes over economic damages can be identified by asking each of these questions to the parties. Of course, some questions, especially in section IV, are only relevant in their specific subject areas. Most of the questions in section III, however, should help sort out areas of contention that may well arise in any dispute involving economic losses.

#### Figure 1





2. In the Appendix, we give an example of a complete damages study in the spreadsheet format often presented by damages experts. Readers who prefer learning from an example may want to read the Appendix before the body of this reference guide.

Reference Manual on Scientific Evidence

# **II.** Expert's Qualifications

Experts who quantify damages come from a variety of backgrounds. Whatever his or her background, however, a damages expert should be trained and experienced in quantitative analysis. For economists, the standard qualification is the Ph.D. Damages experts with business or accounting backgrounds often have MBA degrees or CPA credentials, or both. The specific areas of specialization needed by the expert are dictated by the method used and the substance of the damages claim. In some cases, participation in original research and the authorship of professional publications may add to the qualifications of an expert. The relevant research and publications are less likely to be in damages measurement per se than in topics and methods encountered in damages analysis. For example, a damages expert may need to restate prices and quantities in a market with more sellers than are actually present. Direct participation in research on the relation between market structure and performance would be helpful for an expert undertaking that task.

Statistical *regression analysis* is sometimes used to make inferences in damages studies.<sup>3</sup> Specific training is required to apply regression analysis. As another example, damages studies may involve statistical surveys of customers.<sup>4</sup> In this case, the damages expert should be trained in survey methods or should work in collaboration with a qualified survey statistician. Because damages estimation often makes use of accounting records, most damages experts need to be able to interpret materials prepared by professional accountants. Some damages issues may require assistance from a professional accountant.

Experts benefit from professional training and experience in areas relevant to the substance of the damages claim. For example, in the case of lost earnings, an expert will benefit from training in labor economics; in intellectual property and antitrust, a background in industrial organization will be helpful; and in securities damages, a background in finance will assist the expert.

<sup>3.</sup> For a discussion of regression analysis, see generally Daniel L. Rubinfeld, Reference Guide on Multiple Regression, in this manual.

<sup>4.</sup> For a discussion of survey methods, see generally Shari Seidman Diamond, Reference Guide on Survey Research, in this manual.

# III. Issues Common to Most Damages Studies

Throughout our discussion, we assume that the plaintiff is entitled to compensation for losses sustained from a harmful act of the defendant. The harmful act may be an act whose occurrence itself is wrongful, as in a tort, or it may be a failure to fulfill a promise, as in a breach. In the first instance, damages are generally calculated under the principle that compensation should place the plaintiff in a position economically equivalent to the plaintiff's position absent the harmful event. In applications of this principle, either *restitution damages* or *reliance damages* are calculated. These two terms are essentially synonyms with respect to their economic content. The term restitution is used when the harmful act is an injury or theft, and reliance is used when the harmful act is fraud. In the second instance, breach of a contract or duty, damages are generally calculated under the expectation principle, where the compensation is intended to replace what the plaintiff would have gotten if the promise or bargain had been fulfilled. These types of damages are called *expectations damages*.

In this section, we review the elements of the standard loss measurement in the format of Figure 1. For each element, there are several areas of potential dispute. The sequence of questions posed in section III should identify most if not all of the areas of disagreement between the damages analyses of opposing parties.

### A. Characterization of the Harmful Event

1. How was the plaintiff harmed and what legal principles govern compensation for the harm?

The first step in a damages study is the translation of the legal theory of the harmful event into an analysis of the economic impact of that event. In most cases, the analysis considers the difference between the plaintiff's economic position if the harmful event had not occurred and the plaintiff's actual economic position. The damages study restates the plaintiff's position "but for" the harmful event; this part is often called the *but-for analysis*. Damages are the difference between the but-for value and the actual value.

In cases where damages are calculated under the restitution–reliance principle, the but-for analysis posits that the harmful event did not occur. In many cases—such as injuries resulting from accidents—the but-for analysis presumes no contact at all between the parties. Damages are the difference between the value the plaintiff would have received had there been no contact with the defendant and the value actually received. When the harmful event is misrepresentation by the defendant, resulting in an economically detrimental relation-ship between the defendant and the plaintiff, the but-for analysis will again consider the value the plaintiff would have received in the absence of that relation-ship. Typically, the but-for analysis for fraud will adopt the premise that the plaintiff would have entered into a valuable relationship with an entity other than the defendant. For example, if the defendant's misrepresentations have caused the plaintiff to purchase property unsuited to the plaintiff's planned use, the but-for analysis might consider the value that the plaintiff would have received by purchasing a suitable property from another seller.

Expectations damages generally arise from the breach of a contract or duty. The harmful event is the defendant's failure to perform. Damages are the difference between the value the plaintiff would have received had the defendant performed its obligations and the value the plaintiff actually obtained.

Although the characterization of the harmful event begins with a clear statement of the harmful event and its effect on the plaintiff, that alone is not sufficient. It must also include:

- a statement about the economic situation absent the wrongdoing;
- a characterization of the causal link between the wrongdoing and the harm the plaintiff suffered; and
- a description of the defendant's proper behavior.

In addition, the characterization will resolve such questions as whether to measure damages before or after taxes and the appropriate measure of costs. Many conflicts between the damages experts for the plaintiff and the defendant arise from different characterizations of the harmful event and its effects.

A comparison of the parties' statements about the harmful event and what would have happened in its absence will reveal differences in legal theories that result in potentially large differences in damages.

Example:	Client is the victim of unsuitable investment advice by Broker (all of Client's investments made by Broker are the result of Broker's negligence). Client's damages study mea- sures the sum of the losses of the investments made by Broker, including only the investments that incurred losses. Broker's damages study measures the net loss by including an offset for those investments that achieved gains.
Comment :	Client is considering the harmful event to be the recom- mendation of investments that resulted in losses, whereas Broker is considering the harmful event to be the entire body of investment advice. Under Client's theory, Client

would not have made the unsuccessful investments but would have made the successful ones, absent the unsuitable advice. Under Broker's theory, Client would not have made any investments based on Broker's advice.

A clear statement about the plaintiff's situation but for the harmful event is also helpful in avoiding double counting that can arise if a damages study confuses or combines reliance and expectations damages.

- Example: Marketer is the victim of defective products made by Manufacturer; Marketer's business fails as a result. Marketer's damages study adds together the out-of-pocket costs of creating the business in the first place and the projected profits of the business had there been no defects. Manufacturer's damages study measures the difference between the profit margin Marketer would have made absent the defects and the profit margin he actually made.
- Comment: Marketer has mistakenly added together damages from the reliance principle and the expectations principle. Under the reliance principle, Marketer is entitled to be put back to where he would have been had he not started the business in the first place. Damages are his total outlays less the revenue he actually received. Under the expectations principle, applied in Manufacturer's damages study, Marketer is entitled to the profit on the extra sales he would have received had there been no product defects. Out-of-pocket expenses of starting the business would have no effect on expectations damages because they would be present in both the actual and the but-for cases, and would offset each other in the comparison of actual and but-for value.
- 2. Are the parties disputing differences in the plaintiff's economic environment absent the harmful event?

The analysis of some types of harmful events requires consideration of effects, such as *price erosion*, that involve changes in the economic environment caused by the harmful event. For a business, the main elements of the economic environment that may be affected by the harmful event are the prices charged by rivals, the demand facing the seller, and the prices of inputs. Misappropriation of intellectual property might enable rivals to set lower prices because of their royalty-free use of the technology, for example. In contrast, some harmful events do not change the plaintiff's economic environment. For example, the theft of some of the plaintiff's products would not change the market price of those products, nor would an injury to a worker change the general level of wages in

the labor market. A damages study need not analyze changes in broader markets when the harmful act plainly has minuscule effects in those markets.

For example, the plaintiff may assert that, absent the defendant's wrongdoing, a higher price could have been charged; the defendant's harmful act has eroded the market price. The defendant may reply that the higher price would lower the quantity sold. The parties may then dispute by how much the quantity would fall as a result of higher prices.

- *Example:* Valve Maker infringes patent of Rival. Rival calculates lost profits as the profits actually made by Valve Maker plus a price-erosion effect. The amount of price erosion is the difference between the higher price that Rival would have been able to charge absent Valve Maker's presence in the market and the actual price. The price-erosion effect is the price difference multiplied by the combined sales volume of the Valve Maker and Rival. Defendant Valve Maker counters that the volume would have been lower had the price been higher. Defendant measures damages taking account of lower volume.
- Comment: Wrongful competition is likely to cause some price erosion and, correspondingly, some enlargement of the total market because of the lower price. The actual magnitude of the price-erosion effect could be determined by economic analysis.

We consider price erosion in more detail in section IV.B, in connection with intellectual property damages. However, price erosion may be an issue in many other commercial disputes. For example, a plaintiff may argue that the disparagement of its product in false advertising has eroded its price.

In more complicated situations, the damages analysis may need to focus on how an entire industry would be affected by the defendant's wrongdoing. For example, one federal appeals court held that a damages analysis for exclusionary conduct must consider that other firms beside the plaintiff would have enjoyed the benefits of the absence of that conduct, so prices would have been lower and the plaintiff's profits correspondingly less than those posited in the plaintiff's damages analysis.<sup>5</sup>

*Example:* Photographic Film Maker has used unlawful means to exclude rival film manufacturers. Rival calculates damages on the assumption that it would have been the only additional seller in the market absent the exclusionary conduct, and that Rival would have been able to sell its film at the same

<sup>5.</sup> See Dolphin Tours, Inc. v. Pacifico Creative Servs., Inc., 773 F.2d 1506, 1512 (9th Cir. 1985).

price actually charged by Film Maker. Film Maker counters that other sellers would have entered the market and driven the price down, so Rival has overstated damages.

- *Comment:* Increased competition lowers price in all but the most unusual situation. Again, determination of the number of entrants attracted by the elimination of exclusionary conduct and their effect on the price probably requires a full economic analysis.
- 3. Is there disagreement about the causal effect of the injury?

The plaintiff might argue that the injury has dramatically reduced earnings for many years. The defendant might reply that most of the reduction in earnings that occurred up to the time of trial is the result of influences other than the injury and that the effects of the injury will disappear completely soon after the trial.

Example:	Worker is the victim of a disease caused either by exposure to xerxium or by smoking. Worker sues employer, Xerxium Mine, and calculates damages as all lost wages. Defendant Xerxium Mine, in contrast, attributes most of the losses to smoking and calculates damages as only a fraction of lost wages.
Comment:	The resolution of this dispute will turn on the legal question of comparative or contributory fault. If the law permits the division of damages into parts attributable to exposure to xerxium and to smoking, then medical evidence on the like- lihood of cause may be needed to make that division.
Example:	Real Estate Agent is wrongfully denied affiliation with Broker. Plaintiff Agent's damages study projects past earn- ings into the future at the rate of growth of the previous three years. Broker's study projects that earnings would have declined even without the breach because the real estate market has turned downward.
Comment :	The difference between a damages study based on extrapola- tion from the past, here used by Agent, and a study based on actual data after the harmful act, here used by Broker, is one of the most common sources of disagreement in damages. This is a factual dispute that hinges on the relationship be- tween real estate market conditions and the earnings of agents.

Frequently, the defendant will calculate damages on the premise that the harmful act had little, if any, causal relationship to the plaintiff's losses.

- Example: Defendants conspired to rig bids in a construction deal. Plaintiff seeks damages for subsequent higher prices. Defendants' damages calculation is zero because they assert that the only effect of the bid rigging was to determine the winner of the contract and that prices were not affected.
- *Comment:* This is a factual dispute about how much effect bid rigging has on the ultimate price. The analysis must go beyond the mechanics of the bid-rigging system to consider how the bids would be different had there been no collaboration among the bidders.

The defendant may also argue that the plaintiff has overstated the scope of the injury. Here the legal character of the harmful act may be critical; the law may limit the scope to proximate effects if the harmful act was negligence, but require a broader scope if the harmful act was intentional.

- Example:Plaintiff Drugstore Network experiences losses because defendant Superstore priced its products predatorily. Drugstore<br/>Network reduced prices in all its stores because it has a pol-<br/>icy of uniform national pricing. Drugstore Network's dam-<br/>ages study considers the entire effect of national price cuts<br/>on profits. Defendant Superstore argues that Network low-<br/>ered prices only on the West Coast and its price reductions<br/>elsewhere should not be included in damages.
- Comment: It is a factual question whether adherence to a policy of national pricing is the reasonable response to predatory pricing in only part of the market.
- 4. Is there disagreement about alternative nonharmful conduct of the defendant in projecting the plaintiff's earnings but for the harmful event?

One party's damages analysis may hypothesize the absence of any act of the defendant that influenced the plaintiff, whereas the other's damages analysis may hypothesize an alternative, legal act. This type of disagreement is particularly common in antitrust and intellectual property disputes. Although, generally, disagreement over the alternative scenario in a damages study is a legal question, opposing experts may have been given different legal guidance and therefore made different economic assumptions, resulting in major differences in their damages estimates.

- Example:Defendant Copier Service's long-term contracts with customers are found to be unlawful because they create a barrier to entry that maintains Copier Service's monopoly power. Rival's damages study hypothesizes no contracts between Copier Service and its customers, so Rival would face no contractual barrier to bidding those customers away from Copier Service. Copier Service's damages study hypothesizes medium-term contracts with its customers and argues that these would not have been found to be unlawful. Under Copier Service's assumption, Rival would have been much less successful in bidding away Copier Service's customers, and damages are correspondingly lower.
- Comment: Assessment of damages will depend greatly on the substantive law governing the injury. The proper characterization of Copier Service's permissible conduct involves a mixture of legal and economic issues.

#### 5. Are losses measured before or after the plaintiff's income taxes?

A damages award compensates the plaintiff for lost economic value. In principle, the calculation of compensation should measure the plaintiff's loss after taxes and then calculate the magnitude of pretax award needed to compensate the plaintiff fully, once taxation of the award is considered. In practice, the tax rates applied to the original loss and to the compensation are frequently the same. When the rates are the same, the two tax adjustments are a wash. In that case, the appropriate pretax compensation is simply the pretax loss, and the damages calculation may be simplified by the omission of tax considerations.<sup>6</sup>

In some damages analyses, explicit consideration of taxes is essential, and disagreements between the parties may arise about these tax issues. If the plaintiff's lost income would have been taxed as a capital gain, at a preferential rate, but the damages award will be taxed as ordinary income, the plaintiff can be expected to include an explicit calculation of the extra compensation needed to make up for the loss of the tax advantage. Sometimes tax considerations are paramount in damages calculations.

- *Example:* Trustee wrongfully sells Beneficiary's property, at full market value. Beneficiary would have owned the property until death and avoided all capital gains tax.
- Comment: Damages are the amount of the capital gains tax, even though the property fetched its full value upon sale.

6. There is a separate issue about the effect of taxes on the interest rate for prejudgment interest and discounting. See discussion infra \$\$ III.C, III.E.

In some cases, the law requires different tax treatment of loss and compensatory award. Again, the tax adjustments do not offset each other and consideration of taxes may be a source of dispute.

- *Example:* Driver injures Victim in a truck accident. A state law provides that awards for personal injury are not taxable, even though the income lost as a result of the injury is taxable. Victim calculates damages as lost pretax earnings, but Driver calculates damages as lost earnings after tax. Driver argues that the nontaxable award would exceed actual economic loss if it were not adjusted for the taxation of the lost income.
- *Comment:* Under the principle that damages are to restore the plaintiff to the economic equivalent of the plaintiff's position absent the harmful act, it may be recognized that the income to be replaced by the award would have been taxed. However, case law in a particular jurisdiction may not allow a jury instruction on the taxability of an award.
- Example: Worker is wrongfully deprived of tax-free fringe benefits by Employer. Under applicable law, the award is taxable. Worker's damages estimate includes a factor so that the amount of the award, after tax, is sufficient to replace the lost tax-free value.
- Comment: Again, to achieve the goal of restoring plaintiff to a position economically equivalent absent the harmful act, an adjustment of this type is appropriate. The adjustment is often called "grossing up" damages. To accomplish grossing up, divide the lost tax-free value by one minus the tax rate. For example, if the loss is \$100,000 of tax-free income, and the income tax rate is 25%, the award should be \$100,000 divided by 0.75, or \$133,333.
- 6. Is there disagreement about the costs that the plaintiff would have incurred but for the harmful event?

Where the injury takes the form of lost volume of sales, the plaintiff's lost value is the lost *present value* of profit. Lost profit is lost revenue less the costs avoided by selling a lower volume. Calculation of these costs is a common area of disagreement about damages.

Conceptually, *avoided cost* is the difference between the cost that would have been incurred at the higher volume of sales but for the harmful event and the cost actually incurred at the lower volume of sales achieved. In the format of Figure 1, the avoided-cost calculation is done each year. The following are some of the issues that arise in calculating avoided cost:

- For a firm operating at capacity, expansion of sales is cheaper in the longer run than in the short run; whereas, if there is unused capacity, expansion may be cheaper in the short run.
- The costs that can be avoided if sales fall abruptly are smaller in the short run than in the longer run.
- Avoided costs may include marketing, selling, and administrative costs as well as the cost of manufacturing.
- Some costs are fixed, at least in the shorter run, and are not avoided as a result of the reduced volume of sales caused by the harmful act.

Sometimes it is useful to put cost into just two categories, that which varies in proportion to sales (*variable cost*) and that which does not vary with sales (*fixed cost*). This breakdown is rough, however, and does not do justice to important aspects of avoided costs. In particular, costs that are fixed in the short run may be variable in the longer run. Disputes frequently arise over whether particular costs are fixed or variable. One side may argue that most costs are fixed and were not avoided by losing sales volume, while the other side will argue that many costs are variable.

Certain accounting concepts are related to the calculation of avoided cost. Profit and loss statements frequently report the "cost of goods sold." Costs in this category are frequently, but not uniformly, avoided when sales volume falls. But costs in other categories, called "operating costs" or "overhead costs," also may be avoided, especially in the longer run. One approach to the measurement of avoided cost is based on an examination of all of a firm's cost categories. The expert determines how much of each category of cost is avoided.

An alternative approach uses regression analysis or other statistical methods to determine how costs vary with sales as a general matter within the firm or across similar firms. The results of such an analysis can be used to measure the costs avoided by the decline in sales volume caused by the harmful act.

#### 7. Is there a dispute about the costs of stock options?

In some firms, employee stock options are a significant part of total compensation. The parties may dispute whether the value of options should be included in the costs avoided by the plaintiff as a result of lost sales volume. The defendant might argue that stock options should be included, because their issuance is costly to the existing shareholders. The defendant might place a value on newly issued options and amortize this value over the period from issuance to vesting. The plaintiff, in contrast, might exclude options costs on the grounds that the options cost the firm nothing, even though they impose costs on the firm's shareholders.

### B. Mitigation and Earnings Before Trial

We use the term earnings for almost any dollar receipts that a plaintiff should have received. Earnings could include:

- wages, salary, commissions, bonuses, or other compensation;
- profits of a business;
- cash flow;
- royalties;
- proceeds from sales of property; and
- purchases and sales of securities.

Note that earnings in some of these categories, such as cash flow or purchases of securities, could be negative in some years.

1. Is there a dispute about mitigation?

Normally, the actual earnings of the plaintiff before trial are not an important source of disagreement. Sometimes, however, the defendant will argue that the plaintiff has failed to meet its duty to mitigate. The defendant will propose that the proper offset is the earnings the plaintiff should have achieved, under proper *mitigation*, rather than actual earnings. In some cases the defendant may presume the ability of the plaintiff to mitigate in certain ways unless the defendant has specific knowledge otherwise at the time of a breach. For example, unless the defendant could reasonably foresee otherwise, the defendant may presume that the plaintiff could mitigate by locating another source of supply in the event of a breach of a supply agreement. Damages are limited to the difference between the contract price and the current market price in that situation.

For personal injuries, the issue of mitigation often arises because the defendant believes that the plaintiff's failure to work after the injury is a withdrawal from the labor force or retirement rather than the result of the injury. For commercial torts, mitigation issues can be more subtle. Where the plaintiff believes that the harmful act destroyed a company, the defendant may argue that the company could have been put back together and earned profit, possibly in a different line of business. The defendant will then treat the hypothetical profits as an offset to damages.

Alternatively, where the plaintiff continues to operate the business after the harmful act, and includes subsequent losses in damages, the defendant may argue that the proper mitigation was to shut down after the harmful act.

*Example:* Franchisee Soil Tester starts up a business based on Franchisor's proprietary technology, which Franchisor represents as meeting government standards. During the startup phase, Franchisor notifies Soil Tester that the technology has failed. Soil Tester continues to develop the business but sues Franchisor for profits it would have made from successful technology. Franchisor calculates much lower damages on the theory that Soil Tester should have mitigated by terminating start-up.

*Comment:* This is primarily a factual dispute about mitigation. Presumably Soil Tester believes it has a good case, that it was appropriate to continue to develop the business despite notification of the failure of the technology.

Disagreements about mitigation may be hidden within the frameworks of the plaintiff's and the defendant's damages studies.

- *Example*: Defendant Board Maker has been found to have breached an agreement to supply circuit boards. Plaintiff Computer Maker's damages study is based on the loss of profits on the computers to be made from the circuit boards. Board Maker's damages study is based on the difference between the contract price for the boards and the market price at the time of the breach.
- Comment: There is an implicit disagreement about Computer Maker's duty to mitigate by locating alternative sources for the boards not supplied by the defendant. The Uniform Commercial Code spells out the principles for resolving these legal issues under the contracts it governs.

# C. Prejudgment Interest

1. Do the parties agree about how to calculate prejudgment interest?

The law may specify how to calculate interest for past losses (*prejudgment inter* - est). State law may exclude prejudgment interest, limit prejudgment interest to a statutory rate, or exclude compounding. Table 1 illustrates these alternatives. With simple uncompounded interest, losses from five years before trial earn five times the specified interest, so compensation for a \$100 loss from five years ago is exactly \$135 at 7% interest. With *compound interest* , the plaintiff earns interest on past interest. Compensation is about \$140 for a loss of \$100 five years before trial. The difference between simple and compound interest becomes much larger if the time from loss to trial is greater or if the interest rate is higher. Because, in practice, interest receipts do earn further interest, economic analysis would generally support the use of compound interest.

Years Before Trial	Loss Without Interest	Loss with Compound Interest at 7%	Loss with Simple Uncompounded Interest at 7%
10	100	197	170
9	100	184	163
8	100	172	156
7	100	161	149
6	100	150	142
5	100	140	135
4	100	131	128
3	100	123	121
2	100	114	114
1	100	107	107
0	100	100	100
Total	1,100	1,578	1,485

## Table 1 Calculation of Prejudgment Interest (In Dollars)

Where the law does not prescribe the form of interest for past losses, the experts will normally apply a reasonable interest rate to bring those losses forward. The parties may disagree on whether the interest rate should be measured before or after tax. The before-tax interest rate is the normally quoted rate. To calculate the corresponding after-tax rate, one subtracts the amount of income tax the recipient would have to pay on the interest. Thus, the after-tax rate depends on the tax situation of the recipient, who is the plaintiff in the context of damages. The format for calculation of the after-tax interest rate is shown in the following example:

(1) Interest rate before tax:	9%
(2) Tax rate:	30%
(3) Tax on interest (line (1) times line (2)):	2.7%
(4) After-tax interest rate (line (1) less line (3)):	6.3%

Even where damages are calculated on a pretax basis, economic considerations suggest that the prejudgment interest rate should be on an after-tax basis: Had the plaintiff actually received the lost earnings in the past and invested the earnings at the assumed rate, income tax would have been due on the interest. The plaintiff's accumulated value would be the amount calculated by compounding past losses at the after-tax interest rate.

Where there is economic disparity between the parties, there may be a disagreement about whose interest rate should be used—the borrowing rate of the defendant or the lending rate of the plaintiff, or some other rate. There may also be disagreements about adjustment for risk.<sup>7</sup>

- *Example:* Farmer receives insurance payment one year late from Crop Insurer. Farmer calculates damages as the large amount of interest charged by a personal finance company; no bank was willing to lend to him, given his precarious financial condition. Crop Insurer calculates damages as the interest on the late payment at the normal bank loan rate.
- Comment: The law may limit claims for prejudgment interest, and a court may hold that this situation falls within the limit. Economic analysis does support the idea that delays in payments are more costly to people with higher borrowing rates.

### D. Projections of Future Earnings

1. Is there disagreement about the projection of profitability but for the harm-ful event?

A common source of disagreement about the likely profitability of a business is the absence of a track record of earlier profitability. Whenever the plaintiff is a start-up business, the issue will arise of reconstructing the value of a business with no historical benchmark.

- Example: Plaintiff Xterm is a failed start-up. Defendant VenFund has been found to have breached a venture-capital financing agreement. Xterm's damages study projects the profits it would have made under its business plan. VenFund's damages estimate, much lower, is based on the value of the startup revealed by sales of Xterm equity made just before the breach.
- Comment: Both sides confront factual issues to validate their damages estimates. Xterm needs to show that its business plan was still a reasonable forecast as of the time of the breach. VenFund needs to show that the sale of equity places a reasonable value on the firm; that is, that the equity sale was at arms' length and was not subject to discounts.

7. See generally James M. Patell et al., Accumulating Damages in Litigation: The Roles of Uncertainty and Interest Rates, 11 J. Legal Stud. 341 (1982) (extensive discussion of interest rates in damages calculations).

2. Is there disagreement about the plaintiff's actual earnings after the harmful event?

When the plaintiff has mitigated the adverse effects of the harmful act by making an investment that has not yet paid off at the time of trial, disagreement may arise about the value that the plaintiff has actually achieved.

- *Example:* Manufacturer breaches agreement with Distributor. Distributor starts a new business that shows no accounting profit as of the time of trial. Distributor's damages study makes no deduction for actual earnings during the period from breach to trial. Manufacturer's damages study places a value on the new business as of the time of trial and deducts that value from damages.
- *Comment:* Some offset for economic value created by Distributor's mitigation efforts may be appropriate. Note that if Distributor made a good-faith effort to create a new business, but was unsuccessful because of adverse events outside its control, the issue of the treatment of unexpected subsequent events will arise. (See section III.F.1)
- 3. Do the parties use constant dollars for future losses, or is there escalation for inflation?

Persistent inflation in the U.S. economy complicates projections of future losses. Although inflation rates in the 1990s have been only in the range of 3% per year, the cumulative effect of inflation has a pronounced effect on future dollar quantities. At 3% annual inflation, a dollar today buys what \$4.38 will buy fifty years from now. Under inflation, the unit of measurement of economic values becomes smaller each year, and this shrinkage must be considered if future losses are measured in the smaller dollars of the future. We refer to the calculations of this process as embodying *escalation*. Dollar losses grow into the future because of the use of the shrinking unit of measurement. For example, an expert might project that revenues will rise at 5% per year for the next ten years—3% because of general inflation and 2% more because of the growth of a firm.

Alternatively, the expert may project future losses in *constant dollars* without escalation for future inflation. The use of constant dollars avoids the problems of dealing with a shrinking unit of measurement and often results in more intuitive damages calculations. In the example just given, the expert might project that revenues will rise at 2% per year in constant dollars. Constant dollars must be stated with respect to a base year. Thus a calculation in constant 1995 dollars means that the unit for future measurement is the purchasing power of the dollar in 1995.

# E. Discounting Future Losses

For future losses, a damages study calculates the amount of compensation needed at the time of trial to replace expected future lost income. The result is discounted future losses; it is also sometimes referred to as the present discounted value of the future losses. Discounting is conceptually separate from the adjustment for inflation considered in the previous section. Discounting is typically carried out in the format shown in Table 2.

### Table 2

### Calculation of Discounted Loss at 5% Interest

Years in Future	Loss	Discount Factor	Discounted Loss <sup>a</sup>
0	\$100.00	1.000	\$100.00
1	125.00	0.952	119.00
2	130.00	0.907	118.00
Total			\$337.00

<sup>a</sup>"Discounted Loss" equals "Loss" times "Discount Factor."

"Loss" is the estimated future loss, in either escalated or constant-dollar form. "Discount Factor" is a factor that calculates the number of dollars needed at the time of trial to compensate for a lost dollar in the future year. The discount factor is calculated by applying compound interest forward from the base year to the future year, and then taking the reciprocal. For example, in Table 2, the interest rate is 5%. The discount factor for the next year is calculated as the reciprocal of 1.05. The discount factor for two years in the future is calculated as the reciprocal of 1.05 times 1.05. Future discounts would be obtained by multiplying by 1.05 a suitably larger number of times and then taking the reciprocal. The discounted loss is the loss multiplied by the discount factor for that year. The number of dollars at time of trial that compensates for the loss is the sum of the discounted losses, \$337 in this example.

The interest rate used in discounting future losses is often called the *discount* rate.

1. Are the parties using a discount rate properly matched to the projection in constant dollars or escalated terms?

To discount a future loss projected in escalated terms, one should use an ordinary interest rate. For example, in Table 2, if the losses of \$125 and \$130 are in dollars of those years, and not in constant dollars of the initial year, then the use

of a 5% discount rate is appropriate if 5% represents an accurate measure of the time value of money.

To discount a future loss projected in constant dollars, one should use a *real interest rate* as the discount rate. A real interest rate is an ordinary interest rate less an assumed rate of future inflation. The deduction of the inflation rate from the discount rate is the counterpart of the omission of escalation for inflation from the projection of future losses. In Table 2, the use of a 5% discount rate for discounting constant-dollar losses would be appropriate if the ordinary interest rate was 8% and the rate of inflation was 3%. Then the real interest rate would be 8% minus 3%, or 5%.

The ordinary interest rate is often called the *nominal interest rate* to distinguish it from the real interest rate.

2. Is one of the parties assuming that discounting and earnings growth offset each other?

An expert might make the assumption that future growth of losses will occur at the same rate as the appropriate discount rate. Table 3 illustrates the standard format for this method of calculating discounted loss.

Ta	bl	le	3
		_	_

Calculation of Discounted Loss When Growth and Discounting Offset Each Other

Years in Future	Loss	Discount Factor	Discounted Loss <sup>a</sup>	
0	\$100.00	1.000	\$100.00	
1	105.00	0.952	100.00	
2	110.30	0.907	100.00	
Total			\$300.00	

<sup>a</sup>"Discounted Loss" equals "Loss" times "Discount Factor."

When growth and discounting exactly offset each other, the present discounted value is the number of years of lost future earnings multiplied by the current amount of lost earnings.<sup>8</sup> In Table 3, the loss of \$300 is exactly three times the base year's loss of \$100. Thus the discounted value of future losses can be calculated by a shortcut in this special case. The explicit projection of future losses and the discounting back to the time of trial are unnecessary. However, the par-

<sup>8.</sup> Certain state courts have, in the past, required that the offset rule be used so as to avoid speculation about future earnings growth. In *Beaulieu v. Elliott*, 434 P.2d 665, 671–72 (Alaska 1967), the court ruled that discounting was exactly offset by wage growth. In *Kaczkowki v. Bolubasz*, 421 A.2d 1027, 1036–38 (Pa. 1980), the Pennsylvania Supreme Court ruled that no evidence on price inflation was to be introduced and deemed that inflation was exactly offset by discounting.

ties may dispute whether the assumption that growth and discounting are exactly offsetting is realistic in view of projected rates of growth of losses and market interest rates at the time of trial.

In Jones & Laughlin Steel Corp. v. Pfeifer ,<sup>9</sup> the Supreme Court considered the issue of escalated dollars with nominal discounting against constant dollars with real discounting. It found both acceptable, though the Court seemed to express a preference for the second format. In general, the Court appeared to favor discount rates in the range of 1% to 3% per year in excess of the growth of earnings.

3. Is there disagreement about the interest rate used to discount future lost value?

Discount calculations should use a reasonable interest rate drawn from current data at the time of trial. The interest rate might be obtained from the rates that could be earned in the bond market from a bond of maturity comparable to the lost stream of receipts. As in the case of prejudgment interest, there is an issue as to whether the interest rate should be on a before- or after-tax basis. The parties may also disagree about adjusting the interest rate for risk. A common approach for determining lost business profit is to use the Capital Asset Pricing Model (CAPM) to calculate the risk-adjusted discount rate. The CAPM is the standard method in financial economics to analyze the relation between risk and discounting. In the CAPM method, the expert first measures the firm's "beta"—the amount of variation in one firm's value per percentage point of variation in the value of all businesses. Then the risk-adjusted discount rate is the risk-free rate from a U.S. Treasury security plus the beta multiplied by the historical average risk premium for the stock market. <sup>10</sup> For example, the calculation may be presented in the following format:

(1) Risk-free interest rate:	4.0%
(2) Beta for this firm:	1.2%
(3) Market equity premium:	8.0%
(4) Equity premium for this firm ((2) times (3))	): 9.6%
(5) Discount rate for this firm ((1) plus (4)):	13.6%

#### 4. Is one of the parties using a capitalization factor?

Another approach to discounting a stream of losses uses a market *capitalization factor*. A capitalization factor is the ratio of the value of a stream of continuing income to the current amount of the stream; for example, if a firm is worth \$1 million and its current earnings are \$100,000, its capitalization factor is ten.

<sup>9. 462</sup> U.S. 523 (1983).

<sup>10.</sup> Richard A. Brealey & Stewart C. Myers, Principles of Corporate Finance 181-212 (4th ed. 1991).

The capitalization factor is generally obtained from the market values of comparable assets or businesses. For example, the expert might locate a comparable business traded in the stock market and compute the capitalization factor as the ratio of stock market value to operating income. In addition to capitalization factors derived from markets, experts sometimes use rule-of-thumb capitalization factors. For example, the value of a dental practice might be taken as one year's gross revenue (the capitalization factor for revenue is one). Often the parties dispute whether there is reliable evidence that the capitalization factor accurately measures value for the specific asset or business.

Once the capitalization factor is determined, the calculation of the discounted value of the loss is straightforward: It is the current annual loss in operating profit multiplied by the capitalization factor. A capitalization-factor approach to valuing future losses may be formatted in the following way:

(1) Ratio of market value to current annual earnings	
in comparable publicly traded firms:	13
(2) Plaintiff's lost earnings over past year:	\$200
(3) Value of future lost earnings ((1) times (2)):	\$2,600

cal future year.

The capitalization-factor approach might also be applied to revenue, cash flow, accounting profit, or other measures. The expert might adjust market values for any differences between the valuation principles relevant for damages and those that the market applies. For example, the value in the stock market may be considered the value placed on a business for a noncontrolling interest, whereas the plaintiff's loss relates to a controlling interest. The parties may dispute almost every element of the capitalization calculation.

Example:	Lender is responsible for failure of Auto Dealer. Plaintiff Auto Dealer's damages study projects rapid growth of future profits but for Lender's misconduct. The study uses a dis- count rate calculated as the after-tax interest rate on Treasury bills. The resulting estimate of lost value is \$10 million. Defendant Lender's damages study uses data on the actual sale prices of similar dealerships in various parts of the country. The data show that the typical sales price of a dealership is six times its annual pretax profit. Lender's damages study multiplies the capitalization factor of six by the most recent annual pretax profit of Auto Dealer of \$500,000 to estimate lost value as \$3 million.
Comment:	Part of the difference comes from the lower effective dis- count rate used by Auto Dealer. Another reason may be that the \$500,000 pretax profit may understate profit in the typi-

Reference Manual on Scientific Evidence

13

5. Is one party using the appraisal approach to valuation and the other the discounted-income approach?

The *appraisal* approach places a value on a stream of earnings by determining the value of a similar stream in a market for such earnings streams. For example, to place a value on the stream of earnings from a rental property, the appraisal approach would look at the market values of similar properties. The appraisal approach is suitable for many kinds of real property and some kinds of businesses.

- *Example*: Oil Company deprives Gas Station Operator of the benefits of Operator's business. Operator's damages study projects future profits and discounts them to the time of trial, to place a value of \$5 million on the lost business. Oil Company's damages study takes the average market prices of five nearby gas station businesses with comparable gasoline volume, to place a value of \$500,000 on the lost business.
- Comment: This large a difference probably results from a fundamental difference in assumptions. Operator's damages study is probably assuming that profits are likely to grow, while Oil Company's damages study may be assuming that there is a high risk that the neighborhood will deteriorate and the business will shrink.
- F. Other Issues Arising in General in Damages Measurement
- 1. Is there disagreement about the role of subsequent unexpected events?

Random events occurring after the harmful event can affect the plaintiff's actual loss. The effect might be either to amplify the economic loss from what might have been expected at the time of the harmful event or to reduce the loss.

Example:	Housepainter uses faulty paint, which begins to peel a month after the paint job. Owner measures damages as the cost of repainting. Painter disputes on the grounds that a hurricane that actually occurred three months after the paint job would have ruined a proper paint job anyway.
Comment:	This dispute will need to be resolved on legal rather than economic grounds. Both sides can argue that their approach to damages will, on the average over many applications, re- sult in the right incentives for proper house painting. <sup>11</sup>

11. See Franklin M. Fisher & R. Craig Romaine, Janis Joplin's Yearbook and the Theory of Damages, in Industrial Organization, Economics, and the Law 392, 399–402 (John Monz ed., 1991).

The issue of subsequent random events should be distinguished from the legal principle of supervening events. The subsequent events occur after the harmful act; there is no ambiguity about who caused the damage, only an issue of quantification of damages. Under the theory of a supervening event, there is precisely a dispute about who caused an injury. In the example above, there would be an issue of the role of a supervening event if the paint did not begin to peel until *after* the hurricane.

Disagreements about the role of subsequent random events are particularly likely when the harmful event is fraud.

- *Example:* Seller of property misstates condition of property. Buyer shows that he would not have purchased the property absent the misstatement. Property values in general decline sharply between the fraud and the trial. Buyer measures damages as the difference between the market value of the property at the time of trial and the purchase price. Seller measures damages as the difference between the purchase price and the market value at the time of purchase, assuming full disclosure.
- Comment: Buyer may be able to argue that retaining the property was the reasonable course of action after uncovering the fraud; in other words, there may be no issue of mitigation here. In that sense, Seller's fraud caused not only an immediate loss, as measured by Seller's damages analysis, but also a subsequent loss. Seller, however, did not cause the decline in property values. The dispute needs to be resolved as a matter of law.
- 2. How should damages be apportioned among the various stakeholders?

Usually the plaintiff need not distinguish between the defendant and the beneficiaries of the wrongdoing. In some cases, the law unambiguously determines who should pay for losses. For example, if a corporation increases its own profit through an antitrust violation, the defendant is the corporation and the shareholders are the recipients of the illegal profits. In general, the corporation is sued and current shareholder profits are reduced by the amount of the damages award. A current shareholder who may have purchased shares after the wrongdoing ceased will pay for the plaintiff's injury even though the shareholder did not share in the illegal profits. The shareholder's only recourse is to sue the firm and its officers.

A related issue can arise when a public utility is sued.

*Example:* Electric Utility infringes a patent. Patent Owner seeks compensation for lost royalties. Utility argues that the royalty

would have been part of its rate base, and it would have been allowed higher prices so as to achieve its allowed rate of return had it paid a royalty. It, therefore, did not profit from its infringement. Instead, the ratepayers benefited. Patent Owner argues that Utility stands in for all stakeholders.

*Comment:* In addition to the legal issue of whether Utility does stand in for ratepayers, there are two factual issues: Would a royalty actually have been passed on to ratepayers? Will the award be passed on to ratepayers?

Similar issues can arise in employment law.

