

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

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

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The development of the third edition of the *Reference Manual on Scientific Evidence* was supported by Contract No. B5727.R02 between the National Academy of Sciences and the Carnegie Corporation of New York and a grant from the Starr Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Academies or the organizations that provided support for the project.

International Standard Book Number-13: 978-0-309-21421-6

International Standard Book Number-10: 0-309-21421-1

Library of Congress Cataloging-in-Publication Data

Reference manual on scientific evidence. — 3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-309-21421-6 (pbk.)

ISBN-10: 0-309-21421-1 (pbk.)

1. Evidence, Expert—United States. I. Federal Judicial Center.

KF8961.R44 2011

347.73'67—dc23

2011031458

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Copyright 2011 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Summary Table of Contents

A detailed Table of Contents appears at the front of each chapter.

Introduction, 1	Stephen Breyer
The Admissibility of Expert Testimony, 11	Margaret A. Berger
How Science Works, 37	David Goodstein
Reference Guide on Forensic Identification Expertise, 55	Paul C. Giannelli, Edward J. Imwinkelried, & Joseph L. Peterson
Reference Guide on DNA Identification Evidence, 129	David H. Kaye & George Sensabaugh
Reference Guide on Statistics, 211	David H. Kaye & David A. Freedman
Reference Guide on Multiple Regression, 303	Daniel L. Rubinfeld
Reference Guide on Survey Research, 359	Shari Seidman Diamond
Reference Guide on Estimation of Economic Damages, 425	Mark A. Allen, Robert E. Hall, & Victoria A. Lazear
Reference Guide on Exposure Science, 503	Joseph V. Rodricks
Reference Guide on Epidemiology, 549	Michael D. Green, D. Michal Freedman, & Leon Gordis
Reference Guide on Toxicology, 633	Bernard D. Goldstein & Mary Sue Henifin
Reference Guide on Medical Testimony, 687	John B. Wong, Lawrence O. Gostin, & Oscar A. Cabrera
Reference Guide on Neuroscience, 747	Henry T. Greely & Anthony D. Wagner
Reference Guide on Mental Health Evidence, 813	Paul S. Appelbaum
Reference Guide on Engineering, 897	Channing R. Robertson, John E. Moalli, & David L. Black
Appendix A. Biographical Information of Committee and Staff, 961	

Reference Guide on Survey Research

SHARI SEIDMAN DIAMOND

Shari Seidman Diamond, J.D., Ph.D., is the Howard J. Trienens Professor of Law and Professor of Psychology, Northwestern University, and a Research Professor, American Bar Foundation, Chicago, Illinois.

CONTENTS

- I. Introduction, 361
 - A. Use of Surveys in Court, 363
 - B. Surveys Used to Help Assess Expert Acceptance in the Wake of *Daubert*, 367
 - C. Surveys Used to Help Assess Community Standards: *Atkins v. Virginia*, 369
 - D. A Comparison of Survey Evidence and Individual Testimony, 372
- II. Purpose and Design of the Survey, 373
 - A. Was the Survey Designed to Address Relevant Questions? 373
 - B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey? 374
 - C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced? 375
 - D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced? 375
- III. Population Definition and Sampling, 376
 - A. Was an Appropriate Universe or Population Identified? 376
 - B. Did the Sampling Frame Approximate the Population? 377
 - C. Does the Sample Approximate the Relevant Characteristics of the Population? 380
 - D. What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey? 383
 - E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample? 385
 - F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey? 386

- IV. Survey Questions and Structure, 387
 - A. Were Questions on the Survey Framed to Be Clear, Precise, and Unbiased? 387
 - B. Were Some Respondents Likely to Have No Opinion? If So, What Steps Were Taken to Reduce Guessing? 389
 - C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified? 391
 - 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? 394
 - E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects? 395
 - F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question? 397
 - G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey? 401
 - 1. In-person interviews, 402
 - 2. Telephone interviews, 403
 - 3. Mail questionnaires, 405
 - 4. Internet surveys, 406
- V. Surveys Involving Interviewers, 409
 - A. Were the Interviewers Appropriately Selected and Trained? 409
 - B. What Did the Interviewers Know About the Survey and Its Sponsorship? 410
 - C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias? 411
- VI. Data Entry and Grouping of Responses, 412
 - A. What Was Done to Ensure That the Data Were Recorded Accurately? 412
 - B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately? 413
- VII. Disclosure and Reporting, 413
 - A. When Was Information About the Survey Methodology and Results Disclosed? 413
 - B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics? 415
 - C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents? 417
- VIII. Acknowledgment, 418
- Glossary of Terms, 419
- References on Survey Research, 423

I. Introduction

Sample surveys are used to describe or enumerate the beliefs, attitudes, or behavior of persons or other social units.¹ Surveys typically are offered in legal proceedings to establish or refute claims about the characteristics of those individuals or social units (e.g., whether consumers are likely to be misled by the claims contained in an allegedly deceptive advertisement;² which qualities purchasers focus on in making decisions about buying new computer systems).³ In a broader sense, a *survey* can describe or enumerate the attributes of any units, including animals and objects.⁴ We focus here primarily on sample surveys, which must deal not only with issues of population definition, sampling, and measurement common to all surveys, but also with the specialized issues that arise in obtaining information from human respondents.

In principle, 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). In practice, surveys typically count or measure only a portion of the individuals or other units that the survey is intended to describe (e.g., a sample of jury-eligible citizens, a sample of potential job applicants). In either case, the goal is to provide information on the relevant population from which the sample was drawn. Sample surveys can be carried out using probability or nonprobability sampling techniques. Although probability sampling offers important advantages over nonprobability sampling,⁵ 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.⁶ 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.

1. Sample surveys conducted by social scientists “consist of (relatively) systematic, (mostly) standardized approaches to collecting information on individuals, households, organizations, or larger organized entities through questioning systematically identified samples.” James D. Wright & Peter V. Marsden, *Survey Research and Social Science: History, Current Practice, and Future Prospects*, in *Handbook of Survey Research* 1, 3 (James D. Wright & Peter V. Marsden eds., 2d ed. 2010).

2. *See Sanderson Farms v. Tyson Foods*, 547 F. Supp. 2d 491 (D. Md. 2008).

3. *See SMS Sys. Maint. Servs. v. Digital Equip. Corp.*, 118 F.3d 11, 30 (1st Cir. 1999). For other examples, see notes 19–32 and accompanying text.

4. In *J.H. Miles & Co. v. Brown*, 910 F. Supp. 1138 (E.D. Va. 1995), clam processors and fishing vessel owners sued the Secretary of Commerce for failing to use the unexpectedly high results from 1994 survey data on the size of the clam population to determine clam fishing quotas for 1995. The estimate of clam abundance is obtained from surveys of the amount of fishing time the research survey vessels require to collect a specified yield of clams in major fishing areas over a period of several weeks. *Id.* at 1144–45.

5. *See infra* Section III.C.

6. Fed. R. Evid. 703 recognizes facts or data “of a type reasonably relied upon by experts in the particular field. . . .”

As a method of data collection, surveys have several crucial potential advantages over less systematic approaches.⁷ When properly designed, executed, and described, surveys (1) economically present the characteristics of a large group of respondents or other units and (2) permit an assessment of the extent to which the measured respondents or other units are likely to adequately represent a relevant group of individuals or other units.⁸ All questions asked of respondents and all other measuring devices used (e.g., criteria for selecting eligible respondents) 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. The survey questions should not be the only focus of attention. To make it possible for the court and the opposing party to closely scrutinize the survey so that its relevance, objectivity, and representativeness can be evaluated, the party proposing to offer the survey as evidence should also describe in detail the design, execution, and analysis of the survey. This should include (1) a description of the population from which the sample was selected, demonstrating that it was the relevant population for the question at hand; (2) a description of how the sample was drawn and an explanation for why that sample design was appropriate; (3) a report on response rate and the ability of the sample to represent the target population; and (4) an evaluation of any sources of potential bias in respondents' answers.

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.⁹ 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

7. This does not mean that surveys can be relied on to address all questions. For example, if survey respondents had been asked in the days before the attacks of 9/11 to predict whether they would volunteer for military service if Washington, D.C., were to be bombed, their answers may not have provided accurate predictions. Although respondents might have willingly answered the question, their assessment of what they would actually do in response to an attack simply may have been inaccurate. Even the option of a "do not know" choice would not have prevented an error in prediction if they believed they could accurately predict what they would do. Thus, although such a survey would have been suitable for assessing the *predictions* of respondents, it might have provided a very inaccurate estimate of what an actual response to the attack would be.

8. The ability to quantitatively assess the limits of the likely margin of error is unique to probability sample surveys, but an expert testifying about any survey should provide enough information to allow the judge to evaluate how potential error, including coverage, measurement, nonresponse, and sampling error, may have affected the obtained pattern of responses.

9. See *infra* text accompanying note 31.

what weight, if any, the survey should be given.¹⁰ 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 individuals or other units, 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, whether conducted in-person, on the telephone, or online, raise methodological issues involving survey questions and structure (Section IV) and confidentiality (Section VII.C). Interview surveys introduce additional issues (e.g., interviewer training and qualifications) (Section V), and online surveys raise some new issues and questions that are currently under study (Section VI). The sections of this reference guide are labeled to direct the reader to those topics that are relevant to the type of survey being considered. The scope of this reference guide is necessarily limited, and additional issues might arise in particular cases.

A. Use of Surveys in Court

Fifty years ago the question of whether surveys constituted acceptable evidence still was unsettled.¹¹ Early doubts about the admissibility of surveys centered on their use of sampling¹² and their status as hearsay evidence.¹³ Federal Rule of Evidence

10. Lanham Act cases involving trademark infringement or deceptive advertising frequently require expedited hearings that request injunctive relief, so judges may need to be more familiar with survey methodology when considering the weight to accord a survey in these cases than when presiding over cases being submitted to a jury. Even in a case being decided by a jury, however, the court must be prepared to evaluate the methodology of the survey evidence in order to rule on admissibility. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 589 (1993).

11. Hans Zeisel, *The Uniqueness of Survey Evidence*, 45 Cornell L.Q. 322, 345 (1960).

12. 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*, tried in Los Angeles Superior Court in 1955, is described in R. Clay Sprows, *The Admissibility of Sample Data into a Court of Law: A Case History*, 4 UCLA L. Rev. 222, 226–29 (1956–1957).

13. Judge Wilfred Feinberg's thoughtful analysis in *Zippo Manufacturing 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." In *Schering Corp. v. Pfizer Inc.*, 189 F.3d 218 (2d Cir. 1999), the Second Circuit distinguished between perception surveys designed to reflect the present sense impressions of respondents and "memory" surveys designed to collect information about a past occurrence based on the recollections of the survey respondents. The court in *Schering* suggested that if a survey is offered to prove the existence of a specific idea in the public mind, then the survey does constitute hearsay

703 settled both matters for surveys by redirecting attention to the “validity of the techniques employed.”¹⁴ 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.”¹⁵ For 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?”¹⁶ This focus on the adequacy of the methodology used in conducting and analyzing results from a survey is also consistent with the Supreme Court’s discussion of admissible scientific evidence in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹⁷

Because the survey method provides an economical and systematic way to gather information and draw inferences about a large number of individuals or other units, surveys are used widely in business, government, and, increasingly,

evidence. As the court observed, Federal Rule of Evidence 803(3), creating “an exception to the hearsay rule for such statements [i.e., state-of-mind expressions] rather than excluding the statements from the definition of hearsay, makes sense only in this light.” *Id.* at 230 n.3. *See also* Playtex Prods. v. Procter & Gamble Co., 2003 U.S. Dist. LEXIS 8913 (S.D.N.Y. May 28, 2003), *aff’d*, 126 Fed. Appx. 32 (2d Cir. 2005). Note, however, that when survey respondents are shown a stimulus (e.g., a commercial) and then respond to a series of questions about their impressions of what they viewed, those impressions reflect both respondents’ initial perceptions and their memory for what they saw and heard. Concerns about the impact of memory on the trustworthiness of survey responses appropriately depend on the passage of time between exposure and testing and on the likelihood that distorting events occurred during that interval.

Two additional exceptions to the hearsay exclusion can be applied to surveys. First, surveys may constitute a hearsay exception if the survey data were collected in the normal course of a regularly conducted business activity, unless “the source of information or the method or circumstances of preparation indicate lack of trustworthiness.” Fed. R. Evid. 803(6); *see also* Ortho Pharm. Corp. v. Cosprophar, Inc., 828 F. Supp. 1114, 1119–20 (S.D.N.Y. 1993) (marketing surveys prepared in the course of business were properly excluded because they lacked foundation from a person who saw the original data or knew what steps were taken in preparing the report), *aff’d*, 32 F.3d 690 (2d Cir. 1994). In addition, if a survey shows guarantees of trustworthiness equivalent to those in other hearsay exceptions, it can be admitted if the court determines that the statement is offered as evidence of a material fact, it is more probative on the point for which it is offered than any other evidence that the proponent can procure through reasonable efforts, and admissibility serves the interests of justice. Fed. R. Evid. 807; *e.g.*, *Schering*, 189 F.3d at 232. Admissibility as an exception to the hearsay exclusion thus depends on the trustworthiness of the survey. *New Colt Holding v. RJG Holdings of Fla.*, 312 F. Supp. 2d 195, 223 (D. Conn. 2004).

14. Fed. R. Evid. 703 Advisory Committee Note.

15. Fed. R. Evid. 703.

16. Manual for Complex Litigation § 2.712 (1982). Survey research also is addressed in the Manual for Complex Litigation, Second § 21.484 (1985) [hereinafter MCL 2d]; the Manual for Complex Litigation, Third § 21.493 (1995) [hereinafter MCL 3d]; and the Manual for Complex Litigation, Fourth § 11.493 (2004) [hereinafter MCL 4th]. 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. An assessment of the 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.

17. 509 U.S. 579 (1993); *see also* General Elec. Co. v. Joiner, 522 U.S. 136, 147 (1997).

administrative settings and judicial proceedings.¹⁸ Both federal and state courts have accepted survey evidence on a variety of issues. In a case involving allegations of discrimination in jury panel composition, the defense team surveyed prospective jurors to obtain their age, race, education, ethnicity, and income distribution.¹⁹ Surveys of employees or prospective employees are used to support or refute claims of employment discrimination.²⁰ Surveys provide information on the nature and similarity of claims to support motions for or against class certification.²¹ In ruling on the admissibility of scientific claims, courts have examined surveys of scientific experts to assess the extent to which the theory or technique has received widespread acceptance.²² Some courts have admitted surveys in obscenity cases to provide evidence about community standards.²³ Requests for a change of venue on grounds of jury pool bias often are backed by evidence from a survey of jury-eligible respondents in the area of the original venue.²⁴ 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.²⁵ In a Title IX suit based on allegedly discriminatory scheduling of girls'

18. Some sample surveys are so well accepted that they even may not be recognized as surveys. For example, some 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.

19. *United States v. Green*, 389 F. Supp. 2d 29 (D. Mass. 2005), *rev'd on other grounds*, 426 F.3d 1 (1st Cir. 2005) (evaluating minority underrepresentation in the jury pool by comparing racial composition of the voting-age population in the district with the racial breakdown indicated in juror questionnaires returned to court); *see also* *People v. Harris*, 36 Cal. 3d 36, 679 P.2d 433 (Cal. 1984).

20. *John Johnson v. Big Lots Stores, Inc.*, No. 04-321, 2008 U.S. Dist. LEXIS 35316, at *20 (E.D. La. Apr. 29, 2008); *Stender v. Lucky Stores, Inc.*, 803 F. Supp. 259, 326 (N.D. Cal. 1992); *EEOC v. Sears, Roebuck & Co.*, 628 F. Supp. 1264, 1308 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988).

21. *John Johnson v. Big Lots Stores, Inc.*, 561 F. Supp. 2d 567 (E.D. La. 2008); *Marlo v. United Parcel Service, Inc.*, 251 F.R.D. 476 (C.D. Cal. 2008).

22. *United States v. Scheffer*, 523 U.S. 303, 309 (1998); *United States v. Bishop*, 64 F. Supp. 2d 1149 (D. Utah 1999); *United States v. Varoudakis*, No. 97-10158, 1998 WL 151238 (D. Mass. Mar. 27, 1998); *State v. Shively*, 268 Kan. 573 (2000), *aff'd*, 268 Kan. 589 (2000) (all cases in which courts determined, based on the inconsistent reactions revealed in several surveys, that the polygraph test has failed to achieve general acceptance in the scientific community). *Contra, see* *Lee v. Martinez*, 136 N.M. 166, 179-81, 96 P.3d 291, 304-06 (N.M. 2004). *People v. Williams*, 830 N.Y.S.2d 452 (2006) (expert permitted to testify regarding scientific studies of factors affecting the perceptual ability and memory of eyewitnesses to make identifications based in part on general acceptance demonstrated in survey of experts who study eyewitness identification).

23. *E.g.*, *People v. Page Books, Inc.*, 601 N.E.2d 273, 279-80 (Ill. App. Ct. 1992); *State v. Williams*, 598 N.E.2d 1250, 1256-58 (Ohio Ct. App. 1991).

24. *E.g.*, *United States v. Eagle*, 586 F.2d 1193, 1195 (8th Cir. 1978); *United States v. Tokars*, 839 F. Supp. 1578, 1583 (D. Ga. 1993), *aff'd*, 95 F.3d 1520 (11th Cir. 1996); *State v. Baumruk*, 85 S.W.3d 644 (Mo. 2002); *People v. Boss*, 701 N.Y.S.2d 342 (App. Div. 1999).

25. *Dolphin Tours, Inc. v. Pacifico Creative Servs., Inc.*, 773 F.2d 1506, 1508 (9th Cir. 1985). *See also* *SMS Sys. Maint. Servs., Inc. v. Digital Equip. Corp.*, 188 F.3d 11 (1st Cir. 1999); Benjamin F. King, *Statistics in Antitrust Litigation*, in *Statistics and the Law* 49 (Morris H. DeGroot et al. eds.,

sports, a survey was offered for the purpose of establishing how girls felt about the scheduling of girls' and boys' sports.²⁶ A routine use of surveys in federal courts occurs in Lanham Act²⁷ cases, when the plaintiff alleges trademark infringement²⁸ or claims that false advertising²⁹ has confused or deceived consumers. The pivotal legal question in such cases virtually demands survey research because it centers on consumer perception and memory (i.e., is the consumer likely to be confused about the source of a product, or does the advertisement imply a false or misleading message?).³⁰ 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.*³¹ jointly drafted a discovery survey that was administered

1986). Surveys have long been used in antitrust litigation to help define relevant markets. In *United States v. E.I. du Pont 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. In *Mukand, Ltd. v. United States*, 937 F. Supp. 910 (Ct. Int'l Trade 1996), a survey of purchasers of stainless steel wire rods was conducted to support a determination of competition and fungibility between domestic and Indian wire rod.

26. *Alston v. Virginia High Sch. League, Inc.*, 144 F. Supp. 2d 526, 539–40 (W.D. Va. 1999).

27. Lanham Act § 43(a), 15 U.S.C. § 1125(a) (1946) (amended 2006).

28. *E.g.*, *Herman Miller v. Palazzetti Imports & Exports*, 270 F.3d 298, 312 (6th Cir. 2001) ("Because the determination of whether a mark has acquired secondary meaning is primarily an empirical inquiry, survey evidence is the most direct and persuasive evidence."); *Simon Property Group v. MySimon*, 104 F. Supp. 2d 1033, 1038 (S.D. Ind. 2000) ("Consumer surveys are generally accepted by courts as one means of showing the likelihood of consumer confusion."). *See also* *Qualitex Co. v. Jacobson Prods. Co.*, No. CIV-90-1183HLH, 1991 U.S. Dist. LEXIS 21172 (C.D. Cal. Sept. 3, 1991), *aff'd in part & rev'd in part on other grounds*, 13 F.3d 1297 (9th Cir. 1994), *rev'd on other grounds*, 514 U.S. 159 (1995); *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.

29. *E.g.*, *Southland Sod Farms v. Stover Seed Co.*, 108 F.3d 1134, 1142–43 (9th Cir. 1997); *American Home Prods. Corp. v. Johnson & Johnson*, 577 F.2d 160 (2d Cir. 1978); *Rexall Sundown, Inc. v. Perrigo Co.*, 651 F. Supp. 2d 9 (E.D.N.Y. 2009); *Mutual Pharm. Co. v. Ivax Pharms, Inc.*, 459 F. Supp. 2d 925 (C.D. Cal. 2006); *Novartis Consumer Health v. Johnson & Johnson-Merck Consumer Pharms.*, 129 F. Supp. 2d 351 (D.N.J. 2000).

30. 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. Rohrllich*, 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 that, 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 976–77.

31. 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.

in person by neutral third parties, thus replacing interrogatories and depositions. It resulted in substantial savings in both time and cost.

B. Surveys Used to Help Assess Expert Acceptance in the Wake of *Daubert*

Scientists who offer expert testimony at trial typically present their own opinions. These opinions may or may not be representative of the opinions of the scientific community at large. In deciding whether to admit such testimony, courts applying the *Frye* test must determine whether the science being offered is generally accepted by the relevant scientific community. Under *Daubert* as well, a relevant factor used to decide admissibility is the extent to which the theory or technique has received widespread acceptance. Properly conducted surveys can provide a useful way to gauge acceptance, and courts recently have been offered assistance from surveys that allegedly gauge relevant scientific opinion. As with any scientific research, the usefulness of the information obtained from a survey depends on the quality of research design. Several critical factors have emerged that have limited the value of some of these surveys: problems in defining the relevant target population and identifying an appropriate sampling frame, response rates that raise questions about the representativeness of the results, and a failure to ask questions that assess opinions on the relevant issue.

Courts deciding on the admissibility of polygraph tests have considered results from several surveys of purported experts. Surveys offered as providing evidence of relevant scientific opinion have tested respondents from several populations: (1) professional polygraph examiners,³² (2) psychophysiologicals (members of the Society for Psychophysiological Research),³³ and (3) distinguished psychologists (Fellows of the Division of General Psychology of the American Psychological Association).³⁴ Respondents in the first group expressed substantial confidence in the scientific accuracy of polygraph testing, and those in the third group expressed substantial doubts about it. Respondents in the second group were asked the same question across three surveys that differed in other aspects of their methodology (e.g., when testing occurred and what the response rate was). Although over 60% of those questioned in two of the three surveys characterized the polygraph as a useful diagnostic tool, one of the surveys was conducted in 1982 and the more recent survey, published in 1984, achieved only a 30% response rate. The third

32. See plaintiff's survey described in *Meyers v. Arcudi*, 947 F. Supp. 581, 588 (D. Conn. 1996).

33. Susan L. Amato & Charles R. Honts, *What Do Psychophysiologicals Think About Polygraph Tests? A Survey of the Membership of SPR*, 31 *Psychophysiology* S22 [abstract]; Gallup Organization, *Survey of Members of the Society for Psychological Research Concerning Their Opinions of Polygraph Test Interpretation*, 13 *Polygraph* 153 (1984); William G. Iacono & David T. Lykken, *The Validity of the Lie Detector: Two Surveys of Scientific Opinion*, 82 *J. Applied Psychol.* 426 (1997).