- Example:Plaintiff Sales Representative sues for wrongful denial of a<br/>commission. Sales Representative has subcontracted with<br/>another individual to do the actual selling and pays a por-<br/>tion of any commission to that individual as compensation.<br/>The subcontractor is not a party to the suit. Defendant<br/>Manufacturer argues that damages should be Sales<br/>Representative's lost profit measured as the commission less<br/>costs, including the payout to the subcontractor. Sales<br/>Representative argues that she is entitled to the entire com-<br/>mission.
- *Comment:* Given that the subcontractor is not a plaintiff, and Sales Representative avoided the subcontractor's commission, the literal application of standard damages-measurement principles would appear to call for the lost-profit measure. The subcontractor, however, may be able to claim its share of the damages award. In that case, restitution would call for damages equal to the entire lost commission, so that, after paying off the subcontractor, Sales Representative receives exactly what she would have received absent the breach. Note that the second approach would place the subcontractor in exactly the same position as the Internal Revenue Service in our discussion of adjustments for taxes in section III.A.5.<sup>12</sup>

<sup>12.</sup> This example provoked vehement reactions from our reviewers. All believed the resolution was obvious, but some thought the plaintiff should receive only its anticipated profit, and others thought the plaintiff should receive the entire commission.

# IV. Subject Areas of Economic Loss Measurement

### A. Personal Lost Earnings

A claim for loss of personal earnings occurs as the result of wrongful termination, discrimination, injury, or death. The earnings usually come from employment, but essentially the same issues arise if self-employment or partnership earnings are lost. Most damages studies for personal lost earnings fit the paradigm of Figure 1 quite closely.

1. Is there a dispute about projected earnings but for the harmful event?

The plaintiff seeking compensation for lost earnings will normally include wages or salary; other cash compensation, such as commissions, overtime, and bonuses; and the value of fringe benefits. Disputes about wages and salary before trial are the least likely, especially if there are employees in similar jobs whose earnings were not interrupted. Even so, the plaintiff may make the case that a promotion would have occurred after the time of the termination or injury. The more variable elements of cash compensation are more likely to be in dispute. One side may measure bonuses and overtime during a period when these parts of compensation were unusually high, and the other side may choose a longer period, during which the average is lower.

2. What benefits are part of damages?

Loss of benefits may be an important part of lost personal earnings damages. A frequent source of dispute is the proper measurement of vacation and sick pay. Here the strict adherence to the format of Figure 1 can help resolve these disputes. Vacation and sick pay is part of the earnings the plaintiff would have received but for the harmful event. It would be double counting to include vacation and sick pay in benefits when it has already been included in cash earnings.

The valuation of fringe benefits is frequently a source of important disputes. When benefits take a form other than immediate cash, there are two basic approaches to valuation: (1) the cost to the employer, and (2) the value to the worker. Disputes may arise because of differences between these two approaches or in the application of either one.

- Example: Employee is terminated in breach of an employment agreement. Employee's damages analysis includes the value of Employee's coverage under Employer's company medical plan, estimated by the cost of obtaining similar coverage as an individual. Employee's damages analysis also includes Employer's contribution to Social Security. Employer's opposing study values the medical benefits at the cost of the company plan, which is much less than an individual plan. Employer places a value of zero on Social Security contributions, on the grounds that the Social Security benefit formula would give the same benefits to Employee whether or not the additional employer contributions had been made.
- *Comment:* Although the valuation of benefits from Employer's point of view has theoretical merit, the obstacles are obvious from these two examples. On the value of the medical benefits, if Employee actually has purchased equivalent coverage as an individual, there is a case for using that cost. The valuation of prospective Social Security benefits is forbiddingly complex, and most experts settle for measuring the value as the employer's contribution.

#### 3. Is there a dispute about mitigation?

Actual earnings before trial, although known, may be subject to dispute if the defendant argues that the plaintiff took too long to find a job or the job taken was not sufficiently remunerative. Even more problematic may be the situation where the plaintiff continues to be unemployed.

Parties disputing the length of a job search frequently offer testimony from job placement experts. Testimony from a psychologist also may be offered if the plaintiff has suffered emotional trauma as a result of the defendant's actions. Recovery from temporarily disabling injuries may be the subject of testimony by experts in vocational rehabilitation. Also, data about displaced workers, which can be obtained from the U.S. Bureau of Labor Statistics, provide information about how long others have taken to find jobs.

The defendant may argue that the plaintiff—for reason of illness, injury, or vacation, not related to the liability issues in the case—has chosen not to undertake a serious job search and therefore failed to meet the duty to mitigate. A damages study based on that conclusion will impute earnings to replace the actual earnings (if any) in the box labeled "Actual earnings before trial" in Figure 1.
- Example: Plumber loses two years of work as a result of slipping on ice. His damages claim is for two years of earnings as a plumber. Defendant Hotel Owner calculates damages as the difference between those earnings and one year of earnings as a bartender, on the grounds that Plumber was capable of working as a bartender during the second year of his recovery.
- *Comment:* Employment law may limit the type of alternative job that the plaintiff is obligated to consider.

Resolution of the mitigation issue can also be complicated if the plaintiff has taken a less remunerative job in anticipation of subsequent increases. For example, the plaintiff may have gone back to school to qualify for a better-paying job in the future. Or, the plaintiff may have taken a lower-paying job in which the career path offers more advancement. A common occurrence, particularly for more experienced workers with the appropriate skills, is to become a self-employed businessperson. The problem becomes how to value the plaintiff's activities during the development period of the business. On the one hand, the plaintiff may have made a reasonable choice of mitigating action by starting a business. On the other hand, the defendant is entitled to an offset to damages for the value of the plaintiff's investment in the development of the business.

When damages are computed over the entire remaining work life of the plaintiff, the timing of earnings on the mitigation side is less critical. The economic criterion for judging the adequacy of mitigation is that the present value of the stream of earnings over the plaintiff's work life in the chosen career exceeds the present value of the stream of earnings from alternative careers. In other words, it is appropriate that the defendant should be charged with replacing the entire amount of but-for earnings during a period of schooling or other investment if the defendant is being relieved of even more responsibility in future years as the investment pays off. If, however, the plaintiff appears to have chosen a lower-paying career for noneconomic reasons, then the defendant may argue that the amounts corresponding to the boxes labeled "Actual earnings before trial" and "Projected earnings after trial" in Figure 1 should be based on the plaintiff's highest-paying alternative. The defendant also may argue along these lines if damages are computed over a period shorter than the plaintiff's work life.

4. Is there disagreement about how the plaintiff's career path should be projected?

The issues that arise in projecting but-for and actual earnings after trial are similar to the issues that arise in measuring damages before trial. In addition, the parties are likely to disagree regarding the plaintiff's future increases in compensation. A damages analysis should be internally consistent. For example, the com-

Estimation of Economic Losses in Damages Awards

pensation path for both but-for and actual earnings paths should be based on consistent assumptions about general economic conditions, about conditions in the local labor market for the plaintiff's type of work, and about the plaintiff's likely increases in skills and earning capacity. The analysis probably should project a less successful career on the mitigation side if it is projecting a slow earnings growth absent the harm. Similarly, if the plaintiff is projected as president of the company in ten years absent the harm, the study should probably project similar success in the mitigating career.

- *Example:* Executive suffers wrongful termination. His damages study projects rapid growth in salary, bonus, and options, thanks to a series of likely promotions had he not been terminated. After termination, he looked for work unsuccessfully for a year and then started up a consulting business. Earnings from the consulting business rise, but never reach the level of his projected compensation but for the termination. Damages are estimated at \$3.6 million. His former employer's opposing damages study is based on the hypothesis that he would have been able to find a similar job within nine months if he had searched diligently. Damages are estimated at \$275,000.
- Comment: This example illustrates the type of factual disputes that are typical of executive termination damages. Note that there may be an issue of random subsequent events both in the duration of Executive's job search and in the success of his consulting business.
- 5. Is there disagreement about how earnings should be discounted to present value?

Because personal lost earnings damages may accrue over the remainder of a plaintiff's working life, the issues of predicting future inflation and discounting earnings to present value are particularly likely to generate quantitatively important disagreements. As we noted in section III.D, projections of future compensation can be done in constant dollars or escalated terms. In the first case, the interest rate used to discount future constant-dollar losses should be a real interest rate—the difference between the ordinary interest rate and the projected future rate of inflation. All else being the same, the two approaches will give identical calculations of damages. Under some conditions, future wage growth may be about equal to the interest rate, so that discounted future losses are the same in each future year. Damages after trial are then just the appropriate multiple of the current year's loss. Equivalently, the calculation can be done by projected future wage growth in escalating dollars and discounting by an ordinary interest

rate. Of course, the projected wage growth must be consistent with the expert's conclusion about inflation.

Substantial disagreements can arise about the rate of interest. Even when the parties agree that the interest rate should approximate what the plaintiff can actually earn by investing the award prudently, the parties may dispute the type of investment the plaintiff is likely to make. The plaintiff may argue that the real rate of interest should correspond to the real rate of interest for a money market fund, while the defendant may argue that the plaintiff would be expected to invest in instruments, such as the stock market, with higher expected returns. There may also be a disagreement about whether the discount rate should be calculated before or after taxes.

6. Is there disagreement about subsequent unexpected events?

Disagreements about subsequent unexpected events are likely in cases involving personal earnings, as we discussed in general in section III.F. For example, the plaintiff may have suffered a debilitating illness that would have compelled the resignation from a job even if the termination or injury had not occurred. Or the plaintiff would have been laid off as a result of employer hardship one year after the termination. The plaintiff might respond that the bad times were unexpected as of the time of the termination and so should be excluded from consideration in the calculation of damages.

#### 7. Is there disagreement about retirement and mortality?

For damages after trial, there is another issue related to the issue of unexpected events before trial: How should future damages reflect the probability that the plaintiff will die or decide to retire? Sometimes an expert will assume a work-life expectancy and terminate damages at the end of that period. Tables of work-life expectancy incorporate the probability of both retirement and death. Another approach is to multiply each year's lost earnings by the probability that the plaintiff will be alive and working in that year. That probability declines gradually with age; it can be inferred from data on labor-force participation and mortality by age.

Within either approach, there may be disagreements about how much information to use about the individual. For example, if the plaintiff is known to smoke, should his survival rates be those of a smoker? Similarly, if the plaintiff is a woman executive, should her retirement probability be inferred from data on women in general, or would it be more reasonable to look at data on executives, who are mostly men?

#### B. Intellectual Property Damages

Intellectual property damages are calculated under federal law for patents, trademarks, and copyrights and under state law for trade secrets. Damages may be a combination of the value lost by the intellectual property owner and the value gained by the infringer, with adjustment to avoid double counting. The value lost by the intellectual property owner is lost profits, calculated as in other types of damages analysis. Under patent law, the lost profit includes a reasonable royalty the infringer should have paid the patent owner for the use of the patented invention. The reasonable royalty is generally defined as the amount the defendant would have paid the patent owner as the result of a license negotiation occurring at the time that the infringement began or the patent issued. Patent law does not provide for recovery of value gained by the infringer, except through the reasonable royalty. Under copyright law, the plaintiff is entitled to the revenue received by the infringer as a result of selling the copyrighted work, but the defendant is entitled to deduct the costs of reproducing the infringing work as an offset to damages (the plaintiff's damages case need not include the offset; the defendant typically raises this issue later). Under the Uniform Trade Secrets Law, the concept of value gained by the misappropriator is not limited to a particular formula.

1. Is there disagreement about what fraction of the defendant's sales would have gone to the plaintiff?

Patent law now makes it easier for a patent owner to argue that it would have received a share of the infringer's actual sale.<sup>13</sup> Previously, the presence of a non-infringing product in the market required a lost-profit analysis to show, directly, which sales were lost. The damages analysis may now use some type of market-share model. The simplest model would consider the total market to have a given volume of sales, *S*. If the market shares of the plaintiff and the defendant are *P* and *D*, respectively, this model would predict that the plaintiff's market share, absent the defendant's sales, would be:

This formula corresponds to the assumption that the defendant's sales would have been distributed evenly across the other sellers, including the plaintiff. Then the plaintiff's sales, absent the presence of the infringer in the market, would be:

$$\frac{P}{1-D}$$
 S

13. State Indus., Inc. v. Mor-Flo Indus., Inc., 639 F. Supp. 937 (E.D. Tenn. 1986), aff'd without op., 818 F.2d 875 (Fed. Cir.), cert. denied, 484 U.S. 845 (1987).

Reference Manual on Scientific Evidence

But this model is likely to be disputed. The issues are how large the market would have been, absent the defendant's infringing product, and what share of that market the plaintiff would have enjoyed. The defendant may argue that it enlarged the total market. Its product may appeal to customers who would not buy from any of the other sellers; for example, some of the infringing sales may be to affiliates of the infringer. With respect to the plaintiff's market share but for the infringement, the defendant may demonstrate that the rivals for the defendant's sales rarely included the plaintiff. Either the plaintiff or the defendant may argue that there are actually several different markets, each to be analyzed according to some type of market-share model.

2. Is there disagreement about the effect of infringement or misappropriation on prices as well as quantities (price erosion)?

The plaintiff may measure price erosion directly, by comparing prices before and after infringement, or indirectly, through an economic analysis of the market. The defendant may dispute direct measures of price erosion on the grounds that the drop in prices would have occurred despite the infringement as a result of normal trends or events occurring at the same time, unrelated to the infringement.

The parties may also dispute the relation between the size of the total market and prices. When a plaintiff's analysis projects that prices would have been higher absent infringement, the defendant may point out that higher prices would reduce the volume of total sales and thus reduce the plaintiff's sales. Disagreements about the measurement of lost profit are most likely to be resolved if both parties make their lost-profit calculations in the same format. The preferred format is:

Lost profit = [price but for infringement] × [quantity sold but for infringement] - [actual revenue] - [extra cost of producing the extra quantity]

This format avoids the danger of double counting that arises when the plaintiff makes separate claims for lost sales and price erosion.

3. Is there a dispute about whether the lost-profit calculation includes contributions from noninfringing features of the work or product (apportionment)?

Where the protected work or technology is not the only feature or selling point of the defendant's product, there may be disagreement about apportionment. One approach to quantitative apportionment of damages is to hypothesize that the defendant would have sold a different, noninfringing product containing the other features or selling points. The damages study then measures the plaintiff's losses from the defendant's selling of the actual product rather than the alternative, hypothetical, noninfringing product.

- Example: Camera Maker sells a camera that competes directly with Rival's similar camera. A court has determined that this is an infringement of Rival's autofocus patent. Rival's damages study hypothesizes the absence of Camera Maker's from the market. Camera Maker's damages study hypothesizes that it would have sold the same camera with a different, noninfringing autofocus system. Camera Maker has apportioned lost sales to take account of the other selling points of the camera, whereas Rival is considering all of the lost sales. Rival argues that its approach is correct because the camera would not have been put on the market absent the infringing autofocus system.
- *Comment:* Note that the issue of apportionment here is, in essence, a special case of the more general issue discussed in section III.A, of disagreements about the alternative nonharmful conduct of the defendant. Here the alternative is what type of noninfringing product Camera Maker can hypothesize it would have sold absent infringement.<sup>14</sup>
- 4. Do the parties disagree about whether the defendant could have designed around the plaintiff's patent?

Under patent law, part of the plaintiff's lost profit from infringement is measured as the reasonable royalty the defendant would have paid for a license under the patent. The conceptual basis for the reasonable royalty is the outcome of a hypothetical negotiation occurring at the time the infringement began. Validity of the patent and the defendant's use of the protected technology are presumed in the hypothetical negotiation.

An important source of disagreement about the basis for the reasonable royalty and corresponding quantum of damages is the defendant's ability to design around the patent. A defendant may argue that any but a modest royalty would have caused it to reject the license and choose not to use the technology but to design around it instead.

<sup>14.</sup> In Computer Assocs. Int'l v. Altai, Inc., 982 F.2d 693 (2d Cir. 1992), the appeals court determined that defendant could hypothesize that sales of its noninfringing earlier version of a software package would partially replace the actual sales of its infringing package, thus limiting the extra sales that plaintiff would have enjoyed absent the infringement.

5. Is there disagreement about how much of the defendant's advantage actually came from infringement (apportionment)?

Under patent law, apportionment is implicit in the reasonable-royalty framework; a defendant would not pay more for a patent license than its contribution to profit. Under copyright law, where damages include the defendant's gain measured as its revenue or profit, apportionment may be a major source of disagreement.

Example:	Recording Company's compact disk contains one infringing
	song among twelve. Defendant's damages study is based on
	one-twelfth of the profit from the sales of the disk. Rock
	Composer argues that the infringing song is the main selling
	point of the disk and seeks all of defendant's profit.
Comment:	This is a factual dispute. The parties may use survey evi-
	dence on consumers' reasons for purchasing the disk.

6. Is there disagreement about how to combine the plaintiff's loss and the defendant's gain in a way that avoids double counting?

The calculation normally involves calculation of the profit on the part of the defendant's sales not considered to be the plaintiff's lost sales. For example, if the defendant has sold 100 units and in the process has taken 60 units of sales away from the plaintiff, the damages would consist of the plaintiff's lost profits on the 60 units and the defendant's revenue or profit on the remaining 40 units that were incremental sales not taken from the plaintiff.

Disputes can arise about the elimination of double counting when the plaintiff and the defendant sell their products in different ways. For example, the plaintiff may bundle its product with related products, while the defendant sells a component to be bundled by others.

#### C. Antitrust Damages

Where the plaintiff is the customer of the defendant or purchases goods in a market where the defendant's antitrust misconduct has raised prices, damages are the amount of the overcharge. This amount may exceed the lost profit of the plaintiff, if it is a business, because the plaintiff may pass along part of the effect of the price increase to its own customers.<sup>15</sup> Where the plaintiff is a rival of the defendant, injured by exclusionary or predatory conduct, damages are the lost profits from the antitrust misconduct.

Estimation of Economic Losses in Damages Awards

<sup>15.</sup> Hanover Shoe v. United Shoe Mach. Corp., 392 U.S. 481, 499 (1968) and Illinois Brick Co. v. Illinois, 431 U.S. 720 (1977) established the principle under the federal antitrust laws that, generally, a business plaintiff should not lower its damages claim on account of passing on overcharges to its customers, but rather the plaintiff should stand in for the downstream victims of overcharges.

1. Is there disagreement about the scope of the damages?

The plaintiff might calculate damages affecting all of its business activities, whereas the defendant might calculate damages only in markets where there is a likelihood of adverse impact from the defendant's conduct.

- *Example:* Trucker's exclusionary conduct has monopolized certain routes, but only modestly raised its market share on many other nonmonopolized routes. Shippers seek damages for elevated prices in all affected markets, but Trucker's damages study considers only the routes where monopolization has occurred.
- *Comment* : Here is a mixture of legal and economic issues. The law may set limits on the reach of antitrust damages even if economic analysis could quantify price elevation in all of the markets.
- 2. Is there a dispute about the causal link between the misconduct and the measured damages?

Experts face a particular challenge in making a complete analysis of the economic impact of antitrust misconduct on the relevant market. To overcome the analytical challenge, experts sometimes compare market conditions in a period affected by the misconduct with conditions in another period, during which the misconduct is known to be absent. The plaintiff might take the increase in price from the benchmark period to the affected period as a measure of the price elevation caused by the misconduct. The defendant may argue that the misconduct is not the only difference between the periods—prices rose, for example, because of cost increases or rising demand and not just because of a conspiracy or other misconduct.

- Example: The price of plywood rises soon after a meeting of Plywood Producers. Plywood Purchasers attribute all of the price increase to a price-fixing conspiracy. Plywood Producers argue that increases in timber prices would have compelled increases in plywood prices even without a price-fixing agreement; their damages study attributes only part of the price increase to the conspiracy.
  Comment: Economic analysis is capable, in principle, of inferring how much of a price increase is caused by a cost increase. Ply
  - much of a price increase is caused by a cost increase. Plywood Purchasers' damages analysis could be strengthened in this example by direct evidence on the amount of the price increase determined by the conspirators. In more sophisticated measurements of damages through comparisons of pe-

riods with and without the misconduct, experts may use regression analysis to adjust for influences other than the misconduct. Explanatory variables may include general economic indicators such as the national price level and Gross Domestic Product, and variables specific to the industry.<sup>16</sup>

3. Is there a dispute about how conditions would differ absent the challenged misconduct?

The plaintiff may calculate damages for exclusionary conduct on the basis that prices in the market would have been the same but for that conduct. The defendant may argue that the activities of the plaintiff and other firms, absent exclusion, would have driven prices down, and thus that the plaintiff has overstated the profit it lost from exclusion.

Example: Concert Promoter is the victim of exclusion by Incumbent through Incumbent's unlawful contracts with a ticket agency. Promoter's damages study hypothesizes that Promoter would be the only additional seller in the industry absent the contracts. Incumbent's damages study hypothesizes numerous additional sellers and price reductions sufficient to eliminate almost all profit. Incumbent's estimate of damages is a small fraction of Promoter's.

Comment: The elimination of one barrier to entry in the market—the unlawful contracts—will increase the profit available to potential rivals. On this account, some new rivals to the Concert Promoter might enter the market and share the benefits flowing from the elimination of the unlawful contracts. This is a limiting factor for Concert Promoter's damages. But there may be other barriers to the entry of rivals. For example, it may take an extended period for a new promoter to attract major performers. The plaintiff, already established in the business, might expect to make added profits from the elimination of the unlawful contracts, even though some new competitors would enter. See discussion of Dolphin Tours in section III.A.2.

When the harmful act is a tied sale, the issue of different conditions absent the harmful act is particularly critical. Tying arrangements are attempts by a business to extend its monopoly in one market into a related market. A pur-

Estimation of Economic Losses in Damages Awards

<sup>16.</sup> See Daniel L. Rubinfeld, Reference Guide on Multiple Regression § II.B.3, in this manual.

chaser who wants the "tying" good must also purchase the "tied" good.<sup>17</sup> The plaintiff, if a purchaser, may calculate damages as the price paid for the purchase of the tied product, on the theory that the purchase was unwanted and would not have occurred absent the tie. If the plaintiff is a rival in the market for the tied good, the plaintiff may calculate damages on the theory that it would have enjoyed higher sales absent the tie. In both cases, the defendant may respond that, absent the tie, the price for the tying good would have been higher and the price for the tied good would have been lower. Damages are then lower than those calculated by the purchaser plaintiff to the extent of the higher price for the tying good. Damages are lower than those calculated by the rival plaintiff because the lost sales would occur at a lower price.

- *Example:* Dominant Film Seller has required that purchasers of film also buy processing. Film and processing Purchasers calculate damages on the theory that they could have bought film at the stated price from Dominant Seller but could have bought processing from a cheaper rival, absent the tie. Dominant Seller counters that it would have charged more for film absent the tie. In addition, Independent Processor calculates damages based on the theory that it would have picked up part of Dominant Seller's processing business and enabled it to charge the same price charged by Dominant Seller. Defendant Dominant Seller responds that it would have charged less for processing and more for film, absent the tie, so Independent Processor would be forced to charge a lower price.
- Comment : When there is a strict tie between two products, the economist will be careful in interpreting the separate stated prices for the two products. In this example, all that matters to the customer is the combined price of film and processing. A full factual analysis is needed to restate pricing absent a tie. Eliminating a tie may stimulate entry into the market for the tied product (indeed, there was an upsurge of competition in the independent film processing market when tying was eliminated). Economists sometimes disagree why dominant firms use ties rather than simply extract all of the available monopoly profit from the product in which they are dominant.

<sup>17.</sup> For further explanation, see Stephen H. Knowlton et al., *Antitrust, in* Litigation Services Handbook: The Role of the Accountant as Expert Witness 208–09 (Peter B. Frank et al. eds., 1990).

#### D. Securities Damages

Where the harmful act takes the form of a failure to disclose adverse information about a firm whose securities are publicly traded, damages are typically sought by investors who bought the securities after the information should have been disclosed and before it was actually disclosed. Their losses are the excess value they paid for the securities, provided they did not sell before the adverse information affected the market. The damages study typically measures the excess price by the decline in the price that occurred when the information reached the market. Finance theory provides the framework generally used for this purpose.<sup>18</sup> The effect of the adverse information on the price of the securities is the part of the total price change not predicted by finance theory, considering what happened in similar securities markets at the time the information affected the market.

1. Is there disagreement about when the adverse information affected the market?

The plaintiff might argue that the adverse information reached the market in a number of steps, and thus measure damages as the excess decline in value over a period including all of the steps. Defendant might reply that only one of those steps involved the actual disclosure, and measure damages as the excess decline only on the day of that disclosure. The length and timing of the "window" for measuring the excess decline is probably the most important source of disagreement in securities damages.

2. Is there disagreement about how to take proper account of turnover of the securities?

Frequently, securities damages must be measured before the victims are individually identified. The victims are those who purchased the securities after the time when a disclosure should have been made and still owned them when the disclosure was actually made. In order to estimate the volume of securities for which damages accrued, the pattern of turnover in ownership must be determined. Generally, data on total daily purchases of the securities will be available. These data provide an upper bound on the volume for damages. However, the actual volume will be lower because some of the securities will change hands more than once during the period between proper and actual disclosure. A detailed study of turnover patterns is needed for this purpose. The representatives of the plaintiff class might argue that few shares turned over more than once, while the defendant might reply that the observed transactions were largely the same shares turning over repeatedly.

18. See generally Brealey & Myers, supra note 10.

Estimation of Economic Losses in Damages Awards

- E. Liquidated Damages
- 1. Is there a dispute about the proper application of a provision for liquidated damages?

After parties have entered into a contract with liquidated damages, they may dispute whether the liquidated-damages provision actually should apply to a subsequent harmful event. The parties may disagree on whether the event falls within the class intended by the contract provision, or they may disagree on whether the liquidated damages bear a reasonable relation to actual damages, in the sense required by applicable law. In particular, the defendant may attack the amount of liquidated damages as a penalty that exaggerates the plaintiff's actual loss.

Changes in economic conditions may be an important source of disagreement about the reasonableness of a liquidated-damages provision. One party may seek to overturn a liquidated-damages provision on the grounds that new conditions make it unreasonable.

Example :	Scrap Iron Supplier breaches supply agreement and pays liquidated damages. Buyer seeks to set aside the liquidated- damages provision because the price of scrap iron has risen, and the liquidated damages are a small fraction of actual damages under the expectations principle.
Comment:	There may be conflict between the date for judging the rea- sonableness of a liquidated-damages provision and the date for measurement of expectations damages, as in this exam- ple. Generally, the date for evaluating the reasonableness of liquidated damages is the date the contract is made. In con- trast, the date for expectations damages is the date of the breach. The result is a conundrum for which the economist needs guidance from the law. Enforcement of the liqui- dated-damages provision in this example will induce ineffi- cient breach.

# Appendix: Example of a Damages Study

Plaintiff SBM makes telephone switchboards. Defendant TPC is a telephone company. By denying SBM technical information and by informing SBM's potential customers that SBM's switchboards are incompatible with TPC's network, TPC has imposed economic losses on SBM. TPC's misconduct began in 1992. SBM's damages study presented at trial at the end of 1994 proceeds as follows (see Table 4):

- 1. Damages theory is compensation for lost profit from TPC's exclusionary conduct.
- 2. SBM would have sold more units and achieved a higher price per unit had SBM had access to complete technical information and had SBM not faced disparagement from TPC.
- 3. SBM would have earned profits before tax in 1992–94 in millions of dollars as shown in column 2 of Table 4, based on an analysis of lost business and avoided costs.
- 4. SBM's actual profits before tax are shown in column 3. Column 4 shows lost earnings. Column 5 shows the factor for the time value of money prescribed by law, with 7% annual simple interest without compounding. Column 6 shows the loss including prejudgment interest.
- 5. For the years 1995 through 1999, column 2 shows projected earnings but for TPC's misconduct.
- 6. For the same years, column 3 shows projected actual earnings.
- 7. Column 4 shows SBM's future earnings losses. Column 5 shows the discount factor based on a 4% annual after-tax interest rate, obtained by applying SBM's corporate tax rate to TPC's medium-term borrowing rate. TPC has an AA bond rating. Column 6 shows the discounted future loss. At the bottom of the table is the total loss of economic value, according to SBM's damages study, of \$1.237 billion.

	(2)				
	Earnings	(3)		(5)	(6)
(1)	but for	Actual	(4)	Discount	Discounted
Year	Misconduct	Earnings	Loss	Factor	Loss
1992	187	34	153	1.21	185
1993	200	56	144	1.14	164
1994	213	45	168	1.07	180
1995	227	87	140	1.00	140
1996	242	96	147	0.96	141
1997	259	105	153	0.92	142
1998	276	116	160	0.89	142
1999	294	127	167	0.85	143
Total					1,237

# Table 4 SBM's Damages Analysis (in Millions of Dollars)

# Table 5 TPC's Damages Analysis (in Millions of Dollars)

	(2)	(3)			
	Earnings	Earnings		(5)	(6)
(1)	but for	with	(4)	Discount	Discounted
Year	Misconduct	Mitigation	Loss	Factor	Loss
1992	101	79	22	1.21	27
1993	108	85	23	1.14	26
1994	115	81	34	1.07	36
1995	123	98	25	1.00	25
1996	131	108	23	0.87	20
1997	140	119	21	0.76	16
1998	149	130	19	0.66	12
1999	159	143	16	0.57	9
Total					171

Defendant TPC presents an alternative damages study in the same format (see Table 5). TPC argues that SBM's earnings but for the misconduct, before and after trial, are the lower numbers shown in column 2 of Table 5. TPC believes that the number of units sold would be lower, the price would be lower, and costs of production higher, than in SBM's damages study. TPC further argues that SBM failed to mitigate the effects of TPC's misconduct—SBM could have obtained the technical information it needed from other sources, and SBM could have counteracted TPC's disparagement by vigorous marketing. Column 3 displays the earnings that TPC believes SBM could have achieved with proper mitigation. TPC argues that future losses should be discounted at a 14% rate determined from SBM's cost of equity and debt; SBM is a small, risky corporation with a high cost of funds. According to TPC's damages study, total lost value is only \$171 million.

### **Glossary of Terms**

- Appraisal. A method of determining the value of the plaintiff's claim on an earnings stream by reference to the market values of comparable earnings streams. For example, if the plaintiff has been deprived of the use of a piece of property, the appraised value of the property might be used to determine damages.
- Avoided Cost. Cost that the plaintiff did not incur as a result of the harmful act. Usually it is the cost that a business would have incurred in order to make the higher level of sales the business would have enjoyed but for the harmful act.
- But-for Analysis. Restatement of the plaintiff's economic situation but for the defendant's harmful act. Damages are generally measured as but-for value less actual value received by the plaintiff.
- Capitalization Factor. Factor used to convert a stream of revenue or profit into its capital or property value. A capitalization factor of 10 for profit means that a firm with \$1 million in annual profit is worth \$10 million.
- Compound Interest. Interest calculation giving effect to interest earned on past interest. As a result of compound interest at rate r, it takes  $(1+r)(1+r) = 1+2r+r^2$  dollars to make up for a lost dollar of earnings two years earlier.
- Constant Dollars. Dollars adjusted for inflation. When calculations are done in constant 1995 dollars, it means that future dollar amounts are reduced in proportion to increases in the cost of living expected to occur after 1995.
- Discount Rate. Rate of interest used to discount future losses.
- Discounting. Calculation of today's equivalent to a future dollar, to reflect the time value of money. If the interest rate is *r*, the discount applicable to one year in the future is:

 $\frac{1}{1+r}$ 

Discounts for multiple years are the products of one-year discounts, to achieve compounding.

- *Earnings.* Economic value received by the plaintiff. Earnings could be salary and benefits from a job, profit from a business, royalties from licensing intellectual property, or the proceeds from a one-time or recurring sale of property. Earnings are measured net of costs. Thus, lost earnings are lost receipts less costs avoided.
- *Escalation.* Consideration of future inflation in projecting earnings or other dollar flows. The alternative is to make projections in constant dollars.
- Expectations Damages. Damages measured on the principle that the plaintiff is entitled to the benefit of the bargain originally made with the defendant.
- Fixed Cost. Cost that would not have risen if a business had enjoyed higher sales.
- *Mitigation.* Action taken by the plaintiff to minimize the economic effect of the harmful act. Also often refers to the actual level of earnings achieved by the plaintiff after the harmful act.
- Nominal Interest Rate. Interest rate quoted in ordinary dollars, without adjustment for inflation. Interest rates quoted in markets and reported in the financial press are always nominal interest rates.
- Prejudgment Interest. Interest on losses occurring before trial.
- Present Value. Value today of money due in the past (with interest) or in the future (with discounting).
- *Price Erosion.* Effect of the harmful act on the price charged by the plaintiff. When the harmful act is wrongful competition, as in intellectual property infringement, price erosion is one of the ways that the plaintiff's earnings have been harmed.
- *Real Interest Rate.* Interest rate adjusted for inflation. The real interest rate is the nominal interest rate less the annual rate of inflation.
- Regression Analysis. Statistical technique for inferring stable relationships among quantities. For example, regression analysis may be used to determine how costs typically rise when sales rise.
- Reliance Damages. Damages measured on the principle that the transaction or relationship should not have existed in the first place but was brought into being by the harmful act.
- Restitution Damages. Damages measured on the principle of restoring the economic equivalent of lost property or value.
- Variable Cost. Component of a business's cost that would have been higher if the business had enjoyed higher sales. See also Avoided Cost.

# **References on Damages Awards**

- Richard A. Brealey & Stewart C. Myers, Principles of Corporate Finance (4th ed. 1991).
- Industrial Organization, Economics and the Law: Collected Papers of Franklin M. Fisher (John Monz ed., 1991).
- Litigation Services Handbook: The Role of the Accountant as Expert Witness (Peter B. Frank et al. eds., 1990).
- A. Mitchell Polinsky, An Introduction to Law and Economics (2d ed. 1989).
- W. Kip Viscusi, Reforming Products Liability (1991).

# **Court-Appointed Experts**

Joe S. Cecil Thomas E. Willging

**Joe S. Cecil**, J.D., Ph.D., and **Thomas E. Willging**, J.D., LL.M., are members of the staff in the Research Division of the Federal Judicial Center, Washington, D.C.

# Contents

- I Introduction 529
  - A. Methodology 529
  - B. Overview 530
- II. Authority to Appoint an Expert 531
  - A. Federal Rule of Evidence 706 531
  - B. Inherent Authority to Appoint a Technical Advisor 534
- III. Use and Nonuse of Court-Appointed Experts 535
  - A. Use of Court-Appointed Experts 535
  - B. Satisfaction with Appointed Expert 537
  - C. Receptivity to Appointment of Experts 537
  - D. Reasons for Appointing Experts 537
    - 1. To aid decision making 538
    - 2. To aid settlement 539
  - E. Reasons for Failure to Appoint an Expert 540
    - 1. Infrequency of cases requiring extraordinary assistance 540
    - 2. Respect for the adversarial system 542
- IV. Identification and Appointment of Experts 543
  - A. Timing of the Appointment 543
  - B. Initiation of the Appointment 544
  - C. Selection of the Appointed Expert 544
- V. Communication with the Appointed Expert 547
  - A. Instruction of the Appointed Expert 547
  - B. Ex Parte Communication 549
    - 1. Communication between the judge and the appointed expert 549
    - 2. Communications between the parties and the expert 550
  - C. Pretrial Reports and Depositions 551
  - D. Presentation of Expert Opinion in Court 552
    - 1. Frequency and nature of testimony 552
    - 2. Advising jury of court-appointed status 552

- 3. Effect of the testimony of the appointed expert 553
- VI. Compensation of Court-Appointed Experts 557
  - A. Statutory Basis for Compensation from Public Funds 558
  - B. Payment of Fees by Parties 558
  - C. Compensation of Appointed Experts When One Party Is Indigent 560
  - D. Compensation of Technical Advisors 561
- VII. Procedures for the Effective Use of Court-Appointed Experts 563
  - A. Clarification of Disputed Issues Arising from Complex Evidence 563
    - 1. Early identification of disputed expert testimony 563
    - 2. Attempts to narrow disputes 564
    - 3. Screening of expert testimony 565
  - B. Appointment of an Expert 565
    - 1. Initiation of the appointment 566
    - 2. Communicating with the appointed expert 567
    - 3. Testimony of appointed experts 570
- VIII. Conclusion 571
- Appendix 573

### I. Introduction

Evidence involving complex issues of science and technology plays an increasing role in federal litigation.<sup>1</sup> Appointing an expert is often suggested as a means for the court to enhance its ability to deal with such issues.<sup>2</sup> The Supreme Court has urged judges to "be mindful" of this authority in assessing a proffer of expert testimony.<sup>3</sup> Yet court-appointed experts are infrequently used. This paper summarizes the findings of a study intended to answer the question "Why are court-appointed experts, as authorized by Federal Rule of Evidence 706, employed so infrequently?"<sup>4</sup> In discussing with judges the reasons for infrequent appointments, we also learned of techniques and procedures that may aid judges when considering whether to appoint an expert and when managing an expert who has been appointed. These suggested techniques are collected in section VII.

#### A. Methodology

We gathered information for this report through a mail survey and telephone interviews. First, we sent to each active federal district court judge a cover letter and a one-page questionnaire asking the following questions: "Have you appointed an expert under the authority of Rule 706 of the Federal Rules of Evidence?" <sup>5</sup> and "Are experts appointed under Rule 706 likely to be helpful in certain types of cases?" The questionnaire was intended to determine the extent to which the authority to appoint an expert under Rule 706 had been employed

<sup>1.</sup> The Federal Courts Study Comm., Report of the Federal Courts Study Committee 97 (1990) ("Economic, statistical, technological, and natural and social scientific data are becoming increasingly important in both routine and complex litigation.").

<sup>2.</sup> See, e.g., id.; Jack B. Weinstein & Margaret A. Berger, Weinstein's Evidence Manual: A Guide to the United States Rules Based on Weinstein's Evidence ¶ 13.06[01] (1993); 3 Jack B. Weinstein & Margaret A. Berger, Weinstein's Evidence: Commentary on Rules of Evidence for the United States Courts and State Courts ¶ 706[01] (1993) [hereinafter Weinstein's Evidence]. See also AAAS-ABA Nat'l Conference of Lawyers & Scientists Task Force on Science & Technology in the Courts; Enhancing the Availability of Reliable and Impartial Scientific and Technical Expertise to the Federal Courts: A Report to the Carnegie Commission on Science, Technology, and Government (1991); Carnegie Comm'n on Science, Technology, & Gov't, Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges 37 (1993).

<sup>3.</sup> Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786, 2797-98 (1993).

<sup>4.</sup> For a more detailed report of this study, see Joe S. Cecil & Thomas E. Willging, Court-Appointed Experts: Defining the Role of Experts Appointed Under Federal Rule of Evidence 706 (Federal Judicial Center 1993).

<sup>5.</sup> Judges who answered "yes" were asked about the number of appointments made.

and the extent to which opportunities for Rule 706 appointments exist. Second, we asked those judges who had made Rule 706 appointments to participate in a telephone interview concerning their experiences with court-appointed experts. We sought to identify uses of Rule 706 that judges have found appropriate and, at the same time, identify reasons for nonuse.<sup>6</sup>

In brief, we found that much of the uneasiness with court-appointed experts arises from the difficulty in accommodating such experts in a court system that values, and generally anticipates, adversarial presentation of evidence. More specifically, we found the following:

- Judges view the appointment of an expert as an extraordinary activity that is appropriate only in rare instances in which the traditional adversarial process has failed to permit an informed assessment of the facts. We found no evidence of general disenchantment with the adversarial process by judges who had made such appointments.
- Parties rarely suggest appointing an expert and typically do not participate in the nomination of appointed experts.
- The opportunity to appoint an expert is often hindered by failure to recognize the need for such assistance until the eve of trial.
- Compensation of an expert often obstructs an appointment, especially when one of the parties is indigent.
- Judges report little difficulty in identifying persons to serve as court-appointed experts, largely because of the judges' willingness to use personal and professional relationships to aid the recruitment process.
- Ex parte communication between judges and court-appointed experts occurs frequently, usually with the consent of the parties.
- The testimony or report presented by a court-appointed expert exerts a strong influence on the outcome of litigation.