34. Iacono & Lykken, *supra* note 33.

survey, also conducted in 1984, achieved a response rate of 90% and found that only 44% of respondents viewed the polygraph as a useful diagnostic tool. On the basis of these inconsistent reactions from the several surveys, courts have determined that the polygraph has failed to achieve general acceptance in the scientific community.³⁵ In addition, however, courts have criticized the relevance of the population surveyed by proponents of the polygraph. For example, in *Meyers v. Arcudi* the court noted that the survey offered by proponents of the polygraph was a survey of “practitioners who estimated the accuracy of the control question technique [of polygraph testing] to be between 86% and 100%.”³⁶ The court rejected the conclusions from this survey on the basis of a determination that the population surveyed was not the relevant scientific community, noting that “many of them . . . do not even possess advanced degrees and are not trained in the scientific method.”³⁷

The link between specialized expertise and self-interest poses a dilemma in defining the relevant scientific population. As the court in *United States v. Orians* recognized, “The acceptance in the scientific community depends in large part on how the relevant scientific community is defined.”³⁸ In rejecting the defendants’ urging that the court consider as relevant only psychophysicologists whose work is dedicated in large part to polygraph research, the court noted that *Daubert* “does not require the court to limit its inquiry to those individuals that base their livelihood on the acceptance of the relevant scientific theory. These individuals are often too close to the science and have a stake in its acceptance; i.e., their livelihood depends in part on the acceptance of the method.”³⁹

To be relevant to a *Frye* or *Daubert* inquiry on general acceptance, the questions asked in a survey of experts should assess opinions on the quality of the scientific theory and methodology, rather than asking whether or not the instrument should be used in a legal setting. Thus, a survey in which 60% of respondents agreed that the polygraph is “a useful diagnostic tool when considered with other available information,” 1% viewed it as sufficiently reliable to be the sole determinant, and the remainder thought it entitled to little or no weight, failed to assess the relevant issue. As the court in *United States v. Cordoba* noted, because “useful” and “other available information” could have many meanings, “there is little wonder why [the response chosen by the majority of respondents] was most frequently selected.”⁴⁰

35. *United States v. Scheffer*, 523 U.S. 303, 309 (1998); *United States v. Bishop*, 64 F. Supp. 2d 1149 (D. Utah 1999); *Meyers v. Arcudi*, 947 F. Supp. 581, 588 (D. Conn. 1996); *United States v. Varoudakis*, 48 Fed. R. Evid. Serv. 1187 (D. Mass. 1998).

36. *Meyers v. Arcudi*, 947 F. Supp. at 588.

37. *Id.*

38. 9 F. Supp. 2d 1168, 1173 (D. Ariz. 1998).

39. *Id.*

40. 991 F. Supp. 1199 (C.D. Cal. 1998), *aff'd*, 194 F.3d 1053 (9th Cir. 1999).

A similar flaw occurred in a survey conducted by experts opposed to the use of the polygraph in trial proceedings. Survey respondents were asked whether they would advocate that courts admit into evidence the outcome of a polygraph test.⁴¹ That question calls for more than an assessment of the accuracy of the polygraph, and thus does not appropriately limit expert opinion to issues within the expert's competence, that is, to the accuracy of the information provided by the test results. The survey also asked whether respondents agreed that the control question technique, the most common form of polygraph test, is accurate at least 85% of the time in real-life applications for guilty and innocent subjects.⁴² Although polygraph proponents frequently claim an accuracy level of 85%, it is up to the courts to decide what accuracy level would be required to justify admissibility. A better approach would be to ask survey respondents to estimate the level of accuracy they believe the test is likely to produce.⁴³

Surveys of experts are no substitute for an evaluation of whether the testimony an expert witness is offering will assist the trier of fact. Nonetheless, courts can use an assessment of opinion in the relevant scientific community to aid in determining whether a particular expert is proposing to use methods that would be rejected by a representative group of experts to arrive at the opinion the expert will offer. Properly conducted surveys can provide an economical way to collect and present information on scientific consensus and dissensus.

C. Surveys Used to Help Assess Community Standards: Atkins v. Virginia

In *Atkins v. Virginia*,⁴⁴ the U.S. Supreme Court determined that the Eighth Amendment's prohibition of "cruel and unusual punishment" forbids the execution of mentally retarded persons.⁴⁵ Following the interpretation advanced in *Trop v. Dulles*⁴⁶ that "The Amendment must draw its meaning from the evolving standards of decency that mark the progress of a maturing society,"⁴⁷ the Court examined a variety of sources, including legislative judgments and public opinion polls, to find that a national consensus had developed barring such executions.⁴⁸

41. See Iacono & Lykken, *supra* note 33, at 430, tbl. 2 (1997).

42. *Id.*

43. At least two assessments should be made: an estimate of the accuracy for guilty subjects and an estimate of the accuracy for innocent subjects.

44. 536 U.S. 304, 322 (2002).

45. Although some groups have recently moved away from the term "mental retardation" in response to concerns that the term may have pejorative connotations, mental retardation was the name used for the condition at issue in *Atkins* and it continues to be employed in federal laws, in cases determining eligibility for the death penalty, and as a diagnosis by the medical profession.

46. 356 U.S. 86 (1958).

47. *Id.* at 101.

48. *Atkins*, 536 U.S. at 313–16.

In a vigorous dissent, Chief Justice Rehnquist objected to the use of the polls, arguing that legislative judgments and jury decisions should be the sole indicators of national opinion. He also objected to the particular polls cited in the majority opinion, identifying what he viewed as serious methodological weaknesses.

The Court has struggled since *Furman v. Georgia*⁴⁹ to develop an adequate way to measure public standards regarding the application of the death penalty to specific categories of cases. In relying primarily on surveys of state legislative actions, the Court has ignored the forces that influence whether an issue emerges on a legislative agenda, and the strong influence of powerful minorities on legislative actions.⁵⁰ Moreover, the various members of the Court have disagreed about whether states without any death penalty should be included in the count of states that bar the execution of a particular category of defendant.

The Court has sometimes considered jury verdicts in assessing public standards. In *Coker v. Georgia*,⁵¹ the Court forbade the imposition of the death penalty for rape. Citing *Gregg v. Georgia*⁵² for the proposition that “[t]he jury . . . is a significant and reliable objective index of contemporary values because it is so directly involved,” the Court noted that “in the vast majority of cases [of rape in Georgia], at least 9 out of 10, juries have not imposed the death sentence.”⁵³ In *Atkins*, Chief Justice Rehnquist complained about the absence of jury verdict data.⁵⁴ Had such data been available, however, they would have been irrelevant because a “survey” of the jurors who have served in such cases would constitute a biased sample of the public. A potential juror unwilling to impose the death penalty on a mentally retarded person would have been ineligible to serve in a capital case involving a mentally retarded defendant because the juror would not have been able to promise during voir dire that he or she would be willing to listen to the evidence and impose the death penalty if the evidence warranted it. Thus, the death-qualified jury in such a case would be composed only of representatives from that subset of citizens willing to execute a mentally retarded defendant, an unrepresentative and systematically biased sample.

Public opinion surveys can provide an important supplementary source of information about contemporary values.⁵⁵ The Court in *Atkins* was presented with data from 27 different polls and surveys,⁵⁶ 8 of them national and 19 statewide.

49. 408 U.S. 238 (1972).

50. See *Stanford v. Kentucky*, 492 U.S. 361 (1989), *abrogated by Roper v. Simmons*, 543 U.S. 551 (2005).

51. 433 U.S. 584, 596 (1977).

52. 428 U.S. 153, 181 (1976).

53. *Coker v. Georgia*, 433 U.S. at 596.

54. See *Atkins*, 536 U.S. at 323 (Rehnquist, C.J., dissenting).

55. See *id.* at 316 n.21 (“[T]heir consistency with the legislative evidence lends further support to our conclusion that there is a consensus”).

56. The quality of any poll or survey depends on the methodology used, which should be fully visible to the court and the opposing party. See Section VII, *infra*.

The information on the polling data appeared in an amicus brief filed by the American Association on Mental Retardation.⁵⁷ Respondents were asked in various ways how they felt about imposing the death penalty on a mentally retarded defendant. In each poll, a majority of respondents expressed opposition to executing the mentally retarded. Chief Justice Rehnquist noted two weaknesses reflected in the data presented to the Court. First, almost no information was provided about the target populations from which the samples were drawn or the methodology of sample selection and data collection. Although further information was available on at least some of the surveys (e.g., the nationwide telephone survey of 1000 voters conducted in 1993 by the Tarrance Group used a sample based on voter turnout in the last three presidential elections), that information apparently was not part of the court record. This omission violates accepted reporting standards in survey research, and the information is needed if the decisionmaker is to intelligently evaluate the quality of the survey. Its absence in this instance occurred because the survey information was obtained from secondary sources.

A second objection raised by Chief Justice Rehnquist was that the wording of some of the questions required respondents to say merely whether they favored or were opposed to the use of the death penalty when the defendant is mentally retarded. It is unclear how a respondent who favors execution of a mentally retarded defendant only in a rare case would respond to that question. Some of the questions, however, did ask whether the respondent felt that it was never appropriate to execute the mentally retarded or whether it was appropriate in some circumstances.⁵⁸ In responses to these questions as well, a majority of respondents said that they found the execution of mentally retarded persons unacceptable under any circumstances. The critical point is that despite variations in wording of questions, the year in which the poll was conducted, who conducted it, where it was conducted, and how it was carried out, a majority of respondents (between 56% and 83%) expressed opposition to executing mentally retarded defendants. The Court thus was presented with a consistent set of findings, providing striking reinforcement for the *Atkins* majority's legislative analysis. Opinion poll data and legislative decisions have different strengths and weaknesses as indicators of contemporary values. The value of a multiple-measure approach is that it avoids a potentially misleading reliance on a single source or measure.

57. The data appear as an appendix to the Opinion of Chief Justice Rehnquist in *Atkins*.

58. Appendix to the Opinion of Chief Justice Rehnquist in *Atkins*. "Some people feel that there is nothing wrong with imposing the death penalty on persons who are mentally retarded, depending on the circumstances. Others feel that the death penalty should never be imposed on persons who are mentally retarded under any circumstances. Which of these views comes closest to your own?" The Tarrance Group, *Death Penalty Poll*, Q. 9 (Mar. 1993), citing Samuel R. Gross, *Update: American Public Opinion on the Death Penalty—It's Getting Personal*, 83 *Cornell L. Rev.* 1448, 1467 (1998).

D. 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 survey conducted in anticipation of litigation by a nontestifying expert if it produced unfavorable results,⁵⁹ 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. Although 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 of demonstrating particular facts, presenting the results of a well-done survey through the testimony of an expert 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.⁶⁰ 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.⁶¹

59. *In re FedEx Ground Package System*, 2007 U.S. Dist. LEXIS 27086 (N.D. Ind. April 10, 2007); *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 non-litigation 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).

60. *See, e.g., Morrison Entm't Group v. Nintendo of Am.*, 56 Fed. App'x. 782, 785 (9th Cir. Cal. 2003).

61. *Ortho Pharm. Corp. v. Cosprophar, Inc.*, 32 F.3d 690, 695 (2d Cir. 1994); *Henri's Food Prods. Co. v. Kraft, Inc.*, 717 F.2d 352, 357 (7th Cir. 1983); *Medici Classics Productions LLC v. Medici Group LLC*, 590 F. Supp. 2d 548, 556 (S.D.N.Y. 2008); *Citigroup v. City Holding Co.*, 2003 U.S. Dist. LEXIS 1845 (S.D.N.Y. Feb. 10, 2003); *Chum Ltd. v. Lisowski*, 198 F. Supp. 2d 530 (S.D.N.Y. 2002).

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). Surveys not conducted specifically in preparation for, or in response to, litigation may provide important information,⁶² but they frequently ask irrelevant questions⁶³ or select inappropriate samples of respondents for study.⁶⁴ Nonetheless, surveys do not always achieve their stated goals. Thus, the content and execution of a survey must be scrutinized whether or not the survey was designed to provide relevant data on the issue before the court.⁶⁵ Moreover, 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 provide evidence supporting its side of the issue in controversy.

62. 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., Joe S. Cecil, *Judicially Compelled Disclosure of Research Data*, 1 Cts. Health Sci. & L. 434 (1991); Richard L. Marcus, *Discovery Along the Litigation/Science Interface*, 57 Brook. L. Rev. 381, 393–428 (1991); see also *Court-Ordered Disclosure of Academic Research: A Clash of Values of Science and Law*, Law & Contemp. Probs., Summer 1996, at 1.

63. See *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, while the “single issue at hand . . . [was] whether consumers understood the term ‘Super Glue’ to designate glue from a single source”).

64. 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 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).