#### B. Overview

Section II offers a brief summary of the authority of the court to appoint an expert, either under Rule 706 of the Federal Rules of Evidence or under the inherent authority of the court. In subsequent sections we present the results of our mail survey and discuss our interviews with the judges about the origination, selection, pretrial and trial activity, and compensation of the appointed experts. Finally, in section VII we outline suggestions to facilitate the early identification of disputed issues arising from scientific and technical evidence, to clarify and narrow disputes, and to ease appointment of an expert when an independent source of information is necessary for a principled resolution of a conflict.

<sup>6.</sup> We also contacted judges who had not appointed experts but who had indicated, when responding to the mailed questionnaire, strong feelings regarding such practices. We asked these judges how they responded to a number of the situations that the appointing judges had identified as being suitable for making an appointment. This information is detailed in Cecil & Willging, *supra* note 4, at 67–78.

# II. Authority to Appoint an Expert

Two principal sources of authority permit a court to appoint an expert, each source envisioning a somewhat different role for the expert. Rule 706 of the Federal Rules of Evidence most directly addresses the role of the appointed expert as a testifying witness; the structure, language, and procedures of Rule 706 specifically contemplate the use of appointed experts to present evidence to the trier of fact. Supplementing this authority is the broader inherent authority of the court to appoint experts who are necessary to permit the court to carry out its duties, including authority to appoint a technical advisor to consult with the court during the decision-making process. The narrower testimonial focus and procedural confines of Rule 706 do not envision such a role.<sup>7</sup> The authority to appoint a special master under Rule 53 of the Federal Rules of Civil Procedure is addressed elsewhere in this manual.<sup>8</sup> We found instances in which experts appointed under Rule 706 engaged in fact finding much like a special master, yet were also prepared to offer testimony.<sup>9</sup>

#### A. Federal Rule of Evidence 706

Federal Rule of Evidence 706 specifies a set of procedures governing the appointment, assignment of duties, reporting of findings, testimony, and compensation of experts (for text of Rule 706, see the Appendix). Other questions—such as how to identify the need for a Rule 706 expert, how to shape pretrial procedures to reduce conflicts between the parties' experts, how to compensate experts, and how to reduce interference with the adversarial process—are not addressed by the rule but are discussed in later sections of this paper.

The trial court has broad discretion in deciding whether to appoint a Rule 706 expert. Although it has been suggested that "extreme variation" among the parties' experts is a circumstance suggesting that such an appointment may be

<sup>7.</sup> Reilly v. United States, 863 F.2d 149, 155–56 (1st Cir. 1988) ("Rule 706 . . . was not intended to subsume the judiciary's inherent power to appoint technical advisors.").

<sup>8.</sup> See Margaret G. Farrell, Special Masters § III, in this manual.

<sup>9.</sup> At least one district court has held that a single appointee may serve as both a special master and as a court-appointed expert in the same case. Hart v. Community Sch. Bd., 383 F. Supp. 699, 765–66 (E.D.N.Y. 1974), *aff'd*, 512 F.2d 37 (2d Cir. 1975). Another district court expressly granted a special master the power, subject to approval by the court, to "seek the assistance of court-appointed experts." Young v. Pierce, 640 F. Supp. 1476, 1478 (E.D. Tex. 1986), vacated on other grounds, 822 F.2d 1368 (5th Cir. 1987), order reinstated, 685 F. Supp. 984, 985–86 (E.D. Tex. 1988).

beneficial, <sup>10</sup> the trial court retains discretion to refuse to appoint an expert despite such a circumstance. <sup>11</sup> Such experts should be appointed when they are likely to clarify issues under consideration; it is not an abuse of discretion for a trial court to refuse to appoint an expert under Rule 706 when "additional experts would . . . add more divergence and opinion differences." <sup>12</sup>

Appellate courts on occasion have reminded judges of this authority. Where a trial court has been unaware of its authority to appoint a neutral expert under Rule 706 or its inherent power to do so, a reviewing court may order the trial court to exercise its discretion and decide whether appointment of a neutral expert is justified in the circumstances of the case.<sup>13</sup> Indeed, in a case in which the experts' testimony is especially disparate on an issue of valuation, a trial court should consider the value of "a court-appointed witness [who] would be unconcerned with either promoting or attacking a particular estimate of . . . [plaintiff's] damages." <sup>14</sup> The standard for review of a trial court's appointment of an expert under Rule 706 is whether the appointment constituted an abuse of discretion. <sup>15</sup> One factor to consider in such a review is whether the expert selected by the court had any bias toward one party or one side of an issue.<sup>16</sup>

Two cases demonstrate the range of functions that may be performed by court-appointed experts. *Computer Associates International, Inc. v. Altai, Inc.*<sup>17</sup> offers an example of an expansive role by an appointed expert in difficult technical litigation concerning alleged infringement of a software copyright. The question before the court was how to separate the idea underlying a computer program from its expression, since only the latter is protected by copyright. The parties agreed to the court's appointment of a computer science professor from the Massachusetts Institute of Technology to aid the judge in a nonjury trial in understanding the technical issues of the case. In analyzing and interpreting the facts for the court, the appointed expert also pointed out deficiencies in the legal doctrines and suggested alternative standards that would bring the copyright law protecting computer software into conformity with current practices in computer science. The district court adopted this proposal and assessed the allegedly copied program under this new standard. On appeal one party sought to over-

11. Oklahoma Natural Gas Co. v. Mahan & Rowsey, Inc., 786 F.2d 1004, 1007 (10th Cir.), cert. denied, 479 U.S. 853 (1986); Georgia-Pacific Corp. v. United States, 640 F.2d 328, 333–35 (Ct. Cl. 1980).

12. Georgia-Pacific, 640 F.2d at 334.

13. Fugitt v. Jones, 549 F.2d 1001, 1006 (5th Cir. 1977).

14. Eastern Air Lines, Inc. v. McDonnell Douglas Corp., 532 F.2d 957, 1000 (5th Cir. 1976).

15. Gates v. United States, 707 F.2d 1141, 1144 (10th Cir. 1983).

16. Id.

17. 775 F. Supp. 544, 549, 559–60 (E.D.N.Y. 1991), aff'd, Nos. 91-7893, 91-7935, 1992 U.S. App. LEXIS 14305 (2d Cir. June 22, 1992), vacated in part on other grounds, 982 F.2d 693 (2d Cir. 1992).

<sup>10.</sup> Eastern Air Lines, Inc. v. McDonnell Douglas Corp., 532 F.2d 957, 999 (5th Cir. 1976). In Reilly v. United States, 863 F.2d at 156–57, the court identified "some cognizable judicial need for specialized skills" as a justifiable reason for utilizing an expert as a technical advisor. See also Computer Assocs. Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 713 (2d Cir. 1992) (complicated nature of computer software programming justifies assessment by court-appointed expert if similarities arise to the level of a wrongful appropriation of copyrighted work).

turn the standard, contending that the district court had erred by relying too heavily on the court-appointed expert's opinions. The court of appeals noted that the technical nature of assessments of computer software justified a more expansive role for expert assistance and that the appointed expert's opinion "was instrumental in dismantling the intricacies of computer science so that the court could formulate and apply an appropriate rule of law." <sup>18</sup> Since, in the final analysis, the district court judge exercised judicial authority in reviewing these findings, the court of appeals found the assistance provided by the expert to be appropriate.

In contrast to this expansive role, the court in Renaud v. Martin Marietta Corp.<sup>19</sup> relied on the appointed expert for the more limited purpose of assessing the acceptability within the scientific community of the methodology used by the plaintiffs to measure exposure to a toxic chemical. Residents of a community brought a toxic tort action against a nearby manufacturer; the residents alleged injuries caused by contaminated drinking water. The defendants challenged the admissibility of expert testimony by the plaintiffs concerning the level of exposure to the chemical. Estimates of exposure over an eleven-year period were based on an extrapolation from a single measure of contamination in one place and one time two years after the last alleged exposure. The court appointed an expert in geochemistry and hydrology to assess not the general question of causation, but the narrow question of the scientific acceptability of using a single data point to estimate exposure over such a period. In her report to the court, the appointed expert wrote, "'[i]t is unsound scientific practice to select one concentration measured at a single location and point in time and apply it to describe continuous releases of contaminants over an 11-year period."<sup>20</sup> On this basis the court refused to admit the evidence of exposure and, in the absence of other evidence, granted the defendants' motion for summary judgment. On appeal the plaintiffs challenged the authority of the expert to render such an assessment. The court noted such duties are well within the scope of the authority of an appointed expert.<sup>21</sup> The use of appointed experts to comment on the acceptability of scientific methods that underlie expert opinions may expand as courts assess the scientific validity of expert testimony under the standards estab-

Court-Appointed Experts

<sup>18.</sup> Computer Assocs. Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 713-14 (2d Cir. 1992).

<sup>19. 749</sup> F. Supp. 1545, 1552-53 (D. Colo. 1990), aff'd, 972 F.2d 304 (10th Cir. 1992).

<sup>20. 749</sup> F. Supp. at 1553. See generally E. Donald Elliott, Toward Incentive-Based Procedure: Three Approaches for Regulating Scientific Evidence, 69 B.U. L. Rev. 487, 508 (1989) (suggesting that in cases with "substantial doubt" regarding the scientific integrity of testimony by a party's expert, the court appoint a "peer review expert learned in the relevant fields to testify at trial concerning whether the principles, techniques, and conclusions by the experts for the parties would be generally accepted as valid by persons learned in the field").

<sup>21.</sup> Renaud v. Martin Marietta Corp., 972 F.2d 304, 308 n.8 (10th Cir. 1992). The court of appeals also rejected the plaintiffs' argument that they were wrongly denied the right to depose the appointed expert, noting that the appointed experts were "more technical advisors to the Court than expert witnesses as contemplated by Fed. R. Evid. 706, and accordingly depositions and cross-examination were inappropriate."

lished by the Supreme Court's decision in Daubert v. Merrell Dow Pharmaceuticals, Inc.<sup>22</sup>

#### B. Inherent Authority to Appoint a Technical Advisor

The court's authority under Rule 706 to appoint an expert to offer testimony represents a specific application of its broader inherent authority to invite expert assistance in a broad range of duties necessary to decide a case. The most striking exercise of this broader authority involves appointing an expert as a technical advisor to confer in chambers with the judge regarding the evidence, as opposed to offering testimony in open court and being subject to cross-examination. Although few cases deal with the inherent power of a court to appoint a technical advisor, the power to appoint remains virtually undisputed,<sup>23</sup> tracing a clear line from the 1920 decision of the Supreme Court in Ex parte Peterson<sup>24</sup> to the recent decision of the U.S. Court of Appeals for the First Circuit in Reilly v. United States.<sup>25</sup> Generally, a district court has discretion to appoint a technical advisor, but it is expected that such appointments will be "hen's teeth rare," a "last" or "near-to-last resort."<sup>26</sup> General factors that might justify an appointment are "problems of unusual difficulty, sophistication, and complexity, involving something well beyond the regular questions of fact and law with which judges must routinely grapple."<sup>27</sup> The role of the technical advisor, as the name implies, is to give advice to the judge, not to give evidence and not to decide the case.<sup>28</sup> Compensation of a technical advisor can be especially awkward; this issue is discussed at length in section VI, infra.

22. 113 S. Ct. 2786 (1993). For a discussion of admissibility of expert testimony after *Daubert*, see Margaret A. Berger, Evidentiary Framework, in this manual.

23. In the words of the Advisory Committee on the Rules of Evidence, "[t]he inherent power of a trial judge to appoint an expert of his own choosing is virtually unquestioned." Fed. R. Evid. 706 advisory committee's note; see also United States v. Green, 544 F.2d 138, 145 (3d Cir. 1976), cert. denied sub nom. Tefsa v. United States, 430 U.S. 910 (1977) ("[T]he inherent power of a trial judge to appoint an expert of his own choosing is clear."); Scott v. Spanjer Bros., 298 F.2d 928, 930 (2d Cir. 1962) ("Appellate courts no longer question the inherent power of a trial court to appoint an expert under proper circumstances....").

24. 253 U.S. 300, 312 (1920) (In approving the appointment of an auditor to segregate the claims that were in dispute and to express an opinion on the disputed items, the Court found that "[c]ourts have (at least in the absence of legislation to the contrary) inherent power to provide themselves with appropriate instruments required for the performance of their duties.").

25. 863 F.2d 149, 154 & n.4 (1st Cir. 1988) (In a case involving appointment by the district court of an economist to assist the court in calculating damages to an infant resulting from medical malpractice, the United States (defendant) conceded that "a district court has inherent authority to appoint an expert as a technical advisor." The circuit court agreed that "such power inheres generally in a district court."); see also Bullard Co. v. General Elec. Co., 348 F.2d 985, 990 (4th Cir. 1965) ("Of course, the District Court has the right on an intricate subject of suit, as here [a patent infringement case], to engage an advisor to attend the trial and assist the court in its comprehension of the case."); Friends of the Earth v. Carey, 535 F.2d 165, 173 & n.13 (2d Cir. 1976) (District judge has "power to obtain such expert advice and assistance as may be necessary to guide him" and "to assist him in the performance of his duties."), vacated on other grounds, 552 F.2d 25 (2d Cir.), cert. denied, 434 U.S. 902 (1977).

26. Reilly v. United States, 863 F.2d 149, 157 (1st Cir. 1988).

27. Id.

28. *Id.* ("Advisors . . . are not witnesses and may not contribute evidence. Similarly, they are not judges, so they may not be allowed to usurp the judicial function.").

# III. Use and Nonuse of Court-Appointed Experts

#### A. Use of Court-Appointed Experts

Many have mentioned that the use of court-appointed experts appears to be rare, an impression based on the infrequent references to such experts in published cases.<sup>29</sup> To obtain an accurate assessment of the extent to which court-ap pointed experts have been employed, we sent a one-page questionnaire to all active federal district court judges.<sup>30</sup>

Figure 1 Have You Appointed an Expert Under Rule 706?



As indicated in Figure 1, eighty-six judges, or 20% of those responding to the survey, revealed that they had appointed an expert on one or more occasions.

<sup>29.</sup> Weinstein's Evidence, supra note 2, ¶ 706[01], at 706–13.

<sup>30.</sup> Questionnaires were sent to 537 active federal district court judges; 431 judges responded (a response rate of 80%).

Of the eighty-six judges reporting appointment of an expert, just over half had appointed an expert on only one occasion. Only four judges appointed an expert in ten or more cases, a frequency that suggests a somewhat systematic use of appointed experts to deal with difficult scientific or technical issues.

During the telephone interviews, we asked the judges to describe the cases in which they had appointed experts under authority of Rule 706. Three circumstances accounted for almost two-thirds of the appointments: medical experts appointed in personal injury cases, engineering experts appointed in patent and trade secret cases, and accounting experts appointed in commercial cases. The appointed expert usually served a different function in each type of case.

The expertise most commonly sought by the courts (required in twenty-four cases) was that of medical professionals concerning the nature and extent of injuries. In thirteen of these cases experts were appointed to help assess claims for injuries arising from improper medical care. In eight other cases the appointed expert considered injuries arising from defective products, five of which were tort claims based on injuries caused by exposure to toxic chemical products.

The services of the appointed medical experts varied with the type of personal injury case. In cases arising from claims of improper medical care, the parties' experts usually were in complete opposition, and the appointed expert advised the court on the proper standards of medical care and treatment. During the product liability litigation, the appointed medical expert addressed the cause and extent of injuries. In four of five tort cases about toxic products, the appointed expert addressed the likelihood that the product caused the injuries.

In fifteen cases judges sought experts with skills in engineering. <sup>31</sup> Twelve of these cases raised questions of patentability, patent infringement, or technical issues surrounding trade secret protection. Unlike the personal injury cases in which the expert was appointed to resolve a dispute among the parties' experts, in these cases the expert typically was appointed to interpret technical information for the judge. Almost all of these cases were bench trials, and the parties agreed to the appointment of an expert to enhance the court's ability to understand the technology underlying the dispute.

In twelve cases involving disputes over contracts or failed commercial enterprises, judges sought the assistance of accountants.<sup>32</sup> Often these cases involved complex financial transactions, and the expert was appointed to assist the court in placing a value on a claim. In reaching such an assessment, the appointed expert often functioned like a special master, reviewing records and preparing a report that was submitted as evidence in the case.<sup>33</sup> In several cases the judge

<sup>31.</sup> We include in this category experts who had knowledge of the development of computer hardware and software (accounts for six cases).

<sup>32.</sup> We include in this category those appointed experts who were identified as accountants or described as providing accounting services. Some may have lacked formal training as accountants. We did not inquire about the credentials of the appointed experts.

<sup>33.</sup> Some judges expressed a preference for appointing an expert under Rule 706, as opposed to a special master under Fed. R. Civ. P. 53, so the accountant could testify in court and be cross-examined by the parties.

asked the appointed expert not to place a value on a disputed claim, but to address acceptable standards of accounting that should be followed in making such a determination, or to educate the court regarding acceptable methods for making such a determination. The remainder of the appointments were scattered across a variety of specialties and types of cases.

#### B. Satisfaction with Appointed Experts

The judges who appointed experts were almost unanimous in expressing their satisfaction with the expert: All but two of the sixty-five judges indicated that they were pleased with the services provided. The two judges who did not indicate that they were satisfied remain open to appointing an expert in the future. One judge indicated that he had little basis from which to form a judgment regarding the performance of the two experts he appointed; one expert was called on to do little before the case settled, and the other testified before a visiting judge. The other judge who did not express satisfaction with the process indicated some frustration that the interactions with the expert had been constrained by a need to avoid direct communication with the expert outside the presence of the parties.

### C. Receptivity to Appointment of Experts

The second question asked on the one-page questionnaire ("Are experts appointed under Rule 706 likely to be helpful in certain types of cases?") was intended to assess the extent to which judges consider appointment of an expert to be an acceptable alternative in at least some types of cases.

Few judges fail to see any value in appointment of experts by the court. Eighty-seven percent of the judges responding to the question indicated that court-appointed experts are likely to be helpful in at least some circumstances (see *infra* Figure 2). This openness to appointment of experts extended to judges who had never appointed an expert, 67% of whom indicated that such an appointment might be helpful.

#### D. Reasons for Appointing Experts

Judges who had made a single appointment were asked to describe their reasons for making the appointment. They were also asked in another portion of the interview what concerns led to their decision to appoint an expert. Our interviews revealed two distinct sets of judges who have used Rule 706. One group uses the rule primarily to advance the court's understanding of the merits of the litigation and to enhance the court's ability to reach a reasoned decision on the merits; a smaller group, apparently mostly multiple users, invokes the rule primarily to enhance settlement.

#### 1. To aid decision making

As might be expected, experts are most often appointed to assist in understanding technical issues necessary to reach a decision.<sup>34</sup> The desire for such assistance was attributed by the judges to a lack of knowledge in an essential area, a concern over the technical nature of an issue or issues, or a concern over the need to properly articulate the rationale for a decision. Many judges mentioned more than one of these concerns.

In explaining the reason for the appointments, judges often admitted their need to become better informed on an essential topic of the litigation. Typical comments were "I was aware of the limits of my knowledge of [biochemistry]," and "The experts took almost diametrically opposed positions in areas in which I knew next to nothing." In some contexts, the judge's need for technical expertise was coupled with a first-time exposure to a complex legal specialty area, such as patent law.

The need for assistance in decision making often arose when the parties failed to present credible expert testimony, thereby failing to inform the trier of fact on essential issues. Judges' doubts regarding the credibility of testimony by the parties' experts were common. Twenty-seven of the forty-five judges who appointed an expert on only one occasion described a situation in which both parties employed testifying experts. These judges often described a situation in which each party offered apparently competent expert testimony that was in direct opposition on virtually every issue to the other party's expert testimony. Such total disagreement in areas unfamiliar to the judge invited a general distrust of the experts. This concern over the integrity of testimony of experts was echoed elsewhere in the survey. When judges were asked in a separate question what concerns led them to appoint an expert, in eighteen of thirty-six cases judges indicated that there was a failure by one or both parties to present credible expert testimony to aid in resolving a disputed issue. Appointment of an independent expert enabled access to testimony that was thought to be both impartial and necessary to understand the testimony of the parties' experts.

The second typical circumstance involved appointment of an expert when at least one of the parties failed to offer expert testimony, resulting in what the judge perceived to be an inadequate presentation of issues. This circumstance, reported by thirteen of the forty-five judges who had appointed an expert on one occasion, typically arose because of a party's inability to pay for expert testimony.<sup>35</sup> In many of these cases the judge had heard expert testimony by one party and could have resolved the dispute in favor of that party because of the failure of the opponent to present countervailing expert testimony in support of a

<sup>34.</sup> More than two-thirds of the forty-five judges who had made only one appointment reported that they made the appointment to obtain assistance in understanding technical issues necessary to reach a decision. We did not ask judges who appointed experts on more than one occasion about the reasons for their most recent appointment, focusing instead on the general characteristics of cases in which they appointed experts.

<sup>35.</sup> See discussion of this issue infra notes 99-102 and related text.

critical issue. In discussing such cases the judges made clear their uneasiness in basing their decisions strictly on the adversarial presentations of the parties. Such a resolution would have failed to adequately resolve the disputed issue and may have complicated a fair and accurate resolution of similar issues in the future. These judges were sufficiently concerned about the nature of the proffered expert testimony to undertake the considerable effort necessary to obtain an independent assessment from an appointed expert, thereby obtaining a valid rationale for a decision.

Though circumstances differed in these cases, each reveals a judge's marked dissatisfaction with the parties' experts' presentation of information and the traditional means of resolving such conflicting testimony. In each circumstance an expert was appointed by the court when traditional adversarial presentation by parties failed to provide the court with information necessary to make a reasoned determination of disputed issues of fact.

#### 2. To aid settlement

Some judges suggested that appointment of an expert may bring about settlement, although enhancement of settlement prospects was rarely an articulated purpose of the appointment. Indeed, the judges we interviewed indicated that the prospect of settlement often argued against the appointment of an expert. In the words of a judge who had never made an appointment, judges might be reluctant to "get all dressed up with no place to go."

Judges who have appointed more than one expert are more likely to view settlement as a reason to make an appointment; a majority of those judges reported that when appointing an expert they had in mind enhancing the opportunity for settlement.<sup>36</sup> These judges sometimes appeared to appoint an expert in an effort to change parties' extreme evaluations of a case. In situations in which the experts for the parties are highly qualified, yet give disparate opinions (in the words of one judge "fixed on two equally good positions"), an appointment is intended to resolve the impasse and permit the parties to move on to discussion of other issues.

As with judicial involvement in settlement in general, <sup>37</sup> there is no consensus on the use of court-appointed experts to aid in settlement. The time and expense involved in the process, however, raises the question of whether an appointment for the purpose of improving judicial decision making will be worthwhile if the parties are likely to settle.

<sup>36.</sup> We asked those who had made multiple appointments, "How do the prospects for settlement of the case influence your decision to appoint an expert?" Of the nineteen judges who responded to the question, nine indicated that the possibility of settlement would favor their decisions to appoint experts and two indicated that the prospect of settlement was a secondary consideration supporting appointment. Four of the multiple users said that serious prospects for settlement would lead them to *not* appoint an expert and four more said that the prospects of settlement would have no effect on their decision.

<sup>37.</sup> See generally D. Marie Provine, Settlement Strategies for Federal District Judges (Federal Judicial Center 1986).

#### E. Reasons for Failure to Appoint an Expert

Almost all judges are willing to consider the appointment of an expert in at least some circumstances, so the infrequency of such appointments is not related to a strict opposition to the practice. Our investigation revealed problems in identifying suitable experts, communicating effectively with such appointed experts, and compensating appointed experts. Many of these practical problems can be overcome and are discussed in detail in the following sections. But the two principal reasons for failure to appoint an expert are the infrequency of cases requiring such assistance and the reluctance of judges to intrude into the adversarial process. These two issues set a limit on the opportunities to use such appointed experts, a limit that will not be overcome by improvements in procedures.

#### 1. Infrequency of cases requiring extraordinary assistance.

To better understand the reasons for the infrequent appointment of experts, we asked eighty-one judges why they thought the authority had been exercised so infrequently. Fifty judges indicated that they see the appointment of an expert as an extraordinary action. The importance of reserving appointment of experts for cases involving special needs was especially apparent in the responses of the judges who had made only a single appointment. Thirty-two of the forty-five judges who had appointed an expert on a single occasion indicated that they had not used the procedure more often because the unique circumstances in which they employed the expert had not arisen again. They simply had not found another suitable occasion in which to appoint an expert.

When we asked judges in the mail survey to indicate types of cases in which an appointed expert might be helpful, they usually indicated types of cases that are both rare and unusually demanding, implying that appointed experts should be reserved for cases with extraordinary needs. Figure 2 indicates the types of cases, as identified by the judges, in which the appointment of an expert would be helpful. More than half of the judges mentioned patent cases. Cases involving questions of product liability and antitrust violations also were common candidates for such assistance. It follows that one reason appointments are rare is that the kinds of cases in which judges are likely to require such assistance are themselves rare.

Reference Manual on Scientific Evidence





Number of judges specifying case type

*Note:* Of the 537 judges surveyed, there were 385 respondents to this question. Forty-six of the 431 who answered the first question did not answer this one (all of those judges had answered no to the first question).

In the "Other" category, the most common responses were "Depends on particular case" (twenty-seven judges) and "All cases" (nineteen judges).

Appointments were often made in response to a combination of unusual events, such as a failure by the parties to provide a basis for a reasoned resolution of a technical issue, combined with a perceived need by the court to protect poorly represented parties (such as minors or members of a certified class action). One judge, in a case alleging injuries to a family arising from toxic contamination of a water supply, appointed an expert when the plaintiff's attorney failed to retain an expert witness to establish the occurrence of injury to the children. The judge could have entered a summary judgment in favor of the defendant, and suggested he would have done so but for the presence of children.

**Court-Appointed Experts**
The failure of the plaintiff's attorney to present expert testimony and the presence of children combined to motivate the court to appoint an expert.<sup>38</sup>

A number of judges mentioned the need for an appointed expert when the parties' experts are in complete disagreement, one judge remarking, "One needs a complete divergence in the views of the parties' experts in a technically complex field. Often experts differ, but not in a crazy way." Several of these judges questioned the belief that court-appointed experts were being used too infrequently. While acknowledging that such authority is useful, one judge remarked, "I don't know that [court-appointed experts have] been used too infrequently. It should remain a rare device that is suited for unusual circumstances."

#### 2. Respect for the adversarial system

Respect for the adversarial system was cited as a reason for the infrequent appointment of experts by thirty-nine of the eighty-one judges, including thirteen of the eighteen judges who had not appointed an expert.<sup>39</sup> Many of those who had appointed experts professed commitment to the adversarial process and the ability of juries to assess difficult evidence, and they indicated that they would appoint an expert only where the adversarial process had failed.

A related reason for infrequent appointment of experts is deference by the judge to objections by the parties. Several judges alluded to such resistance, one stating "The parties resist, saying that they have their own experts." Similarly, another judge said that generally "the plaintiffs or their attorneys do not want such an expert because it will reduce the value of their case. I don't appoint experts without consent of the parties." Judges who favored other alternatives over the use of court-appointed experts cited deference to the parties as an important consideration.<sup>40</sup>

<sup>38.</sup> See discussion infra § VI.C.

<sup>39.</sup> Judges were permitted to offer more than one reason, and many of the judges who cited the unique circumstances in which such an appointment would be appropriate also stressed the importance of the judge not intruding on the adversarial system where it appears to be functioning.

<sup>40.</sup> See also Manual for Complex Litigation, Third, § 21.51 ("Although the appointment is made by the court, every effort should be made to select a person acceptable to the litigants; in fact, the parties should first be asked to submit a list of proposed experts and may be able, with the assistance of their own expert, to agree on one or more candidates.") (forthcoming 1995) [hereinafter MCL 3d].

### **IV. Identification and Appointment of Experts**

#### A. Timing of the Appointment

One of the impediments to broader use of court-appointed experts mentioned earlier is the difficulty in identifying the need for an expert in time to make the appointment without delaying the trial.<sup>41</sup> Thirteen judges indicated that effective appointment of an expert requires the court's awareness of the need for such assistance early in the litigation. Since the parties rarely suggest that the court appoint an expert, judges sometimes don't realize that they need assistance until the eve of trial—when there is not sufficient time to identify and appoint an expert. Several judges indicated that they had learned of the need for such assistance when it was too late.

Procedures specified in Rule 706 imply that the appointment process "will ordinarily be invoked considerably before trial" to allow time for hearings on the appointment, consent of the expert, notification of duties, research by the expert, and communication of the expert's findings to the parties in sufficient time for the parties to conduct depositions of the expert and prepare for trial.<sup>42</sup> For example, one authority has suggested that identification of the need for a neutral expert should begin at a pretrial conference held pursuant to Federal Rule of Civil Procedure 16.<sup>43</sup> However, specific procedures for identifying such a need are left to the trial judge.<sup>44</sup>

Timing of the appointment was discussed regarding fifty-two cases. A majority of the experts were appointed at an early point in the litigation, but a sizable minority were appointed on the eve of trial.<sup>45</sup> A few judges even appointed experts

<sup>41.</sup> The role of timing of the appointment is discussed in greater detail in Cecil & Willging, supra note 4, at 26–29.

<sup>42.</sup> Weinstein's Evidence, supra note 2, ¶ 706[02], at 706–14; see also United States v. Weathers, 618 F.2d 663, 664 n.1 (10th Cir.), cert. denied, 446 U.S. 956 (1980).

<sup>43.</sup> Weinstein's Evidence, supra note 2, ¶ 706[02], at 706-14 to -15.

<sup>44.</sup> For example, a court may want to time the neutral expert's testimony and final report to allow that expert to hear and comment on the testimony of the parties' experts. See, e.g., Leesona Corp. v. Varta Batteries, Inc., 522 F. Supp. 1304, 1311–12 (S.D.N.Y. 1981).

<sup>45.</sup> In discussing the timing of the appointment, the term *trial* is used in a broad sense to indicate the anticipated evidentiary hearing before the court in which the opinion of the appointed expert would be solicited. Usually this will be a formal trial before a judge or jury. Sometimes, however, the court invited the assistance of an expert to aid in resolving an issue to be addressed in a pretrial hearing. In this circumstance the timing of

during or after bench trials. Often, judges who acted immediately before, during, or after trial indicated that an earlier appointment would have been helpful. Thirty-one of the judges reported that they appointed the expert early in the pretrial process, usually at the close of discovery, leaving time to recruit an expert and permit the expert to prepare a report.

Asked if it would have been helpful to appoint the expert at an earlier point in the litigation, those who made an appointment shortly after discovery generally expressed satisfaction with the timing of the appointment. By contrast, most of those judges who appointed the expert immediately before or during the trial indicated that appointment earlier in the process would have been helpful.<sup>46</sup> Often they noted the need to reschedule the proceeding to permit time to appoint and employ the expert. Another judge mentioned that an earlier appointment would have been helpful in recruiting more skilled experts, remarking, "Only one of the potential experts was available. With more time it may have been possible to choose among several experts."

#### B. Initiation of the Appointment

Our interviews revealed that the initial suggestion to appoint an expert almost always comes from the judge, not the parties. When asked who had initiated the appointment, almost all of the judges who responded (fifty-four of sixty-one judges) indicated that they had. In only seven instances did the initial suggestion come from the parties—twice from the plaintiff, twice from the defendant, and three times from both parties. In one instance the plaintiff's suggestion for appointment of a panel of experts <sup>47</sup> appeared to be part of a broader litigation strategy, since the plaintiff had recommended such appointments in related litigation in other districts.

### C. Selection of the Appointed Expert

Identification and selection of a neutral expert by the court is a critical step in ensuring the fairness of the proceeding.<sup>48</sup> When we asked why experts are ap-

48. By neutral expert we mean an expert who can respond to the technical or scientific issue in a manner consistent with generally accepted knowledge in an area, without regard to the interests advanced by either

the appointment was examined with reference to the hearing rather than to the trial itself. For convenience, this pretrial hearing is referred to as a trial.

<sup>46.</sup> It is worth noting that all but one of these instances in which an appointment was made immediately before or during trial involved a judge rather than a jury serving as the finder of fact. One judge remarked that a bench trial permits such flexibility because the judge can schedule the proceedings without having to accommodate the need for a continuous period of service by jurors.

<sup>47.</sup> Panels of experts also may be appointed by the court. Rule 706 uses the plural term *expert witnesses* to indicate that more than one expert may be appointed in a case. See Gates v. United States, 707 F.2d 1141, 1144 (10th Cir. 1983); Fund for Animals, Inc. v. Florida Game & Fresh Water Fish Comm'n, 550 F. Supp. 1206, 1208 (S.D. Fla. 1982); Lightfoot v. Walker, 486 F. Supp. 504, 506 (S.D. Ill. 1980), *later proceeding*, 619 F. Supp. 1481 (S.D. Ill. 1985), *affd* 797 F.2d 505 (7th Cir. 1986); *In re* Repetitive Stress Injury Cases Pending in the U.S. Dist. Court., 142 F.R.D. 584 (E.D.N.Y. 1992), vacated on other grounds sub nom. Debruyne v. National Semiconductor Corp. (*In re* Repetitive Stress Injury Litig.), 11 F.3d 368 (2d Cir. 1993).

pointed infrequently, the difficulty in identifying a suitable neutral expert to serve the court was mentioned by fourteen judges. Some judges spoke of the difficulty in recruiting unbiased experts with the knowledge demanded in litigation. Some didn't know where to turn to initiate the process. And expressed repeatedly in the interviews was the distrust of expert testimony in general. Several judges doubted that such testimony would be truly neutral, even if the expert was invited to testify by the court.

Those judges who actually appointed experts did not seem to encounter such difficulty. Only six of sixty-six judges reported difficulty finding a neutral expert willing to serve.<sup>49</sup> Those six judges cited either difficulty in finding a skilled person who could be considered neutral (some had ties with the parties while others had previously taken positions on the technical issues that were the object of the dispute), or difficulty in finding a neutral expert who would consent to serve.

Perhaps one reason judges who made such appointments found little difficulty in identifying experts is that they often appointed experts with whom they were familiar. We found that it is far more common for judges to appoint experts that they have identified and recruited, often based on previous personal or professional relationships, than for judges to appoint experts nominated by the parties. <sup>50</sup>

In forty-one of the sixty-six appointments, the judge appointed an expert without suggestions by the parties. In twenty-nine of these cases, the judge used preexisting personal or professional contacts to identify an expert. The extent to which judges relied on their informal networks of friends and acquaintances raises concerns about the extent to which such networks can be relied on to provide skilled and neutral experts to inform the deliberations of the trier of fact. While such persons may be "disinterested" with regard to the issues of the specific case, there is little assurance that such acquaintances bring an unbiased, or even a well-informed, perspective to the disputed technical issues. Personal associations formed while practicing law may reflect a narrow spectrum of professional opinion that was suited to the interests of the judges' former clients and colleagues. Even if such an appointment results in the selection of a suitable expert, the parties may perceive such an expert as biased.<sup>51</sup>

party. This would rule out experts with significant ideological, financial, or professional interests in debatable normative issues related to the issue in dispute. *Cf. In re* Philadelphia Mortgage Trust, 930 F.2d 306, 309 (3d Cir. 1991) (comparing "neutral" court-appointed expert with accountants appointed to assist a trustee in bankruptcy).

<sup>49.</sup> Some judges may have encountered difficulty in finding a neutral expert and abandoned their efforts to appoint such a person, thereby eluding our investigation.

<sup>50.</sup> Judges are afforded great discretion under Rule 706 in designating a procedure for appointing such an expert. Gates v. United States, 707 F.2d 1141, 1144 (10th Cir. 1983). Rule 706(a) provides that "[t]he court may appoint any expert witnesses agreed upon by the parties, and may appoint expert witnesses of its own selection."

<sup>51.</sup> We should note that while our interview with judges raised the possible dangers of such appointments, we found no indication that such harms have resulted.

Judges did not always rely on friends and associates to suggest experts; in nine instances in which an appointment was made without suggestions by the parties, judges contacted nearby institutions for assistance in identifying suitable experts to serve the court. <sup>52</sup> These were almost all instances in which medical expertise was needed and the judges contacted nearby medical schools or associations for suggestions of candidates. Such a procedure, while more burdensome and not foolproof, <sup>53</sup> is likely to be more effective than using informal contacts to identify skilled, neutral experts.

In eighteen instances the expert was selected from a list of experts provided by one or more of the parties.<sup>54</sup> Published cases commonly suggest that a court direct the parties to seek agreement on an appointment and for the court to exercise its discretion only if the parties fail to agree.<sup>55</sup> Sometimes the parties agreed on an expert with little or no involvement from the judge. Normally each party submitted a slate of experts that would be acceptable to them. Occasionally one or more names would appear on each list, making selection easy. Often the parties identified one or more suitable experts with little or no involvement by the judge. When the parties could not agree, the judge often chose the expert from the slates after listening to objections from each of the parties.

In summary, the identification of a need for and the selection of a court-appointed expert appears to be a process in which the parties infrequently play an active role. The judge typically identifies the need for assistance and raises the possibility of such an appointment, sometimes very late in the pretrial process. The judge is usually responsible for identifying suitable candidates and often relies on informal recommendations from friends and associates. Such unsystematic approaches to identifying needs and recruiting experts raise doubts about the extent to which the procedure provides the timely and neutral assistance warranted by the central importance of the expert's task.

<sup>52.</sup> The selection procedure suggested in the *Manual for Complex Litigation*, Third, is for the court to "call on professional organizations and academic groups to provide a list of qualified and available persons...." MCL 3d, *supra* note 40, § 21.51; see also 1 McCormick on Evidence § 17, at 71 (John William Strong ed., 4th ed. 1992) (recommends "establishing panels of impartial experts designated by groups in the appropriate fields, from which panel court-appointed experts would be selected ...").

<sup>53.</sup> Professional associations and academic groups also may have skewed approaches to a specific issue, perhaps giving subconscious, or even conscious, priority to the impact of a rule or ruling on their professional autonomy. Medical malpractice cases, for example, may test the ability of medical schools or professional associations to assist in identifying neutral experts.

<sup>54.</sup> The few reported cases dealing with selection of experts tend to emphasize nomination by the parties. See, e.g., Gates v. United States, 707 F.2d 1141, 1144 (10th Cir. 1983); Fund for Animals, Inc. v. Florida Game & Fresh Water Fish Comm'n, 550 F. Supp. 1206, 1208 (S.D. Fla. 1982); Leesona Corp. v. Varta Batteries, Inc., 522 F. Supp. 1304, 1311 (S.D.N.Y. 1981); Lightfoot v. Walker, 486 F. Supp. 504, 506 (S.D. Ill. 1980), *later proceeding*, 619 F. Supp. 1481 (S.D. Ill. 1985), *aff'd*, 797 F.2d 505 (7th Cir. 1986); United States v. Ridling, 350 F. Supp. 90, 99 (E.D. Mich. 1972).