65. See *Merisant Co. v. McNeil Nutritionals, LLC*, 242 F.R.D. 315 (E.D. Pa. 2007).

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 survey interviews be “conducted independently of the attorneys in the case.”⁶⁶ Some courts interpreted this to mean that any evidence of attorney participation is objectionable.⁶⁷ A better interpretation is that the attorney should have no part in carrying out the survey.⁶⁸ However, some attorney involvement in the survey design is necessary to ensure that relevant questions are directed to a relevant population.⁶⁹ The 2009 amendments to Federal Rule of Civil Procedure 26(a)(2)⁷⁰ no longer allow an inquiry into the nature of communications between attorneys and experts, and so the role of attorneys in constructing surveys may become less apparent. The key issues for the trier of fact concerning the design of the survey are 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.⁷¹

66. Judicial Conference of the United States, Handbook of Recommended Procedures for the Trial of Protracted Cases 75 (1960).

67. See, e.g., *Boehringer Ingelheim G.m.b.H. v. Pharmadyne Lab.*, 532 F. Supp. 1040, 1058 (D.N.J. 1980).

68. *Upjohn Co. v. American Home Prods. Corp.*, No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at *42 (W.D. Mich. Apr. 5, 1996) (objection that “counsel reviewed the design of the survey carries little force with this Court because [opposing party] has not identified any flaw in the survey that might be attributed to counsel’s assistance”). For cases in which attorney participation was linked to significant flaws in the survey design, see *Johnson v. Big Lots Stores, Inc.*, No. 04-321, 2008 U.S. Dist. LEXIS 35316, at *20 (E.D. La. April 29, 2008); *United States v. Southern Indiana Gas & Elec. Co.*, 258 F. Supp. 2d 884, 894 (S.D. Ind. 2003); *Gibson v. County of Riverside*, 181 F. Supp. 2d 1057, 1069 (C.D. Cal. 2002).

69. See 6 J. Thomas McCarthy, McCarthy on Trademarks and Unfair Competition § 32:166 (4th ed. 2003).

70. www.uscourts.gov/News/TheThirdBranch/10-11-01/Rules_Recommendations_Take_Effect_December_1_2010.aspx.

71. *Gibson*, 181 F. Supp. 2d at 1068.

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 (especially social, cognitive, or consumer psychology), sociology, political science, 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 teaching or conducting and publishing survey research may provide the requisite background. In all cases, the expert must demonstrate an understanding of foundational, current, and best practices in survey methodology, including sampling,⁷² instrument design (questionnaire and interview construction), and statistical analysis.⁷³ Publication in peer-reviewed journals, authored books, fellowship status in professional organizations, faculty appointments, consulting experience, research grants, 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, some surveys involving highly technical subject matter (e.g., the particular preferences of electrical engineers for various pieces of electrical equipment and the bases for those preferences) or special populations (e.g., developmentally disabled adults with limited cognitive skills) may require experts to have some further specialized knowledge. Under these conditions, 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.⁷⁴ The secondary expert who gives an opinion

72. The one exception is that sampling expertise would be unnecessary if the survey were administered to all members of the relevant population. See, e.g., McGovern & Lind, *supra* note 31.

73. If survey expertise is being provided by several experts, a single expert may have general familiarity but not special expertise in all these areas.

74. See Margaret A. Berger, *The Admissibility of Expert Testimony*, Section III.A, in this manual.

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. Purpose of the survey;
2. Survey methodology,⁷⁵ 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. Results, including rates and patterns of missing data; and
4. Statistical analyses used to interpret the results.

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).⁷⁶ The target population consists of all elements (i.e., individuals or other units) 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 past purchasers of the plaintiff's goods or services and all prospective and past purchasers of the defendant'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. In a community survey designed to provide evidence for a motion for a change of venue, the relevant population consists of all jury-eligible citizens in the community in which the trial is to take place.⁷⁷

75. See *A & M Records, Inc. v. Napster, Inc.*, 2000 U.S. Dist. LEXIS 20668 (N.D. Cal. Aug. 10, 2000) (holding that expert could not attest credibly that the surveys upon which he relied conformed to accepted survey principles because of his minimal role in overseeing the administration of the survey and limited expert report).

76. Identification of the proper target population or 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 66; MCL 4th, *supra* note 16, § 11.493; see also 3 McCarthy, *supra* note 69, § 32:166; Council of Am. Survey Res. Orgs., Code of Standards and Ethics for Survey Research § III.A.3 (2010).

77. A second relevant population may consist of jury-eligible citizens in the community where the party would like to see the trial moved. By questioning citizens in both communities, the survey can test whether moving the trial is likely to reduce the level of animosity toward the party requesting the change of venue. See *United States v. Haldeman*, 559 F.2d 31, 140, 151, app. A at 176–79 (D.C. Cir. 1976) (court denied change of venue over the strong objection of Judge MacKinnon, who cited survey evidence that Washington, D.C., residents were substantially more likely to conclude, before

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 a sample from the target population of prospective and actual adult purchasers, it would exclude a crucial relevant population. The appropriate population in this instance would include children as well as parents.⁷⁸

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 the target population. Frequently, however, the target population includes members who are inaccessible or who cannot be identified in advance. As a result, reasonable compromises are sometimes required in developing the sampling frame. The survey report should contain (1) a description of the target population, (2) a description of the sampling frame from which the sample is to be drawn, (3) a discussion of the difference between the target population and the sampling frame, and, importantly, (4) an evaluation of the likely consequences of that difference.

A survey that provides information about a wholly irrelevant population is itself irrelevant.⁷⁹ Courts are likely to exclude the survey or accord it little

trial, that the defendants were guilty); *see also* *People v. Venegas*, 31 Cal. Rptr. 2d 114, 117 (Cal. Ct. App. 1994) (change of venue denied because defendant failed to show that the defendant would face a less hostile jury in a different court).

78. *See, e.g., Warner Bros., Inc. v. Gay Toys, Inc.*, 658 F.2d 76 (2d Cir. 1981) (surveying children users of the product rather than parent purchasers). Children and some other populations create special challenges for researchers. For example, very young children should not be asked about sponsorship or licensing, concepts that are foreign to them. Concepts, as well as wording, should be age appropriate.

79. A survey aimed at assessing how persons in the trade respond to an advertisement should be conducted on a sample of persons in the trade and not on a sample of consumers. *See Home Box Office v. Showtime/The Movie Channel*, 665 F. Supp. 1079, 1083 (S.D.N.Y.), *aff'd in part and vacated in part*, 832 F.2d 1311 (2d Cir. 1987); *J & J Snack Food Corp. v. Earthgrains Co.*, 220 F. Supp. 2d 358, 371–72 (N.J. 2002). *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

weight.⁸⁰ Thus, when the plaintiff submitted the results of a survey to prove that the green color of its fishing rod had acquired a secondary meaning, the court gave the survey little weight in part because the survey solicited the views of fishing rod dealers rather than consumers.⁸¹ More commonly, however, the sampling frame and the target population have some overlap, but the overlap is imperfect: The sampling frame excludes part of the target population, that is, it is underinclusive, or the sampling frame includes individuals who are not members of the target population, that is, it is overinclusive relative to the target population. Coverage error is the term used to describe inconsistencies between a sampling frame and a target population. If the coverage is underinclusive, the survey's value depends on the proportion of the target population that has been excluded from the sampling frame and the extent to which the excluded population is likely to respond 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 sampling frame 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.⁸²

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 coverage.⁸³ With the growth in individual cell phone use over time, noncoverage error would be an even greater problem for this survey today.

of the lamp, whether consciously or not, must have believed reasonably that the consuming public would be likely to rely on the salespersons' inaccurate statements about the name of the company that manufactured the lamp they were selling.

80. See *Wells Fargo & Co. v. WhenU.com, Inc.*, 293 F. Supp. 2d 734 (E.D. Mich. 2003).

81. See *R.L. Winston Rod Co. v. Sage Mfg. Co.*, 838 F. Supp. 1396, 1401–02 (D. Mont. 1993).

82. See *Brooks Shoe Mfg. Co. v. Suave Shoe Corp.*, 533 F. Supp. 75, 80 (S.D. Fla. 1981), *aff'd*, 716 F.2d 854 (11th Cir. 1983); see also *Hodgdon Power Co. v. Alliant Techsystems, Inc.*, 512 F. Supp. 2d 1178 (D. Kan. 2007) (excluding survey on gunpowder brands distributed at plaintiff's promotional booth at a shooting tournament); *Winning Ways, Inc. v. Holloway Sportswear, Inc.*, 913 F. Supp. 1454, 1467 (D. Kan. 1996) (survey flawed in failing to include sporting goods customers who constituted a major portion of customers). *But see* *Thomas & Betts Corp. v. Panduit Corp.*, 138 F.3d 277, 294–95 (7th Cir. 1998) (survey of store personnel admissible because relevant market included both distributors and ultimate purchasers).

83. See *Gencom, Inc.*, 56 Rad. Reg. 2d (P&F) 1597, 1604 (1984). This position was affirmed on appeal. See *Gencom, Inc. v. FCC*, 832 F.2d 171, 186 (D.C. Cir. 1987); see also *Beacon Mut. Ins. Co. v. Onebeacon Ins. Corp.*, 376 F. Supp. 2d 251, 261 (D.R.I. 2005) (sample included only defendant's insurance agents and lack of confusion among those agents was “nonstartling”).

In some cases, it is difficult to determine whether a sampling frame that omits some members of the population 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.⁸⁴ The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the target population should have included users of other analgesics, "so that the full range of potential customers for whom plaintiff and defendants would compete could be studied."⁸⁵ 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 defendants' product or users of a third analgesic.⁸⁶

An overinclusive sampling frame generally presents less of a problem for interpretation than does an underinclusive sampling frame.⁸⁷ If the survey expert can demonstrate that a sufficiently large (and representative) subset of respondents in the survey was drawn from the appropriate sampling frame, the responses obtained from that subset can be examined, and inferences about the relevant population can be drawn based on that subset.⁸⁸ If the relevant subset cannot be identified, however, an overbroad sampling frame will reduce the value of the survey.⁸⁹ If the sampling frame does not include important groups in the target population, there is generally no way to know how the unrepresented members of the target population would have responded.⁹⁰

84. See *American Home Prods. Corp. v. Barr Lab., Inc.*, 656 F. Supp. 1058 (D.N.J.), *aff'd*, 834 F.2d 368 (3d Cir. 1987).

85. *Id.* at 1070.

86. See also *Craig v. Boren*, 429 U.S. 190 (1976).

87. See *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992, 1134–35 (E.D.N.Y. 2006) ("Studies evaluating broadly the beliefs of low tar smokers generally are relevant to the beliefs of "light" smokers more specifically.").

88. See *National Football League Props. Inc. v. Wichita Falls Sportswear, Inc.* 532 F. Supp. 651, 657–58 (W.D. Wash. 1982).

89. See *Leelanau Wine Cellars, Ltd. v. Black & Red, Inc.*, 502 F.3d 504, 518 (6th Cir. 2007) (lower court was correct in giving little weight to survey with overbroad universe); *Big Dog Motorcycles, L.L.C. v. Big Dog Holdings, Inc.*, 402 F. Supp. 2d 1312, 1334 (D. Kan. 2005) (universe composed of prospective purchasers of all t-shirts and caps overinclusive for evaluating reactions of buyers likely to purchase merchandise at motorcycle dealerships). See also *Schieffelin & Co. v. Jack Co. of Boca*, 850 F. Supp. 232, 246 (S.D.N.Y. 1994).

90. See, e.g., *Amstar Corp. v. Domino's Pizza, Inc.*, 615 F.2d 252, 263–64 (5th Cir. 1980) (court found both plaintiff's and defendant's surveys substantially defective for a systematic failure to include parts of the relevant population); *Scott Fetzer Co. v. House of Vacuums, Inc.*, 381 F.3d 477 (5th Cir. 2004) (universe drawn from plaintiff's customer list underinclusive and likely to differ in their familiarity with plaintiff's marketing and distribution techniques).

C. Does the Sample 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.⁹¹ 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 sampling designs that use stratification, clustering of population elements into various groupings, or both. In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample.⁹² In simple random sampling, the most basic type of probability 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.⁹³ 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 elements are sampled in groups or clusters, rather than on an individual basis.⁹⁴ Note that selection probabilities do not need to be the same for all population elements; however, if the probabilities are unequal, compensatory adjustments should be made in the analysis.