<sup>55.</sup> United States v. Michigan, 680 F. Supp. 928, 957 (W.D. Mich. 1987); Unique Concepts, Inc. v. Brown, 659 F. Supp. 1008, 1011 (S.D.N.Y. 1987), *later proceeding*, 735 F. Supp. 145 (S.D.N.Y. 1990), *affd*, 939 F.2d 1558 (Fed. Cir. 1991); Hatuey Prods. v. United States Dep't of Agric., 509 F. Supp. 21, 23 (D.N.J. 1980). *See also* Pamela Louise Johnston, *Court-Appointed Scientific Expert Witnesses: Unfettering Expertise*, 2 High Tech. L.J. 249, 267–68 (1988) (suggesting that Rule 706 be amended to require parties to submit a list of proposed experts suitable for appointment by the court for each area of disputed scientific testimony).

### V. Communication with the Appointed Expert

#### A. Instruction of the Appointed Expert

Rule 706(a) of the Federal Rules of Evidence specifies two options for instructing the expert in his or her duties—both options ensure that the parties will be aware of the assignment. The court may communicate with the expert either in writing (filing a copy with the clerk) or at a conference in which the parties have an opportunity to participate. In practice, judges instructed experts by conference call (involving the judge, the expert, and the parties), informal conferences in chambers, formal hearings in open court, and letters and written orders, sometimes with accompanying documents and exhibits. In only two instances did judges instruct experts outside the presence of the parties.<sup>56</sup>

Judges' instructions were used to meet multiple needs, including (1) establishing a record of the terms and conditions of the appointment, including the terms of payment; (2) defining the legal and technical issues in the case and identifying the technical issues the expert was to address; (3) clarifying the role of the expert in relation to the role of the judge; and (4) establishing procedures for assembling information, communicating with the parties, and reporting findings and opinions. The following discussion summarizes how judges met those needs in the cases we encountered.<sup>57</sup>

Regarding terms of payment, judges included in the order of appointment the rate of payment,<sup>58</sup> any ceiling on the total amount of work and payment, the al-location of payment among the parties, the timing of installment payments, the amount of an initial payment, the court's role, if any, in reviewing the bills and serving as a conduit for payments, and reallocation of payments upon taxation of costs.

<sup>56.</sup> Direct instructions from the judge outside the presence of the parties occurred in an emergency situation (appointment of a doctor to review medical records on the day of trial) and in a nonadversarial situation in which the expert functioned like a special master in preparing a report to assist the judge in formulating the distribution of a settlement fund.

<sup>57.</sup> For an example of an order appointing an expert, see *In re* Swine Flu Immunization Prods. Liab. Litig., 495 F. Supp. 1185 (E.D. Okla. 1980) (comprehensive order appointing panel of medical experts to review swine flu cases, detailing the areas of inquiry, the duties of the panel, the content and timing of the reports, the deposition process, exchange of information by counsel, and the charges and method of claiming compensation).

<sup>58.</sup> Issues regarding compensation of experts are discussed in § VI, infra.

Judges also used the order of appointment to define the role of the court-appointed expert in relation to the judicial role, distinguishing between the expert's duty to provide technical expertise and the judge's duty to decide the case.

The form of the expert's report should also be defined. By detailing the formalities of reporting, the court may prevent unnecessary confusion regarding ex parte communication between the expert and the court.<sup>59</sup>

In addition to defining the roles of the judge and expert, the court also must define the issues for the expert to consider. This may be as straightforward as directing a panel of physicians to determine a plaintiff's injuries, prognosis, and the treatment required.<sup>60</sup> In other cases, defining the technical issues for the expert may require an explanation of legal issues as well. For example, in a case dealing with conditions of confinement at a correctional facility, the court used the appointment of an expert to articulate the applicable legal standards.<sup>61</sup>

Defining the issues to be considered by the expert seems to serve multiple purposes. For the expert, a written definition will serve as an essential guide to the generally unfamiliar world of litigation and the role of the appointed expert. For the parties and counsel, the use of court-appointed experts is so rare that a clear definition of the issues and the process should enhance understanding and allay concerns. For the court itself, defining the issues may help clarify the roles of the court and expert. In one of the few cases in which a party contested an appointment, the court asked the parties to propose instructions to the expert. After reviewing them, the court formulated its own instructions, addressing issues raised by the parties' proposals.<sup>62</sup>

Finally, judges frequently use the order of appointment as a way to define the process of assembling information for the expert.<sup>63</sup> This process permitted easy assembly of a record of the basis for the expert's opinions. In other cases, the court established a way for the parties to convey information to the expert without the court's participation.

<sup>59.</sup> See discussion infra notes 64-71 and related text.

<sup>60.</sup> See, e.g., In re Swine Flu Immunization Prods. Liab. Litig., 495 F. Supp. 1185, 1186 (E.D. Okla 1980); see also In re Asbestos Litig. (S.D. Ohio Apr. 29, 1987) (order issuing instructions to court-appointed expert witnesses—"render an objective medical diagnosis of the presence or absence of asbestosis or other as bestos-related diseases").

<sup>61.</sup> Stickney v. List, 519 F. Supp. 617, 619 (D. Nev. 1981); see also United States v. Michigan, 680 F. Supp. 928, 983-84, 986-88 (W.D. Mich. 1987).

<sup>62.</sup> Students of the Cal. Sch. for the Blind v. Riles, Civ. No. S-80-473-MLS (E.D. Cal. Mar. 31, 1982) (order appointing expert witness). See also Leesona Corp. v. Varta Batteries, Inc., 522 F. Supp. 1304, 1311–12 & n.18 (S.D.N.Y. 1981) (parties asked to prepare a statement of the technical issues for inclusion in written instructions to the expert).

<sup>63.</sup> In one reported case, the court invited the parties to bring their own experts to participate in the conference at which the judge instructed the court-appointed expert. United States v. Articles . . . Provimi, 74 F.R.D. 126, 127 (D.N.J. 1977) (supplementing 425 F. Supp. 228 (D.N.J. 1977)). A joint meeting of the experts at that stage could initiate a process of assembling common information for all of the experts.

#### B. Ex Parte Communication

1. Communication between the judge and the appointed expert

Rule 706 does not explicitly address the issue of whether the judge and the appointed expert may communicate ex parte during the course of the litigation. Case law and canons of judicial ethics discourage off-the-record contacts between a judge and an expert witness. Reacting to ex parte communication between the district court and an expert, one appeals court ruled that "if any experts are . . . [appointed] to advise the district court on any further matters in this litigation, they shall prepare written reports, copies of which shall become part of the record and shall be made available to all parties or their attorneys."<sup>64</sup> Another appellate tribunal recommended that all communications with an expert be conducted in either an on-the-record conference in chambers or an on-the-record conference call.<sup>65</sup> The norm, as stated in the *Code of Conduct for United States Judges*, is that a judge should not consider "ex parte or other communications on the merits . . . of a pending or impending proceeding."<sup>66</sup> The scope of the term *ex parte* is not defined further. Whether this concept is applicable to court-appointed experts is unclear.

A broad prohibition of ex parte communications between a judge and a courtappointed expert would impede necessary communication when the expert is appointed to serve as a technical advisor to the court, <sup>67</sup> a role analogous to that of a judicial clerk. In such cases, either the parties consented to off-the-record discussions between the judge and the expert or the court relied on its broader inherent power to appoint the expert as a technical advisor. In either event, the very purpose of the appointment was to secure an expert who would "act as a sounding board for the judge—helping the jurist to educate himself in jargon and theory disclosed by the testimony and to think through the critical technical problems." <sup>68</sup> That educational function seems to contemplate ex parte communication, albeit with procedural safeguards.<sup>69</sup>

**Court-Appointed Experts** 

<sup>64.</sup> Bradley v. Miliken, 620 F.2d 1143, 1158 (6th Cir.), cert. denied, 449 U.S. 870 (1980).

<sup>65.</sup> United States v. Green, 544 F.2d 138, 146 n.16 (3d Cir. 1976), cert. denied sub nom. Tefsa v. United States, 430 U.S. 910 (1977); cf. Leesona Corp., 522 F. Supp. at 1312 & n.18.

<sup>66.</sup> Canon 3(A)(4) of the Code of Conduct for United States Judges provides that

<sup>[</sup>a] judge should accord to every person who is legally interested in a proceeding, or the person's lawyer, full right to be heard according to law, and, except as authorized by law, neither initiate nor consider ex parte communications on the merits, or procedures affecting the merits, of a pending or impending proceeding.

Judicial Conference of the U.S., Code of Conduct for United States Judges, in 2 Guide to Judiciary Policies and Procedures, Canon 3(A)(4), at I-9 (rev. Nov. 1993).

<sup>67.</sup> For illustrations of the contexts in which such discussions took place and for a description of some safeguards short of prohibition, see discussion at note 71 infra and related text.

<sup>68.</sup> Reilly v. United States, 863 F.2d 149, 158 (1st Cir. 1988).

<sup>69.</sup> *Id.* at 158, 159–60 (ground rules included advising parties if expert ranged into area not discussed in briefs; appellate court recommends inclusion of a comprehensive job description on the record and submission of an affidavit of the expert's compliance with the ground rules at the end of the appointment).

Our interviews revealed considerable ex parte communication between judges and experts as well as some confusion concerning the proper standard. More than half of the judges who responded to the question "Did you communicate directly with the expert outside of the presence of the parties?" answered "yes." <sup>70</sup> About half of those judges limited their ex parte discussion to procedural aspects of the expert's service—including matters of availability. Lengthy ex parte communicated extensively with . . . [the prospective expert] in chambers prior to the appointment to convince him to accept it."

The remaining judges communicated with the court-appointed experts on at least some occasions to elicit technical advice outside the presence of the parties. In most of these situations the very purpose of the appointment was to provide the judge with one-to-one technical advice. We did not systematically ask about consent, but some judges indicated that the parties expressly consented to the ex parte communications. In all other cases it appeared from the context of the interviews that the parties were generally aware of the arrangements and either expressly consented or failed to object.

Several judges devised procedures to subject their contact with a technical advisor to some of the checks and balances of the adversary system. <sup>71</sup> For example, one judge communicated ex parte with the expert, but made a record of the discussions and disclosed the exact contents to the parties. Another judge indicated that the parties' agreement to ex parte discussion was conditioned on his reporting the substance of such discussions to the parties. These procedures inform the parties of the content of the judge's information about a case and allow them an opportunity to clarify, rebut, or even reinforce the expert's statements.

#### 2. Communications between the parties and the expert

Rule 706 also fails to address the question of whether ex parte communication should be permitted between the expert and the parties.<sup>72</sup> Some judges apply the same rules to court-appointed experts that they would apply to themselves.<sup>73</sup> This would seem especially apt for cases in which the expert, as a technical advisor, is intimately involved in the decision-making process. Even in the

<sup>70.</sup> Two-thirds of the multiple users of the Rule 706 process reported ex parte communication with an expert in at least one case.

<sup>71.</sup> See, e.g., Jackson v. Fort Stanton Hosp. & Training Sch., 757 F. Supp. 1231, 1234 (D.N.M. 1990) (judge kept a record of the discussions with the appointed expert and made these available to the parties), *rev'd in part*, 964 F.2d 980 (10th Cir. 1992).

<sup>72.</sup> During the original consideration of the Federal Rules of Evidence, a committee from the American Bar Association suggested that a direct prohibition on ex parte communication by a party with a court-appointed expert should be added to Rule 706. While the suggested procedure was not adopted, Weinstein & Berger suggest that such a prohibition "may prove useful to the court and parties in using [the appointment] procedure." Weinstein's Evidence, *supra* note 2, ¶ 706[02], at 706–20 n.21.

<sup>73.</sup> See, e.g., Leesona Corp. v. Varta Batteries, Inc., 522 F. Supp. 1304, 1312 n.18 (S.D.N.Y 1981) (parties were not permitted to communicate directly with the court's expert; materials selected by the parties for the expert to use were transmitted through the court and entered in the court's docket).

absence of an explicit order, however, attorneys should be aware that "ex parte attempts to influence the expert are improper."<sup>74</sup>

We found that about half of the judges who responded permitted direct, separate communication between the expert and one or more parties. Often, the nature of the appointment and the role of the expert led naturally, if not inexorably, to that practice. The clearest example was the medical examination of a party by an expert to determine the extent of injuries. Normally such examinations are conducted in private (i.e., technically ex parte) with a copy of the report furnished to the parties and the court.<sup>75</sup> Adversarial participation would invade the privacy of the party and might compromise the expert's ability to obtain information on which to base a diagnosis.

#### C. Pretrial Reports and Depositions

Unless the parties agree otherwise, the court-appointed expert must advise the parties of any findings, submit to a deposition by any party, and respond to cross-examination of his or her testimony, if any, at trial. <sup>76</sup> Findings may be presented in a written report, by deposition, in testimony in open court, or through some combination of the above.<sup>77</sup>

We found that, except when used as a technical advisor, <sup>78</sup> the expert invariably reports findings to the parties. In several cases the parties met informally with the expert to discuss his or her report. Generally, the findings are in the form of a written report furnished to the court and the parties. We were told of two instances in which the expert reported orally to the parties, once by deposition, and once in a meeting in the judge's conference room. In the few cases where the expert was appointed immediately before or during trial, the expert

<sup>74.</sup> Weinstein's Evidence, supra note 2, ¶ 706[02], at 706–20 n.21. See also Model Code of Professional Responsibility DR 7-110(B), at 39 (1982) ("a lawyer shall not communicate . . . as to the merits of the cause with a judge or an official before whom the proceeding is pending . . ." (emphasis added)). Presumably, the expert is an "official" agent of the court. Cf. Model Rules of Professional Conduct Rule 3.5 (1983) ("A lawyer shall not: (a) seek to influence a judge . . . by means prohibited by law; (b) communicate ex parte with . . . [a judge] except as permitted by law . . .").

<sup>75.</sup> Cf. Fed. R. Civ. P. 35, which provides for a physical examination of a party and production of a report. Presumably the party who calls for the examination is not entitled to be present during it. The plain language of Rule 35 does not confer such a right. In any event, the practice under Rule 35 could serve as a guide regarding physical or mental examinations under Rule 706. The ABA exempted medical examinations from their proposed restriction on ex parte communication between a party and a court-appointed expert. Weinstein's Evidence, *supra* note 2, ¶ 706[02], at 706–20 n.21.

<sup>76.</sup> Fed. R. Evid. 706(a). See also Unique Concepts, Inc. v. Brown, 659 F. Supp. 1008, 1011 (S.D.N.Y. 1987), *later proceeding*, 735 F. Supp. 145 (S.D.N.Y. 1990), *aff'd*, 939 F.2d 1558 (Fed. Cir. 1991). Cf. Reilly v. United States, 863 F.2d 149, 159 (1st Cir. 1988) ("If . . . the advisor was not an evidentiary source, there was neither a right to cross-question him as to the economics of the situation nor a purpose in doing so."). Weinstein and Berger observe that the right of a party to depose the court-appointed expert in a criminal case "goes considerably further than any other rule or statute in authorizing depositions in a criminal case." Weinstein's Evidence, *supra* note 2, ¶ 706[02], at 706–21.

<sup>77.</sup> Leesona Corp. v. Varta Batteries, Inc., 522 F. Supp. 1304, 1312 (S.D.N.Y. 1981).

<sup>78.</sup> As noted above in the discussion of ex parte communication between the judge and the expert (see discussion *supra* notes 67–71 and related text), in several cases the expert reported directly to the judge without any report to the parties.

reported by way of testimony at the trial or hearing. One judge reported the practice of using the report of the expert as the equivalent of direct testimony at the trial.

Three judges, all of whom had appointed experts more than once, asked the expert for a preliminary report, then permitted the expert to modify this report after reviewing the reports of the parties' experts. The use of a preliminary report "serve[s] to give [the judge] an independent report" and allows "an opportunity to take into account the reports of other experts." Formal depositions are relatively infrequent, occurring in about one case in four.<sup>79</sup>

#### D. Presentation of Expert Opinion in Court

#### 1. Frequency and nature of testimony

Although Rule 706 seems to anticipate that court-appointed experts will testify at trial, our earlier review of reported decisions found that court-appointed experts can serve a range of nontestimonial functions during different stages of the litigation.<sup>80</sup>

Our interviews revealed more testimonial use of experts than suggested by published opinions. Roughly half of the cases discussed by judges involved court-appointed experts' testimony presented in court, usually at a trial, less frequently at a pretrial evidentiary hearing. On the other hand, settlement was less frequent than commentary on Rule 706 led us to expect. Relatively few (approximately one in five) of the testimonial uses of court-appointed experts occurred in jury trials.

#### 2. Advising jury of court-appointed status

One of the controversial aspects of Rule 706 is that it explicitly grants the trial judge discretion whether to inform the jury that the expert was appointed by the court.<sup>81</sup> Some commentators have opposed informing the jury of the expert's status, fearing that knowledge that the court appointed the expert will undermine the adversarial system and dominate the jury decision-making process.<sup>82</sup> The trial court retains discretion, however, to decline to place a judicial impri-

<sup>79.</sup> See Renaud v. Martin Marietta Corp., 972 F.2d 304, 308 n.8 (10th Cir. 1992) (depositions and crossexamination found to be inappropriate where expert appointed under authority of Rule 706 in fact functioned as a technical advisor).

<sup>80.</sup> Although published opinions reveal instances of court-appointed experts presenting testimony at trial, references to nontestimonial functions were more frequent. Thomas E. Willging, Court-Appointed Experts 18–23 (Federal Judicial Center 1986).

<sup>81.</sup> Fed. R. Evid. 706(c).

<sup>82.</sup> See, e.g., Nicholas J. Bua, Experts—Some Comments Relating to Discovery and Testimony Under New Federal Rules of Evidence, 21 Trial Law. Guide 1 (1977); Weinstein's Evidence, supra note 2, ¶ 706[02], at 706–26.

matur on a witness if concerned that the jury will give undue weight to a courtappointed expert's testimony.<sup>83</sup>

Only seven jury trials were identified from the interviews in which the courtappointed expert offered testimony in court. In all but one of these cases, the judge or the party calling the witness informed the jury of the expert's court-appointed status. In the only exception, it appears that neither party was sufficiently advantaged by the report to want to underscore its source. At the other extreme, one judge reported that the advantaged party called the expert "with great flourish," had the order appointing the expert read to the jury, and asked a series of questions emphasizing neutrality, the source of the appointment, and the method of payment.

#### 3. Effect of the testimony of the appointed expert

Our interviews revealed that juries and judges alike tend to decide cases consistent with the advice and testimony of court-appointed experts. We asked, "Was the disputed issue resolved in a manner consistent with the advice or testimony of the 706 expert?" Of fifty-eight responses, only two indicated that the result was not consistent with the guidance given by the expert. Both of those cases involved bench trials in which the judge pursued a legal analysis that was independent of the technical issues. In one, the judge decided about an appropriate remedy but found it useful to have the expert's analysis of the strengths and weaknesses of an alternative proposal. In the other, the judge ruled that the plaintiff had not met its legal burden of proof. Two of the fifty-eight judges indicated that the expert did not give any advice, but simply had explained the technical issues and the testimony of the parties' experts. Three judges indicated that the information provided by the expert was used in conjunction with other information to shape a resolution of the issue.

In the remaining fifty-one cases, including seven jury trials, the outcome was consistent with the expert's advice or testimony. Note that we asked only if the outcome was consistent with the advice of the appointed expert. Twenty-one of the judges who indicated outcomes consistent with the appointed experts' testimony also volunteered the information that the experts' opinions were not the exclusive, or even the most important, factor in determining the outcome of their cases. Seven of the twenty-one cases settled following the submission of the expert's report or testimony, and the judges believed that the resolution was consistent with the report of the appointed expert. In the remaining fourteen cases the judge indicated that the report or testimony of the appointed expert provided

Court-Appointed Experts

<sup>83.</sup> Weinstein's Evidence, supra note 2, ¶ 706[02], at 706–27. See also Tahirih V. Lee, Court-Appointed Experts and Judicial Reluctance: A Proposal to Amend Rule 706 of the Federal Rules of Evidence, 6 Yale L. & Pol'y Rev. 480, 500 (1988) (suggesting that Rule 706 be amended to include a duty of the court to caution the jury against excessive reliance on the testimony of the expert appointed by the court).

a context for understanding and evaluating other evidence presented by the parties.  $^{\rm 84}$ 

If the case involved testimony by an appointed expert at a jury trial, we asked, "Did the testimony of the court-appointed expert appear to overwhelm the expert testimony offered by the parties?" In a dozen jury cases, <sup>85</sup> it appears that the testimony of court-appointed experts dominated the proceedings. In general, the testimony of the court's expert affirmed the testimony of one of the parties' experts, thereby overcoming contrary evidence.

When viewed in the light of the circumstances leading to an appointment, perhaps it should come as no surprise that the outcome of a case is greatly influenced by the testimony of an appointed expert. Since the absence of an impartial factual basis to decide the case was a prerequisite to the appointment, it follows that the testimony of the appointed expert is likely to be influential. The primary reasons for appointment of an expert were either a failure of the parties to offer credible expert testimony or an actual or anticipated conflict in the testimony of the parties' experts that defied resolution through traditional means. Regarding the failure of advocacy cases, we reported (in section II *supra*) that in eighteen of the thirty-six cases involving judges who had used Rule 706 only once, the judges indicated that there was a failure by one or both parties to present credible expert testimony. In many of these cases there was no credible evidence at all on the technical issue. Given a void of evidence on a critical issue, the court-appointed expert's testimony would necessarily be influential.

Similarly, in cases with an unresolvable conflict among the parties' experts, the equipoise in the evidence prior to appointment renders the court-appointed expert likely to tip the scale to one side or another. Any other result would raise significant questions about whether there had been a need for an outside expert. These reasons tend to explain and qualify our findings. Nevertheless, the central finding is clear: Judges who appointed an expert indicated that the final outcome on the disputed issue was almost always consistent with the testimony of the appointed expert.

In summary, the concerns of judges and commentators that court-appointed experts will exert a strong influence on the outcome of litigation seem to be well founded. Whether such influence is appropriate is a different question. In almost all cases, the jury was aware of the expert's court-appointed status and seemed influenced by the expert's apparent neutrality. Some judges think that it is important for the jury to know the status as an aid in assessing credibility. Some judges who presided over jury trials, however, expressed misgivings about permitting revelation of court-appointed status because it seemed to have led to automatic reliance on the expert by the jury. Potential controls, such as impos-

<sup>84.</sup> A more detailed analysis of these cases appears in Cecil & Willging, supra note 4, at 52-56.

<sup>85.</sup> In addition to the seven cases elicited in our discussions with judges who had appointed an expert a single time, five additional cases were uncovered when we asked judges who were multiple users if they had ever presided at a jury trial at which a court-appointed expert testified.

ing limited restrictions on lawyers and camouflaging the source of a witness, remain untested.

Judges were, of course, always aware of the experts' status. In their instructions to experts and in the course of work with them, judges frequently showed a conscious effort to maintain control of the legal and policy analysis and decision making, while limiting technical information and advice to a subsidiary, instrumental role. Nevertheless, our interviews reveal a high degree of consistency between the outcome of litigation and the testimony and advice of court-appointed experts.

### VI. Compensation of Court-Appointed Experts

Payment of court-appointed experts presents an awkward problem for judges. Although judges appoint the experts, judges usually must turn to the parties for compensation. Furthermore, because an expert may serve long before the case is resolved, a means must be found to provide prompt payment while retaining the option of reallocating the expenses among the parties based on the resolution of the issues. Parties may resist compensating experts they did not retain and who offer testimony that is damaging to their interests. If the parties balk at payment, the judge must either enforce payment by means of a formal order and a hearing, thereby disrupting the litigation and increasing the level of acrimony between the parties, or postpone payment, thereby leaving the expert uncompensated for an indefinite period.

Interviews with judges suggest that such problems in providing compensation can thwart the appointment of an expert. Judges expressed concerns regarding payment when describing how the experts were compensated and at a number of other points in the interviews. When asked why more judges do not use courtappointed experts, fourteen judges focused on the difficulties in providing compensation. Reliance on the parties for payment of fees was cited by several judges as the principal reason for restricting appointment of experts to cases in which the parties consent to an appointment. As one judge who had never appointed an expert stated, the lawyers find the process "hard to justify to their clients when the client is paying for expert testimony already," particularly when the courtappointed expert may "hurt the client's case, making the client even angrier." When asked what changes in the rule would make court-appointed experts more useful, the most common suggestion from judges was for clarification of the means of compensating the expert.<sup>86</sup> While appointment of an expert poses many practical problems, providing a mechanism ensuring the prompt compensation for appointed experts appears to be one of the more serious ones.

Rule 706, supplemented by statutory authority and case law, grants judges broad discretion in allocating the costs of appointed experts among the parties but allows little opportunity to turn elsewhere for compensation. The following subsections address four different circumstances that affect the manner of com-

<sup>86.</sup> This suggestion was mentioned by ten of the nineteen judges who suggested changes in the rule. See also Weinstein's Evidence, supra note 2,  $\P$  706[03], at 706–27 to –29.

pensation: special instances of land condemnation actions and criminal cases in which the rule permits the expert to be compensated from public funds; matters involving general civil litigation (in which the court must rely on the parties for compensation); general civil litigation when one of the defendants is indigent; and occasions when the court wishes to employ a technical advisor as opposed to a testifying expert.

#### A. Statutory Basis for Compensation from Public Funds

In two circumstances—land condemnation cases and criminal cases—Rule 706 and related statutes authorize payment of the appointed expert from public funds. In land condemnation cases, all costs, including fees for an appointed expert to testify regarding compensation for the taking of property, are assessed against the government, not the property owner.<sup>87</sup> In the few instances we encountered in which an expert was appointed to assist in a condemnation proceeding, the fee was paid by the Department of Justice with little difficulty.

Obtaining payment for experts in criminal cases follows a similar process. Again, the rule and related statutes<sup>88</sup> permit payment of the experts' fees from public funds. The Criminal Justice Act authorizes payment of experts' expenses when such assistance is needed for effective representation of indigent individuals in federal criminal proceedings.<sup>89</sup> In criminal cases in which the United States is a party, the Comptroller General has ruled that the source of payment is to be the Department of Justice, not the Administrative Office of the U.S. Courts.<sup>90</sup> Four judges revealed that they had appointed experts to aid in assessing the physical or mental condition of a defendant; three of these judges indicated no difficulty in obtaining payment, while one indicated some initial reluctance by the Department of Justice followed by prompt payment.

#### B. Payment of Fees by Parties

In the most common litigation context, the court appoints an expert with the expectation that the expert will offer testimony at a trial or hearing or produce a pretrial report that will facilitate settlement. Except for criminal and land condemnation cases, under Rule 706(b) "the compensation shall be paid by the parties in such proportion and at such time as the court directs, and thereafter charged in like manner as other costs."<sup>91</sup> The flexibility of the rule permits the

<sup>87.</sup> Fed. R. Evid. 706(b); Fed. R. Civ. P. 71A(1) advisory committee's note.

<sup>88.</sup> See, e.g., Fed. R. Evid. 706(b); 18 U.S.C. § 3006A(e) (1988).

<sup>89. 18</sup> U.Š.C. § 3006A(e) (1988). See generally John F. Decker, Expert Services in the Defense of Criminal Cases: The Constitutional and Statutory Rights of Indigents, 51 U. Cin. L. Rev. 574 (1982).

<sup>90.</sup> In re Payment of Court-Appointed Expert Witness, 59 Comp. Gen. 313 (1980) (expert appraisal of property to be forfeited in a criminal case; same rule applies to land condemnation proceedings).

<sup>91.</sup> By statute, payments to court-appointed experts are taxable as costs to the losing party. 28 U.S.C. § 1920(6) (1988). Cf. Miller v. Cudahy, 656 F. Supp. 316, 338–39 (D. Kan. 1987), aff'd in part and rev'd in part, 858 F.2d 1449 (10th Cir. 1988), cert. denied, 492 U.S. 926 (1989) (costs of what the district court had incor-

court to rely on the parties to compensate the expert when service is rendered rather than waiting until the conclusion of the litigation. The court may order the advance payment of a reasonable fee<sup>92</sup> for a court-appointed expert and defer the final decision on costs assessment until the outcome of the litigation is known.<sup>93</sup> The court may allocate the fees among the parties as it finds appropriate both as an interim measure and in the final award. One court has held that the "plain language of Rule 706(b) . . . permits a district court to order one party or both to advance fees and expenses for experts that it appoints."<sup>94</sup> In brief, the court has discretion to order a single party to prepay the full cost of the appointment.<sup>95</sup>

Rule 706(b) also provides that, at the conclusion of the litigation, the expert's "compensation shall be . . . charged in like manner as other costs." This means that "costs . . . shall be allowed as of course to the prevailing party unless the court otherwise directs." <sup>96</sup> Courts sometimes have apportioned fees among the parties, in some cases simply splitting the costs equally <sup>97</sup> and in other cases basing the apportionment on the outcome of the litigation. <sup>98</sup> Of course, if the parties settle short of a resolution of the merits of the dispute, allocation of the expert's fees may be part of such a settlement agreement.

Most judges require the parties to split the expert's fee, with the party prevailing at trial being reimbursed for its portion. Often the parties arrive at this arrangement without judicial involvement. In other instances, especially those in which the parties are reluctant to endorse the court's appointment of an expert, the judge may issue an order that requires the parties to pay a fixed amount to cover the expert's fees. In several cases in which an appointed expert served for a lengthy period, the court required the parties to make periodic payments into an

rectly characterized as a court-appointed expert could not be taxed, beyond the statutory allowance, to the party ordered by the court to use the expert).

92. Rule 706(b) states that court-appointed experts "are entitled to reasonable compensation in whatever sum the court may allow."

93. See United States v. Articles . . . Provimi, 425 F. Supp. 228, 231 (D.N.J. 1977) (assessing "one-half of the cost of the expert's services . . . with further decision on the expert's cost to abide the event"), *supplemental op.*, 74 F.R.D. 126 (D.N.J. 1977). *Cf.* Baker Indus. v. Cerberus, Ltd., 570 F. Supp. 1237, 1248 (D.N.J. 1983) (85% of costs were assessed against defendant and 15% against plaintiff who prevailed on almost all issues), *aff'd*, 764 F.2d 204 (3d Cir. 1985).

94. United States Marshals Serv. v. Means, 741 F.2d 1053, 1058 (8th Cir. 1984) (en banc), *cert. denied*, 492 U.S. 910 (1989); *see also* Webster v. Sowders, 846 F.2d 1032, 1039 (6th Cir. 1988) (allocation of Rule 706 costs, at least temporarily, to the party against whom a preliminary injunction is granted is permitted when the parties obtaining the relief were impecunious).

95. McKinney v. Anderson, 924 F.2d 1500, 1510 (9th Cir. 1991), vacated on other grounds sub nom. Helling v. McKinney, 112 S. Ct. 291 (1991).

96. Fed. R. Civ. P. 54(d)(1).

97. See United States v. Michigan, 680 F. Supp. 928, 956–57 (W.D. Mich. 1987); Unique Concepts, Inc. v. Brown, 659 F. Supp. 1008, 1011 (S.D.N.Y. 1987), *later proceeding*, 735 F. Supp. 145 (S.D.N.Y. 1990), *affd*, 939 F.2d 1558 (Fed. Cir. 1991).

98. See, e.g., In re Fleshman, 82 B.R. 994, 996 (Bankr. W.D. Mo. 1987) (court stated that parties would have to pay for an appraiser's services "according to a ratio determined by comparing the final finding on value to their initial contention"); cf. Baker Indus. v. Cerberus, Ltd., 570 F. Supp. 1237, 1248 (D.N.J. 1983) (assessment of 85% of special master costs against defendant and 15% against plaintiff who prevailed on almost all issues was approved), aff'd, 764 F.2d 204 (3d Cir. 1985).

account from which the court then compensated the expert. Judicial participation in the payment process varied greatly. Some judges permitted the expert to bill the parties directly; other judges had the expert submit the bill directly to the judge with copies to the parties and required the parties to pay a proportional amount unless they objected to the bill.

Obtaining payment for the expert from the parties proved to be troublesome in several instances. As one judge noted, "It [is] a bitter pill for the disadvantaged party to have to pay for harmful testimony." Occasionally one of the parties would simply refuse to pay. Then the judge generally held a hearing and, when necessary, demanded that the payment be made. In several instances the court had to impose injunctive relief as a means of ensuring that the payment was made. In discussing these instances the judges repeatedly indicated their great uneasiness at the prospect of incurring the services of an expert and then being unable to pay for those services in a timely manner. Concerns about securing payment moved several judges to employ a court-appointed expert only with the consent of the parties.

### C. Compensation of Appointed Experts When One Party Is Indigent

As a practical matter, the indigent status of one or more of the parties restricts the ability of a court to allocate the expense of the expert among the parties. The court has the authority to order the nonindigent party to advance the entire cost of the expert.<sup>99</sup> However, the judges indicated a great reluctance to employ such experts when the expense cannot be shared. We asked a number of the judges, including those who had not appointed experts, what they would do if one of the parties was indigent. Often they responded that they would proceed with the evidence at hand and decide the case to the best of their abilities, since forcing one party to bear the full expense of the court-appointed expert was a step they were unwilling to take.

We found six instances in which a judge appointed an expert when one or more of the parties were indigent. In each case, the indigent status of the party limited the extent to which the party could present expert testimony, limited the effectiveness of the adversarial examination of the opponent's contentions, and raised concerns that the judge sought to address by appointment of an expert. Three of these cases involved prisoners proceeding pro se and challenging the conditions of their incarceration. In each circumstance there was reason to believe that there was merit in the prisoner's complaint, <sup>100</sup> and the court appointed

<sup>99.</sup> United States Marshals Serv. v. Means, 741 F.2d 1053, 1058 (8th Cir. 1984) (en banc), cert. denied, 492 U.S. 910 (1989).

<sup>100.</sup> In each of these cases the fact that the defendant was the state and that some preliminary investigation revealed the complaint to be of merit appeared to weigh heavily in the court's decision to appoint the expert and impose the costs on the defendant.

an expert with the expectation that the expert would be compensated by the state.

The most difficult circumstance identified concerned the appointment of an expert in a suit by an indigent family contending that exposure to toxic chemicals caused a number of physical injuries as well as emotional harm. The indigent status of the plaintiffs limited the amount of expert testimony they offered. The judge doubted the integrity of the defendants' expert testimony and appointed an expert to testify about whether the chemicals had carcinogenic properties. The judge indicated that the presence of children as plaintiffs in the case caused him to be especially reluctant to decide the case without additional expert testimony, since the children as well as the parents would be barred by an adverse judgment from raising future claims. In this case, much of the difficulty was avoided when the defendant agreed to pay the expense of the court-appointed expert.

These few instances suggest the difficulties that may be encountered when added expert assistance is required and one or more of the parties are indigent. Although Rule 706 supports the imposition of the expenses on the nonindigent party, <sup>101</sup> judges seem willing to impose one-sided expenses only when the indigent party's claim shows some merit, or when the nonindigent party has agreed to assume the cost of the expert. The difficulties in providing payment in such circumstances suggest that the few instances recounted above may be far overshadowed by instances in which no appointment was made because of an inability to find a means of fairly compensating an appointed expert. <sup>102</sup>

#### D. Compensation of Technical Advisors

Finally, it also proves difficult to compensate an expert appointed as a "technical advisor" who may confer in private with the judge and who is not expected to offer testimony. Through our interviews we identified several instances in which a Rule 706 expert advised the court on the interpretation of evidence submitted by the parties rather than present evidence as a witness. Payment in these circumstances was simplified by the fact that the parties apparently consented to the appointment and agreed to share the cost of the expert. However, in a limited number of circumstances the Administrative Office of the U.S. Courts has been willing to assume the costs of such services. The Administrative Office has denied requests for such services where appointment of such an expert would be appropriate under Rule 706 of the Federal Rules of Evidence or under Rule 53 of the Federal Rules of Civil Procedure.

Court-Appointed Experts

<sup>101.</sup> See supra note 95 and related text.

<sup>102.</sup> David Medine, The Constitutional Right to Expert Assistance for Indigents in Civil Cases, 41 Hastings L.J. 281, 298 (1990) ("court appointment of expert witnesses [under Fed. R. Evid 706] does not provide adequate assistance to indigent civil litigants").

In Reilly v. United States,<sup>103</sup> the U.S. Court of Appeals for the First Circuit addressed the district court's use of a technical advisor and payment of the technical advisor's fees and expenses by the Administrative Office. Citing statutory authority that permits the judiciary to employ consultants and experts, <sup>104</sup> the district judge petitioned the Director of the Administrative Office for permission to appoint and compensate a technical advisor.<sup>105</sup> The judge expressly disavowed appointment under authority of Rule 706 because he wanted the expert to advise him in chambers regarding interpretation of evidence presented at trial, and not to present additional evidence or testimony. Permission to appoint the technical expert was granted and the expert was compensated from the funds appropriated to the judiciary. We are aware of only one other instance in which the Administrative Office has agreed to pay the expenses of a technical advisor. <sup>106</sup> In both cases the payment was at the behest of a plaintiff who suffered childhood injuries. In one case, the proceedings were nonadversarial; in the other, the presentation on a highly technical issue was one-sided.<sup>107</sup> It seems that this form of payment is available only in very unusual circumstances in which the expert is to provide technical assistance to the judge rather than to present evidence to the court, and in which the Director of the Administrative Office has approved such an expenditure prior to the appointment.

107. In the words of the court of appeals, the case "involved esoterica: complex economic theories, convoluted by their nature, fraught with puzzlement in their application." *Reilly*, 863 F.2d at 157.

<sup>103. 682</sup> F. Supp. 150 (D. R.I.), aff'd in part and rev'd in part, 863 F.2d 149 (1st Cir. 1988).

<sup>104. 5</sup> U.S.C. § 3109 (1988 & Supp. 1993) and 28 U.S.C. § 602(c) (1988).

<sup>105.</sup> Reilly, 682 F. Supp. at 152–55. The court also secured the permission of the Chief Judge of the First Circuit Court of Appeals and the Circuit Council. The court of appeals did not address which of these permissions would be necessary in order to appoint a technical expert. *Reilly*, 863 F.2d at 154 n.2.

<sup>106.</sup> Letter from L. Ralph Mecham, Director, Administrative Office of the U.S. Courts, to Gary J. Golkiewicz, Chief Special Master, U.S. Claims Court (Oct. 10, 1989) (on file with author) (approving a request to hire an economic expert to assist a special master in a case brought under the National Vaccine Injury Compensation Program). No similar authority exists for appointment of a technical advisor to serve the court of appeals. See E. I. du Pont de Nemours & Co. v. Collins, 432 U.S. 46, 57 (1977).

## VII. Procedures for the Effective Use of Court-**Appointed Experts**

Effective use of court-appointed experts must be grounded in a pretrial procedure that enables a judge to consider the possibility of an appointment in a timely manner and to anticipate problems in expert testimony.<sup>108</sup> Such a pretrial process is discussed in the paper on case management of this manual and is summarized here to provide a context for suggested improvements in the use of court-appointed experts.

The pretrial procedure described in the paper on case management will be useful in a wide variety of cases involving expert testimony—this procedure need not culminate in the appointment of an expert by the court. It is intended to permit recognition of difficulties at an early point in the litigation and allow the judge to narrow disputed issues by encouraging the parties and experts to specify their assumptions and designate areas of agreement and disagreement. If questions of admissibility are raised, the suggested procedure would enable the judge to conduct in limine hearings to resolve such questions and to enter summary judgment where disputed issues are not supported by admissible evidence.

In those extraordinary cases in which the court requires the assistance of an appointed expert, the additional procedures specified in this section will enable an appointment early enough to avoid delay in the litigation and difficulties in securing the effective services of an expert.