Probability sampling offers two important advantages over other types of sampling. First, the sample can provide an unbiased estimate that summarizes the responses of all persons in the population from which the sample was drawn; that is, the expected value of the sample estimate is the population value being estimated. Second, the researcher can calculate a confidence interval that describes explicitly how reliable the sample estimate of the population is likely to be. If the sample is unbiased, the difference between the estimate and the exact value is called the sampling error.⁹⁵ Thus, suppose a survey collected responses from a simple random sample of 400 dentists selected from the population of all dentists

91. MCL 4th, *supra* note 16, § 11.493. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.B, in this manual.

92. The exception is that population elements omitted from the sampling frame have a zero probability of being sampled.

93. 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 unusual situations when periodicities occur, systematic samples and simple random samples generally produce the same results. Thomas Piazza, *Fundamentals of Applied Sampling*, in Handbook of Survey Research, *supra* note 1, at 139, 145.

94. *Id.* at 139, 150–63.

95. See David H. Kaye & David A. Freedman, *supra* note 91, Glossary, for a definition of sampling error.

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 could properly 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.⁹⁶ In this example, the margin of error is $\pm 4\%$, and so the confidence interval is the range between 16% and 24%, that is, the estimate (20%) plus or minus 4%.

All sample surveys produce estimates of population values, not exact measures of those values. Strictly speaking, the margin of 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 primary characteristics:

1. Size of the sample (the larger the sample, the narrower the interval);
2. Variability of the response being measured; and
3. Confidence level the researcher wants to have.⁹⁷

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 be expected to include the true population value.⁹⁸

Stratified probability sampling can be used to obtain more precise response estimates by using what is known about characteristics of the population that are likely to be associated with the response being measured. Suppose, for example, we anticipated that more-experienced and less-experienced dentists might respond differently to Goldgate toothpaste, and we had information on the year in which each dentist in the population began practicing. By dividing the population of dentists into more- and less-experienced strata (e.g., in practice 15 years or more versus in practice less than 15 years) and then randomly sampling within experience stratum, we would be able to ensure that the sample contained precisely

96. Actually, because 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. The expert should be prepared to discuss possible sources of bias due to, for example, an address list that is not current.

97. When the sample design does not use a simple random sample, the confidence interval will be affected.

98. To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%) without increasing the sample size, 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*, Section IV.A, in this manual.

proportionate representation from each stratum, in this case, more- and less-experienced dentists. That is, if 60% of dentists were in practice 15 years or more, we could select 60% of the sample from the more-experienced stratum and 40% from the less-experienced stratum and be sure that the sample would have proportionate representation from each stratum, reducing the likely sampling error.⁹⁹

In proportionate stratified probability sampling, as in simple random sampling, each individual member of the population has an equal chance of being selected. Stratified probability sampling can also disproportionately sample from different strata, a procedure that will produce more precise estimates if some strata are more heterogeneous than others on the measure of interest.¹⁰⁰ Disproportionate sampling may also be used to enable the survey to provide separate estimates for particular subgroups. With disproportionate sampling, sampling weights must be used in the analysis to accurately describe the characteristics of the population as a whole.

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 members of the target population are rare. A majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability convenience samples.¹⁰¹ 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 considerable consequence.”¹⁰² 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.¹⁰³ In addition, quantitative values computed from such samples (e.g., percentage of respondents indicating confusion) should be viewed as rough

99. See *Pharmacia Corp. v. Alcon Lab.*, 201 F. Supp. 2d 335, 365 (D.N.J. 2002).

100. Robert M. Groves et al., *Survey Methodology, Stratification and Stratified Sampling*, 106–18 (2004).

101. Jacob Jacoby & Amy H. Handlin, *Non-Probability Sampling Designs for Litigation Surveys*, 81 Trademark Rep. 169, 173 (1991). For probability surveys conducted in trademark cases, see *James Burrough, Ltd. v. Sign of Beefeater, Inc.*, 540 F.2d 266 (7th Cir. 1976); *Nightlight Systems, Inc., v. Nite Lights Franchise Sys.*, 2007 U.S. Dist. LEXIS 95565 (N.C. Ga. July 17, 2007); *National Football League Props., Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651 (W.D. Wash. 1982).

102. *National Football League Props., Inc. v. New Jersey Giants, Inc.*, 637 F. Supp. 507, 515 (D.N.J. 1986). A survey of 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 independent contacts in studies done in 1985 took place in malls or shopping centers. Jacoby & Handlin, *supra* note 101, at 172–73, 176. More recently, surveys conducted over the Internet have been administered to samples of respondents drawn from panels of volunteers; see *infra* Section IV.G.4 for a discussion of online surveys. Although panel members may be randomly selected from the panel population to complete the survey, the panel population itself is not usually the product of a random selection process.

103. See *infra* Sections III.D–E.

indicators rather than as precise quantitative estimates.¹⁰⁴ Confidence intervals technically should not be computed, although if the calculation shows a wide interval, that may be a useful indication of the limited value of the estimate.

D. 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 is distributed randomly, valid inferences about the population can be drawn with assurance using the measures obtained from 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.¹⁰⁵ Efforts to increase response rates include making several attempts to contact potential respondents, sending advance letters,¹⁰⁶ and providing financial or nonmonetary incentives for participating in the survey.¹⁰⁷

The key to evaluating the effect of nonresponse in a survey is to determine as much as possible the extent to which nonrespondents differ from the respondents in the nature of the responses they would provide if they were present in the sample. That is, the difficult question to address is the extent to which nonresponse has biased the pattern of responses by undermining the representativeness of the sample and, if it has, the direction of that bias. It is incumbent on the expert presenting the survey results to analyze the level and sources of nonresponse, and to assess how that nonresponse is likely to have affected the results. On some occasions, it may be possible to anticipate systematic patterns of nonresponse. For example, a survey that targets a population of professionals may encounter difficulty in obtaining the same level of participation from individuals with high-volume practices that can be obtained from those with lower-volume practices. To enable the researcher to assess whether response rate varies with the volume of practice, it may be possible to identify in advance potential respondents

104. The court in *Kinetic Concept, Inc. v. Bluesky Medical Corp.*, 2006 U.S. Dist. LEXIS 60187, *14 (W.D. Tex. Aug. 11, 2006), found the plaintiff's survey using a nonprobability sample to be admissible and permitted the plaintiff's expert to present results from a survey using a convenience sample. The court then assisted the jury by providing an instruction on the differences between probability and convenience samples and the estimates obtained from each.

105. 2 Gastwirth, *supra* note 64, at 501. This volume contains a useful discussion of sampling, along with a set of examples. *Id.* at 467.

106. Edith De Leeuw et al., *The Influence of Advance Letters on Response in Telephone Surveys: A Meta-analysis*, 71 *Pub. Op. Q.* 413 (2007) (advance letters effective in increasing response rates in telephone as well as mail and face-to-face surveys).

107. Erica Ryu et al., *Survey Incentives: Cash vs. In-kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error*, 18 *Int'l J. Pub. Op. Res.* 89 (2005).

with varying years of experience. Even if it is not possible to know in advance the level of experience of each potential member in the target population and to design a sampling plan that will produce representative samples at each level of experience, the survey itself can include questions about volume of practice that will permit the expert to assess how experience level may have affected the pattern of results.¹⁰⁸

Although high response rates (i.e., 80% or higher)¹⁰⁹ are desirable because they generally eliminate the need to address the issue of potential bias from nonresponse,¹¹⁰ such high response rates are increasingly difficult to achieve. Survey nonresponse rates have risen substantially in recent years, along with the costs of obtaining responses, and so the issue of nonresponse has attracted substantial attention from survey researchers.¹¹¹ Researchers have developed a variety of approaches to adjust for nonresponse, including weighting obtained responses in proportion to known demographic characteristics of the target population, comparing the pattern of responses from early and late responders to mail surveys, or the pattern of responses from easy-to-reach and hard-to-reach responders in telephone surveys, and imputing estimated responses to nonrespondents based on known characteristics of those who have responded. All of these techniques can only approximate the response patterns that would have been obtained if nonrespondents had responded. Nonetheless, they are useful for testing the robustness of the findings based on estimates obtained from the simple aggregation of answers to questions given by responders.

To assess the general impact of the lower response rates, researchers have conducted comparison studies evaluating the results obtained from surveys with

108. In *People v. Williams*, *supra* note 22, a published survey of experts in eyewitness research was used to show general acceptance of various eyewitness phenomena. See Saul Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research: A New Survey of the Experts*, 56 *Am. Psychologist* 405 (2001). The survey included questions on the publication activity of respondents and compared the responses of those with high and low research productivity. Productivity levels in the respondent sample suggested that respondents constituted a blue ribbon group of leading researchers. *Williams*, 830 N.Y.S.2d at 457 n.16. See also *Pharmacia Corp. v. Alcon Lab., Inc.*, 201 F. Supp. 2d 335 (D.N.J. 2002).

109. Note that methods of computing response rates vary. For example, although response rate can be generally defined as the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample, decisions on how to treat partial completions and how to estimate the eligibility of nonrespondents can produce differences in measures of response rate. *E.g.*, American Association of Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (rev. 2008), available at www.Aapor.org/uploads/Standard_Definitions_07-08_Final.pdf.

110. Office of Management and Budget, *Standards and Guidelines for Statistical Surveys* (Sept. 2006), Guideline 1.3.4: Plan for a nonresponse bias analysis if the expected unit response rate is below 80%. See *Albert v. Zabin*, 2009 Mass. App. Unpub. LEXIS 572 (July 14, 2009) reversing summary judgment that had excluded surveys with response rates of 27% and 31% based on a thoughtful analysis of measures taken to assess potential nonresponse bias.

111. *E.g.*, Richard Curtin et al., *Changes in Telephone Survey Nonresponse Over the Past Quarter Century*, 69 *Pub. Op. Q.* 87 (2005); *Survey Nonresponse* (Robert M. Groves et al. eds., 2002).

varying response rates.¹¹² Contrary to earlier assumptions, surprisingly comparable results have been obtained in many surveys with varying response rates, suggesting that surveys may achieve reasonable estimates even with relatively low response rates. The key is whether nonresponse is associated with systematic differences in response that cannot be adequately modeled or assessed.

Determining whether the level of nonresponse in a survey seriously impairs inferences drawn from the results of a survey 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 is used to chart changes in the proportion of the population that knows someone with HIV, the survey would underestimate the population value if some groups more likely to know someone with HIV (e.g., intravenous drug users) are 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 surveys that are self-administered, some respondents may refuse to provide answers or may provide incomplete answers (i.e., item rather than unit nonresponse).¹¹³ 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 (i.e., imputation).¹¹⁴

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 to reduce the likelihood of biased selection. For example, in many studies the target population consists of all consumers or purchasers of a product. Because it is impractical to randomly sample from that population, research is often conducted in shopping malls where some members of the target population may not shop. Mall locations, however, can be sampled randomly from a list of possible mall sites. By administering the survey at several different

112. E.g., Daniel M. Merkle & Murray Edelman, *Nonresponse in Exit Polls: A Comprehensive Analysis*, in *Survey Nonresponse*, *supra* note 111, at 243–57 (finding minimal nonresponse error associated with refusals to participate in in-person exit polls); see also Jon A. Krosnick, *Survey Research*, 50 *Ann. Rev. Psychol.* 537 (1999).

113. See Roger Tourangeau et al., *The Psychology of Survey Response* (2000).

114. See Paul D. Allison, *Missing Data*, in *Handbook of Survey Research*, *supra* note 1, at 630; see also *Survey Nonresponse*, *supra* note 111.

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 onsite interview operations, it is less likely that idiosyncrasies of sample selection or administration can account for the results.¹¹⁵ Similarly, because 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.¹¹⁶

In mall intercept surveys, the organization that manages the onsite 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 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 *n*th person entering the facility through a particular door.¹¹⁷

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 members of the target population 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 appeal to or deter specific groups within the target population, or convey information that will influence the respondent's

115. Note, however, that differences in results across sites may arise from genuine differences in respondents across geographic locations or from a failure to administer the survey consistently across sites.

116. Seymour Sudman, *Improving the Quality of Shopping Center Sampling*, 17 J. Marketing Res. 423 (1980).

117. In the end, even if malls are randomly sampled and shoppers are randomly selected within malls, results from mall surveys technically can be used to generalize only to the population of mall shoppers. The ability of the mall sample to describe the likely response pattern of the broader relevant population will depend on the extent to which a substantial segment of the relevant population (1) is not found in malls and (2) would respond differently to the interview.

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 6 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 content of a screening questionnaire (or screener) can also set the context for the questions that follow. In *Pfizer, Inc. v. Astra Pharmaceutical Products, Inc.*,¹¹⁸ physicians were asked a screening question to determine whether they prescribed particular drugs. The survey question that followed the screener asked “Thinking of the practice of cardiovascular medicine, what first comes to mind when you hear the letters XL?” The court found that the screener conditioned the physicians to respond with the name of a drug rather than a condition (long-acting).¹¹⁹

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.¹²⁰

118. 858 F. Supp. 1305, 1321 & n.13 (S.D.N.Y. 1994).

119. *Id.* at 1321.

120. Floyd J. Fowler, Jr., *How Unclear Terms Affect Survey Data*, 56 *Pub. Op. Q.* 218, 225–26 (1992).

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.¹²¹ If the crucial question is sufficiently ambiguous or unclear, it may be the basis for rejecting the survey. For example, a survey was designed to assess community sentiment that would warrant a change of venue in trying a case for damages sustained when a hotel skywalk collapsed.¹²² The court found that the question “Based on what you have heard, read or seen, do you believe that in the current compensatory damage trials, the defendants, such as the contractors, designers, owners, and operators of the Hyatt Hotel, should be punished?” could neither be correctly understood nor easily answered.¹²³ The court noted that the phrase “compensatory damages,” although well-defined for attorneys, was unlikely to be meaningful for laypersons.¹²⁴

A variety of pretest activities may be used to improve the clarity of communication with respondents. Focus groups can be used to find out how the survey population thinks about an issue, facilitating the construction of clear and understandable questions. Cognitive interviewing, which includes a combination of think-aloud and verbal probing techniques, may be used for questionnaire evaluation.¹²⁵ Pilot studies involving a dress rehearsal for the main survey can also detect potential problems.

Texts on survey research generally recommend pretests as a way to increase the likelihood that questions are clear and unambiguous,¹²⁶ and some courts have recognized the value of pretests.¹²⁷ In many pretests or pilot tests,¹²⁸ the proposed survey is administered to a small sample (usually between 25 and 75)¹²⁹ of the

121. *See id.* at 219.

122. *Firestone v. Crown Ctr. Redevelopment Corp.*, 693 S.W.2d 99 (Mo. 1985) (en banc).

123. *See id.* at 102, 103.

124. *See id.* at 103. When there is any question about whether some respondents will understand a particular term or phrase, the term or phrase should be defined explicitly.

125. Gordon B. Willis et al., *Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques*, in *Cognitive and Survey Research* 136 (Monroe G. Sirken et al. eds., 1999). *See also* Tourangeau et al., *supra* note 113, at 326–27.

126. *See* Jon A. Krosnick & Stanley Presser, *Questions and Questionnaire Design*, in *Handbook of Survey Research*, *supra* note 1, at 294 (“No matter how closely a questionnaire follows recommendations based on best practices, it is likely to benefit from pretesting. . .”). *See also* Jean M. Converse & Stanley Presser, *Survey Questions: Handcrafting the Standardized Questionnaire* 51 (1986); Fred W. Morgan, *Judicial Standards for Survey Research: An Update and Guidelines*, 54 *J. Marketing* 59, 64 (1990).

127. *See e.g.*, *Zippo Mfg. Co. v. Rogers Imports, Inc.*, 216 F. Supp. 670 (S.D.N.Y. 1963); *Scott v. City of New York*, 591 F. Supp. 2d 554, 560 (S.D.N.Y. 2008) (“[s]urvey went through multiple pretests in order to insure its usefulness and statistical validity.”).

128. The terms *pretest* and *pilot test* are sometimes used interchangeably to describe pilot work done in the planning stages of research. When they are distinguished, the difference is that a pretest tests the questionnaire, whereas a pilot test generally tests proposed collection procedures as well.

129. Converse & Presser, *supra* note 126, at 69. Converse and Presser suggest that a pretest with 25 respondents is appropriate when the survey uses professional interviewers.

same type of respondents who would be eligible to participate in the full-scale survey. 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.¹³⁰ Attorneys who commission surveys for litigation sometimes are reluctant to approve 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 is a standard and valuable way to improve the quality of a survey¹³¹ 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.¹³²

B. Were Some Respondents Likely to Have No Opinion? If So, What Steps Were Taken 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 they 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 1-year guarantee, a 60-day guarantee, or a 30-day guarantee?”). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an

130. Methods for testing respondent understanding include concurrent and retrospective think-alouds, in which respondents describe their thinking as they arrive at, or after they have arrived at, an answer, and paraphrasing (asking respondents to restate the question in their own words). Tourangeau et al., *supra* note 113, at 326–27; *see also* Methods for Testing and Evaluating Survey Questionnaires (Stanley Presser et al. eds., 2004).

131. *See* OMB Standards and Guidelines for Statistical Survey, *supra* note 110, Standard 1.4, Pre-testing Survey Systems (specifying that to ensure that all components of a survey function as intended, pretests of survey components should be conducted unless those components have previously been successfully fielded); American Association for Public Opinion Research, Best Practices (2011) (“Because it is rarely possible to foresee all the potential misunderstandings or biasing effects of different questions or procedures, it is vital for a well-designed survey operation to include provision for a pretest.”).

132. *See infra* Section VII.B for a discussion of obligations to disclose pilot work.

answer even if (in this example) 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 imprecision introduced by this approach 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?”).¹³³ 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 choose 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 distribution of the respondents selecting the *listed* choices is unlikely to change dramatically, presentation of an explicit “don’t know” or “no opinion” alternative commonly leads to a 20% to 25% increase in the proportion of respondents selecting that response.¹³⁴

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?”).¹³⁵ The interviewer then asks the substantive question only of those respondents who have indicated that they have an opinion on the issue.

Which of these three approaches is used and the way it is used can affect the rate of “no opinion” responses that the substantive question will evoke.¹³⁶ Respondents are more likely to say that they do not have an opinion on an issue if a full filter is used than if a quasi-filter is used.¹³⁷ 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 respondents can avoid difficult followup questions by saying that they have no opinion.¹³⁸

133. Norbert Schwarz & Hans-Jürgen Hippler, *Response Alternatives: The Impact of Their Choice and Presentation Order*, in *Measurement Errors in Surveys* 41, 45–46 (Paul P. Biemer et al. eds., 1991).

134. Howard Schuman & Stanley Presser, *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context* 113–46 (1981).

135. See, e.g., *Johnson & Johnson–Merck Consumer Pharmas. Co. v. SmithKline Beecham Corp.*, 960 F.2d 294, 299 (2d Cir. 1992).

136. 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. Op. Q.* 528 (1983).

137. Schwarz & Hippler, *supra* note 133, at 45–46.

138. *Id.* at 46.

In general, then, a survey that uses full filters provides a conservative estimate of the number of respondents holding an opinion, while a survey that uses neither full filters nor quasi-filters may overestimate the number of respondents with opinions, if some respondents offering opinions are guessing. The strategy of including a “no opinion” or “don’t know” response 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?”¹³⁹ 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, or that they are made differently, or don’t you have an opinion about the way they are made?”

Recent research on the effects of including a “don’t know” option shows that quasi-filters as well as full filters may discourage a respondent who would be able to provide a meaningful answer from expressing it.¹⁴⁰ The “don’t know” option provides a cue that it is acceptable to avoid the work of trying to provide a more substantive response. Respondents are particularly likely to be attracted to a “don’t know” option when the question is difficult to understand or the respondent is not strongly motivated to carefully report an opinion.¹⁴¹ One solution that some survey researchers use is to provide respondents with a general instruction not to guess at the beginning of an interview, rather than supplying a “don’t know” or “no opinion” option as part of the options attached to each question.¹⁴² Another approach is to eliminate the “don’t know” option and to add followup questions that measure the strength of the respondent’s opinion.¹⁴³

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, closed-ended, 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

139. 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). See also *Procter & Gamble Pharms., Inc. v. Hoffmann-La Roche, Inc.*, 2006 U.S. Dist. LEXIS 64363 (S.D.N.Y. Sept. 6, 2006).

140. Jon A. Krosnick et al., *The Impact of “No Opinion” Response Options on Data Quality: Non-Attitude Reduction or Invitation to Satisfice?* 66 Pub. Op. Q. 371 (2002).

141. Krosnick & Presser, *supra* note 126, at 284.

142. *Anheuser-Busch, Inc. v. VIP Prods, LLC*, No. 4:08cv03558, 2008 U.S. Dist. LEXIS 82258, at *6 (E.D. Mo. Oct. 16, 2008).

143. Krosnick & Presser, *supra* note 126, at 285.

in these waters?”¹⁴⁴). Closed-ended questions provide the respondent with an explicit set of responses from which to choose; the choices may be as simple as *yes* or *no* (e.g., “Is Colby College coeducational?”¹⁴⁵) or as complex as a range of 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.”¹⁴⁶). When a survey involves in-person interviews, the interviewer may show the respondent these choices on a showcard that lists them.

Open-ended and closed-ended questions may elicit very different responses.¹⁴⁷ Most responses are less likely to be volunteered by respondents who are asked an open-ended question than they are to be chosen by respondents who are presented with 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.¹⁴⁸

The advantage of open-ended questions is that they give the respondent fewer hints about expected or preferred answers. Precoded responses on a closed-ended question, in addition to reminding respondents of options that they might not otherwise consider,¹⁴⁹ 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 ratings than

144. A relevant example from *Wilhoite v. Olin Corp.* is described in McGovern & Lind, *supra* note 31, at 76.

145. *Presidents & Trustees of Colby College v. Colby College—N.H.*, 508 F.2d 804, 809 (1st Cir. 1975).

146. This question is based on one asked in *American Home Products 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 Products 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?”

147. Howard Schuman & Stanley Presser, *Question Wording as an Independent Variable in Survey Analysis*, 6 Soc. Methods & Res. 151 (1977); Schuman & Presser, *supra* note 134, at 79–112; Converse & Presser, *supra* note 126, at 33.

148. 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 134, at 104–07. An open-ended question presents the respondent with a free-recall task, whereas 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.

149. Schwarz & Hippler, *supra* note 133, at 43.

other brands.¹⁵⁰ 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?”¹⁵¹ Instead, the survey asked a closed-ended question; respondents were given the choice of “one,” “two,” “three,” “four,” or “five or more.” The fact that four of the five choices in the closed-ended question provided a response that was greater than one implied that the correct answer was probably more than one.¹⁵² 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. Similarly, an open-ended question that asks, “[W]hich company or store do you think puts out this shirt?” indicates to the respondent that the appropriate answer is the name of a company or store. The question would be leading if the respondent would have considered other possibilities (e.g., an individual or Webstore) if the question had not provided the frame of a company or store.¹⁵³ Thus, the wording of a question, open-ended or closed-ended, can be leading or non-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 chosen will be meaningful only if the list of choices is exhaustive—that is, if the choices cover all possible answers a respondent might give to 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.¹⁵⁴ Moreover, if respondents are told explicitly that they are

150. See *Vidal Sassoon, Inc. v. Bristol-Myers Co.*, 661 F.2d 272, 273 (2d Cir. 1981).

151. This was the wording of the closed-ended question in the survey discussed in *Vidal Sassoon*, 661 F.2d at 275–76, without the closed-ended options that were supplied in that survey.

152. 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., Roger Tourangeau & Tom W. Smith, *Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context*, 60 *Pub. Op. Q.* 275, 292 (1996).

153. *Smith v. Wal-Mart Stores, Inc.*, 537 F. Supp. 2d 1302, 1331–32 (N.D. Ga. 2008).