#### A. Clarification of Disputed Issues Arising from Complex Evidence

#### Early identification of disputed expert testimony 1.

All but the simplest techniques for addressing problems arising from difficult expert testimony require early awareness of disputed scientific and technical issues. One of the major impediments to the appointment of experts, according to our survey, is that judges are often unaware of a trial's difficulty until it is too late to make an appointment.<sup>109</sup> Even if a judge decides to invoke none of the extraordinary procedures intended to address problems with expert testimony (e.g.,

<sup>108.</sup> Such a pretrial procedure is described in William W Schwarzer, Management of Expert Evidence § II, in this manual. See also Margaret A. Berger, Procedural and Evidentiary Mechanisms for Dealing with Experts in Toxic Tort Litigation: A Critique and Proposal (1991); and Cecil & Willging, supra note 4, chap. 7. 109. See discussion supra notes 41-46 and related text.

appointment of an expert or special master), knowledge of especially difficult disputed issues prior to trial will enable a more informed consideration of such issues and related motions when they arise. If extraordinary procedures are to be invoked, awareness of looming difficulties may be critical if the full range of pre-trial devices are to be considered.

Recent amendments to Rule 26(a)(2) of the Federal Rules of Civil Procedure increase the information to be disclosed by experts that are to testify at trial, thereby easing early identification of disputed issues. Not less than ninety days before the trial, each party must disclose written reports prepared by the testifying witnesses that include, among other things, "a complete statement of all opinions to be expressed and the basis and reasons therefor; [and] the data or other information considered by the witness in forming the opinions." <sup>110</sup> Failure to make such disclosures will bar testimony by the expert at trial. <sup>111</sup> The *Manual for Complex Litigation* also encourages early identification of difficult or complex litigation and early intervention by the judge to ensure the efficient conduct of the litigation. <sup>112</sup>

#### 2. Attempts to narrow disputes

Rule 16 of the Federal Rules of Civil Procedure encourages efforts to narrow disputes between parties before trial, a mandate that can extend to disputes between parties' experts. One subject appropriate for discussion at the pretrial conference is "the possibility of obtaining admissions of fact and of documents which will avoid unnecessary proof . . . ."<sup>113</sup> Efforts to narrow disputes among experts may be especially useful where identification of disputed issues suggests that the experts' testimony will be in direct and complete opposition. Interviews with judges revealed that early indications of complete and thorough disagreement between experts often foreshadowed greater difficulties at trial.

A variety of devices can be used to explore the differences among experts, determine the extent of their disagreement, and clarify issues that underlie the dispute. Identifying the differences in assumptions that drive the more general disagreements will permit the trier of fact to focus on the assumptions rather than attempt to sort through the consequences of such disagreements. Some judges approach this task by asking experts to stipulate to those issues on which they agree and disagree, much like the factual stipulations that parties are often asked to provide.<sup>114</sup> Or the parties may be asked to submit a joint report, setting forth areas of agreement and disagreement. Some judges present the parties with a list of issues that they should respond to in preparing such a report. The reference guides in this manual, when supplemented by the parties, should offer an effec-

110. Fed. R. Civ. P. 26(a)(2)(B).
111. Fed. R. Civ. P. 37(c)(1).
112. MCL 3d, *supra* note 40, § 20.1.
113. Fed. R. Civ. P. 16(c)(3).
114. Fed. R. Civ. P. 36(a).

tive means for structuring consideration of such issues in these particular areas of science. When faced with especially demanding expert testimony, some judges convene a joint conference with counsel and the key experts and engage in a formal or informal colloquy concerning the experts' differences.<sup>115</sup>

#### 3. Screening of expert testimony

Identifying and narrowing disputed issues may lead to doubts concerning the admissibility of some of the proffered expert testimony. Questions may arise concerning the qualifications of those likely to be called as experts, or the validity of the information on which the experts base their testimony.<sup>116</sup> As part of the "gatekeeping" role recognized by the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,<sup>117</sup> the judge may wish to conduct a separate pretrial hearing to determine the admissibility of proposed expert testimony. Such a hearing may dispose of questionable testimony, thereby providing the parties with a better understanding of the evidence to be presented at trial. If the court finds that there is no admissible evidence to support essential elements of a claim, the court may dispose of the action by summary judgment.<sup>118</sup>

#### B. Appointment of an Expert

When a pretrial procedure based on the above elements fails to reveal information necessary to permit a reasoned resolution of the disputed issues, a judge may wish to consider appointing an expert. Our interviews suggested that such cases will be infrequent and will be characterized by (1) evidence that is particularly difficult to comprehend, (2) credible experts who find little basis for agreement, and (3) a profound failure of the adversarial system to provide the information necessary to sort through the conflicting claims and interpretations. Judges who had appointed experts emphasized the extraordinary nature of such a procedure and showed no willingness to abandon the adversarial process before it had failed to provide the information necessary to understand the issues and resolve the dispute.

Cases involving unrepresented or poorly represented parties may also merit appointment of an expert, although such cases are rare. When one or more of the parties are unable to or choose not to present expert testimony, a court may be uneasy resolving the issue on the basis of expert testimony provided by a sin-

<sup>115.</sup> Jack B. Weinstein, Role of Expert Testimony and Novel Scientific Evidence in Proof of Causation, Address at ABA Annual Meeting, Panel Discussion on Managing Mass Torts 22 (Aug. 9, 1987) (on file with authors) (describing an occasional practice of swearing in all the experts, seating them at a table with counsel, and engaging them in recorded colloquy under court direction). Other techniques for clarifying and narrowing issues are found in MCL 3d, *supra* note 40, § 21.33.

<sup>116.</sup> These issues are addressed in Margaret A. Berger, Evidentiary Framework §§ I, III, in this manual.

<sup>117. 113</sup> S. Ct. 2786, 2795 & n.7 (1993).

<sup>118.</sup> Celotex Corp. v. Catrett, 477 U.S. 317 (1986); In re "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223, 1239 (E.D.N.Y. 1985), aff'd on other grounds, 818 F.2d 187 (2d Cir. 1987), cert. denied sub nom. Lombardi v. Dow Chem. Co., 487 U.S. 1234 (1988).

gle party. If the court doubts the credibility or competence of the testifying experts, it may have to choose between appointing an expert and proceeding without competent and credible testimony on a critical issue. Several judges, in describing the issues that caused them to consider an appointment, mentioned the interests of minors or a public interest that was not adequately represented. In such cases the importance of reaching a correct resolution of disputed evidentiary issues may be especially great, and appointing an expert may be the most practical means of obtaining information.

The pretrial procedure outlined above and described in greater detail elsewhere in this manual should ensure that every effort has been made to obtain the necessary information short of appointing an expert. Where appointment of an expert appears to be the only means of obtaining necessary information, an effective pretrial procedure also provides an early indication of the problem, permitting the appointment to be undertaken in a timely manner without disrupting or postponing the anticipated trial. The proposed procedure also will develop material that will aid in instruction of the appointed expert. While we do not advocate appointment of an expert to encourage settlement, early awareness by the parties that such an appointment is being considered will permit them to engage in settlement negotiations with an awareness of that prospect.

Appointing an expert increases the burden on the judge, increases the expense to the parties, and raises unique problems concerning the presentation of evidence. These added costs will be worth enduring only if the information provided by the expert is critical to the resolution of the disputed issues. An effective pretrial procedure will identify cases that can be resolved in an expeditious manner without appointing an expert, as well as cases that require such assistance.

#### 1. Initiation of the appointment

Our interviews suggest that the appointment process will have to be initiated by the judge; rarely do the parties raise the idea of the court appointing an expert. Again, an effective pretrial procedure is intended to inform the judge of the nature of the underlying evidentiary disputes so that the judge is less reliant on the parties to inform the court of such disputes. The possibility of appointing an expert may be raised at pretrial conferences.<sup>119</sup> The court can initiate the appointment process on its own by entering an order to show cause why an expert witness or witnesses should not be appointed.<sup>120</sup>

<sup>119.</sup> Although Rule 16 does not specifically address court-appointed experts as a topic to be considered at a pretrial conference, the rule does recognize that it may be necessary to inquire into "the need for adopting special procedures for managing potentially difficult or protracted actions that may involve complex issues, multiple parties, difficult legal questions, or unusual proof problems." Fed. R. Civ. P. 16(c)(12).

<sup>120.</sup> Fed. R. Evid. 706(a). See also In re Joint E. & S. Dists. Asbestos Litig., 830 F. Supp. 686, 694 (E.D.N.Y. 1993) (parties are entitled to be notified of the court's intention to use an appointed expert and be given an opportunity to review the expert's qualifications and work in advance).

In responding to the order, parties should address a number of issues that may prove troublesome as the appointment process proceeds. Parties should be asked to nominate candidates for the appointment and give guidance concerning characteristics of suitable candidates. Those judges who encouraged both parties to create a list of candidates and permitted the parties to strike nominees from each other's list found this to be a useful method for increasing party involvement and developing a list of acceptable candidates.

Greater party involvement in identifying suitable candidates diminishes the judge's reliance on friends and colleagues for recommendations. When parties fail to recommend a suitable candidate, the judge may find it difficult to identify a candidate who is both knowledgeable in the relevant specialties and disinterested with respect to the outcome of the litigation. Academic departments and professional organizations may be a source of such expertise.

Compensation of the expert also should be discussed with the parties during initial communications concerning the appointment. Unless the expert is to testify in a criminal case or a land condemnation case, the judge should inform the parties that they must compensate the appointed expert for his or her services. Typically, each party pays half of the expense, with the prevailing party being reimbursed by the losing party at the conclusion of the litigation. Raising this issue at the outset will indicate that the court seriously intends to pursue an appointment and may help avoid subsequent objections to compensation. If difficulty in securing compensation is anticipated, the parties may be ordered to contribute a portion of the expected expense to an escrow account prior to the selection of the expert. Objections to payment should be less likely to impede the work of the expert once the appointment is made.

Finally, the court should make clear in its initial communications the anticipated procedure for interaction with the expert. The court should describe the assistance sought and the anticipated manner of interaction. If ex parte communication between the court and the expert is expected, the court should outline the specific nature of such communications, the extent to which the parties will be informed of the content of such communications, and the parties' opportunities to respond. Each of these issues is discussed in greater detail below. This initial communication may be the best opportunity to raise such considerations, entertain objections, and inform the parties of the court's expectations of the practices to be followed regarding the appointed expert.

#### 2. Communicating with the appointed expert

Conversations with judges revealed that communications with experts is one of the most troubling areas when dealing with court-appointed experts. Several judges mentioned the need for guidance regarding ex parte communications with experts. Complete avoidance of ex parte communication seems impractical in light of the judge's obligation to contact the expert, explain the general nature of the task, and determine the expert's willingness to undertake the assignment. While an initial letter inviting participation may be drafted with the assistance of the parties, there are likely to be telephone inquiries and other incidental communications (e.g., concerning time of hearing, details of compensation) in which full participation by the parties is unnecessary.

Once the expert has agreed to serve and seeks more specific information regarding the nature of the task, concerns over communications between the judge and experts outside the presence of the parties become more acute. Participation of the parties in the instruction of the expert offers an early opportunity to ease such concerns and ensure that the parties are fully aware of the services being sought of the expert. Since appointment of an expert is a rare event, the parties and the expert are likely to require clear guidance regarding the expectations of the court.

A common practice is to instruct the expert at a conference with the parties present, then formalize the instructions with a written order filed with the clerk. This practice permits easy interaction with the expert at the initial conference, ensures that the parties and the expert understand the nature of the task, and avoids misunderstanding and disagreements over the initial instructions. The instructions themselves can be based on the materials prepared by the parties as part of the pretrial process, which should set forth areas of disagreement and confusion. A written order also will help the expert focus his or her inquiry and will serve as a reminder of the limitations of the expert's role in relation to the judge's.

If an appointed expert has questions regarding his or her duties, the parties should be informed of the nature of the inquiry. <sup>121</sup> In most cases this should pose no difficulty. A written request for clarification from the expert and a written response by the court, with copies to all interested parties, will permit parties to remain informed of the proceedings and offer objections or clarifications to the response. If the judge and the expert expect to confer in person, several options are available. Representatives of the parties can be invited to attend the conference or, if this proves impractical, a record of the discussion can be forwarded to the parties. In any event, we believe that parties should be informed of the nature of those communications. This will permit a party to challenge the substance of the expert's advice or object to inquiries and information that exceed the expert's agreed-upon duties.

The "technical advisor" who provides a judge with instruction and advice outside the presence of the parties poses a more difficult problem.<sup>122</sup> While the need for such assistance should be diminished by the pretrial procedure out-

<sup>121.</sup> There may be questions concerning nonsubstantive issues, such as the timing of a report or hearing, or conditions of compensation, that do not require the participation of the parties.

<sup>122.</sup> Although such an appointment does not require the authority of Rule 706, several of the judges invoked this rule and obtained consent of the parties in retaining a technical advisor.

lined above, our interviews suggested that in a very few circumstances such an appointment may be essential for a reasoned resolution of a dispute.<sup>123</sup> The difficulty is in providing such assistance while preserving the effective participation of the parties in presenting and refuting evidence.

The U.S. Court of Appeals for the First Circuit affirmed the inherent authority of the court to appoint a technical advisor and offered a number of suggestions for diminishing the concerns that arise when such an appointment is made.<sup>124</sup> Before making the appointment, the court should inform the parties of its intention to appoint a technical advisor, identify the person to be appointed, and give the parties an opportunity to object to the appointee on the basis of bias or inexperience. The expert should be instructed on the record and in the presence of the parties, or the duties of the expert should be recorded in a written order. And at the conclusion of his or her service, the technical advisor should file an affidavit attesting to his or her compliance with these instructions. Some judges have gone further, making a record of discussions and disclosing the record to the parties. These safeguards may do little to comfort those who see in the technical expert an unforgivable intrusion into the adversarial system, but such safeguards will permit the parties to remain informed of the nature of the technical assistance and raise objections when the intended form of assistance encroaches on the duties of the judge. At the same time, information about the expert's advice will permit parties to challenge misplaced factual assumptions and debatable opinions.

Ex parte communication between the appointed expert and representatives of the parties poses a separate but manageable set of problems.<sup>125</sup> Ex parte communication between experts and parties will rarely be necessary—the most common instance occurs during the physical examination of a party. The expert can notify the opposing party of the intended nature of the examination and then report the findings, giving the opposing party an opportunity to raise objections. Ex parte communication may also be necessary when an expert must learn a trade secret in order to advise the court regarding a motion for a protective order. The exparte communication serves the same purpose as an in camera examination of claims of privilege and should be equally permissible.

In most other occasions ex parte communication seems unnecessary. Even in the instance where the expert must seek clarification of the position of a party, the opposing party can be notified and may participate by conference call. In such circumstances it is likely that many parties will consent to ex parte communication between the expert and the opposing party. When an expert is deposed, representatives of all parties can be invited to attend.

Court-Appointed Experts

<sup>123.</sup> See Reilly v. United States, 863 F.2d 149, 156–57 (1st Cir. 1988); MCL 3d, supra note 40, § 21.54. 124. Reilly, 863 F.2d at 159–61.

<sup>125.</sup> Some judges apply the same restrictions on parties' ex parte communications as they impose on themselves and their law clerks. When the appointed expert is serving as a technical advisor, such restrictions would be especially appropriate.

#### 3. Testimony of appointed experts

We found that almost all appointed experts, other than those serving as technical advisors, presented a written report of their findings. In approximately half of the appointments, experts concluded their service with the presentation of a report. In the remaining instances the appointed experts also presented their findings in court, either at trial or in a pretrial evidentiary hearing.

Presentation of expert testimony presents few problems where the judge acts as the finder of fact. In such a case the judge is obviously aware of the expert's court-appointed status and is sensitive to the role of the appointed expert and the duties of the judge. The judge and the parties will have reviewed the report prior to the proceeding, and testimony can be presented in a less formal manner. <sup>126</sup> In at least one case the expert was permitted to adopt the report as his direct testimony after being sworn in.

When an appointed expert testifies before a jury, the court must decide how the appointed expert will be presented. The court may, in its discretion, decide whether to disclose to the jury that the expert was appointed by the court.<sup>127</sup> In six of the seven instances we discovered, the court advised the jury or permitted the parties to advise the jury that the expert was appointed by the court. Still, we found no consensus among the judges about whether the court's sponsorship of such an expert should be mentioned. Those who favor acknowledging the court's sponsorship note that the purpose of appointing an expert often is to provide a credible witness for the jury to rely on, and independence from the parties is an important indicator of credibility. Those opposed cite the influence of such testimony, and question whether it is necessary to so discredit the testimony of the parties' experts in order for the appointed expert to serve effectively.

We believe that in almost all cases the court's sponsorship of the expert should be explicitly acknowledged, along with whatever limiting instructions are thought to be appropriate regarding the weight to be given the expert's testimony relative to the testimony of the parties' experts. If experts are appointed where doubts about the credibility of the parties' experts persist and other efforts to provide a basis for a reasoned decision have failed, knowledge of the independence of the appointed expert will be relevant to achieving the goals of the appointment. There may be instances in which the appointed expert offers testimony that serves as background information for the jury, or serves as a context for the interpretation of the testimony by the parties' experts—in these cases the court's sponsorship is less relevant to the task of the jury. But in such cases acknowledging sponsorship should disadvantage neither party. In other cases, if the need for independent testimony is sufficiently great to appoint an expert, this same need argues that such an action should be explicitly acknowledged.

<sup>126.</sup> Formal depositions of appointed experts proved to be infrequent, although on occasion an appointed expert met informally with the parties to discuss the report.

<sup>127.</sup> Fed. R. Evid. 706(c).

### **VIII.** Conclusion

Appointment of an expert by the court represents a striking departure from the adversarial process of presenting information for the resolution of disputes. But such an appointment should not be regarded as a lack of faith in the adversarial system. We learned that judges who appointed experts appear to be as devoted to the adversarial system as those who made no such appointments. Most appointments were made after extensive efforts failed to find a means within the adversarial system to gain the information necessary for a reasoned resolution of the dispute. Appointment of an expert was rarely considered until the parties had been given an opportunity and failed to provide such information. We find it hard to fault judges for failing to stand by a procedure that had proved incapable of meeting the court's need for information; to insist, in such a circumstance, that the court limit its inquiry to inadequate presentations by the parties is a poor testament to the adversarial system and the role of the courts in resolving disputes in a principled and thoughtful manner.

A better approach is to encourage the parties to present information that is responsive to the concerns of the court, inform the parties of the manner in which their presentations fall short, encourage the development of more useful testimony, and appoint an expert only when no other means is available for reaching a reasoned decision. An effective pretrial procedure will enable the development of such information, thereby strengthening the presentations of the parties and facilitating the appointment of an expert when such efforts have failed.

Appointment of an expert will undoubtedly remain a rare and extraordinary event, suited only to the most demanding cases. Regardless, Rule 706 remains an important alternative source of authority to deal with some of the most demanding evidentiary issues that arise in federal courts.

### Appendix

#### Federal Rule of Evidence Rule 706. Court Appointed Experts

(a) Appointment. The court may on its own motion or on the motion of any party enter an order to show cause why expert witnesses should not be appointed, and may request the parties to submit nominations. The court may appoint any expert witnesses agreed upon by the parties, and may appoint expert witnesses of its own selection. An expert witness shall not be appointed by the court unless the witness consents to act. A witness so appointed shall be informed of the witness' duties by the court in writing, a copy of which shall be filed with the clerk, or at a conference in which the parties shall have opportunity to participate. A witness so appointed shall advise the parties of the witness' findings, if any; the witness' deposition may be taken by any party; and the witness may be called to testify by the court or any party. The witness shall be subject to cross-examination by each party, including a party calling the witness.

(b) Compensation. Expert witnesses so appointed are entitled to reasonable compensation in whatever sum the court may allow. The compensation thus fixed is payable from funds which may be provided by law in criminal cases and civil actions and proceedings involving just compensation under the fifth amendment. In other civil actions and proceedings the compensation shall be paid by the parties in such proportion and at such time as the court directs, and thereafter charged in like manner as other costs.

(c) Disclosure of appointment. In the exercise of its discretion, the court may authorize disclosure to the jury of the fact that the court appointed the expert witness.

(d) Parties' experts of own selection. Nothing in this rule limits the parties in calling expert witnesses of their own selection.

# **Special Masters**

Margaret G. Farrell

**Margaret G. Farrell**, J.D., is Associate Professor of Law, Benjamin N. Cardozo School of Law, Yeshiva University, New York, New York. During 1992–1993 she was a judicial fellow in the Research Division of the Federal Judicial Center.

### Contents

- I. Introduction 579
- II. How Special Masters Can Serve Courts in Cases That Involve Scientific and Technical Evidence 583
  - A. Reasons for Appointing Special Masters 583
  - B. Preliability-Stage Appointments 584
    - 1. Assisting in pretrial proceedings 584
    - 2. Providing case management 585
    - 3. Facilitating settlement 586
  - C. Liability-Stage Appointments 588
  - D. Remedial-Stage Appointments 590
    - 1. Framing remedial decrees 590
    - 2. Monitoring remedial decrees 591
    - 3. Assessing damages 593
- III. Legal Authority for the Appointment of a Special Master 595
  - A. Appointments Under Federal Rule of Civil Procedure 53(b) 596
  - B. Limits on Broad-Scale Delegation 596
  - C. Powers of Masters Under Federal Rule of Civil Procedure 53 598
  - D. Appealing Appointments Under Federal Rule of Civil Procedure 53 598
- IV. Issues to Consider When Appointing a Special Master 601
  - A. Selection and Qualification 601
  - B. Avoiding Conflict-of-Interest and Ethical Problems 603
  - C. Orders of Reference 605
  - D. Ex Parte Communications 606
    - 1. With the parties 607
    - 2. With the judge 608
  - E. Potential Liability for Malfeasance 609
  - F. Type of Hearings 610
  - G. Payment 611

- H. Avoiding Delay and Inertia 614
- V. Use of Special Master's Report and Review Given 617
  - A. Jury Trials 617
  - B. Nonjury Trials 618
  - C. On Appeal 618
  - D. Standards for Determining Weight 619
- VI. Conclusion 621
# I. Introduction

The federal judicial system relies on generalist judges and lay juries to make findings of fact and conclusions about liability. In the traditional model, the generalist judge is a neutral, passive judge who receives evidence and hears argument only from the parties, applies legal rules and principles, and awards victory to one side or the other. As a result of advances in science and technology, the number of cases in which judges and juries must understand scientific and technological issues to find facts responsibly has increased. Judges have sought to meet the need for greater understanding of scientific and technological issues in several ways, one of which is through the appointment of special masters.

Historically, special masters were appointed to assist chancery judges by reporting on matters of evidence and accounting before, during, and after trials.<sup>1</sup> While explicitly permitted by Federal Rule of Civil Procedure 53, which was modeled on the superseded equity rules, use of special masters in nonjury trials has been restricted by the rule's requirement that, except in matters of accounting and difficulty in computing damages, "a reference [to a special master] shall be made only upon a showing that some exceptional condition requires it," or in jury trials, "only when the issues are complicated."<sup>2</sup> Nevertheless, increases in the caseload of the federal courts, <sup>3</sup> in the scientific and technological complexity of the subject matters presented,<sup>4</sup> in the vast amounts of data available (often as a

<sup>1.</sup> Irving R. Kaufman, Masters in the Federal Courts: Rule 53, 58 Colum. L. Rev. 452 (1958); Linda J. Silberman, Masters and Magistrates, Part I: The English Model, 50 N.Y.U. L. Rev. 1070, 1075–79, 1321–32 (1975) [hereinafter Silberman, The English Model]; Linda J. Silberman, Judicial Adjuncts Revisited: The Proliferation of Ad Hoc Procedure, 137 U. Pa. L. Rev. 2131, 2134 (1989) [hereinafter Silberman, Judicial Adjuncts]; David I. Levine, The Authority for the Appointment of Remedial Special Masters in Federal Institutional Reform Litigation: The History Reconsidered, 17 U.C. Davis L. Rev. 753 (1984). Authority for the appointment of a special master dates from Parliament's passage of the Superior Courts Officer Act in 1837. 7 Will. 4 & 1 Vict., ch. 30; Silberman, The English Model, supra, at 1078. However, the appointment of persons acting as masters may go back to the time of Henry VIII. Kaufman, *supra*, at 452. Until 1983, Fed. R. Civ. P. 53 provided for the appointment to fobth standing and special masters. See Fed. R. Civ. P. 53 advisory committee's note (relating to the 1983 amendment of subdivision (a)).

<sup>2.</sup> Fed. R. Civ. P. 53(b).

See Federal Courts Study Comm., Report of the Federal Courts Study Committee 4–10, 109–11 (1990).
 Carnegie Comm'n on Science, Technology, and Gov't, Science and Technology in Judicial Decision

Making: Creating Opportunities and Meeting Challenges 12–13 (1993). Federal Courts Study Comm., supra note 3, at 97 (citing economic, statistical, technological, and natural and social scientific data as becoming increasingly important in routine and complex litigation).

result of computer technology<sup>5</sup>), and in the numbers of claimants and corresponding amounts of money involved <sup>6</sup> have prompted judges to seek assistance through the appointment of special masters under Rule 53.<sup>7</sup>

It is unclear whether highly technical studies and controversial scientific expert testimony offered as evidence alone constitute an exceptional condition justifying appointment of a master under Rule 53. <sup>8</sup> However, the Notes of the Advisory Committee on Rules relating to the 1983 amendments to Rule 53 acknowledge that "masters may prove useful when some special expertise is desired or when a magistrate is unavailable for lengthy and detailed supervision of a case," and in many cases, the expectation or proffer of scientific or highly technical, but critical, evidence has been an important factor supporting such appointments.<sup>9</sup> Commentators generally approve of such use of special masters, <sup>10</sup> although some see potential for abuse in the flexibility permitted under the rule.<sup>11</sup>

As discussed in section III, critics of the liberal use of masters argue that delegation of adjudicatory authority to masters violates the constitutional requirement that civil cases brought in federal courts be tried and decided by Article III

7. E.g., In re "Agent Orange" Prod. Liab. Litig., 94 F.R.D. 173, 174 (E.D.N.Y. 1982) (appointing master to rule on discovery motions involving production of 4 million documents and 2,000 privilege claims, discovery of expert testimony, and protective orders); Jenkins v. Raymark Indus., Inc., 109 F.R.D. 269, 289 (E.D. Tex. 1985) (appointing master to profile the characteristics of the claims of a 1,000-member class for the jury in asbestos litigation), *affd*, 782 F.2d 468 (5th Cir. 1986); McLendon v. Continental Group, Inc., 749 F. Supp. 582, 612 (D.N.J. 1989) (appointing master to assist the parties in post-liability settlement of damages due 5,600 ERISA claimants), *affd sub nom*. McLendon v. Continental Can Co., 908 F.2d 1171 (3d Cir. 1990). See also Managing Complex Litigation: A Practical Guide to the Use of Special Masters (Wayne D. Brazil et al. eds., 1983) [hereinafter Managing Complex Litigation]; Ronald E. McKinstry, Use of Special Masters in Major Complex Cases, in Federal Discovery in Complex Cases: Antitrust, Securities and Energy 213, 225 (1980).

8. Several commentators have supported the appointment of masters in cases presenting scientific or technical questions of unusual complexity. See Judicial Conference of the United States, Procedure in Anti-trust and Other Protracted Cases (Prettyman Committee Report) (1951), *reprinted in* "Short Cuts" in Long Cases, 13 F.R.D. 41, 79–81 (1951) [hereinafter Prettyman Comm. Report].

9. See, e.g., United States v. Conservation Chem. Co., 106 F.R.D. 210, 216–17 (W.D. Mo. 1985) (appointing a master to conduct discovery and make findings on claims for inclusion in a request for injunctive relief in a chemical waste cleanup suit involving 250 parties; the court found exceptional circumstances in the imminent danger to public health presented, the analysis of voluminous scientific and technical data, the number of parties, and the vast amount of evidence necessary to litigate the case); Costello v. Wainwright, 387 F. Supp. 324, 325 (M.D. Fla. 1973) (prima facie showing of inadequate health care in state prison based on medical testimony was grounds for appointment of a physician as special master to survey conditions in the prison health system); Wayne D. Brazil, Special Masters in Complex Cases: Extending the Judiciary or Reshaping Adjudication?, 53 U. Chi. L. Rev. 394, 395 (1986).

10. See, e.g., Prettyman Comm. Report, *supra* note 8; Silberman, *Judicial Adjuncts, supra* note 1. Judge Jack B. Weinstein has noted, "Use of technical masters to supervise discovery and preparation of expert testimony is also possible. We will, I believe, see an expansion of the use of special masters for this purpose as well as for settlement and control of general discovery." Jack B. Weinstein, *Improving Expert Testimony*, 20 U. Rich. L. Rev. 473, 490 (1986).

11. See, e.g., James S. DeGraw, Note, Rule 53, Inherent Powers, and Institutional Reform: The Lack of Limits on Special Masters, 66 N.Y.U. L. Rev. 800, 803–04 (1991).

<sup>5.</sup> See Silberman, Judicial Adjuncts, supra note 1, at 2144 (discussing computerized data-collection process used in Ohio asbestos litigation and case-management techniques used in AT&T antitrust litigation).

<sup>6.</sup> See, e.g., Cimino v. Raymark Indus., Inc., 751 F. Supp. 649, 652–53 (E.D. Tex. 1990) (stating that the challenge presented to the court is to provide a fair and cost-effective means of trying large numbers of asbestos cases).

judges. Those who favor their use argue that the caseloads and expanded responsibilities of modern federal judges require the use of masters. They argue that Article III is not contravened so long as judges bind their masters' authority and supervise the performance of their duties—retaining final adjudicatory authority.

To the extent that such appointments fragment responsibility for fact finding and adjudication and permit judicial agents to take a more active, assertive role in seeking out and evaluating scientific and technical information, they depart from the traditional model. Yet, under Rule 53, generalist judges continue to exercise ultimate decision-making authority over the matters referred to masters. Moreover, the argument can be made that without masters' assistance in gaining access to specialized information, judges, because of their own lack of scientific knowledge and expertise, ultimately would have to abdicate their authority to partisan scientists and third-party experts.

How can the appointment of special masters ease the burden of judges when they confront complicated cases that involve scientific data, technological advances, and computerized data? What tasks do judges ask masters to perform? How do judges select masters to deal with scientific and technical evidence, and what authority do they give them? How do they communicate with their masters? What functions can masters perform effectively, and what obstacles inhibit their more effective use? How is the rate of compensation for masters set, and how are masters paid? How do courts limit the cost and delay occasioned by the appointment of masters?

To help answer these questions, the Research Division of the Federal Judicial Center (FJC) conducted a study on the use of special masters in 1992–1993. The study sought to identify special masters who had been appointed to deal with technically difficult matters or cases in which unusual scientific or technical expertise was required. <sup>12</sup> An effort was made to identify and interview masters who had participated at various stages of litigation and who had performed a variety of functions. Magistrate judges appointed as special masters were not included in the study. The issues arising from the appointment of full-time, government-paid U.S. magistrate judges under Rule 53 are not identical to those resulting from the appointment of part-time, party-paid, lay masters. However, where appropriate, some issues relating to the appointment of a magistrate judge as special master are discussed.

The FJC formulated a protocol of questions seeking information about the tasks assigned in specific cases and on several issues pertaining especially to the use of masters to deal with scientific evidence. After the masters were interviewed, the judges who appointed them were contacted and asked to identify ob-

<sup>12.</sup> An empirical study of the incidence and prevalence of Fed. R. Civ. P. 53 appointments as a whole, and appointments for these purposes in particular, was not possible because the district courts and the Administrative Office of the U.S. Courts do not collect data on such appointments uniformly. In the absence of such data, the FJC interviewed twelve masters who were identified in legal literature and by their colleagues as experienced and knowledgeable about the use of masters.

stacles to and benefits of their appointments. Although these participants do not constitute a representative sample of all masters, or the judges who appoint them, they did provide many valuable insights into the more effective use of masters under Federal Rule of Civil Procedure 53(b). This paper is based on information gained in the FJC's interviews, as well as on existing case law and law review commentary.

# II. How Special Masters Can Serve Courts in Cases That Involve Scientific and Technical Evidence

## A. Reasons for Appointing Special Masters

Judges have indicated that they need scientific or technical knowledge during litigation for at least four purposes:

- 1. to evaluate scientific and technical evidence presented by the parties<sup>13</sup> or by other experts<sup>14</sup> and to rule on objections to evidence;
- to assess claims and facilitate settlement, such as claims of loss in product liability cases;
- to educate the fact finder—judge or jury—in the subject matter of particular controversies, such as patent disputes;<sup>15</sup> and
- 4. to scientifically analyze and evaluate other evidence, such as evidence of discrimination.<sup>16</sup>

13. For instance, scientific evidence of injury and causation figured prominently in the litigation of pharmaceutical injuries, e.g., Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S. Ct. 2786 (1993); DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941 (3d Cir. 1990); Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307 (5th Cir.), modified , 884 F.2d 166 (5th Cir. 1989), cert. denied , 494 U.S. 1046 (1990); product liability injuries, e.g., In re A. H. Robins Co., 88 B.R. 742, 746 (Bankr. E.D. Va. 1988) (Dalkon Shield), affd, 880 F.2d 694 (4th Cir.), cert. denied, 493 U.S. 959 (1989); and toxic tort injury suits against both private entities, e.g., Cimino v. Raymark Indus, Inc., 751 F. Supp. 649 (E.D. Tex. 1990), and the government, e.g., In re "Agent Orange" Prod. Liab. Litig., 611 F. Supp. 1223 (E.D.N.Y. 1985) (veterans suit against the Defense Department as well as chemical manufacturers), aff'd, 818 F.2d 187 (2d Cir. 1987), cert. denied, 487 U.S. 1234 (1988).

14. See, e.g., Brazil, *supra* note 9, at 410–12; Fox v. Bowen, 656 F. Supp. 1236, 1253–54 (D. Conn. 1986) (master would be appointed to hire experts and conduct studies necessary to the framing of a remedial order).

15. See, e.g., In re Newman, 763 F.2d 407, 409 (Fed. Cir. 1985). Courts also appoint expert witnesses for this purpose under Fed. R. Evid. 706. See, e.g., Computer Assocs. Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 712 (2d Cir. 1992) (court recognized the need for expert assistance but appointed an expert under Fed. R. Evid. 706 to hear testimony of parties' experts and to testify in court under cross-examination); Gates v. United States, 707 F.2d 1141, 1144 (10th Cir. 1983); see generally Joe S. Cecil & Thomas E. Willging, Court-Appointed Experts, in this manual.

16. Knowledge of scientific methodology and statistical techniques is needed to manage scientific and nonscientific information relevant to evaluating and trying claims, facilitating negotiations, distributing judgments, and deciding appeals in complex litigation, as in a Texas asbestos case (Cimino v. Raymark Indus., Inc., 751 F. Supp. 649 (E.D. Tex. 1990)) and a Dalkon Shield bankruptcy proceeding (*Inre* A.H. Robins Co., 88 B.R. 742, 746 (Bankr. E.D. Va. 1988)). See generally B. Thomas Florence & Judith Gurney, *The Computerization of Mass Tort Settlement Facilities*, Law & Contemp. Probs., Autumn 1990, at 189 (describing the computer systems necessary to process 60,000 asbestos claims and 200,000 Dalkon Shield claims); The Evolving Role of In illustrating ways in which masters have helped meet judges' needs for special assistance, the following discussion characterizes masters according to their tasks and the stage of litigation at which they are appointed. Sometimes the tasks assigned and the purpose of the appointment are the same (e.g., settlement masters are appointed to facilitate settlement by the parties). Sometimes tasks are assigned for several purposes (e.g., masters are appointed to conduct case management, the purpose of which is to make preliminary findings of fact, facilitate settlements, and advise the court on technical matters).

## B. Preliability-Stage Appointments

## 1. Assisting in pretrial proceedings

In complex cases, special masters sometimes are appointed during discovery to limit massive discovery requests, to rule on claims of privilege, and to make factual determinations necessary to rule on the admissibility of evidence.<sup>17</sup> Where information sought in discovery is scientific, highly technical, or complex in nature, even stronger reason exists to seek the appointment of a master under Rule 53.<sup>18</sup>

Discovery masters sometimes hold formal hearings on nondispositive motions and preliminary facts, but often they proceed more informally to make findings based on their own knowledge or on information received from the parties outside of evidentiary hearings. When discovery motions involve the production of technical information in trademark, patent, copyright, and product liability cases, courts often appoint special masters who have expertise in the subject matter of the case. These masters sometimes will examine scientific and technical evidence, assess the potential qualifications of expert witnesses, or determine the admissibility of scientific studies.<sup>19</sup> Discovery masters who do not have special knowledge often develop considerable expertise in the technical or scientific subject matter of the suit after devoting a large amount of their time to particularly complex cases.

Statistical Assessments as Evidence in the Courts (Stephen E. Fienberg ed., 1989); Richard Lempert, Statistics in the Courtroom: Building on Rubinfeld, 85 Colum. L. Rev. 1098 (1985).

<sup>17.</sup> E.g., United States v. International Business Machs. Corp., 76 F.R.D. 97, 98 (S.D.N.Y. 1977); United States v. American Tel. & Tel. Co., 461 F. Supp. 1314, 1347–49 (D.D.C. 1978). For a discussion of the legal history of the authority to appoint masters to supervise discovery, see Wayne D. Brazil, Authority to Refer Discovery Tasks to Special Masters: Limitations on Existing Sources and the Need for a New Federal Rule, in Managing Complex Litigation, supra note 7, at 305.

<sup>18.</sup> In re "Agent Orange" Prod. Liab. Litig., 94 F.R.D. 173, 174–75 (E.D.N.Y. 1982). But see, e.g., Cadwell Indus., Inc. v. New York Hospital-Cornell Medical Ctr., No. 88-C7307, 1993 U.S. Dist. LEXIS 2263, at \*8 n.1 (S.D.N.Y. Feb. 26, 1993) (court denied motion for appointment of master where it found parties' counsel had demonstrated that they were quite capable of explaining difficult medical and scientific materials and theories to an audience unfamiliar with such subjects).

<sup>19.</sup> See, e.g., Brazil, supra note 9, at 410–12; Report and Recommendation from George L. Priest, Special Master, regarding the Fairness of the Settlement and of the Proposed Plan of Distribution to The Honorable H. Lee Sarokin, *filed in* McLendon v. Continental Group, Inc., C.A. No. 83-1340 (SA), 5, 21 (D.N.J. July 13, 1992) [hereinafter Report of George L. Priest, Special Master].

Judges' need for assistance during pretrial proceedings to handle proffers of scientific expert testimony or rule on motions in limine may increase, given the Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*<sup>20</sup> In *Daubert*, the Court clarified the trial court's obligation under Rule 702 of the Federal Rules of Evidence to determine the reliability, as well as relevance, of scientific evidence upon which expert opinion is based. The Court held that "[i]n a case involving scientific evidence, evidentiary reliability will be based upon scientific validity."<sup>21</sup> To make these determinations, trial judges must determine whether the expert is proposing to testify to scientific knowledge that will assist the trier of fact.<sup>22</sup>

The determination of these factors in a Rule 104(a) hearing may take a substantial amount of a judge's already limited time<sup>23</sup> and may also require the judge to educate himself or herself in the science underlying the evidence offered. In a pre- *Daubert* product liability case, for example, scientific issues raised by a motion for summary judgment required a district court to hold five days of hearings and consider extensive post-hearing submissions in order to determine the validity of the epidemiological data and the methods an expert was willing to give his opinion on regarding causation.<sup>24</sup> Appointing special masters to conduct Rule 104(a) hearings may be one way for district courts to hold the type of hearing suggested by *Daubert* without burdening the courts' resources. A special master can devote more time to becoming familiar with the evidence submitted. Moreover, the court can select a master who has expertise in the science involved. Thus, Rule 53(b) may permit the appointment of a pretrial special master to hold Rule 104(a) hearings and make recommendations regarding the conditions necessary for the admissibility of expert testimony.