154. See, e.g., *American Home Prods. Corp. v. Johnson & Johnson*, 654 F. Supp. 568, 581 (S.D.N.Y. 1987).

not limited to the choices presented, most respondents nevertheless will select an answer from among the listed ones.¹⁵⁵

One form of closed-ended question format that typically produces some distortion is the popular agree/disagree, true/false, or yes/no question. Although this format is appealing because it is easy to write and score these questions and their responses, the format is also seriously problematic. With its simplicity comes acquiescence, “[T]he tendency to endorse any assertion made in a question, regardless of its content,” is a systematic source of bias that has produced an inflation effect of 10% across a number of studies.¹⁵⁶ Only when control groups or control questions are added to the survey design can this question format provide reasonable response estimates.¹⁵⁷

Although many courts prefer open-ended questions on the ground that they tend to be less leading, the value of any open-ended or closed-ended question depends on the information it conveys in the question and, in the case of closed-ended questions, in the choices provided. 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 more suitable for assessing choices between well-identified options or obtaining ratings on a clear set of alternatives.

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, or may ask for clarification. In such instances, 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. They may also be instructed what clarification they can provide. In all of these situations, interviewers should record verbatim both what the respondent says and what the interviewer says in the attempt to get or provide clarification. Failure to record every part of the 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.

155. See Howard Schuman, *Ordinary Questions, Survey Questions, and Policy Questions*, 50 Pub. Opinion Q. 432, 435–36 (1986).

156. Jon A. Krosnick, *Survey Research*, 50 Ann. Rev. Psychol. 537, 552 (1999).

157. See *infra* Section IV.F.

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.¹⁵⁸ 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?” Or “Could you say that another way?”) are relatively innocuous and non-controversial in content, but persistent continued requests for further responses to the same or nearly identical questions may convey the idea to the respondent that he or she has not yet produced the “right” answer.¹⁵⁹ Interviewers should be trained in delivering probes to maintain a professional and neutral relationship with the respondent (as they should during the rest of the interview), 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 interviewers so that they 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.¹⁶⁰ For example, although asking a general question before a more specific 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 move from the general (e.g., “What do you recall being discussed

158. Floyd J. Fowler, Jr. & Thomas W. Mangione, *Standardized Survey Interviewing: Minimizing Interviewer-Related Error* 41–42 (1990).

159. *See, e.g., Johnson & Johnson–Merck Consumer Pharms. Co. v. Rhone-Poulenc Rorer Pharms., Inc.*, 19 F.3d 125, 135 (3d Cir. 1994); *American Home Prods. Corp. v. Procter & Gamble Co.*, 871 F. Supp. 739, 748 (D.N.J. 1994).

160. *See Schuman & Presser, supra* note 134, at 23, 56–74. Krosnick & Presser, *supra* note 126, at 278–81. 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 that disclosed 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. *See also E. & J. Gallo Winery v. Pasatiempos Gallo, S.A.*, 905 F. Supp. 1403, 1409–10 (E.D. Cal. 1994) (court recognized that earlier questions referring to playing cards, board or table games, or party supplies, such as confetti, increased the likelihood that respondents would include these items in answers to the questions that followed).

in the advertisement?”) 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?”).¹⁶¹

The mode of questioning can influence the form that an order effect takes. When respondents are shown response alternatives visually, as in mail surveys and other self-administered questionnaires or in face-to-face interviews when respondents are shown a card containing response alternatives, they are more likely to select the first choice offered (a primacy effect).¹⁶² In contrast, when response alternatives are presented orally, as in telephone surveys, respondents are more likely to choose the last choice offered (a recency effect).¹⁶³ 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 depend on 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 respondents to choose the first product mentioned,¹⁶⁴ which order of presentation will produce the more accurate response?¹⁶⁵ To control for order effects, the order of the questions and the order of the response choices in a survey should be rotated,¹⁶⁶ 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 orders¹⁶⁷ are distributed randomly among respondents, no response alternative will have an inflated chance of being selected because of its position, and the average of the three will provide a reasonable estimate of response level.¹⁶⁸

161. 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).

162. Krosnick & Presser, *supra* note 126, at 280.

163. *Id.*

164. Similarly, candidates in the first position on the ballot tend to attract extra votes. J.M. Miller & Jon A. Krosnick, *The Impact of Candidate Name Order on Election Outcomes*, 62 Pub. Op. Q. 291 (1998).

165. See *Rust Env't & Infrastructure, Inc. v. Teunissen*, 131 F.3d 1210, 1218 (7th Cir. 1997) (survey did not pass muster in part because of failure to incorporate random rotation of corporate names that were the subject of a trademark dispute).

166. See, e.g. *Winning Ways, Inc. v. Holloway Sportswear, Inc.*, 913 F. Supp. 1454, 1465–67 (D. Kan. 1996) (failure to rotate the order in which the jackets were shown to the consumers led to reduced weight for the survey); *Procter & Gamble Pharms., Inc. v. Hoffmann-La Roche, Inc.*, 2006 U.S. Dist. LEXIS 64363, 2006-2 Trade Cas. (CCH) P75465 (S.D.N.Y. Sept. 6, 2006).

167. Actually, there are six possible orders of the three alternatives: ABC, ACB, BAC, BCA, CAB, and CBA. Thus, the optimal survey design would allocate equal numbers of respondents to each of the six possible orders.

168. Although rotation is desirable, many surveys are conducted with no attention to this potential bias. Because it is impossible to know in the abstract whether a particular question suffers much, little, or not at all from an order bias, lack of rotation should not preclude reliance on the answer to the question, but it should reduce the weight given to that answer.

F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question?

Many surveys are designed not simply to describe attitudes or beliefs or reported behaviors, but to determine the source of those attitudes or beliefs or behaviors. That is, the purpose of the survey is to test a causal proposition. For example, how does a trademark or the content of a commercial affect respondents' perceptions or understanding of a product or commercial? Thus, the question is not merely whether consumers hold inaccurate beliefs about Product A, but whether exposure to the commercial misleads the consumer into thinking that Product A is a superior pain reliever. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that simply records consumers' impressions after they view the commercial may reflect those preexisting beliefs rather than impressions produced by the commercial.

Surveys that merely 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.¹⁶⁹ 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, "[W]hich [brand] would you say uses these stripes on their package?"¹⁷⁰ The court recognized that the high percentage of respondents selecting "Mennen" from an array of brand names may have represented "merely a playback of brand share";¹⁷¹ 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.¹⁷²

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 (e.g., "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

169. See, e.g., *Procter & Gamble Co. v. Ultreo, Inc.*, 574 F. Supp. 2d 339, 351–52 (S.D.N.Y. 2008) (survey was unreliable because it failed to control for the effect of preexisting beliefs).

170. *Mennen Co. v. Gillette Co.*, 565 F. Supp. 648, 652 (S.D.N.Y. 1983), *aff'd*, 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.*

171. *Id.*

172. See also *Upjohn Co. v. American Home Prods. Corp.*, No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at *42–44 (W.D. Mich. Apr. 5, 1996).

source of their impressions.¹⁷³ The more routine the idea being examined in the survey (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 (1) preexisting impressions; (2) general expectations about what commercials typically say (e.g., the product being advertised is better than its competitors); or (3) 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 one or more appropriate control groups, the survey expert can test directly the influence of the stimulus.¹⁷⁴ In the simplest version of such a survey experiment, respondents are assigned randomly to one of two conditions.¹⁷⁵ For example, respondents assigned to the experimental condition view an allegedly deceptive commercial, and respondents assigned to the control condition either view a commercial that does not contain the allegedly deceptive material or do not view any commercial.¹⁷⁶ Respondents in both the experimental and control groups answer the same set of questions about the allegedly deceptive message. The effect of the commercial's allegedly deceptive message is evaluated by comparing the responses made by the experimental group members with those of the control group members. If 40% of the respondents in the experimental group responded indicating that they received the deceptive message (e.g., the advertised product has fewer calories than its competitor), whereas only 8% of the respondents in the control group gave that response, the difference between 40% and 8% (within the limits of sampling error¹⁷⁷) can be attributed only to the allegedly deceptive message. Without the control group, it is not possible to determine how much of the 40% is attributable to respondents' preexisting beliefs

173. See Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 *Psychol. Rev.* 231 (1977).

174. See Shari S. Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 *Law & Hum. Behav.* 239, 244–46 (1989); Jacob Jacoby & Constance Small, *Applied Marketing: The FDA Approach to Defining Misleading Advertising*, 39 *J. Marketing* 65, 68 (1975). See also David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.A, in this manual.

175. 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* Section III.C.

176. This alternative commercial could be a “tombstone” advertisement that includes only the name of the product or a more elaborate commercial that does not include the claim at issue.

177. For a discussion of sampling error, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.A, in this manual.

or other background noise (e.g., respondents who misunderstand the question or misstate their responses). Both preexisting beliefs and other background noise should have produced similar response levels in the experimental and control groups. In addition, if respondents who viewed the allegedly deceptive commercial respond differently than respondents who viewed the control commercial, the difference cannot be merely the result of a leading question, because both groups answered the same question. The ability to evaluate the effect of the wording of a particular question makes the control group design particularly useful in assessing responses to closed-ended questions,¹⁷⁸ which may encourage guessing or particular responses. Thus, the focus on the response level in a control group design is not on the absolute response level, but on the difference between the response level of the experimental group and that of the control group.¹⁷⁹

In designing a survey-experiment, the expert should select a stimulus for the control group that shares as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed.¹⁸⁰ Although a survey with an imperfect control group may provide better information than a survey with no control group at all, the choice of an appropriate control group requires some care and should influence the weight that the survey receives. For example, a control stimulus should not be less attractive than the experimental stimulus if the survey is designed to measure how familiar the experimental stimulus is to respondents, because attractiveness may affect perceived familiarity.¹⁸¹ Nor should the control stimulus share with the experimental stimulus the feature whose impact is being assessed. If, for example, the control stimulus in a case of alleged trademark infringement is itself a likely source of consumer confusion, reactions to the experimental and control stimuli may not

178. The Federal Trade Commission has long recognized the need for some kind of control for closed-ended questions, although it has not specified the type of control that is necessary. See *Stouffer Foods Corp.*, 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *31 (Sept. 26, 1994).

179. See, e.g., *Cytosport, Inc. v. Vital Pharms., Inc.*, 617 F. Supp. 2d 1051, 1075–76 (E.D. Cal. 2009) (net confusion level of 25.4% obtained by subtracting 26.5% in the control group from 51.9% in the test group).

180. See, e.g., *Skechers USA, Inc. v. Vans, Inc.*, No. CV-07-01703, 2007 WL 4181677, at *8–9 (C.D. Cal. Nov. 20, 2007) (in trade dress infringement case, control stimulus should have retained design elements not at issue); *Procter & Gamble Pharms., Inc. v. Hoffman-LaRoche, Inc.*, No. 06-Civ-0034, 2006 U.S. Dist. LEXIS 64363, at *87 (S.D.N.Y. Sept. 6, 2006) (in false advertising action, disclaimer was inadequate substitute for appropriate control group).

181. See, e.g., *Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club L.P.*, 34 F.3d 410, 415–16 (7th Cir. 1994) (court recognized that the name “Baltimore Horses” was less attractive for a sports team than the name “Baltimore Colts.”); see also *Reed-Union Corp. v. Turtle Wax, Inc.*, 77 F.3d 909, 912 (7th Cir. 1996) (court noted that one expert’s choice of a control brand with a well-known corporate source was less appropriate than the opposing expert’s choice of a control brand whose name did not indicate a specific corporate source); *Louis Vuitton Malletier v. Dooney & Bourke, Inc.*, 525 F. Supp. 2d 576, 595 (S.D.N.Y. 2007) (underreporting of background “noise” likely occurred because handbag used as control was quite dissimilar in shape and pattern to both plaintiff and defendant’s bags).

differ because both cause respondents to express the same level of confusion.¹⁸² In an extreme case, an inappropriate control may do nothing more than control for the effect of the nature or wording of the survey questions (e.g., acquiescence).¹⁸³ That may not be enough to rule out other explanations for different or similar responses to the experimental and control stimuli. Finally, it may sometimes be appropriate to have more than one control group to assess precisely what is causing the response to the experimental stimulus (e.g., in the case of an allegedly deceptive ad, whether it is a misleading graph or a misleading claim by the announcer; or in the case of allegedly infringing trade dress, whether it is the style of the font used or the coloring of the packaging).