#### 2. Providing case management

In some complex cases, judges have required assistance in addition to the supervision of discovery, especially when the claims of class action plaintiffs have had to be evaluated for the purpose of settlement negotiations and trial preparation. This more comprehensive assistance has been obtained by appointing masters to carry out overall management of the case in its pretrial stage and to advise judges on scientific and technical issues.<sup>25</sup>

Special Masters

<sup>20. 113</sup> S. Ct. 2786 (1993).

<sup>21.</sup> ld. at 2795 n.9.

<sup>22.</sup> See Margaret A. Berger, Evidentiary Framework § III, in this manual.

<sup>23.</sup> McCormick on Evidence § 53, at 137 n.8 (Edward W. Cleary ed., 3d ed. 1984) (citing Fed. R. Evid. 104(c) advisory committee's note observing that hearing preliminary matters out of the hearing of the jury is a time-consuming procedure); Bourjaily v. United States, 483 U.S. 171 (1987) (offering party must prove preliminary facts necessary for admission of evidence).

<sup>24.</sup> DeLuca v. Merrell Dow Pharmaceuticals, Inc., 911 F.2d 941 (3d Cir. 1990) (hearings held on motion for summary judgment). At a Rule 104(a) hearing, rules of evidence need not apply, except those with respect to privileges, and therefore, such hearings may differ from those held by the district court in *DeLuca*.

<sup>25.</sup> In re Ohio Asbestos Litig., No. 83-06 (N.D. Ohio Dec. 16, 1983) (order incorporating by reference the case-management plan and evaluation and apportionment process developed by special masters Francis E.

In a case consolidating thousands of asbestos claims, the district court appointed two special masters to develop a case-management plan for resolving all pending cases (eventually numbering more than 9,000) within a two-year period.<sup>26</sup> In addition to supervising discovery, these masters devised a plan for obtaining information on the outcome of similar cases, gathering information about outcome-determinative variables among the members of the class, and developing a system of computerized case-matching that permitted the parties to bargain within estimated settlement ranges.<sup>27</sup> Rather than simply conducting discovery or making recommended findings of fact, these masters provided technical advice to the court, largely about techniques for gathering and analyzing empirical data.

Thus, in some complex suits, judges have needed expert and technical assistance not to understand the subject matter of the suit or issues of causation, but to handle massive amounts of nontechnical information.<sup>28</sup>

#### 3. Facilitating settlement

Masters appointed to supervise discovery sometimes try to promote joint stipulations regarding undisputed scientific facts or techniques. In doing so, they become mediators of differences between or among the parties regarding either scientific information offered in evidence or scientific facts necessary to findings of liability.<sup>29</sup>

For example, in a suit involving multiple plaintiffs, defendants, and amici over fishing rights of Native Americans in Lake Michigan, a law school professor was appointed master both to provide case management and to facilitate settle-

27. For a full discussion of the case-management plan, see Francis E. McGovern, Toward a Functional Approach for Managing Complex Litigation, 53 U. Chi. L. Rev. 440, 478–91 (1986).

28. See, e.g., Burgess v. Williams, 302 F.2d 91, 93–94 (4th Cir. 1962) (denial of writ of mandamus to remove a master appointed to make factual findings regarding 1,500 individuals); Wattleton v. Ladish Co., 520 F. Supp. 1329, 1350 (E.D. Wis. 1981) (master appointed to determine damages owed to each class member in an employment discrimination suit), *aff'd sub nom*. Wattleton v. International Bhd. of Boiler Makers, Local 1509, 686 F.2d 586 (7th Cir. 1982), *cert. denied*, 459 U.S. 1208 (1983).

29. Comment, A Federal Judge's ADR Techniques: Mini-Trials with Special Masters, 5 Alternatives to High Cost Litig. 9 (1987) (discussing Judge Sherman G. Finesilver's use of masters in copyright, trademark, and other disputes requiring expertise). McGovern, supra note 27, at 456–68 (discussion of fishing rights cases in Michigan). See generally Troyen A. Brennan, Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts, 51 U. Pitt. L. Rev. 1 (1989). In a recent case, Judge Jack B. Weinstein observed that where cases turning on medical and scientific evidence of causation are consolidated, a magistrate judge (or master) can be used to facilitate the sharing of data and experts, thereby reducing redundant discovery requests. In re Repetitive Stress Injury Cases, 142 F.R.D. 584, 586–87 (E.D.N.Y. 1992), appeal dismissed, order vacated sub nom. In re

McGovern and Eric D. Green); Jenkins v. Raymark Indus., Inc., 109 F.R.D. 269, 288 (E.D. Tex. 1985), aff/d, 782 F.2d 468 (5th Cir. 1986). See also In re United States Dep't of Defense, 848 F.2d 232 (D.C. Cir. 1988) (master appointed to evaluate classified nature of thousands of documents in freedom of information suit; rather than undertake in camera review, court charged expert master with selecting a scientifically sound representative sample of withheld documents and summarizing contentions regarding their privileged nature). See generally Francis E. McGovern, Resolving Mature Mass Tort Litigation, 69 B.U. L. Rev. 659 (1989).

<sup>26.</sup> In re Ohio Asbestos Litig., No. 83-03 (N.D. Ohio July 14, 1983) (order appointing special masters to develop a case-management plan for the pretrial and trial phases).

ment.<sup>30</sup> The master was authorized to, among other things, receive, seek out, and consider scientific evidence about the distribution of different species of fish in the lake. The master facilitated a settlement on the merits through an integrative bargaining process that added resources to be distributed among the parties and sought to maximize the interests of all the parties.<sup>31</sup> The master obtained an agreement among the parties on important scientific facts about the spawning, environmental habitat, and migration of fish in the lakes and then facilitated an agreement among the parties to pool their scientific information and direct their experts to make consensus recommendations. Finally, the parties agreed to employ a neutral expert in decision modeling to assist their biologists and the special master in creating a computer model that would assess proposed settlement plans in terms of five critical variables.

Recently, courts have made this mediation function more explicit and have appointed special masters expressly to achieve settlements in complex litigation, especially mass tort cases.<sup>32</sup> The 1983 amendments to Federal Rule of Civil Procedure 16 permit federal judges to "take appropriate action, with respect to ... settlement and the use of special procedures to assist in resolving the dispute."<sup>33</sup>

While courts can appoint masters to promote settlement at any stage of litigation, the appointment of masters at the pretrial stage permits judges to use firm, strict trial dates to remind the parties of the expense of litigation and create incentives to settle the case if possible. In addition, the appointment of pretrial settlement masters allows courts to delegate more assertive tasks to the master in order to minimize judicial contacts with the parties and eliminate the apparent bias and prejudgment these contacts suggest.

Some courts choose settlement masters for their particular scientific or technical expertise (usually at the remedial stage of litigation). However, other courts that have found such skills important to settlement negotiations appoint experts under Federal Rule of Evidence 706 to advise the parties and the court on settlements and the framing of consent decrees. Such appointments remain rare, however.<sup>34</sup>

<sup>30.</sup> Brazil, supra note 9, at 410–12 (discussing United States v. Michigan, 471 F. Supp. 192 (W.D. Mich. 1979), remanded , 623 F.2d 448 (6th Cir. 1980), as modified , 653 F.2d 277 (6th Cir.), cert. denied, 454 U.S. 1124 (1981)).

<sup>31.</sup> McGovern, supra note 27, at 459–64. See also Gates v. United States, 707 F.2d 1141, 1142 (10th Cir. 1983) (appointment of a panel of experts in a swine flu vaccine case); Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts,* 43 Hastings L.J. 301 (1992).

<sup>32.</sup> See, e.g., In re Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735, 737 (E.D.N.Y. 1990); In re DES Cases, 789 F. Supp. 552, 558 (E.D.N.Y. 1992), appeal dismissed, 7 F.3d 20 (2d Cir. 1993).

<sup>33.</sup> Fed. R. Civ. P. 16(c)(9). See generally D. Marie Provine, Settlement Strategies for Federal District Judges (Federal Judicial Center 1986). See also Silberman, Judicial Adjuncts, supra note 1, at 2158–59 (discussing use of masters for settlement purposes).

<sup>34.</sup> See, e.g., San Francisco NAACP v. San Francisco Unified Sch. Dist., 576 F. Supp. 34, 39 (N.D. Cal. 1983) (appointment of a "settlement team" of experts pursuant to Rule 706, nominated by the parties and the court to draft consent decree); cf. Gates, 707 F.2d at 1142 (appointment of a panel of experts under Rule 706

The master may be instructed to effect settlements on both evidentiary matters and liability issues (as in the Michigan fishing rights case discussed earlier). In product liability cases, special masters have been appointed to provide case management and expertise in evaluating thousands of claims before trial in an effort to facilitate settlement negotiations. For instance, a law professor was appointed master in several asbestos and toxic tort cases to profile the claims characteristic of class representatives based on evaluation of medical evidence provided by expert consultants, sound questionnaire methodologies, sampling techniques, and statistical analysis.<sup>35</sup> Evaluation of pretrial claims often involves making factual determinations regarding elements of the plaintiffs' case, such as the cause of plaintiffs' losses.

Although most settlement masters fulfill their function through informal procedures, some hold formal evidentiary hearings in the form of mini-trials to evaluate claims for purposes of negotiation.<sup>36</sup> Factual findings of specific causation based on scientific evidence must be made when a master is appointed to evaluate individual claims for purposes of negotiating settlements and distributing awards.<sup>37</sup> The evaluation of scientific evidence and expert witnesses to make findings of causation as a matter of fact occurs at the liability stage as well as the pretrial stage of litigation.

## C. Liability-Stage Appointments

Rule 53(b) anticipates the appointment of masters to make recommended factual findings going to the merits of the dispute before the court. <sup>38</sup> As provided in Rule 53(c), in actions involving complicated issues tried before a jury or exceptional conditions in bench trials, masters may require the production of evidence, hold formal hearings in which the rules of evidence apply, administer oaths, and create a record for review. Although courts use special masters more frequently in the pretrial and remedial stages of litigation, special masters are also appointed to make recommendations with regard to facts that are necessary to determine liability.<sup>39</sup>

to assist the trial court in understanding complex neurological and epidemiological issues in a swine flu vaccine case).

<sup>35.</sup> Jenkins v. Raymark Indus., Inc., 109 F.R.D. 269, 289 (E.D. Tex. 1985) (Francis E. McGovern appointed special master to assess asbestos injury claims), *affd*, 782 F.2d 468 (5th Cir. 1986); Brazil, *supra* note 9, at 399–402, 404–06 (account of McGovern's collection and computerization of information on about 9,000 claimants in a toxic tort case involving DDT through negotiated/mediated survey questionnaire to be used in evaluating claims before liability determination and a description of Masters McGovern and Eric Green's use of collected data and computer models in Ohio asbestos cases).

<sup>36.</sup> For a discussion of this technique, see Michael J. Saks & Peter David Blanck, Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts, 44 Stan. L. Rev. 815 (1992).

<sup>37.</sup> Kenneth R. Feinberg, The Dalkon Shield Claimants Trust, Law & Contemp. Probs., Autumn 1990, at 79.

<sup>38.</sup> Brazil, supra note 17, at 332-44.

<sup>39.</sup> These recommendations are subject to Fed. R. Civ. P. 53(e). See *infra* § V.D for further discussion of the weight given to a special master's report.

In a large, complex class action suit claiming that defendants' layoffs and plant closings violated the Employee Retirement Income Security Act (ERISA), for example, an economics expert knowledgeable about computers was appointed special master to determine factual issues of causation necessary to make findings of liability.<sup>40</sup> Similarly, where scientific medical evidence was anticipated in a trial on the merits of an injunctive action against a prison, the court appointed a special master to aid it in evaluating the quality of medical services and to conduct a medical survey of all correctional institutions.<sup>41</sup> In many instances, masters appointed to try issues of fact find themselves becoming mediators of the dispute, facilitating settlements, as well as finding facts. The parties in the ERISA case ultimately agreed to a settlement facilitated by the master's shuttle diplomacy.<sup>42</sup>

Nevertheless, courts may be more reluctant to recognize exceptional circumstances supporting an appointment under Rule 53 when the master will be recommending findings of fact and conclusions of law on dispositive liability issues, rather than ruling on nondispositive, pretrial motions or monitoring post-decree compliance.<sup>43</sup> Moreover, some courts have found constitutional limitations bounding Rule 53 that would prohibit the appointment of a master to hear the merits, even where exceptional conditions seem to be present.

Rule 53 authority was also used by judges to appoint masters who were experts in the subject matter of litigation to act as neutral advisers to the court during the liability stage of litigation.<sup>44</sup> Rather than acting as fact finders or mediators between the parties, evaluating the parties' scientific evidence, or facilitating

40. McLendon v. Continental Group, Inc., 749 F. Supp. 582, 612 (D.N.J. 1989), aff'd sub nom. McLendon v. Continental Can Co., 908 F.2d 1171 (3d Cir. 1990).

41. Costello v. Wainwright, 387 F. Supp. 324, 325 (M.D. Fla. 1973).

42. Report of George L. Priest, Special Master, supra note 19.

43. See Silberman, Judicial Adjuncts, supra note 1, at 1151–53, 2174. See discussion infra § III and, e.g., In re Bituminous Coal Operators' Ass'n, 949 F.2d 1165, 1166 (D.C. Cir. 1991) (writ of mandamus granted voiding reference to master in nonjury trial "virtually for all purposes," including "trial of the issues of liability"); Stauble v. Warrob, Inc., 977 F.2d 690, 695–97 (1st Cir. 1992) (holding that referring fundamental issues of liability to special master for adjudication over objection is impermissible); Prudential Ins. Co. of Am. v. United States Gypsum Co., 991 F.2d 1080 (3d Cir. 1993) (writ of mandamus issued overturning appointment of master to hear merits of a claim for cost of testing, monitoring, and removing asbestos-containing products at thirtynine Prudential properties); Cadwell Indus., Inc. v. New York Hospital-Cornell Medical Ctr., No. 88-C7307, 1993 U.S. Dist. LEXIS 2263, at \*8 n.1 (S.D.N.Y. Feb. 26, 1993) (motion for appointment of master to hear merits of a claim for damages resulting from destruction of two prototype brain-wave monitors denied where court found complicated questions were insufficient to justify master making preliminary examination of the dispositive issues). See also In re United States, 816 F.2d 1083, 1090–91 (6th Cir. 1987); Manual for Complex Litigation, Third, § 21.52 (forthcoming 1995) [hereinafter MCL 3d].

44. United States v. Conservation Chem. Co., 106 F.R.D. 210, 220 (W.D. Mo. 1985) (expert on environmental law appointed master to prepare case for trial of liability issues in suit to enforce mandatory cleanup of chemical waste disposal site, finding "[R]ule [53] is broad enough to allow appointment of expert advisors."); *Inre* United States Dep't of Defense, 848 F.2d 232, 234–36 (D.C. Cir. 1988) (affirming appointment of a security-cleared intelligence expert in national security matters to advise the court on the sensitive nature of 2,000 Defense Department documents sought in a freedom of information suit, but not to make recommendations); Patricia M. Wald, "*Some Exceptional Condition*"—*The Anatomy of a Decision Under Federal Rule of Civil Procedure* 53(b), 62 St. John's L. Rev. 405 (1988); Danville Tobacco Ass'n v. Bryant-Buckner Assocs., 333 F.2d 202 (4th Cir. 1964) (expert in tobacco marketing appointed special master to provide guidance to the court). their agreement about scientific facts, these masters functioned more as courtappointed experts than as traditional masters.<sup>45</sup> However, unlike court-appointed experts, these masters were not subject to cross-examination by the parties and could be granted more case-management authority.

### D. Remedial-Stage Appointments

### 1. Framing remedial decrees

Masters are often appointed to help formulate remedial decrees and supervise compliance with institutional reform orders because courts need assistance in evaluating technical and scientific evidence submitted by the parties regarding the treatment of prisoners, mentally retarded persons, and mentally ill persons.<sup>46</sup> Persons with specialized knowledge are more likely to be appointed masters at the remedial stage than at other stages of litigation.<sup>47</sup>

Expert masters appointed after a finding of liability in environmental and institutional reform litigation often advise the court by making recommendations for detailed remedial orders or amendments to such orders in periodic reports based on their own expertise.<sup>48</sup> For example, in a New York desegregation suit, a special master who was an expert in government-housing law and educational administration was appointed to develop an integration plan for a particular school area.<sup>49</sup> The master was authorized to solicit the views of community groups, receive evidence, consult with the parties, engage experts, and gather relevant data from the parties.

Although federal judges may appoint expert masters to recommend remedial orders, in some cases studied, the judges instead appointed special masters with the authority to employ experts.<sup>50</sup> In a suit in which the Department of Health and Human Services was found to violate due process in making eligibility de-

48. See generally Timothy G. Little, Court-Appointed Special Masters in Complex Environmental Litigation: City of Quincy v. Metropolitan District Commission, 8 Harv. Envtl. L. Rev. 435 (1984); Alberti v. Klevenhagen, 660 F. Supp. 605, 607–09 (S.D. Tex. 1987) (three expert prison masters appointed to monitor compliance in prison reform suit).

49. Hart v. Community Sch. Bd., 383 F. Supp. 699, 758 (E.D.N.Y.), appeal dismissed, 497 F.2d 1027 (2d Cir. 1974), and aff'd, 512 F.2d 37 (2d Cir. 1975).

50. See, e.g., Puerto Rican Legal Defense & Educ. Fund, Inc. v. Gantt, 796 F. Supp. 681, 684–85 (E.D.N.Y. 1992) (retired judge appointed special master and had authority under Fed. R. Civ. P. 53 to hire experts familiar with redistricting to advise him on creating a new districting plan).

<sup>45.</sup> Commentators have noted that such "experts" who do not resolve factual disputes, take evidence, or make rulings of law, but are appointed to provide guidance to the court under Fed. R. Civ. P. 53, are not true masters. 9 Charles A. Wright & Arthur R. Miller, Federal Practice and Procedure § 2602, at 779 n.16 (1971).

<sup>46.</sup> Reed v. Cleveland Bd. of Educ., 607 F.2d 737, 741–44 (6th Cir. 1979); Halderman v. Pennhurst State Sch. & Hosp., 446 F. Supp. 1295, 1326 (E.D. Pa. 1977), aff'd in part, rev'd in part, 612 F.2d 84 (3d Cir. 1979), cert. granted , 447 U.S. 904, and stay granted in part, 448 U.S. 905 (1980), and rev'd , 451 U.S. 1 (1981); Ruiz v. Estelle, 679 F.2d 1115, 1159, 1172 (5th Cir.), amended in part, vacated in part, 688 F.2d 266 (5th Cir. 1982), cert. denied, 460 U.S. 1042 (1983). For further description of remedial masters, see Silberman, Judicial Adjuncts, supra note 1, at 2161–68.

<sup>47.</sup> See, e.g., Final Report of the Special Master filed in Fox v. Sullivan, No. H-78-541 (JAC) at 7–8 (D. Conn. Aug. 20, 1992) (special master appointed in Fox v. Bowen, 656 F. Supp. 1236 (D. Conn. 1986), gives final report on the outcome of defendant's implementation of the relief ordered) [hereinafter Final Report].

terminations under the Medicare statute, the judge appointed a law professor to assess medical, epidemiological, and other technical information submitted by the parties concerning appropriate procedures for determining when beneficiaries were entitled to care in nursing home facilities.<sup>51</sup> Confronted with the task of framing a remedial decree that would mandate constitutional procedures for determining Medicare benefit coverage, the master also was authorized to identify and consult with expert physical therapists, epidemiologists, and physicians regarding the appropriate procedures for assessing the nursing home needs of elderly patients.<sup>52</sup> In another case, a special master hired an independent expert to review proposed mental health service plans to be ordered by the court.<sup>53</sup> In their capacity as remedial masters, these special masters were engaged in fact finding, facilitating settlement around remedial issues, providing expert advice to the court, and in the Medicare suit, aggregating and analyzing empirical data for the court.

#### 2. Monitoring remedial decrees

Masters are appointed to monitor compliance with remedial decrees when the defendant has been unwilling or unable to comply with the decree. The defendant often opposes such an appointment. In these instances, some judges select special masters on the basis of their specific knowledge of the subject matter of the suit and often without the approval of both parties.<sup>54</sup>

In institutional reform and other reform litigation, such as suits involving school systems, prisons, nursing homes, and mental hospitals, remedial masters often must make findings of fact based on expert testimony about medical, mental health, and penal practices of defendants. <sup>55</sup> As court monitors, these masters are required to find facts regarding defendant compliance, settle disputes over refinement and amendment of remedial orders, and advise the court through periodic reports and accountings.

In addition, some remedial masters are authorized to seek out scientific and technical experts and make findings of fact based on their personal observations of facilities and on ex parte interviews. These masters function more as investigators than as experts or judges. <sup>56</sup> Masters who are appointed to monitor compli-

<sup>51.</sup> Fox v. Bowen, 656 F. Supp. 1236, 1253-54 (D. Conn. 1986); Final Report, supra note 47.

<sup>52.</sup> Final Report, *supra* note 47. The master was authorized to commission a two-year study of defendants' practice regarding the diagnosis and prescription of physical therapy as an element in Medicare eligibility determinations. *Id.* at 6.

<sup>53.</sup> United States v. Michigan, 680 F. Supp. 928, 956-57 (W.D. Mich. 1987).

<sup>54.</sup> Where parties do agree on the need for a monitor, provision for a monitor usually is included in a consent decree and thus is not made pursuant to Fed. R. Civ. P. 53.

<sup>55.</sup> See generally Susan P. Sturm, A Normative Theory of Public Law Remedies, 79 Geo. L.J. 1355 (1991); DeGraw, supra note 11, at 803 nn.23–25.

<sup>56.</sup> Judge Jack B. Weinstein has observed, "An important distinction exists in the relative ability of trial and appellate courts to induce the production of technical evidence superior to that provided by the parties. Trial courts are able, through the means just discussed [the use of 'technical masters'], to develop new evidence." Weinstein, *supra* note 10, at 490.

ance with remedial decrees in institutional reform suits, although often experts themselves, commonly employ other experts to evaluate the defendants' performance of remedial obligations in specialized areas. For instance, in a suit brought to reform the Puerto Rican prison system, the master hired experts, with court approval, to evaluate compliance with constitutionally required safe and sanitary physical conditions, medical treatment, and protection.<sup>57</sup> In this role, the master employed experts to advise both him and ultimately the court.<sup>58</sup> Similarly, in a celebrated case involving pollution of the Boston harbor, the court appointed a law professor as a master to investigate the history and functions of the city's sewage system, consult experts, and propose remedial plans.<sup>59</sup>

Some expert masters, like some lay masters, see themselves as knowledgeable facilitators, not decision makers, who move the parties to find areas of agreement about scientific and technical facts and to develop agreed upon procedures for settling their factual disputes. Generally worded orders of reference give such masters authority to devise creative approaches to these tasks.

Masters sometimes are appointed because they have expertise stemming from prior experience with a particular case or similar cases.<sup>60</sup> Thus, some masters were expert witnesses in the same litigation, <sup>61</sup> and some were institutional administrators. These appointments raised questions about potential conflicts of interest but were sustained where the parties agreed to the appointment. In some instances, these conflicts were placed on the record and expressly waived by the parties. In other cases, a waiver was implied by the parties' agreeing to the appointment of the master or to a subsequently negotiated settlement.<sup>62</sup> See the discussion in sections IV.A and IV.D on the mechanisms used to deal with conflict-of-interest issues.

57. Morales Feliciano v. Romero Barcelo, 672 F. Supp. 591, 623-24 (D.P.R. 1986).

58. Judge Jack B. Weinstein observed in a repetitive stress injury case that courts required to be proactive should consider the appointment of panels of experts to establish protocols for future safe action by defendants. *In re* Repetitive Stress Injury Cases, 142 F.R.D 589, 587 (E.D.N.Y. 1992), *appeal dismissed, order vacated sub. nom. In re* Repetitive Stress Injury Litig., 11 F.3d 368 (2d Cir. 1993).

59. See Brazil, supra note 9, at 414–17; Little, supra note 48, at 473–75 (discussing Quincy v. Metropolitan Dist. Comm'n, Civ. No. 138,477 (Mass. Super. Ct., Norfolk County filed Dec. 17, 1982)).

60. United States v. Conservation Chem. Co., 106 F.R.D. 210 (W.D. Mo. 1985) (in a chemical waste cleanup case, motion denied to revoke the appointment of a master appointed to conduct discovery and prepare a report and recommendations on the issues presented in a suit involving more than 250 parties, including 154 third-party defendants, 14 government defendants, and 16 insurance companies, where master had already served in the pretrial stage, and judge reserved authority to make the ultimate determination on all is sues); United States v. Suquamish Indian Tribe, 901 F.2d 772 (9th Cir. 1990) (upholding appointment of a master who was a former magistrate judge and had previously served as a master in several subproceedings of the litigation begun more than fifteen years earlier to determine the fishing rights of certain Native American tribes).

61. See, e.g., Costello v. Wainwright, 387 F. Supp. 324, 325 (M.D. Fla. 1973) (physician who had testified as expert witness appointed special master).

62. In one case, a party that subsequently objected to a conflict known at the time of the appointment was estopped from doing so. *In re* Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735, 742–44 (E.D.N.Y. 1990).

#### 3. Assessing damages

At the post-liability stage, masters are also appointed to develop statistically sound, technically complex means of evaluating the damages of thousands of claimants from a limited pool of funds, such as the funding established in the Dalkon Shield and Manville asbestos cases.<sup>63</sup> Similar techniques were used by the special master in a suit brought by 5,600 terminated employees claiming damages from firings and plant closings in violation of ERISA. The special master in that case described in his report to the court the necessity for a sophisticated statistical multiple regression analysis in developing a plan to distribute a settlement fund to 5,600 class members who fell into four complicated vesting categories and five award categories, and had wide-ranging individual earning capacities and consequential losses.<sup>64</sup>

Similar needs arose in other cases after a finding of liability when the damages of thousands of successful claimants had to be determined.<sup>65</sup> Knowledge of sound empirical methods, statistical techniques, and computer technology was needed to perform these tasks.<sup>66</sup> Most often such expertise was provided by independent experts hired by the master.<sup>67</sup>

Thus, each appointment should be examined to determine its underlying purpose and the functions to be performed by the master in order to resolve issues concerning selection, ethics, ex parte communication, liability, compensation, and efficiency, which are discussed in section IV. Nevertheless, the generic issues raised in section IV should be resolved in light of the purposes of the appointments rather than the stage of litigation in which appointments are made or the tasks that have been assigned to special masters.

66. E.g., McLendon, 749 F. Supp. at 612.

<sup>63.</sup> Kane v. Johns-Manville Corp., 843 F.2d 636 (2d Cir. 1988). A. H. Robins Co. v. Piccinin, 788 F.2d 994 (4th Cir. 1986), cert. denied, 479 U.S. 876 (1986). A. H. Robins Co. v. Mabey, 880 F.2d 694 (4th Cir.), later proceeding, 880 F.2d 769 (4th Cir.), cert. denied, 493 U.S. 959 (1989). See generally Symposium, Claims Resolution Facilities and the Mass Settlement of Mass Torts, Law & Contemp. Probs., Autumn 1990, at 1; Saks & Blanck, supra note 36; The Evolving Role of Statistical Assessments as Evidence in the Courts, supra note 16.

<sup>64.</sup> Report of George L. Priest, Special Master, *supra* note 19, at 32–40. McLendon v. Continental Group, Inc., 749 F. Supp. 582 (D.N.J. 1989), *aff'd sub nom*. McLendon v. Continental Can Co., 908 F.2d 1171 (3d Cir. 1990). See also Gavalik v. Continental Can Co., 812 F.2d 834, 865–66 (3d Cir.), *cert. denied*, 484 U.S. 979 (1987).

<sup>65.</sup> In re DES Cases, 789 F. Supp. 552, 558 (E.D.N.Y. 1992), appeal dismissed, motion denied, 7 F.3d 20 (2d Cir. 1993).

<sup>67.</sup> E.g., Final Report, supra note 47, at 9–10 (epidemiological study regarding hundreds of applicants for Medicare benefits).

# III. Legal Authority for the Appointment of a Special Master

There are at least four sources of legal authority for the appointment of special masters by federal district court judges—the consent of the parties,<sup>68</sup> the court's inherent powers, <sup>69</sup> legislation providing for the appointment of magistrate judges as masters, <sup>70</sup> and Rule 53(b) of the Federal Rules of Civil Procedure. Rule 53(b) provides:

A reference to a master shall be the exception and not the rule. In actions to be tried by a jury, a reference shall be made only when the issues are complicated; in actions to be tried without a jury, save in matters of account and of difficult computation of damages, a reference shall be made only upon a showing that some exceptional condition requires it. Upon the consent of the parties, a magistrate may be designated to serve as a special master without regard to the provisions of this subdivision.

Title 28 § 636(b)(2) of the United States Code governs the use of U.S. magistrate judges as special masters. It provides:

68. Heckers v. Fowler, 69 U.S. (2 Wall.) 123, 126 (1864); Kimberly v. Arms, 129 U.S. 512, 524 (1889); Peretz v. United States, 111 S. Ct. 2661, 2668–69 (1991). See also Mobil Oil Corp. v. Altrech Indus., Inc., 117 F.R.D. 650 (C.D. Cal. 1987). Commentators seem to agree that references based on the consent of the litigants should not be subject to the same requirements that apply to references made without their consent, although they may not contravene applicable legislation or public policy. See, e.g., Peter G. McCabe, The Federal Magistrate Act of 1979, 16 Harv. J. on Legis. 343, 374–75 (1979); Linda J. Silberman, Masters and Magistrates, Part II: The American Analog, 50 N.Y.U. L. Rev. 1297, 1354 (1975); Brazil, supra note 17, at 312–14. To the extent that Article III and the doctrine of separation of powers limits references under Fed. R. Civ. P. 53(b), the parties cannot consent to measures that violate the separation of powers, though they may waive fairness objections. Commodity Futures Trading Comm'n v. Schor, 478 U.S. 833, 850–55 (1986). See also Erwin Chemerinsky, Federal Jurisdiction § 4.5 (1989).

69. Courts have inherent power to provide themselves with appropriate instruments for the performance of their duties; this power includes the authority to appoint persons unconnected with the court, such as special masters, auditors, examiners, and commissioners, with or without consent of the parties, to simplify and clarify issues and to make tentative findings. *In re* Peterson, 253 U.S. 300, 312–14 (1920). Reilly v. United States, 863 F.2d 149, 154–55 n.4 (1st Cir. 1988) (medical malpractice case). The court's inherent authority to appoint nonjudicial personnel to assist it in discharging its judicial responsibilities is limited, of course, by the boundaries of Article III. *See also* Burlington N.R.R. v. Department of Revenue, 934 F.2d 1064, 1073 (9th Cir. 1991); *In re* United States, 816 F.2d 1083, 1092 (6th Cir. 1987); *In re* Bituminous Coal Operators' Ass'n, 949 F.2d 1165, 1168 (D.C. Cir. 1991); *Kimberly*, 129 U.S. at 524. *But see In re* Armco, Inc., 770 F.2d 103, 105 (8th Cir. 1985) (dictum).

70. See also MCL 3d, supra note 43, §§ 21.52, 21.53. The Civil Rights Act of 1964 also provides that the court may use a magistrate judge as a master whenever a district court judge cannot schedule a case for trial within 120 days after issue has been joined. 42 U.S.C. § 2000e-5(f)(5) (1992).

A judge may designate a magistrate to serve as a special master pursuant to the applicable provisions of this title and the Federal Rules of Civil Procedure for the United States district courts. A judge may designate a magistrate to serve as a special master in any civil case, upon consent of the parties, without regard to the provisions of rule 53(b) of the Federal Rules of Civil Procedure for the United States district courts.

The following discussion is primarily confined to the appointment of nonmagistrate masters under Rule 53.

# A. Appointments Under Federal Rule of Civil Procedure 53(b)

Enacted as part of the Federal Rules of Civil Procedure in 1938, Rule 53(b) authorizes the appointment of special masters in jury cases only when the issues are complicated and in nonjury cases only when the matter is one of accounting or difficulty in computing damages, or one in which some exceptional condition requires it.<sup>71</sup> Therefore, the appointment of a special master should be the exception and not the rule. In *La Buy v. Howes Leather Co.*,<sup>72</sup> the Supreme Court held that the assignment of a complex antitrust case to a special master was improper under Rule 53 because the complicated legal issues, the complex nature of the proof, a congested court docket, and the possibility of a lengthy trial did not amount to exceptional conditions within the meaning of the rule.<sup>73</sup> However, it should be noted that in that case, the special master was assigned the full fact-finding function on the merits.<sup>74</sup> A more limited reference of nondispositive, pretrial, or remedial matters to the master might have been justified under the rule in those circumstances.<sup>75</sup>

## B. Limits on Broad-Scale Delegation

The decision in *La Buy* did not hold that a reference to the master under the circumstances of that case violated Article III of the Constitution, only that it was not warranted under the provision of Rule 53. Nevertheless, the Court has indicated that the delegation of essential judicial functions to personnel who are not judges appointed under Article III, with life tenure and protected salaries, violates the separation of powers doctrine and perhaps the due process clause unless the benefits of such delegation—efficiency and expertise—outweigh the diminution of Article III values—neutrality, independence, and adjudication.<sup>76</sup>

75. In re United States, 816 F.2d 1083, 1091 (6th Cir. 1987).

Reference Manual on Scientific Evidence

<sup>71.</sup> Fed. R. Civ. P. 53(b).

<sup>72. 352</sup> U.S. 249 (1957).

<sup>73.</sup> *Id.* at 259–60.

<sup>74.</sup> See Silberman, Judicial Adjuncts, supra note 1, at 2135.

<sup>76.</sup> Commodity Futures Trading Comm'n v. Schor, 478 U.S. 833, 850–51 (1986); Thomas v. Union Carbide Agric. Prods. Co., 473 U.S. 568, 584 (1985); Northern Pipeline Constr. Co. v. Marathon Pipe Line Co., 458 U.S. 50, 84 (1982).

Thus, the boundaries within which Rule 53 authority must be contained are established by Article III and the due process clause of the Constitution.

Special master appointments can be compared to the appointment of magistrate judges. Magistrate judges decide pretrial, nondispositive motions, <sup>77</sup> try civil cases with the consent of the parties, <sup>78</sup> and recommend decisions on dispositive motions. Nonconsensual references to magistrate judges have been sustained against constitutional attack where they were performed under a "district court's total control and jurisdiction." <sup>79</sup> Such references are adjunct in the sense that the magistrate judge has no independent authority to enforce orders, and dispositive decisions on the law and the facts are reviewed de novo. Special masters may perform some of these same functions.

However, several circuit courts of appeals have reversed appointments of special masters who were assigned to conduct formal evidentiary hearings on the merits of a case, finding that the appointments violated Article III. These courts found that at the stage of litigation, that is, the liability stage, Article III limitations controlled the scope of Rule 53, regardless of whether exceptional circumstances pertained. <sup>80</sup>

In Stauble v. Warrob,<sup>81</sup> the U.S. Court of Appeals for the First Circuit reversed a judgment rendered on the basis of a report by a special master. Mandamus had previously been denied. The First Circuit found that it could not "forge an 'exceptional condition' test for cases of blended liability and damages . . . . [T]he Constitution prohibits us from allowing the nonconsensual reference of a fundamental issue of liability to an adjudicator who does not possess the attributes that Article III demands."<sup>82</sup> Distinguishing the delegation of authority over remedy-related issues, the First Circuit held that where the fundamental determinations of liability are not heard and determined by the district court, the appointment is not within the constitutional limitations that bind Rule 53. The appeals court held that the district court lacked authority to refer the case without a provision for de novo review of the master's report.<sup>83</sup>

Other courts seem to require a greater showing of exceptional conditions to satisfy the requirements of Rule 53 where the appointment is made at the liability stage.<sup>84</sup> Thus, perhaps where liability is at issue, the need for expert assistance

81. 977 F.2d 690 (1st Cir. 1992).

83. Id. at 696.

<sup>77.</sup> Nondispositive motions decided by magistrate judges are reviewed on a "clearly erroneous" standard in accordance with the terms of 28 U.S.C. § 636(b)(1)(A).

<sup>78.</sup> Geras v. Lafayette Display Fixtures, 742 F.2d 1037, 1044–45 (7th Cir. 1984); Pacemaker Diagnostic Clinic v. Instromedix, Inc., 725 F.2d 537, 544 (9th Cir.), cert. denied , 469 U.S. 824 (1984); Caprera v. Jacobs, 790 F.2d 442, 444 (5th Cir. 1986), cert. denied, 484 U.S. 944 (1987).

<sup>79.</sup> United States v. Raddatz, 447 U.S. 667, 681 (1980) (reference of suppression motion in a criminal case to a magistrate judge did not violate Article III so long as the "ultimate decision is made by the district court"). *Id.* at 683.

<sup>80.</sup> E.g., In re Bituminous Coal Operators' Ass'n, 949 F.2d 1165, 1168-69 (D.C. Cir. 1991).

<sup>82.</sup> Id. at 695.

<sup>84.</sup> E.g., Burlington N.R.R. v. Department of Revenue, 934 F.2d 1064, 1070–73 (9th Cir. 1991) (no exceptional circumstances to support Fed. R. Civ. P. 53 reference of the entire case where reference was made

in dealing with complex evidence must be more clearly demonstrated. For example, in *Prudential Insurance Co. of America v. United States Gypsum Co.*, <sup>85</sup> the U.S. Court of Appeals for the Third Circuit granted mandamus withdrawing the appointment of a master who was to rule on nondispositive discovery motions but also hear dispositive legal motions (motions to dismiss and summary judgment motions) and report to the court "all relevant facts and conclusions of law." The Third Circuit found that the district court had not cited any exceptional conditions in the case or specific reasons for the appointment of a master beyond the district court's statement that the volume and breadth of documents and the inherent complexity of an asbestos litigation case merited the appointment. The court of appeals relied strongly on *La Buy* for its holding that neither the volume of work generated by the case nor the complexity of that work sufficed to meet the exceptional condition standards promulgated by Rule 53. The court also observed that a magistrate judge was available at no cost to the parties.

# C. Powers of Masters Under Federal Rule of Civil Procedure 53

Special masters appointed under Rule 53 have many of the same powers that a district court judge has to receive and evaluate scientific and technical evidence submitted by the parties. Unless the order of reference specifies otherwise, a special master has broad powers under Rule 53(c) "to regulate all proceedings in every hearing before the master and to do all acts and take all measures necessary or proper for the efficient performance of the master's duties under the order." <sup>86</sup> The master may require the production of documents and other evidence, rule on the admissibility of evidence, subpoena witnesses, place them under oath, and examine them.<sup>87</sup>

It is unclear what other powers, not enumerated in the rule, can be given to masters expressly or are assumed to be given if they are not limited by the order. For instance, it is unclear whether the powers granted remedial masters to gain access to documents and other information held by defendants can be exercised by a master appointed under Rule 53 if the order of reference does not permit or prohibit it.

# D. Appealing Appointments Under Federal Rule of Civil Procedure 53

The appointment of special masters under Rule 53(b) may be appealed through the extraordinary writ of mandamus brought immediately upon appointment in

Reference Manual on Scientific Evidence

<sup>&</sup>quot;in the interest of judicial economy" and the master's reported recommendations were affirmed in a onesentence order); *In re* Armco, Inc., 770 F.2d 103 (8th Cir. 1985) (in an environmental suit, circumstances sufficient to support a Fed. R. Civ. P. 53 appointment for pretrial duties, including disposition of summary judg ment and dismissal motions, were held insufficient to support master's authority to preside at trial).

<sup>85. 991</sup> F.2d 1080 (3d Cir. 1993).

<sup>86.</sup> Fed. R. Civ. P. 53(c).

<sup>87.</sup> ld.

the court of appeals<sup>88</sup> or through objection to and a general appeal of the district court's final judgment. <sup>89</sup> When the appointment is challenged by way of mandamus, the appellant must establish that the district court abused its discretion in making the appointment and that challenging it in an appeal will not adequately protect the interests at risk.<sup>90</sup>

After judgment, an appeal of reference to a master is treated as presenting a question of law, and plenary review will be exercised. <sup>91</sup> Because the standards of review are different, denial of a motion for mandamus setting aside a reference does not preclude a subsequent appeal which raises the issue again. <sup>92</sup>

88. La Buy v. Howes Leather Co., 352 U.S. 249 (1957). See, e.g., *In re* Bituminous Coal Operators' Ass'n, 949 F.2d 1165 (D.C. Cir. 1991) (reversing reference to special master of nonjury trial of civil case involving multiemployer trust fund where district court failed to reserve decision-making authority over motions dispositive of the merits of the case). A district court's refusal to appoint a master is not a final order and is not appeal-able until after final judgment. See 5A James W. Moore et al., Moore's Federal Practice ¶ 53.05[3], at 53-69, 53-71 to 53-73 & nn.7–11 (2d ed. 1992).

89. E.g., Stauble v. Warrob, 977 F.2d 690 (1st Cir. 1992) (special master appointment reversed on an appeal of the judgment on grounds that trial court exercised insufficient review over the master's findings in a commercial case); Liptak v. United States, 748 F.2d 1254, 1257 (8th Cir. 1984) (found reference to special master not supported by exceptional circumstances upon review of appeal from summary judgment). The party objecting to the appointment of a master must usually make a timely objection either at the time of appointment or promptly thereafter to preserve the assignment of error. See, e.g., Martin Oil Serv., Inc. v. Koch Refining Co., 718 F. Supp. 1334, 1337 (N.D. Ill. 1989); First Iowa Hydro Elec. Coop. v. Iowa-Illinois Gas & Elec. Co., 245 F.2d 613, 628 (8th Cir.), *cert. denied*, 355 U.S. 871 (1957); United States v. Conservation Chem. Co., 106 F.R.D. 210, 216 (W.D. Mo. 1985). See 5A Moore et al., *supra* note 88, ¶ 53.05[3], at 53-69 to 53-71 & nn.1–6.

90. Stauble, 977 F.2d at 693. See In re Fibreboard Corp., 893 F.2d 706, 707 (5th Cir. 1990) ("We are to issue the writ of mandamus only 'to remedy a clear usurpation of power or abuse of discretion' when 'no other adequate means of obtaining relief is available.'" (citations omitted)).

91. Stauble, 977 F.2d at 693.

92. Id. See also United States v. Shirley, 884 F.2d 1130, 1135 (9th Cir. 1989); Key v. Wise, 629 F.2d 1049, 1054–55 (5th Cir. 1980), cert. denied, 454 U.S. 1103 (1981).

# IV. Issues to Consider When Appointing a Special Master

The FJC's interviews with special masters and the judges who appointed them identified the following important issues to be considered when a judge contemplates the appointment of a master to deal with scientific or technical evidence:

- selection and qualification;
- avoiding conflict-of-interest and ethical problems;
- orders of reference—length and specificity;
- ex parte communications;
- potential liability for malfeasance;
- type of hearings;
- payment; and
- avoiding delay and inertia.

## A. Selection and Qualification

Judges seem to use three patterns of selection. In the first, judges simply select a master from among professional acquaintances, persons whose professional skills they admire and whose integrity and loyalty they trust. Most judges and special masters interviewed agreed that the most important qualification for a special master is the complete trust of the judge. Thus, where masters were expected to rule on scientific evidence presented by the parties in formal hearings and make recommended findings of fact at any stage of litigation, judge-like qualifications were sought and usually found in retired judges, former magistrate judges, or experienced hearing masters with whom the judge was acquainted. While seeking these qualifications opens judges to criticisms of favoritism, it may be difficult for judges to ensure the integrity and trustworthiness of masters by other means.

In the second pattern of selection, particularly when making pretrial appointments where settlement seemed possible, judges selected one or more candidates and sought the parties' approval. Although most of the masters interviewed were satisfied with the judicial nomination of master candidates, several settlement masters felt they could not be effective mediators unless the parties, at least, had agreed to their selection.<sup>93</sup> Even when making post-trial appointments, some judges hoped that the remedial master would be able to effect a settlement on outstanding damage and compliance issues, and they sought the parties' approval of the masters they selected. Although a hearing is not a prerequisite to appointing a master, judges will often ask the parties to interview several candidates for master or comment on the judge's proposed appointment of a nominee for master.<sup>94</sup> As will be discussed later, these interviews provide a useful forum in which to explore conflict-of-interest questions.

In the third pattern of selection, where scientific or technical expert assistance was needed to provide case management, investigate facts, hire experts, evaluate claims, and help the parties arrive at settlement, judges were more likely to permit parties to participate in the selection of a master by nominating candidates with particular skills.<sup>95</sup> These nominations were not treated as restricting the judge's discretion, but were respected, and judges were satisfied to select a candidate named by both sides.

Finally, where courts have sought recognized experts in their fields to observe and make findings regarding scientific facts, judges have relied less on their personal acquaintances and nominations from the parties, and more on referrals from other judges or the scientific community. In institutional reform litigation, judges sometimes sought experts through informal networks of judges and other experts and, at other times, through referrals from recognized professional societies.<sup>96</sup>

Judges seemed as satisfied with these selections as those of nonexperts with whom they were previously acquainted, usually because trust and respect between the judge and the expert quickly developed as the case progressed. Parties involved in the selection process felt invested in the choice and perhaps more willing to cooperate. Special masters interviewed who were selected through professional referrals or party nomination said they established good relations with the judge and did not believe that a lack of prior acquaintance was a disadvantage.

<sup>93.</sup> Some states require the consent of the parties to a master appointed to determine a controversy, but not one appointed to assist the parties in settlement. *In re* Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735, 741 (E.D.N.Y. 1990).

<sup>94.</sup> Gary W. v. Louisiana, 601 F.2d 240, 244 (5th Cir. 1979).

<sup>95.</sup> E.g., BIEC Int'l, Inc. v. Global Steel Servs., Ltd., 791 F. Supp. 489, 542 (E.D. Pa. 1992) (trade-secrets determinations referred to a Fed. R. Civ. P. 53 master). Each party proposed ten names of qualified persons to the court for selection.

<sup>96.</sup> There has been some interest in maintaining lists of persons qualified and interested in serving as masters in cases requiring scientific, technical, and other kinds of expertise. For example, the U.S. District Court for the District of Columbia has amended a provision to its local rules which directs the clerk of court to maintain "a list of special masters with experience in this Court and in other courts as a reference source." See Order of November 30, 1993 adopting Civil Justice Expense and Delay Reduction Plan and incorporating plan into the local rules of the court.

## B. Avoiding Conflict-of-Interest and Ethical Problems

U.S. judges are constrained by standards collectively known as judicial ethics, <sup>97</sup> which have a number of legal sources, including the *Code of Conduct for United States Judges*, <sup>98</sup> federal disqualification statutes, <sup>99</sup> financial disclosure requirements, <sup>100</sup> and the judicial oath of office. It is unclear which of these restrictions apply or should apply to special masters. <sup>101</sup> Although it is appropriate to disqualify candidates from serving as masters where they cannot provide neutral, objective determinations, it may be inappropriate to apply all judicial canons of ethics to special masters.

Some courts have reasoned that since masters are subject to control by the court and are needed for their expertise in particular subject matters, they should not be held to the strict standards of impartiality that apply to judges.<sup>102</sup> Other courts have concluded that because the "clearly erroneous" standard of review required by Rule 53 does not provide the district court with plenary control over a special master, the master's conduct must be held to the same high standards applicable to the conduct of judges.<sup>103</sup>

Several of the judges interviewed believed that masters are subject to the same ethical constraints as judges and entitled to the same judicial immunity, without qualification.<sup>104</sup> In the final analysis, the applicability of judicial ethical proscriptions to special masters may depend on what specific functions the masters perform.<sup>105</sup> Indeed, it has been proposed that a special code of ethics for special masters be developed to govern the particular relationships between judges, parties, and masters.<sup>106</sup>

The fact that appointment of special masters under Rule 53 assigns judicial tasks to people who are not full-time judges raises particular conflict-of-interest

101. The Code of Conduct applies in part to special masters and commissioners, as indicated in the section titled "Compliance with the Code of Conduct." 1993 Code of Conduct, *supra* note 98.

102. Morgan v. Kerrigan, 530 F.2d 401, 426 (1st Cir.), cert. denied, 426 U.S. 935 (1976).

<sup>97.</sup> See generally Beth Nolan, The Role of Judicial Ethics in the Discipline and Removal of Federal Judges n 1 Research Papers of the National Commission on Judicial Discipline & Removal 867 (1993).

<sup>98.</sup> Code of Conduct for United States Judges (November 1993) [hereinafter 1993 Code of Conduct].
99. 28 U.S.C. §§ 144, 455 (1988).

<sup>100. 5</sup> U.S.C. app. 6 §§ 101-112 (1992).

<sup>103.</sup> Jenkins v. Sterlacci, 849 F.2d 627, 630 (D.C. Cir. 1988); Belfiore v. New York Times Co., 826 F.2d 177, 185 (2d Cir. 1987), *cert. denied*, 484 U.S. 1067 (1988). *See also In re* Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735, 739 (E.D.N.Y. 1990) ("In general a special master or referee should be considered a judge for purposes of judicial ethics rules.") (citing Code of Judicial Conduct for United States Judges, 69 F.R.D. 273, 286 (1975)).

<sup>104.</sup> See also In re Gilbert, 276 U.S. 6, 9 (1928) (special masters assume the duties and obligations of a judicial officer); Jenkins, 849 F.2d at 630–31 (Code of Conduct for United States Judges applied to special master).

<sup>105.</sup> Jenkins, 849 F.2d at 630 n.1 ("[I]nsofar as special masters perform duties functionally equivalent to those performed by a judge, they must be held to the same standards as judges for purposes of disqualification.").

<sup>106.</sup> Jack B. Weinstein, Ethical Dilemmas in Mass Tort Litigation, 88 Nw. U. L. Rev. 469, 558 & n.352 (1994).

issues.<sup>107</sup> Practicing attorneys (and their firms) who are appointed masters have an interest in maintaining their professional reputations, sometimes as members of a plaintiffs' or defendants' bar, and in obtaining future employment. Such attorneys may have represented one of the parties in the past<sup>108</sup> or have litigated against lawyers who appear before them as masters.<sup>109</sup> Retired judges have an interest in being appointed to future cases; some also maintain private law practices. Law professors may have ideological positions and academic credentials that can affect, or be affected by, their performance as masters.<sup>110</sup> Nonlegal experts, such as prison experts, sometimes have been hired as expert witnesses in previous litigation involving the parties whom they monitor as special masters,<sup>111</sup> or they hope to be hired by such parties in the future. Finally, a small group of "repeat players" has developed—masters who have served in many cases<sup>112</sup> and are invested in their reputations as successful settlement masters.

What steps can be taken to ensure that conflicts of interest do not affect the performance of the master? Courts have dealt with conflict-of-interest issues in several ways. First, most of the judges interviewed indicated that they provide some opportunity, either at a formal hearing on a motion to appoint a master or at a more informal conference with attorneys, for the parties to question the master about possible conflicts of interest and to raise any objections they might have before the appointment is made.

Some judges and masters want any suggestion of conflicts fully disclosed on the record so that parties who do not object will be estopped from complaining later.<sup>113</sup> In some cases, waiver is implied by the parties' agreeing to the appointment or to a subsequent party settlement with knowledge of the alleged conflict. In one case, a party that subsequently objected was estopped from doing so.

110. E.g., Prudential Ins. Co. of Am. v. United States Gypsum Co., 991 F.2d 1080, 1088 n.13 (3d Cir. 1993) (court noted allegations of bias based on law school dean's academic writings).

111. E.g., Lister v. Commissioners Court, Navarro County, 566 F.2d 490, 493 (5th Cir. 1978) (appointment of a special master, who had testified as an expert witness for the plaintiffs in the same suit, to devise a reapportionment plan held improper) (citing *In re* Gilbert, 276 U.S. 6 (1928)).

112. Weinstein, supra note 106, at 558.

113. Where a master is appointed to facilitate a settlement, parties who object to a conflict of interest after appointment of the master may withhold their agreement to a settlement by way of objection. Continued participation is seen as a continuing waiver of any objection.

<sup>107.</sup> See, e.g., United States v. Lewis, 308 F.2d 453, 457 (9th Cir. 1962) ("[T]hose qualified to act as ... [masters] in a particular area are likely to have had prior association with those qualified ... as expert witnesses from that area .... [T]he test should be whether [actual] abuse appears."). See generally Silberman, Judicial Adjuncts, supra note 1, at 2159–61.

<sup>108.</sup> In *In re* Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735 (E.D.N.Y. 1990), a motion to disqualify a special master was denied where the master was appointed to act as a settlement master in cases involving asbestos exposure, and where the master and his firm had acted on behalf of the moving defendant in connection with legislative efforts in the past. The court observed, "As an officer of the court the special master remains bound to respect the confidentiality of and refrain from using to [defendant's] disadvantage any information imparted to him under seal of confidentiality by that company in the course of his legislative or mediation efforts." *Id.* at 742.

<sup>109.</sup> Cf. Id.; Mister v. Illinois Cent. Gulf R.R., 790 F. Supp. 1411, 1417 (S.D. Ill. 1992) (special master in this case was plaintiff's attorney in another case in which the same expert appeared for the defense as appeared before him in this case).

Some courts take steps to eliminate conflicts by restricting the master's subsequent employment by either party or the master's concurrent representation (or that of the master's firm) of other parties with conflicting interests. Addressing conflict-of-interest matters expressly at the time of appointment avoids later controversy.

# C. Orders of Reference

Orders of reference to the master reviewed in the FJC study varied from very short orders that made general assignments to lengthy, detailed orders. A correlation did not seem to exist between the length and specificity of the order and the complexity of the tasks assigned. In fact, some of the most complex tasks (e.g., those involving case management of a mass tort case) were assigned in short, general orders.

Some of the issues that are addressed in orders of reference are the following:

- scope and limitations on authority (i.e., functions assigned and specific authority to carry them out);<sup>114</sup>
- scope of the master's investigative authority;<sup>115</sup>
- discovery rights to evidence supporting the master's findings;
- disclosure of conflicts of interest;
- scope of review;
- periodic reporting requirements;
- duration of the appointment;
- standards of performance;
- periodic accountings—approval by the court;
- compensation—rate and manner of payment;
- ex parte communications with the judge;
- ex parte communications with the parties;
- ex parte communications with the experts and third parties;
- liability and immunity of the master (insurance and bonds); and
- expiration of the appointment.

Sometimes the parties negotiate the terms of orders of reference and propose them to the court by motion or in conference. The court must consider whether

114. Fed. R. Civ. P. 53(c) provides:

The order of reference to the master may specify or limit the master's powers and may direct the master to report only upon particular issues or to do or perform particular acts or to receive and report evidence only and may fix the time and place for beginning and closing the hearings and for the filing of the master's report.

When a U.S. magistrate judge is appointed to be a special master, the order of reference should refer to Fed. R. Civ. P. 53.

<sup>115.</sup> Particularly in the remedial stage of litigation, some courts have granted masters broad access to information held by the parties, whereas others have disapproved of such grants. *Compare* United States v. Parma, 504 F. Supp. 913, 925 (N.D. Ohio 1980), *aff'd in part, rev'd in part,* 661 F.2d 562 (6th Cir. 1981), *cart. denied*, 456 U.S. 926 (1982) *with* Reed v. Cleveland Bd. of Educ., 607 F.2d 737, 741, 743–44 (6th Cir. 1979).

the parties will be allowed to determine the extent of the master's authority and the procedures to be used, or whether the court and the master will determine those matters initially or as litigation progresses.

Some judges believed that detailed orders worked well where the case was especially contentious and disputes over particular issues could be anticipated and resolved in advance, or where the case involved basic issues of fairness, such as ex parte communications. Many masters, however, felt that, where possible, it was wise to give the master flexibility in resolving disputes as they arise. One discovery master attributed his success in avoiding disputes in a large, complex case to the fact that he engaged the parties in negotiating informal procedures, which reduced the need for paper exchanges and motions. Such procedures included involving the master in depositions—either by attending particularly difficult ones or being available to settle disputes by phone. It might have been difficult to include such detailed, but successful, techniques in an order of reference.

Thus, issues which go to the propriety of the appointment itself—conflicts of interest, ex parte communications, scope of authority—might well be addressed expressly in the order of reference, whereas procedural issues—the discovery process, the appointment of experts, formal hearing procedures—might be left to negotiation between the master and the parties after the appointment. Express terms in the order of reference place the parties on notice with regard to essential characteristics of the appointment and permit them to object and seek mandamus if they choose. Procedural matters determined by the master can be appealed to the judge after the appointment.

## D. Ex Parte Communications

The question whether ex parte communications between a special master and the parties and between the special master and the judge should be permitted under Rule 53 is highly debatable.<sup>116</sup> The issue may be important, especially in cases involving sophisticated scientific and technical evidence, because masters, as well as judges, are more likely to seek education from experts outside the adversary process. Nevertheless, the positions with regard to these issues are the same, whether raised in a scientifically complex case or otherwise.

Rule 53 expressly permits a master to proceed ex parte when a party fails to appear for certain meetings. Such explicit permission may be interpreted to prohibit other ex parte proceedings.<sup>117</sup> Thus, some judges hold that ex parte com-

If a party fails to appear at the time and place appointed [for meeting with the parties], the master may proceed ex parte or, in the master's discretion, adjourn the proceedings to a future day, giving notice to the absent party of the adjournment.

<sup>116.</sup> See DeGraw, supra note 11, at 816–20 & nn.95–124.

<sup>117.</sup> Fed. R. Civ. P. 53(d)(1) provides:

munications are improper for both the master and the judge<sup>118</sup> and must be prohibited whether or not the parties consent to them.<sup>119</sup> Others maintain that masters may communicate ex parte with the parties and the judge if the order of reference expressly permits it.<sup>120</sup> Still others hold such communication proper only if expressly consented to by the parties.

#### 1. With the parties

Those who find exparte communication between the master and the parties improper under any circumstances adhere to a traditional, adversary justice model in which a judge passively receives information solely from the parties and decides the dispute on the basis of that information.<sup>121</sup> To the extent that special masters function as judges, some judges interviewed felt it was improper for masters to consult the parties separately and to communicate the information thus gained to the judge outside the presence of the parties because to do so deprives them of the opportunity to raise challenges. Thus, it was understood that in many cases the judge did not want to hear information the master obtained from the parties ex parte.

The proponents of ex parte communication argue that many judicial functions, such as case management and settlement facilitation, require a judge to play a more active, nontraditional litigation role in which exparte communication is appropriate.<sup>122</sup> Therefore, when mediating disputes before trial or facilitating the settlement of damage issues after determinations of liability, judges often individually consult the lawyers, parties, insurance companies, and others to gain information necessary to their task.<sup>123</sup> When masters perform these same functions, it is believed they, too, may engage properly in ex parte communications. 124

121. Fleming James, Jr., & Geoffrey C. Hazard, Jr., Civil Procedure § 1.2 (3d ed. 1985); Judith Resnik, Managerial Judges, 96 Harv. L. Rev. 374 (1982); Peter H. Schuck, The Role of Judges in Settling Complex Cases: The Agent Orange Example, 53 U. Chi. L. Rev. 337 (1986).

<sup>118.</sup> Model Code of Judicial Conduct Canon 3B(7) (1990) provides: "A judge shall not initiate, permit, or consider ex parte communications, or consider other communications made to the judge outside the presence of the parties concerning a pending or impending proceeding . . . ." 119. See, e.g., D'Acquisto v. Washington, 640 F. Supp. 594, 621–22 (N.D. Ill. 1986) ("[A]n ex parte com-

munication is a communication about a case which an adversary makes to the decision maker without notice to an affected party.").

<sup>120.</sup> See, e.g., Alberti v. Klevenhagen, 660 F. Supp. 605, 610 (S.D. Tex. 1987); Ruiz v. Estelle, 679 F.2d 1115, 1170 (5th Cir.), amended in part, vacated in part, 688 F.2d 266 (5th Cir. 1982), cert. denied, 460 U.S. 1042 (1983); Thompson v. Enomoto, 815 F.2d 1323, 1326 n.7 (9th Cir. 1987).

<sup>122.</sup> For example, Fed. R. Civ. P. 16(a) permits judges to discuss settlement at pretrial conferences. 123. In *In re* Joint E. & S. Dists. Asbestos Litig., 737 F. Supp. 735, 739 (E.D.N.Y. 1990), Judge Jack B. Weinstein wrote that "[i]t is standard practice for the presiding judge or magistrate to meet separately with each of the parties for a candid discussion of strategy and the needs of the party." He also stated that "[i]nformation revealed to the mediator should include—absent a confidential communication privilege—relationships to insurers, overall strategy, corporate politics and the like . . . . The role of the mediator is often that of the honest broker . .

<sup>124.</sup> United States v. Conservation Chem. Co., 106 F.R.D. 210, 234-35 (W.D. Mo. 1985) (master appointment was not revoked where master engaged in settlement negotiations ex parte with the parties and the record was void of any evidence suggesting that the master's impartiality might reasonably be questioned).

In fact, some masters interviewed stated that assignment of settlement and mediation to a master insulates the judge from ex parte discussions and permits the judge to subsequently try the case without prejudgment based on those communications. However, these masters indicated that they mitigated the effect of exparte communications in several ways. First, they notified other parties when important information was imparted ex parte by one party. Furthermore, they circulated their findings of fact based on informal, ex parte information in draft to the parties before they reported them to the judge. Thus, if a party wanted to challenge ex parte facts, it could do so, usually in writing, before the master reported to the court. In addition, parties who renewed their objections to the master's report when it was submitted to the court were given a de novo hearing on the disputed facts, rather than the court applying the "clearly erroneous" test.<sup>125</sup> At these hearings, parties were permitted to cross-examine witnesses relied on by the master in making his or her findings and present their own witnesses. Thus, the prejudice that might result from ex parte communication was eliminated.<sup>126</sup> Finally, some judges stated in the order of reference that exparte communications between the master and the parties would be permitted.<sup>127</sup> When no objections were made in these cases, the parties arguably had agreed to such procedures.

To avoid the appearance of impropriety and subsequent objections of the parties to informal fact-finding procedures, the judge should address ex parte communications between the master and the parties directly in the order of reference. Where objections are made to the order of reference, limitations on ex parte communications may be necessary to avoid charges that the order violates due process and the guarantees of Article III of the Constitution.

#### 2. With the judge

Most masters stated that they believed they could not perform their assigned duties effectively if they were prohibited from discussing scheduling, strategies, and procedures with the judge outside the presence of the parties. Yet, several judges indicated that they felt uncomfortable meeting with masters without the parties being present. Most judges met with their appointed master less than once a

<sup>125.</sup> See Alberti v. Klevenhagen, 660 F. Supp. 605, 611 (S.D. Tex. 1987); Ruiz v. Estelle, 679 F.2d 1115, 1163 (5th Cir.), amended in part, vacated in part, 688 F.2d 266 (5th Cir. 1982), cert. denied, 460 U.S. 1042 (1983); Special Project, The Remedial Process in Institutional Reform Litigation, 78 Colum. L. Rev. 784, 829 – 30 (1978) [hereinafter The Remedial Process ].

<sup>126.</sup> For example, in Morales Feliciano v. Romero Barcelo, 672 F. Supp. 591, 626 (D.P.R. 1986), the master and experts toured the prison, spoke to inmates and staff, and made findings of fact about the adequacy of the medical care inmates received. Often, ex parte communication between the defendants and the master was the basis for the findings. Only if the plaintiffs objected to the findings, recounted in the master's report to the judge, were the findings reviewed by the court in any formal way. In those instances where a party objected to informal fact finding, though in a nonjury case, the court did not use a "clearly erroneous" standard of review. Instead, the court held hearings de novo on the factual finding contested. But see Gary W. v. Louisiana Dep't of Health & Human Resources, 861 F.2d 1366, 1368–69 (5th Cir. 1988).

<sup>127.</sup> See supra note 120 and accompanying text.

month and conveyed to the master that they were not interested in hearing any ex parte information from the parties. One judge allowed the master to communicate with him only through monthly reports, written as letters, which were also provided to the parties.

The propriety of ex parte communication can depend on whether the master is characterized as a judicial agent or as an outside adjunct. If the master is viewed as an agent of the court, it is proper for the judge, as principal, to discuss with the master, as agent, the performance of his or her duties. If the master is viewed as an adjunct, it is improper for the judge, as ultimate decision maker, to receive from the master, a non-party, evidence and information that could influence the judge's decisions, or that reasonably might be thought to do so.

It may be more useful, however, to consider the purposes for which the appointment was made to determine whether ex parte communication is consistent with, or will further, the execution of the master's functions. Where, for example, the purpose of the appointment is to obtain the master's recommended findings of fact, ex parte communication with the judge seems inappropriate because the judge will review those findings and the record upon which they are based to determine whether they are clearly erroneous. Information received off the record could prejudice that review. Similarly, most of the masters and judges interviewed indicated that where the master's role is one of mediator and facilitator, information relating to the substance of proposed settlements and the facts of the case should not be communicated to the judge ex parte during the settlement process.

However, a master appointed as an expert to advise the court might appropriately communicate with the judge privately in order to provide the one-on-one education some judges desire, so long as the master does not provide the judge with personal opinions on facts and evidence upon which the judge will rule. Masters who bring their expertise in quantitative analysis to bear on the presentation of data seem to serve a similar role; a judge's private discussions with a master regarding quantitative analyses do not seem to prejudice the judge's independence or the parties' abilities to present their case, so long as the judge permits the parties an opportunity to contest his or her acceptance of rejection of the data.

## E. Potential Liability for Malfeasance

Most of the masters and judges interviewed believed that by being appointed under Rule 53, masters acquired judicial immunity and were not exposed to much, if any, legal liability for dereliction of duty. <sup>128</sup> Some masters appointed to deal with complex scientific and technical issues were involved in cases in

<sup>128.</sup> One judge interviewed for the FJC's study appointed a person under Fed. R. Civ. P. 53 to execute a contract necessary to carry out the court's remedial decree solely to provide that person with judicial immunity.

which large sums of money were at stake, and they did not believe they could obtain enough insurance coverage for all of the potential liability.<sup>129</sup> Others worried about the expense of defending such suits, even if they thought they could not be found personally liable.<sup>130</sup>

Again, the question of a master's liability may hinge on the nature of the functions the master performs. To the extent that masters perform functions that are essentially juridical, they may enjoy judicial immunity. Yet, absolute judicial immunity does not extend to all the functions performed by judges. In suits against judges for personal liability, courts have taken a functional approach, distinguishing the judges' adjudicatory functions from their administrative, managerial, and executive functions.<sup>131</sup> Although judges have qualified, good-faith immunity for actions that do not violate clearly established statutory or constitutional rights a reasonable person would have known of,<sup>132</sup> they may be held liable for administrative actions that do not meet that standard. Presumably, the court cannot provide a master with more immunity than a judge would have in performing the same tasks.

Special masters appointed to deal with scientific and technical issues, particularly in suits involving many claimants and large sums of money, may want to investigate their potential liability and explore the possibility of purchasing malpractice insurance to cover it. The cost of such insurance is a legitimate cost of the appointment and possibly could be included in the expenses for which parties are liable under the rules. Where special masters administer large settlement funds, fiduciary bonding requirements may apply.

## F. Type of Hearings

While Rule 53 anticipates the appointment of a master to hear evidence and make factual findings, many masters are appointed to perform other tasks that require informal fact finding. Thus, although masters usually are not authorized to conduct private investigations into the matters referred, they often are expected to use their personal expertise and knowledge and obtain other expert opinion in evaluating evidence outside of formal proceedings.<sup>133</sup>

Almost all of the masters interviewed engaged in informal fact finding of one kind or another and conducted no formal hearings at all. Most felt that formal

<sup>129.</sup> See discussion in Brazil, supra note 9, at 409.

<sup>130. 1</sup>C Guide to Judiciary Policies and Procedures, ch. XI, pt. E, § 2.1, at 31 (1991), provides that a judge who desires legal representation in a suit for personal liability for official acts performed within the scope of his or her employment may request Department of Justice representation or reimbursement for private representation. However, it does not provide the same for masters.

<sup>131.</sup> Forrester v. White, 484 U.S. 219, 223–24 (1988). The appointment of a special master pursuant to Fed. R. Civ. P. 53 seems to be a judicial function, not an administrative one.

<sup>132.</sup> Harlow v. Fitzgerald, 457 U.S. 800, 818 (1982).

<sup>133.</sup> See Ruiz v. Estelle, 679 F.2d 1115, 1162 (5th Cir.) (master is not precluded from conducting a viewing such as that permitted for judges and juries), *amended in part*, vacated in part, 688 F.2d 266 (5th Cir. 1982), cert. denied, 460 U.S. 1042 (1983).

hearings were expensive and time-consuming and reduced the collegial relationships they often wanted to develop with the parties. Retired judges serving as masters were more likely to hold formal hearings than law professors or practitioners. Hearings apparently were held more often to try scientific facts going to the merits of a case than to resolve factual issues that arose in connection with discovery or the monitoring of compliance with court orders. Some masters used the possibility of formal hearings as an incentive for the parties to cooperate and make more informal procedures work efficiently. Others let the parties decide whether they wanted particular issues tried in formal proceedings.

Masters are more likely to use informal fact finding in the remedial stages of litigation than in discovery.<sup>134</sup> This may be due to the need at the remedial stage for expert evaluation of the ongoing performance of defendants, which can be perceived best as it occurs, rather than as it is related in a courtroom after the fact. As mentioned earlier, when informal fact-finding procedures (i.e., reports from experts, viewings, and ex parte information from parties and other witnesses) were used as the basis for the findings the master reported to the court, parties were permitted to challenge those findings in de novo hearings before the judge.<sup>135</sup>

Particularly with the consent of parties, such informal proceedings, along with an opportunity for later de novo review of findings of scientific fact, seem to provide an efficient and fair means of assisting courts in processing scientific and technical information. The order of reference could address explicitly the weight to be accorded recommended findings of scientific fact based on informal factfinding procedures. Furthermore, the order could expressly grant the master authority to either hold formal hearings or proceed informally, with or without the consent of the parties.

## G. Payment

Besides delay, the reason most often given for limiting the appointment of masters is the added expense for the litigants.<sup>136</sup> Rule 53 provides that "compensation to be allowed to a master shall be fixed by the court, and shall be charged upon such of the parties or paid out of any fund or subject matter of the action, which is in the custody and control of the court as the court may di-

134. Stephen E. Taylor & Howard A. Herman, Using Special Masters in Complex Civil Litigation, 4 Fed. Litig. Guide Rep. (MB) No. 1, at 3 (Jan. 1992).

Special Masters

<sup>135.</sup> See discussion infra § V.D.1.

<sup>136.</sup> E.g., Fraver v. Studebaker Corp., 11 F.R.D. 94, 95 (W.D. Pa. 1950) (motion for appointment of master in patent suit denied because of burdensome cost to plaintiff). In reviewing challenges to appointments of a special master, appellate courts are much influenced by the fact that similar services may be available from a magistrate judge at no cost to the parties. Particularly in the context of scientific and technical evidence, therefore, justification for the appointment of a lay special master may need to be based on the special expertise the lay master alone can supply.

rect."<sup>137</sup> One master interviewed suggested that reference to a master sometimes is regarded as punishment for parties who are uncooperative: The parties must now pay for what they could have received free.

The master's compensation is set by the court, and the judge allocates it to the parties as a cost. In some suits, where one party was impecunious or the other was blameworthy, judges allocated the entire cost to one party or divided it among several defendants<sup>138</sup> or amici. <sup>139</sup> Considerable variation exists in the standards judges use to determine the rate of the master's compensation. The Supreme Court has adopted a "liberal but not exorbitant" standard for compensation of masters,<sup>140</sup> which leaves room for interpretation.

Most often, the master's rate is set in relation to the market in which the master—as a private practitioner, retired judge, academic, or scientific or technical expert consultant—could otherwise sell his or her services.<sup>141</sup> Where the master to be appointed was a private attorney, some judges have considered the specialized area in which the master had a private practice. Others have considered the usual hourly rate for private practitioners in the area of specialty at issue in the suit and in the locality where the suit is brought, regardless of whether the practitioner was practicing there. Still others have discounted such commercial rates for the "public service" nature of the case.

Where a master has skills as a legal practitioner and a technical expert, as do some of the expert prison masters, the court must decide which of the two markets to use as a basis for the master's fee. If the master will use both sets of skills and they can only be procured by others at the higher rate, the master should be paid the higher rate. Although academics sometimes were given their usual consulting rate or the prevailing practice rate, more often they were given a rate reflecting the fact that they did not regularly sell their services in a private market and that they enjoyed the low- or no-risk position of full-time, tenured professors whose law schools paid their overhead. Some judges interviewed said that they asked masters to discount their fees to "subsidize" justice in the public interest.

137. Fed. R. Civ. P. 53(a). See generally David I. Levine, Calculating Fees of Special Masters, 37 Hastings L.J. 141 (1985).

138. See, e.g., Hart v. Community Sch. Bd., 383 F. Supp. 699, 767 (E.D.N.Y. 1974) (holding that court had broad discretion to allocate to the defendant whose action necessitated the school desegregation suit the costs of the master and his required supportive services), *aff'd*, 512 F.2d 37 (2d Cir. 1975).

139. Nebraska v. Wyoming, 112 S. Ct. 2267 (1992) (Supreme Court approved the one-time assessment of special master costs to intervenors/amici where no party or intervenor/amici objected to the propriety of including nonobjecting amici in the assessment, and the proceedings were longer and more costly because of their participation. Justice Stevens dissented, finding nonobjection problematic because it was an interim payment and citing judicial code sections limiting circumstances in which parties may waive judicial disqualification.).

140. Newton v. Consolidated Gas Co., 259 U.S. 101, 105 (1922). The Court recognized that "while salaries prescribed by law for judicial officers performing similar duties are valuable guides, a higher rate of compensation is generally necessary in order to secure ability and experience in an exacting and temporary employment which often seriously interferes with other undertakings."

141. See Reed v. Cleveland Bd. of Educ., 607 F.2d 737, 746 (6th Cir. 1979) (considerations applicable to awarding attorneys' fees apply to setting fees for masters: rate set at the outset at one-half the highest rate of local law firms and two-thirds the average rate of experienced local trial attorneys). See *also* General Motors Corp. v. Circulators & Devices Mfg. Corp., 67 F. Supp. 745, 747–48 (S.D.N.Y. 1946).

A few judges stated that they did not determine the rate, but allowed the parties and the master to negotiate a rate and report back to the court.<sup>142</sup> Both the masters and parties in those cases apparently were satisfied with that process, although it would raise questions about possible bias of the master in cases in which only one party compensated the master.

Apart from the master's rate, expenses usually are billed separately. The law professors interviewed tended to use paid assistants and billed their services separately as well. Some professors hired small support staffs housed in quarters outside the law school to help them carry out their duties, for which they billed separately. Other professors seemed to manage huge cases with little or no assistance or special space. Nevertheless, the separate expenses of these masters (particularly masters who compiled empirical, statistical data themselves in an effort to evaluate and settle claims) were significant and need to be considered in determining whether the aggregate expenses of the master's efforts are costeffective.

Most judges interviewed kept a watchful eye on the compensation paid to masters, but others felt the costs were a matter between the master and the parties. Masters varied in the timing, detail, and frequency of their statements of fees and expenses. Some provided extremely detailed descriptions of meetings, telephone calls, research, and reading; others provided a summary statement of hours, rates, and expenses only. Most masters submitted accounts monthly or quarterly. Some courts required that payments be approved before they were incurred or before they were disbursed, eliminating the possibility of later objections to the amount or purpose of particular expenses. Because of the large amount of money often paid to masters in some complex cases, a detailed breakdown of a master's expenses on the record provides assurance that the expenses are justified in the view of the court.

Masters were paid in several ways. In some cases, payment was made to masters through the court registry (i.e., after submission of the accounting, the charged party paid the approved amount into the court registry, whose clerk made out a check to the master). This procedure makes payment part of the court record and may promote a perception that the master is the court's agent and not the agent of one of the parties.

In other cases, the court ordered the creation of a pool, funded by the parties charged, from which the master's compensation and expenses were paid upon approval by the court. The location of such funds, whether in a court account or outside of the court, may have a bearing on whether the interest on the fund is taxed. Some masters maintained that the pool tied up the parties' funds unnecessarily; others felt that it removed any leverage that responsibility for payment might give the parties.

<sup>142.</sup> This approach has been disapproved by at least one court. Finance Comm. v. Warren, 82 F. 525, 528 (7th Cir. 1897).

Some masters were paid directly by the charged party upon submission of accountings for the court's approval. In one case, an institutional defendant who was paying the full cost of the master simply put the master on the institution's regular payroll. Such an arrangement, however, may present questions of bias by the master and an appearance of partiality.

It is difficult to determine the total cost of many references to masters or whether the use of masters is cost-effective. The expenses of some expert masters in institutional reform litigation have amounted to more than \$1 million a year, <sup>143</sup> whereas some masters have agreed to cap their fees at a nominal amount to make it clear that their work was done in the public interest.<sup>144</sup>

Although there are no costs imposed on the parties directly when discovery, settlement, matters of account, remedial decrees, and the monitoring of orders are not assigned to a master, much higher costs may be incurred by taxpayers when these functions are performed by judges. Furthermore, greater costs may be imposed on the parties indirectly through protracted litigation and unrealized settlements that masters might have effected.<sup>145</sup> Nevertheless, some judges viewed the use of masters as a mechanism that puts a price on justice—a price too high for some litigants to bear. Thus, they were reluctant to appoint masters except where the magnitude of the litigation, the amount at issue, and the financial position of the parties warranted it. Some attorneys privately objected to the fees and expenses charged by masters but were reluctant to voice their objections for fear of alienating an important decision maker in their case and possibly the judge who appointed him or her.

## H. Avoiding Delay and Inertia

Along with costs, delay is cited most often as a reason why masters should not be appointed except in extraordinary circumstances. Particularly in institutional reform litigation and desegregation suits, masters have been known to serve for decades. Some masters interviewed believed that all the parties—masters, attorneys, and judges-become invested in the arrangement and consciously or unconsciously perpetuate their employment. However, examples can be found of some institutional reform masters who have dispatched their responsibilities quickly and efficiently.

One method of avoiding delay is to specify a termination date in the order of reference. Some masters and judges interviewed felt that time-limited appoint-

<sup>143.</sup> See Levine, supra note 137, at 143 n.8; Reed v. Rhodes, 691 F.2d 266 (6th Cir. 1982) (interim attorneys' fees of \$511,261.98 charged by special masters). 144. Fox v. Bowen, 656 F. Supp. 1236, 1254 (D. Conn. 1987) (The special master's fees were capped at

<sup>\$5,000.)</sup> Professor Robert A. Burt served as master for five years. See Final Report, supra note 47.

<sup>145.</sup> Judge Jack B. Weinstein noted that failure to consolidate cases and take other steps to control litigation leads to increased transaction costs. In re Repetitive Stress Injury Cases, 142 F.R.D. 584, 586-87 (E.D.N.Y. 1992), appeal dismissed, order vacated sub nom. In re Repetitive Stress Injury Litig., 11 F.3d 368 (2d Cir. 1993).
ments, particularly before liability is determined, help promote early negotiations and settlement, since the parties are aware that failure to settle will result in the expense of a trial. Other judges felt that keeping abreast of a case through frequent reports and occasional hearings, and imposing pretrial discovery deadlines help move a case along, without limiting the tenure of the master.

When the appointment is made at the remedial stage, some masters feared that imposing a termination date would promote intentional delay on the part of recalcitrant defendants, who would wait for the master's appointment to end. Furthermore, a deadline in the order usually places the burden of justifying an extension of the master's tenure on the successful party in the litigation to ensure that recalcitrant defendants are monitored. Nevertheless, remedial masters' tenures last for years. The establishment of time-limited goals in the order of reference, and the consequences of not meeting those goals, may be a more fruitful means of securing compliance with court-ordered remedial action than the open-ended appointment of a master. Where masters are appointed to provide scientific expertise, their function may be performed earlier than the conclusion of the entire matter, and their tenure can be limited accordingly.

# V. Use of Special Master's Report and Review Given

Special masters appointed under Federal Rule of Civil Procedure 53 are required by the rule to file with the clerk of court a report setting forth findings of fact and conclusions of law as required by the order of reference. <sup>146</sup> The weight given to a master's report depends on whether it is rendered in a jury proceeding or a nonjury proceeding. In a trial before a jury, the master is treated as a source of evidence, and the master's report is to be considered by the jury in its deliberations. In a nonjury trial, the master is treated as a preliminary decision maker, whose recommended findings of fact and conclusions of law are accepted unless they are clearly erroneous.

## A. Jury Trials

Where masters make findings of fact based on scientific and technical evidence in a jury case, the master's report is admitted for the jury's consideration much like the testimony of an expert witness who has heard evidence bearing on the matter, but much of the scientific evidence upon which the report is based will be excluded from the record unless the parties introduce it independently at trial.<sup>147</sup> Thus, in a jury trial, the master's findings, but not the scientific evidence on which they are based, are admissible as evidence and may be read to the jury, subject to objections.<sup>148</sup> Having participated in the evidentiary hearing before the master, parties objecting to the master's report in a jury trial are not entitled to discover the evidence on which it is based.<sup>149</sup> Evidence at variance with the report may be introduced at trial for the jury's consideration, at least if the offering party can show good cause why it was not presented to the master.<sup>150</sup>

<sup>146.</sup> Fed. R. Civ. P. 53(e)(1). Effective December 1991, the master must file the report with the clerk of court and serve a copy on all parties. Fed. R. Civ. P. 53 advisory committee's note relating to 1991 amendment.

<sup>147. 5</sup> Moore et al., supra note 88, § 53.14[4], at 53-136.

<sup>148.</sup> Burgess v. Williams, 302 F.2d 91, 94 (4th Cir. 1962) (master's report given prima facie effect). See generally 5A Moore et al., supra note 88, ¶ 53.10[1], at 53-90 & ¶ 53.14[4], at 53-136.

<sup>149.</sup> The master's findings "are to be received in evidence at the trial by means of a written report. The parties have access to the master's report prior to the trial and limited rights to present objections to the court; they may not, however, conduct other pretrial discovery with respect to the master's findings or examine the master at the trial." MCL 3d, *supra* note 43, § 21.52. *Cf.* United States v. Cline, 388 F.2d 294 (4th Cir. 1968) (master se lected for his expertise as a surveyor to conduct a boundary survey functioned as an expert rather than a common-law master and was subject to questioning by both parties regarding his reported findings).

<sup>150.</sup> Cf. Gay v. United States, 118 F.2d 160, 162 (7th Cir. 1941).

Thus, an objecting party may introduce expert scientific testimony from a witness who has testified before the master. Unlike a court-appointed expert, a special master may not be cross-examined on his or her report. Yet, unlike other testimonial evidence which a jury may find incredible, the findings of a master constitute prima facie evidence which, standing alone, is sufficient to sustain a directed verdict.<sup>151</sup>

## B. Nonjury Trials

In nonjury trials, masters' findings must be accepted by district courts unless they are clearly erroneous, <sup>152</sup> whether the master made the findings in the course of settling a discovery dispute, <sup>153</sup> recommending action on a motion for injunctive relief, <sup>154</sup> or supervising implementation of court-ordered relief. <sup>155</sup>

When a master's report is submitted in a nonjury trial, the master must file a transcript of the proceedings and of the evidence as well as the original exhibits on which the report is based, so that the court can review the evidence and decide whether the master's findings are not clearly erroneous and, therefore, must be sustained.<sup>156</sup> In a recent case, failure to provide this kind of review of the written record provided by the master, who held thirty-nine days of hearings on the issue of liability, was the basis for the reversal of a district court judgment and a remand for retrial.<sup>157</sup>

## C. On Appeal

In appealing a district court's ruling adopting, modifying, or rejecting a master's recommended findings of fact, the appellant has the usual burden of persuading the court of appeals that the district court erred. Where the objection is to a factual finding by the district court in a jury trial, the appellant must show that the finding was not supported by a preponderance of the evidence. In a nonjury trial, a district court's finding based on recommendations by the master will be sustained if the district court did not abuse its discretion in determining that the master's report was not clearly erroneous.<sup>158</sup> If the district court rejected the master's report was not clearly erroneous.

<sup>151.</sup> Fed. R. Civ. P. 53(e)(4). See also 5A Moore et al., supra note 88,  $\P$  53.15. The weight given to the master's findings are the same regardless of whether the parties have consented to the appointment of the master. However, if the parties have agreed to accept the master's findings as final, only questions of law raised by the report may be considered by the court.

<sup>152.</sup> Fed. R. Civ. P. 53(e)(2).

<sup>153.</sup> American Honda Motor Co. v. Vickers Motors, Inc., 64 F.R.D. 118 (W.D. Tenn. 1974).

<sup>154.</sup> United States v. Conservation Chem. Co., 106 F.R.D. 210 (W.D. Mo. 1985).

<sup>155.</sup> Chicago Hous. Auth. v. Austin, 511 F.2d 82 (7th Cir. 1975).

<sup>156.</sup> Within ten days of being served with notice of the filing, any party in a nonjury trial may file objections to the report or may, by motion, request that the court adopt, modify, or reject the report. Fed. R. Civ. P. 53(e)(2).

<sup>157.</sup> Stauble v. Warrob, 977 F.2d 690, 696 (1st Cir. 1992).

<sup>158.</sup> Williams v. Lane, 851 F.2d 867, 884-85 (7th Cir. 1988).

ter's report as clearly erroneous, an appellant seeking review of that ruling must show that it was an unreasonable exercise of discretion.

## D. Standards for Determining Weight

In practice, the weight accorded masters' findings based on scientific and technical evidence is less formally defined than it is in theory. In many cases, expert masters, or masters advised by scientific and technical experts, made findings of fact through informal procedures.<sup>159</sup> Courts that receive such scientific and technical findings have little basis for determining whether they are "clearly erroneous." Because of the informality of the fact-finding procedures used by these expert masters, a slim evidentiary record usually exists for the court to scrutinize, although the reports of experts hired by masters are sometimes appended to the master's report. Perhaps for this reason, some courts treat such findings as advisory.<sup>160</sup>

Nevertheless, there would be little evidentiary support for judicial notice of these same facts based on treatises and other public documents. In effect, the opinion of the special master expressed in his or her fact finding is the evidentiary basis for the findings. Where parties object to such facts, fairness may dictate that the parties be afforded an opportunity to present their own experts at a formal hearing.

159. See, e.g., Little, supra note 48. 160. Chicago Housing Auth. v. Austin, 511 F.2d 82, 83 (7th Cir. 1975)

Special Masters

# VI. Conclusion

The appointment of special masters to handle complex scientific and technical issues and voluminous information has several advantages and disadvantages. The advantages of the use of masters include that masters

- may be able to bring special expertise to bear on the issues referred that would be difficult to secure by other means;
- can spend more time on a case and become better acquainted with the technical facts and the parties than a judge can with a full caseload;
- when permitted, may engage in ex parte discussions with the parties and facilitate settlement without prejudging the merits;
- can provide immediate resolution of technical, pretrial discovery issues through informal conferences;
- can conduct efficient, informal, as well as formal, hearings;
- can engage in intensive, sometimes round-the-clock, efforts to bring about settlement or ready a case for trial; and
- may be able to save the parties time and expense in the long run by developing efficient informal procedures and promoting settlements.

The disadvantages of the use of masters include that the use of masters

- adds immediate litigation expense for the parties;
- can cause delay, particularly when the judge must conduct reviews of voluminous masters' reports; and
- can distance the court from the case before trial. <sup>161</sup>

More fundamentally, the appointment may represent a deviation from the traditional adversary model of justice by interjecting a neutral, but not passive, specialized decision maker into the judicial system, which otherwise depends on more passive, generalist judges. Unlike court-appointed experts, masters are not witnesses to be examined and cross-examined by the parties, nor are they fulltime, government-paid jurists, like magistrate judges. Regardless of the judicial model envisaged, the appointment of masters in some circumstances is viewed as an improper delegation of judicial authority, violating Article III, the due process clause, or the authority granted in Rule 53(b).

<sup>161.</sup> See also MCL 3d, supra note 43, § 20.14.

To date, Rule 53 has provided a flexible, if not unbounded, mechanism through which courts can fashion procedures for dealing with scientific and technical information that are suited to the special needs of an individual case. The decision whether to appoint a special master to assist the court in handling scientific evidence requires consideration of the kind of expertise needed, the purposes to be served, and the relative ability of court-appointed experts, masters, and magistrate judges to further those purposes fairly and efficiently.

# **List of Reviewers**

We sent a draft of the evidentiary framework paper and each reference guide for comment and review to many individuals. We received a range of comments that informed our revision of the material into final form. However, inclusion on the list that follows does not necessarily represent a reviewer's approval of the final product. The names of several reviewers have been omitted from the list at their request. Any errors or shortcomings in the final product are our responsibility.

We are grateful for assistance received from the following people.

Orley C. Ashenfelter Princeton University Princeton, New Jersey

Barbara A. Bailar American Statistical Association Alexandria, Virginia

David C. Baldus University of Iowa College of Law Iowa City, Iowa

David W. Barnes University of Denver College of Law Denver, Colorado

David Z. Beckler Carnegie Commission on Science, Technology, and Government Washington, D.C.

Margaret A. Berger Brooklyn Law School Brooklyn, New York Joan E. Bertin School of Public Health Columbia University New York, New York

Sheila L. Birnbaum Skadden, Arps, Slate, Meagher & Flom New York, New York

Bert Black Weinberg and Green Baltimore, Maryland

JoAnn J. Bordeaux Torts Branch, Civil Division U.S. Department of Justice Washington, D.C.

Allen F. Breed National Council on Crime and Delinquency San Francisco, California

Troyen A. Brennan Department of Health Policy and Management Harvard School of Public Health Boston, Massachusetts Robert A. Burt Yale Law School New Haven, Connecticut

Edward J. Calabrese School of Public Health and Health Sciences University of Massachusetts at Amherst Amherst, Massachusetts

C. Thomas Caskey Department of Molecular and Human Genetics Baylor College of Medicine Houston, Texas

Joel E. Cohen The Rockefeller University New York, New York

Victor Cohn Reports on Health Washington, D.C.

Phillip A. Cole Brigham and Women's Hospital Boston, Massachusetts

Carl F. Cranor College of Humanities & Social Sciences University of California at Riverside Riverside, California

Don A. Dillman Social and Economic Sciences Research Center Washington State University Pullman, Washington

David L. Faigman University of California Hastings College of the Law San Francisco, California

Kenneth R. Feinberg Kenneth R. Feinberg & Associates Washington, D.C.

David Ferleger Philadelphia, Pennsylvania Stephen E. Fienberg Department of Statistics Carnegie Mellon University Pittsburgh, Pennsylvania

Michael O. Finkelstein Lord Day & Lord, Barrett Smith New York, New York

Franklin M. Fisher Department of Economics Massachusetts Institute of Technology Cambridge, Massachusetts

Peter B. Frank Price Waterhouse Los Angeles, California

David Freedman Department of Statistics University of California at Berkeley Berkeley, California

Arthur Furst Institute of Chemical Biology University of San Francisco San Francisco, California

Steven G. Gallagher American Arbitration Association Washington, D.C.

Joseph L. Gastwirth Department of Statistics The George Washington University Washington, D.C.

Paul C. Giannelli Case Western Reserve University Law School Cleveland, Ohio

Michael D. Green University of Iowa College of Law Iowa City, Iowa

Philip S. Guzelian University of Colorado Health Sciences Center Denver, Colorado

Jeffrey L. Harrison University of Florida College of Law Gainesville, Florida

Mitchell M. Holland Armed Forces Institute of Pathology Washington, D.C.

Edward J. Imwinkelried University of California at Davis School of Law Davis, California

Jacob Jacoby Leonard N. Stern School of Business New York University New York, New York

David H. Kaye Arizona State University College of Law Tempe, Arizona

Jonathan J. Koehler College and Graduate School of Business The University of Texas at Austin Austin, Texas

Mary Rose Kornreich Reston, Virginia

Sandra Kornstein RL Associates Princeton, New Jersey

Paul Lavrakas Survey Laboratory Northwestern University Evanston, Illinois

Stanley Lemeshow School of Public Health University of Massachusetts at Amherst Amherst, Massachusetts

Richard O. Lempert The University of Michigan Law School Ann Arbor, Michigan Edward Lloyd Environmental Law Clinic Rutgers, The State University of New Jersey Newark, New Jersey

Lee Loevinger Hogan & Hartson Washington, D.C.

Robert C. MacCallum Department of Psychology Ohio State University Columbus, Ohio

Francis E. McGovern The University of Alabama School of Law Tuscaloosa, Alabama

Nancy A. Mathiowetz Division of Statistics and Research Methodology Agency for Health Care Policy and Research Rockville, Maryland

Scott E. Maxwell Department of Psychology University of Notre Dame Notre Dame, Indiana

Neal Miller Institute for Law and Justice Alexandria, Virginia

Victor Miller Center for Research Information Silver Spring, Maryland

David S. Moore Department of Statistics Purdue University West Lafayette, Indiana

Norval Morris University of Chicago Law School Chicago, Illinois W. Alfred Mukatis College of Business Oregon State University Corvallis, Oregon

Vincent M. Nathan Nathan & Roberts Toledo, Ohio

Peter Neufeld O'Dwyer & Bernstien New York, New York

Gilbert S. Omenn School of Public Health and Community Medicine University of Washington Seattle, Washington

Arthur H. Patterson FTI Jury Analysts Group State College, Pennsylvania

Judyth W. Pendell Aetna Life and Casualty Hartford, Connecticut

David W. Peterson PRI Associates, Inc. Durham, North Carolina

Diana B. Petitti Southern California Kaiser Permanente Medical Care Program Pasadena, California

Stanley Presser Department of Sociology University of Maryland College Park, Maryland

George L. Priest Yale Law School New Haven, Connecticut

Myrna S. Raeder Southwestern University School of Law Los Angeles, California Michael Rappeport RL Associates Princeton, New Jersey

Elaine S. Reiss Council of Better Business Bureaus, Inc. New York, New York

Paul D. Rheingold Rheingold & McGowan, P.C. New York, New York

Paul R. Rice The American University Washington College of Law Washington, D.C.

Patricia A. Riley Office of the U.S. Attorney U.S. Department of Justice Washington, D.C.

Marc Rosenblum Office of General Counsel Equal Employment Opportunity Commission Washington, D.C.

Kenneth J. Rothman Epidemiology Resources, Inc. Newton Lower Falls, Massachusetts

Paul F. Rothstein Georgetown University Law Center Washington, D.C.

Jerome Sacks National Institute of Statistical Sciences Research Triangle Park, North Carolina

Stephen A. Saltzburg The George Washington University National Law Center Washington, D.C.

Joseph Sanders University of Houston Law Center Houston, Texas

Nora Cate Schaeffer Department of Sociology University of Wisconsin at Madison Madison, Wisconsin

Barry C. Scheck Yeshiva University Benjamin N. Cardozo School of Law New York, New York

George Schreiber Westat, Inc. Rockville, Maryland

Sol Schreiber Milberg, Weiss, Bershad, Hynes & Lerach New York, New York

Susan Scott Riker, Danzig, Scherer, Hyland & Perretti Morristown, New Jersey

Ellen Silbergeld Department of Epidemiology and Preventive Medicine University of Maryland Medical School Baltimore, Maryland

Linda J. Silberman New York University School of Law New York, New York

Eleanor Singer Institute for Social Research The University of Michigan Ann Arbor, Michigan

William K. Slate, II American Arbitration Association New York, New York

Miron L. Straf Committee on National Statistics National Academy of Sciences Washington, D.C. William P. Studer Oppenheimer, Wolff & Donnelly Minneapolis, Minnesota

John T. Sylvester Federal Bureau of Investigation Washington, D.C.

Judith M. Tanur Department of Sociology State University of New York at Stony Brook Stony Brook, New York

William C. ThompsonDepartment of Criminology, Law and SocietyUniversity of California at Irvine Irvine, California

Carlton A. Varner Sheppard, Mullin, Richter & Hampton Los Angeles, California

Jon S. Vernick School of Hygiene and Public Health Johns Hopkins University Baltimore, Maryland

W. Kip Viscusi Department of Economics Duke University Durham, North Carolina

Victor W. Weedn Armed Forces Institute of Pathology Washington, D.C.

Noel S. Weiss School of Public Health and Community Medicine University of Washington Seattle, Washington

Sandy Zabell Department of Statistics Northwestern University Evanston, Illinois

# Index

admissibility of scientific evidence and discovery process, 441-42 DNA, 98-102 Federal Rules' liberal thrust, 46; 49 gatekeeping function of trial judge, 14; 46:49 and related procedural issues, 49-53 judicial (in limine) screening, 29-30; 50-51 statistical evidence, 441-43 surveys, 227-28 versus legal sufficiency, 51-53 visual evidence, 115-17 antitrust and use of survey research, 228 appellate review of expert testimony, 53-54 Bayesian approach, 386; 387 Carnegie Commission on Science, Technology, and Government, 1-2 case-control studies, 136-38 case management assessment, 13 issue definition, 15 considerations in disputes over scientific evidence, 16-17 disclosure under Fed. R. Civ. P. 26(a)(2), 17 narrowing the issues, 17; 33 and use of reference guides, 18-19 presentation of evidence, 35 trial procedures for efficiency, 34-35 causation general (see under toxicology) specific (see under toxicology) census undercount cases and multiple regression analysis, 421 clinical ecology, scientific validity of, 73-75

cohort studies, 134-36 collateral estoppel as limit on expert evidence, 20 confidentiality (see under discovery) court-appointed experts authority to appoint under Fed. R. Evid. 706, 22; 531-34 compensation of from public funds, 558; 562 in general, 547; 557-62 technical advisors, 561 when a party is indigent, 560-61 ex parte communication with, 549-51; 567-68; 569 in general, 21-22; 25; 529-34; 571 initiation of appointment, 544-45; 566-67 overlap with special master, 22 reasons for appointing an expert difficult technical issues, 536–37; 538; 541; 548; 550; 565 disputes between parties' experts, 538-39; 542; 554; 564-65 failure of a party to present credible testimony, 538-39; 541; 554; 565-66 in general, 536-39; 541 fig. 2; 541-42; 563 - 66involvement of children, 541-42; 561; 566 settlement, improved chances for, 537; 539; 566 reasons for not appointing an expert difficulty finding an expert, 540; 544-45 difficulty identifying need for expert at early stage, 543; 563 infrequency of extraordinary circumstances, 540

court-appointed experts (continued) reasons for not appointing an expert (continued) in general, 532; 540-42 objection of parties, 542 payment of expert, 540; 557; 560 respect for adversarial system, 540; 542; 565 report of, 23; 24 selecting, 544-46; 567 special masters, employment by, 591-92 support for in Fed. R. Civ. P. 16(c)(12), 21 technical advisors, 534; 549; 550; 551; 558; 561-62; 568-69 testimony of advising jury of court-appointed status, 552-53; 554-55; 570 effect of experts' testimony, 553-54; 570 in general, 551-55; 570 pretrial reports and depositions 551-52; 558: 570 timing of appointment, 543-44; 563-67 use of in general, 535 fig. 1; 535-49; 567-70 judges' satisfaction with appointments, 537 criminal cases, 53 damages antitrust damages, 511-14 causation, 512 exclusionary conduct, 513

lost profits, 511 scope, 511–12 "tying" arrangements, 513–14 apportionment, 500–01; 509–10; 510–11 avoided cost, 488–89 causal effect of injury, disputes over, 485– 86 characterization of harmful event, 481–89 "but-for" analysis, 481–83 and costs, 488–89 disputes over economic effects, 483–85 compensation stock options, 489 tax treatment of, 487–88 damages study, 477–78; 478 fig. 1; 517– 19; 518 tables 4 & 5 double-counting, avoiding, 509; 511 earnings, what constitutes, 490 employment law, 501 expectation, 481 expert qualifications, 479 future earnings, projection of, 493-94 actual earnings of plaintiff after harmful event, 494 profitability of business, 493-94 future losses, discounting, 495-98; 495 table 2 appraisal approach, 499 capitalization factor, 497-98 interest rate, 495-96; 497 offset by growth in earnings, 496-97; 496 table 3 future losses, projection of, 494 in general, 477-78 intellectual property damages apportionment of, 509-10; 510-11 in general, 507-11 market-share analysis (sales), 508-09 price erosion, 484; 509 "reasonable royalty" and designing around the patent, 510 liquidated damages, 515-16 lost profit, 488 measuring losses, tax considerations, 487-88 mitigation, 490-91; 504-05 patent infringement by public utility, 500-01 personal lost earnings, 503-07 benefits, 503 discounting, 506-07 mitigation, 504-05 projected earnings, 503; 505-06 retirement and mortality, 507 prejudgment interest, calculation of, 491-93; 492 table 1 price erosion, 483; 484; 509 and regression analysis, 479 reliance, 481 restitution, 481 securities damages, 514-15 market effect of adverse information, 515 turnover patterns in ownership, 515 subsequent unexpected events, 499-500;

Reference Manual on Scientific Evidence

507

and surveys. 479 Daubert v. Merrell Dow Pharmaceuticals, 2; 14; 29; 30; 43, n.1; 44; 45-47; 47-54; 69; 70-78; 82-89; 91-92; 101; 103-05; 111-12; 113 court-appointed experts and, 534; 565 forensic DNA evidence and, 278; 285-86; 300; 305-06; 307 special masters and, 585 death penalty and multiple regression analysis, 421 observational studies of, 350 statistical studies of. 342 disclosure (see under discovery) discovery confidentiality, 26-27 control and management, 23-26 data and methodology of multiple regression analysis, 441-42 depositions, videotape, 27-28 disclosure and expert witnesses, 49-50 nonretained experts, 27 preservation of documents, order for, 24 protective orders, 26-27 surveys, 236 DNA (see forensic DNA identification analysis) ecological studies, 132-33 economic losses (see damages) employment discrimination and multiple regression analysis, 419-20; 421; 427 use of statistics in assessing disparate impact, 373 and use of survey research, 227 epidemiology association between exposure and disease general causation, 157-66 confounding factors, 158-60 guidelines for determining, 160-64 types of, 164-66 specific (individual) causation, 167-70 admissibility of evidence, 167-68 sufficiency of evidence, 168-69 false results (erroneous association) biases, sources of, 131-56 false negative error, 155-56 false positive error, 152-55 in general, 150-51

power, 156 random sampling error, 151-56 in general, 147-56 measurement of attributable proportion of risk, 149-50: 161-62 in general, 147–50 odds ratio, 149 relative risk, 147-49; 161 putative agent, measuring exposure to, 143 - 46in general, 125-28 research methods animal studies (in vivo), 129-30 extrapolation, 130 health effects, defining and measuring, 146 in general, 129-31 observational methods, 129 research design, 131-38 study populations and samples, 138-43 expert evidence admissibility of, 29-31 and Fed. R. Civ. P. 16(c)(4), 19 lay opinion testimony, 64-67 legal sufficiency of, 51-53 limitations on, 19-20 presentation of, 35 expert testimony appellate review of, 53-54 and Daubert v. Merrell Dow Pharmaceuticals, 45-47 exclusion of, under Fed. R. Evid. 403, 45; 113-17 governed by Fed. R. Evid. 702-705, 43-44 limitations and restrictions, 63-64 opinion supported by reliable data Fed. R. Evid. 703, 103-12 fit, 104 methodology, 104 "reasonably relied upon," 106–12 standard of proof, 104-05 relevancy, or "fit", 47-49 scientific validity, challenges to and Daubert v. Merrell Dow Pharmaceuticals, 71–72

expert testimony (continued) scientific validity, challenges to (continued) discipline, 73-82 clinical medicine, 80-82 clinical ecology, 73-75 forensic techniques, 75-76 and Frye v. United States, 70-71 probative value, 77-82 extrapolation problems (animal studies), 78-80 statistical estimates, 92-102 DNA. 98-102 theory, 82-91 psychological syndrome evidence, 87-88 skewed methodology, 88-91 social sciences, 84-87 scope, 63 sources of, 13-14 and survey research, 232-33 expert witnesses adversarial nature of, 15 availability of, 56 court-appointed (see court-appointed experts) disagreement of, 16-17 disclosure of and their opinions, 17; 23-28 discretion of court in ruling on, 58 engineers, 61 physicians, 59-61 qualifications availability, 56 determined by judge, 64 education or experience, 56-57 expertise (see also secondhand expert), 57 in general, 55-64 meaning of, 58 in multiple regression, 439 professional witness, 62-63 restrictions on (see expert testimony, limitations) secondhand expert, 62 specialization, 58-61 actual knowledge more important than credentials, 61 "two-expert" cases (statistical evidence), 337

two-pronged test, 55-56 extrapolation animal studies, 78-80; 130; 201-02 experiments in statistics, 349-50 Federal Courts Study Committee, 2 Federal Torts Claim Act, 109 forensic DNA identification analysis amplified fragment length polymorphism (AMP-FLP) technique, 288 Daubert v. Merrell Dow Pharmaceuticals and, 278 Frye standard and, 278; 285 in general, 277-79 polymerase chain reaction (PCR) based analysis techniques, 277; 287-88 probability estimates, 297-307 proficiency testing, 101-02 sampling uncertainties, 98-101 sequence-specific oligonucleotide (SSO) probes, 288 see also National Research Council report see also RFLP analysis forensics, scientific validity of, 75-76 Frye v. United States and forensic DNA evidence, 75; 278; 285; 306: 308-09 need for test abolished, 43; 70 in-court demonstrations, 115 Lanham Act cases and survey research, 228; 238; 251 lay opinion on scientific issues, 64-67 authentication testimony, 65-66 experience of lay witness, 65 hypothetical questions (prohibition of), 65 personal knowledge of lay witness, 65 testimony on causation, 66 testimony on economics, 66-67 local rules limitations on expert evidence, 20 magistrate judges in general, 21-22 implied authority for appointment under Fed. R. Civ. P. 16(c)(8), 21 special masters, use as, 595-96

Reference Manual on Scientific Evidence

masters (see special masters)

medicine (clinical) and methodology, 80-82 motion practice motions in limine, 29-30 summary judgment, 30-31; 51-53 multiparty litigation and emerging scientific issues, 14 multiple regression analysis causality, 422 census undercount cases, questionable use in, 421 computer output of, 459-60; 459 table 1 correlation, 446; 446 fig. 3 death penalty cases, questionable use in observational studies of, 350 statistical studies of, 342 dependent variable, choosing, 419; 424; 432 employment discrimination, 419-20; 421; 427 scatterplot, 445 fig. 2 use of statistics in assessing disparate impact, 373 and use of survey research, 227 expert, qualification of, 439 explanatory variables, 419; 424-26; 432-34 feedback, 432-34; 434 fig. 1 forecasting, 460-62 standard error of, 461-62; 462 fig. 9 in general, 419-22; 445-62 growth of use in court, 420 interpreting results, 429-38; 452 correlation versus causality, 421 error in measuring variables, 437-38 practical significance versus statistical significance, 429-32 regression slope, 453 fig. 6 robustness, 432-38 statistical significance, 429-31 linear regression model, 448-51 measurement error, 437-38 model specification (choosing a model), 423 - 28errors in model, 435-36 nonlinear models, 451 patent infringement, 420 precision of results, 453-59 goodness-of-fit, 456-57

least-squares regression, 458-59; 458 fig. 8 standard error, 453-56; 457 fig. 7; 462 fig. 9 regression line, 448 fig. 4; 449-51 goodness-of-fit, 450 fig. 5; 456-57 regression residuals, 451 research design, 423-28 formulating the question for investigation, 423 spurious correlation, 421-22; 433 statistical evidence, 441-43 statistical significance hypothesis test, 430-31 p-value, 431 National Research Council report and comparison of DNA profiles, 296 and modified ceiling principle technique, 99; 300; 301-02; 308-09 and probability estimation techniques, 100; 306-09 and proficiency-testing standards for lab procedures, 291 and quality-control standards for lab procedures, 291 and scientific validity and reliability of

DNA analysis, 278–79 nonexperts (see lay opinion)

#### orders court scheduling order, 25 final pretrial order, 33 order of reference to special master, 605– 06; 608; 611; 614 preclusion order, 33 for preservation and nondestruction of documents, 24 remedial order, 590–92 umbrella order, 26

patent infringement and multiple regression analysis, 420; 421; 423 post-traumatic stress disorder, admissibility in rape cases, 87 power calculations epidemiology, determining appropriate size for study population, 141–43 statistics, 381–82 pretrial conference and adversarial expert witnesses, 15 final conference, 33–35 initial conference, 13–20 to manage expert evidence, 13 and settlement, 34 pretrial management magistrate judge versus trial judge, 21 prosecutor's fallacy and presentation of statistical evidence, 97 protective orders (*see* under discovery) psychological evidence (*see* under social sciences evidence)

rape trauma syndrome, admissibility of, 87 reference guides, purpose and use of, 119 relevance ("fit"), 47-49 remedial orders, 590-92 res judicata as limit on expert evidence, 20 **RFLP** analysis coincidental DNA profile match, estimating probability of fixed-bin method, 298; 305 modified ceiling principle technique, 99; 301-02; 307-309 product rule technique, 300-01; 304-06 Daubert v. Merrell Dow Pharmaceuticals and, 285-86; 300; 305-06; 307 and expert testimony, 297 Frye test and, 306; 308-09 in general, 297-98 National Research Council report and, 301-02; 304-09 and population substructure, 305 and selection of appropriate comparison population, 303-09 Fed. R. Evid. 702, admission of probability estimates under, 306; 307 population genetics, 304 comparison of DNA profiles DNA profile match, declaration of, 295 - 96measurement standard used to declare match, objectivity of, 296

DNA crime sample, suitability of sample quality, 288 sample size, 287-88 sources of DNA, 288-89 in general, 281-84 laboratory procedures and admissibility, 294 appropriate documentation, 291-92 crime sample, 293 Daubert test and, 285-86 deviations from standards, 294 Frye test and, 285 proficiency-testing programs, laboratory's participation in, 292-93 reliability test and, 285 Technical Working Group on DNA Analysis Methods (TWGDAM) Guidelines, 291; 292; 294; 315-21 validity and reliability of, 285-86; 292 molecular biologists, 285; 286 population geneticists, 286 relevant scientific communities, 286 steps in, 282-84; 311 theory underlying, validity of, 285 settlement, 34 social sciences evidence hard versus soft science, 85-86 and jury, 85-87 psychological evidence, scientific validity of, 84-87 psychological syndrome evidence, 87-88 special masters advantages of using, 621 appellate courts' view of, 21 appointment of appealing, 598-99 arguments for and against, 580-81 authority for, 21; 595-99 conflict-of-interest problems in, avoiding, 603-05 delegation, limits on, 596-98 ethical problems in, avoiding, 603-05 "exceptional condition" requirement, 579-80; 596; 597; 598 issues to consider, 601-15 liability stage, 588-90 objections to, 604 preliability stage, 584-88

reasons for. 583-84 remedial stage, 590-93; 611; 615 termination date, specifying, 614-15 time limit on, establishing, 614-15 authority of, 605-06 avoiding delay and inertia, 614-15 case management, providing, 585-86 compensation of, 611-14 damages assessment, providing, 593 Daubert v. Merrell Dow Pharmaceuticals and, 585 discovery, use in, 584 ex parte communications in general, 606 with the judge, 608-09 with the parties, 607-08 expenses of, 613 expert testimony determining admissibility of, 585 handling proffers of, 585 experts employing, 591-92 overlap with, 22 Federal Judicial Center study of, 581-82 Fed. R. Evid. 104(a) hearings, conducting, 585 fees, 611-14 in general, 21-22; 579-82 hearings, type of, 610-11 history of use, 579 judicial ethics, applicability of, 603 judicial immunity and, 610 liability for malfeasance, 609-10 magistrate judges used as, 595-96 mediation function, 597; 608; 609 orders of reference, 605-06; 608; 611; 614 parties' approval of, 601-02; 604 powers under Fed. R. Civ. P. 53, 598 pretrial proceedings, assisting in, 584-85 qualifications, 601-02 referrals from other judges or scientific community, 602 reports, 608; 617-19 scientific and technical issues, advising judges on, 585-86 scope of authority, 605-06 selection of, 601-02 settlement, facilitating, 586-88 staff, expenses of, 613

statement of fees and expenses, 613 statistics association income and education, 360 fig. 5 average, 360-61 Bayesian approach, 386; 387 confidence intervals, 376-78 confounders (third variables), 367 correlation coefficients, 365-67 data, collection of censuses, 343 experiments, 346-50 individual measurements, 341-43 observational studies, 349; 350-52 proper recording, 342-43 reliability, 341-42 surveys, 343-46 validity, 342 data, inferences drawn from estimation, 374-78 in general, 373-74 hypothesis tests, 378-85 p-values, 378-85; 391; 392 fig. 13 posterior probabilities, 374; 386-87 data, presentation and analysis of association between two variables, 362-71 center of distribution, 360-61 completeness, 353-55 crime statistics, 353-54 graphs, 356-60 interpreting rates or percentages, 355-56 misleading data, 353-55 percentages, 363-65 variability, 361-62 discrimination cases, 364; 373 enhancing statistical testimony, 337-39 narrative testimony, 339 sequential testimony, 339 expertise in, 336-67 applied statistics, 336 probability theory, 336 theoretical statistics, 336 two-expert cases, 337 in general, 335-56 graphs association, 359-60 distribution of batch of numbers, 356; 357 fig. 1

statistics (continued) graphs (continued) histograms, 356-57; 390 fig. 2 scatter diagrams, 359 fig. 4; 360 fig. 5; 365 fig. 6; 366 fig. 7 trends, 357-58 linear association. 365-66 mean, 360-61 median, 360-61 mode, 360 normal curve, 390 fig. 12; 391; 392; 393 odds ratio, 364 outliers, 366; 367 fig. 8 pass rates, 378-80; 384; 389 percentages, 363 table 1; 364 table 2 power, 381-82 calculation of, 392-93; 393 fig. 14 random error, 373 range, 361 regression lines, 368-71; 368 fig. 9; 369 fig. 10; 371 fig. 11 intercept, 368; 371 slope, 368; 369-70; 371 unit of analysis, 370-71 and voting rights cases, 370; 371 standard deviation, 361-62 standard error, 375-78 calculation of, 391-92 statistical significance, determining, 374; 380-81 interval estimates, 384-85 multiple testing, 383-84 surveys convenience samples, 344 probability sampling, 345 sampling frame, 343-44 trends, 357-58; 358 figs. 2 & 3 sufficiency, legal, versus admissibility of scientific evidence, 51-53 summary judgment (see motion practice) survey research admissibility, 227 advantages of, 225-26 attorney participation in survey, 232 causal inferences, 249-52 comparing survey evidence to individual testimony, 228-29 confidentiality ethical obligation of survey research organization, 265

professional standards for survey researchers, 265 protecting identities of individual respondents, 265-66 surveyor-respondent privilege, not recognized, 266 consumer impressions, 249-50 data entry, 261 design of survey, 231-33 disclosure of methodology and results, 263 - 64in general, 225-27 in-person interviews, 252-53 interviewer surveys, 257-59 objective administration of survey procedures to minimize error and biases, 259 sponsorship disclosure, 258-59 selecting and training interviewers, 257 - 58mail surveys, 254-55 objectivity of, 232 pilot testing, 265 population definition and sampling, 235-41 bias, 239; 240-41 confidence interval, 237-38; 239 nonresponse, 239-40 probability sampling, 237-38 random sampling, 237 representativeness of sample, 239 response rates, 239-40 sampling frame (or universe), 235-37 screening potential respondents, 241 selecting the sample population, 237-39 target population, 235 purpose of survey, 231-33 questions, 243-55 ambiguous responses, use of probes to clarify, 248 clarity of, 243-44 consumer impressions, 250 control group or question, 249-52 filter questions to reduce guessing, 244 - 46open-ended versus closed-ended guestions, 246-47 order of questions, effect of, 248-49 pretests, 243-44

reporting, 264-65 responses, grouping of, 261 survey expertise, 232-33 telephone surveys, 253-54 and torts, 228 use of surveys in court, 227-28 Toxic Substances Control Act, 203; 205 toxicology acute toxicity testing, 188 additive effect, 211 animal research, extrapolation from, 191-92 antagonism, 211 association (see general and specific causation in this entry) chemical structure of compound, 203 confounding factors, 210 dose-response relationship, 188 and epidemiology, 194-95 expert qualifications advanced degree, 197-98 basis of toxicologist's expert opinion, 197 board certification, 198-99 other indicia of expertise, 199-200 physician, 197-98 professional organization, membership in, 198-99 in general, 185-95 general causation, 201-04 animal testing, extrapolation from, 201-02 biological plausibility, 204 chemical structure of compound, 203 in general, 201 in vitro tests of compound, 203 organ specificity of chemical, 202-03 good laboratory practices, 192-93 patient's medical history competing causes (confounding factors) of disease, 210 different susceptibilities to compound, 211 - 12effect of multiple agents, 211 evidence of interaction with other chemicals, 211 in general, 209-12

relevance of survey, 231

laboratory tests as indication of exposure to compound, 210 when data contradicts expert's opinion, 212 potentiation, 211 regulatory proceedings, 186 research design in general, 186-91 in vitro, 191 in vivo, 187-91 maximum tolerated dose, 189-91 no observable effect level, 188-89 no threshold model. 189 safety and risk assessments, 192-94 specific causation, 205-08 absorption of compound into body, 206 excretory route of compound, 207 exposure, 206 metabolism. 207 no observable effect level, 208 regulatory standards, 205-06 synergistic effect, 211 torts, 186 trademark litigation and survey research, 235; 236; 237; 263 videotape evidence, exclusion of, 115 visual evidence, admissibility of, 115-17

witnesses (see expert witnesses; see also lay opinion)