Explicit attention to the value of control groups in trademark and deceptive-advertising litigation is a relatively recent phenomenon, but courts have increasingly come to recognize the central role the control group can play in evaluating claims.¹⁸⁴ A LEXIS search using *Lanham Act* and *control group* revealed only 4 federal district court cases before 1991 in which surveys with control groups were discussed, 16 in the 9 years from 1991 to 1999, and 46 in the 9 years between 2000 and 2008, a rate of growth that far exceeds the growth in Lanham Act litigation. In addition, courts in other cases have described or considered surveys using control group designs without labeling the comparison group a control group.¹⁸⁵ Indeed, one reason why cases involving surveys with control groups may be underrepresented in reported cases is that a survey with a control group produces

182. See, e.g., *Western Publ'g Co. v. Publications Int'l, Ltd.*, No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917, at *45 (N.D. Ill. May 2, 1995) (court noted that the control product was “arguably more infringing than” the defendant’s product) (emphasis omitted). See also *Classic Foods Int'l Corp. v. Kettle Foods, Inc.*, 2006 U.S. Dist. LEXIS 97200 (C.D. Cal. Mar. 2, 2006); *McNeil-PPC, Inc. v. Merisant Co.*, 2004 U.S. Dist. LEXIS 27733 (D.P.R. July 29, 2004).

183. See text accompanying note 156, *supra*.

184. See, e.g., *SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck*, 2001 U.S. Dist. LEXIS 7061, at *37 (S.D.N.Y. June 1, 2001) (survey to assess implied falsity of a commercial not probative in the absence of a control group); *Consumer American Home Prods. Corp. v. Procter & Gamble Co.*, 871 F. Supp. 739, 749 (D.N.J. 1994) (discounting survey results based on failure to control for participants’ preconceived notions); *ConAgra, Inc. v. Geo. A. Hormel & Co.*, 784 F. Supp. 700, 728 (D. Neb. 1992) (“Since no control was used, the . . . study, standing alone, must be significantly discounted.”), *aff’d*, 990 F.2d 368 (8th Cir. 1993).

185. *Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club L.P.*, No. 94727-C, 1994 U.S. Dist. LEXIS 19277, at *10–11 (S.D. Ind. June 27, 1994), *aff’d*, 34 F.3d 410 (7th Cir. 1994). In *Indianapolis Colts*, the district court described a survey conducted by the plaintiff’s expert in which half of the interviewees were shown a shirt with the name “Baltimore CFL Colts” on it and half were shown a shirt on which the word “Horses” had been substituted for the word “Colts.” *Id.* The court noted that the comparison of reactions to the horse and colt versions of the shirt made it possible “to determine the impact from the use of the word ‘Colts.’” *Id.* at *11. See also *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). See also *Simon Prop. Group L.P. v. MySimon, Inc.*, 104 F. Supp. 2d 1033 (S.D. Ind. 2000) (court criticized the survey design based on the absence of a control that could show that results were produced by legally relevant confusion).

less ambiguous findings, which may lead to a resolution before a preliminary injunction hearing or trial occurs.

A less 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 at issue. In a trademark dispute, for example, a survey indicated that 7.2% of respondents believed that “The Mart” and “K-Mart” were 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.¹⁸⁶

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.¹⁸⁷

Every measure of opinion or belief in a survey reflects some degree of error. Control groups and, as a second choice, 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 have traditionally been used to collect survey data: (1) in-person interviews, (2) telephone interviews, and (3) mail questionnaires.¹⁸⁸ Recently, in the wake of increasing use of the Internet, researchers have added Web-based surveys to their arsenal of tools. Surveys using in-person and telephone interviews, too, now regularly rely on computerized data collection.¹⁸⁹

186. *S.S. Kresge Co. v. United Factory Outlet, Inc.*, 598 F.2d 694, 697 (1st Cir. 1979). Note that the aggregate percentages reported here do not reveal how many of the same respondents were confused by both names, an issue that may be relevant in some situations. See Joseph L. Gastwirth, *Reference Guide on Survey Research*, 36 *Jurimetrics J.* 181, 187–88 (1996) (review essay).

187. *E.I. du Pont de Nemours & Co. v. Yoshida Int’l, Inc.*, 393 F. Supp. 502, 526–27 & n.54 (E.D.N.Y. 1975); see also *Donchez v. Coors Brewing Co.*, 392 F.3d 1211, 1218 (10th Cir. 2004) (respondents evaluated eight brand and generic names in addition to the disputed name). A similar approach is used in assessing secondary meaning.

188. 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.

189. Wright & Marsden, *supra* note 1, at 13–14.

The interviewer conducting a computer-assisted interview (CAI), whether by telephone (CATI) or face-to-face (CAPI), follows the computer-generated script for the interview and enters the respondent's answers as the interview proceeds. A primary advantage of CATI and other CAI procedures is that skip patterns can be built into the program. If, for example, the respondent answers *yes* when asked whether she has ever been the victim of a burglary, the computer will generate further questions about the burglary; if she answers *no*, the program will automatically skip the followup burglary questions. Interviewer errors in following the skip patterns are therefore avoided, making CAI procedures particularly valuable when the survey involves complex branching and skip patterns.¹⁹⁰ CAI procedures also can be used to control for order effects by having the program rotate the order in which the questions or choices are presented.¹⁹¹

Recent innovations in CAI procedures include audio computer-assisted self-interviewing (ACASI) in which the respondent listens to recorded questions over the telephone or reads questions from a computer screen while listening to recorded versions of them through headphones. The respondent then answers verbally or on a keypad. ACASI procedures are particularly useful for collecting sensitive information (e.g., illegal drug use and other HIV risk behavior).¹⁹²

All CAI procedures require additional planning to take advantage of the potential for improvements in data quality. When a CAI protocol is used in a survey presented in litigation, the party offering the survey should supply for inspection the computer program that was used to generate the interviews. Moreover, CAI procedures do not eliminate the need for close monitoring of interviews to ensure that interviewers are accurately reading the questions in the interview protocol and accurately entering the respondent's answers.

The choice of any 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.¹⁹³ When the questions are complex and the interviewers are skilled, in-person interviewing provides the maximum opportunity to

190. Willem E. Saris, *Computer-Assisted Interviewing* 20, 27 (1991).

191. See, e.g., *Intel Corp. v. Advanced Micro Devices, Inc.*, 756 F. Supp. 1292, 1296–97 (N.D. Cal. 1991) (survey designed to test whether the term *386* as applied to a microprocessor was generic used a CATI protocol that tested reactions to five terms presented in rotated order).

192. See, e.g., N. Galai et al., *ACASI Versus Interviewer-Administered Questionnaires for Sensitive Risk Behaviors: Results of a Cross-Over Randomized Trial Among Injection Drug Users* (abstract, 2004), available at <http://gateway.nlm.nih.gov/MeetingAbstracts/ma?f=102280272.html>.

193. A mail survey also can include limited visual materials but cannot exercise control over when and how the respondent views them.

clarify or probe. Unlike a mail survey, both in-person and telephone interviews have the capability 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. Interviewers also can directly verify who is completing the survey, a check that is unavailable in mail and Web-based surveys. As described *infra* Section V.A, appropriate interviewer training, as well as monitoring of the implementation of interviewing, 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. In-person interview quality in recent years has been assisted by technology. Using computer-assisted personal interviewing (CAPI), the interviewer reads the questions off the screen of a laptop computer and then enters responses directly.¹⁹⁴ This support makes it easier to follow complex skip patterns and to promptly submit results via the Internet to the survey center.

2. Telephone interviews

Telephone surveys offer a comparatively fast and lower-cost alternative to in-person surveys and are particularly useful when the population is large and geographically dispersed. Telephone interviews (unless supplemented with mailed or e-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.¹⁹⁵ 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."¹⁹⁶ In contrast, when words are not sufficient because, for example, the survey is assessing reactions to the trade

194. Wright & Marsden, *supra* note 1, at 13.

195. See, e.g., State v. Baumruk, 85 S.W.3d 644 (Mo. 2002). (overturning the trial court's decision to ignore a survey that found about 70% of county residents remembered the shooting that led to the trial and that of those who had heard about the shooting, 98% believed that the defendant was either definitely guilty or probably guilty); State v. Erickstad, 620 N.W.2d 136, 140 (N.D. 2000) (denying change of venue motion based on media coverage, concluding that "defendants [need to] submit qualified public opinion surveys, other opinion testimony, or any other evidence demonstrating community bias caused by the media coverage"). 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); National Jury Project, *Jurywork: Systematic Techniques* (2d ed. 2008).

196. McDonald's Corp. v. McBagel's, Inc., 649 F. Supp. 1268, 1278 (S.D.N.Y. 1986).

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.¹⁹⁷

In evaluating the sampling used in a telephone survey, the trier of fact should consider:

1. Whether (when prospective respondents are not business personnel) some form of random-digit dialing¹⁹⁸ was used instead of or to supplement telephone numbers obtained from telephone directories, because a high percentage of all residential telephone numbers in some areas may be unlisted;¹⁹⁹
2. Whether any attempt was made to include cell phone users, particularly the growing subpopulation of individuals who rely solely on cell phones for telephone services;²⁰⁰
3. 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;²⁰¹ and
4. Whether interviewers were required to call back multiple times at several different times of the day and on different days to increase the likelihood of contacting individuals or businesses with different schedules.²⁰²

197. See *Thompson Med. Co. v. Pfizer Inc.*, 753 F.2d 208 (2d Cir. 1985); *Incorporated Publ'g 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).

198. Random-digit dialing provides coverage of households with both listed and unlisted telephone numbers by generating numbers at random from the sampling 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).

199. Studies comparing listed and unlisted household characteristics show some important differences. *Id.* at 76.

200. According to a 2009 study, an estimated 26.5% of households cannot be reached by landline surveys, because 2.0% have no phone service and 24.5% have only a cell phone. Stephen J. Blumberg & Julian V. Luke, *Wireless Substitution: Early Release of Estimates Based on the National Health Interview Survey, July–December 2009* (2010), available at <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201005.pdf>. People who can be reached only by cell phone tend to be younger and are more likely to be African American or Hispanic and less likely to be married or to own their home than individuals reachable on a landline. Although at this point, the effect on estimates from landline-only telephone surveys appears to be minimal on most topics, on some issues (e.g., voter registration) and within the population of young adults, the gap may warrant consideration. Scott Keeter et al., *What's Missing from National RDD Surveys? The Impact of the Growing Cell-Only Population*, Paper presented at the 2007 Conference of AAPOR, May 2007.

201. This is a consideration 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.

202. This applied equally to in-person interviews.

Telephone surveys that do not include these procedures may not provide precise measures of the characteristics of a representative sample of respondents, but may be adequate for providing rough approximations. The vulnerability of the survey depends on the information being gathered. More elaborate procedures are advisable for achieving a representative sample of respondents if the survey instrument requests information that is likely to differ for individuals with listed telephone numbers versus individuals with unlisted telephone numbers, individuals rarely at home versus those usually at home, or groups who are more versus less likely to rely exclusively on cell phones.

The report submitted by a survey expert who conducts a telephone survey should specify:

1. The procedures that were used to identify potential respondents, including both the procedures used to select the telephone numbers that were called and the procedures used to identify the qualified individual to question),
2. The number of telephone numbers for which no contact was made; and
3. The number of contacted potential respondents who refused to participate in the survey.²⁰³

Like CAPI interviewing,²⁰⁴ computer-assisted telephone interviewing (CATI) facilitates the administration and data entry of large-scale surveys.²⁰⁵ A computer protocol may be used to generate and dial telephone numbers as well as to guide the interviewer.

3. Mail questionnaires

In general, mail surveys tend to be substantially less costly than both in-person and telephone surveys.²⁰⁶ Response rates tend to be lower for self-administered mail surveys than for telephone or face-to-face surveys, but higher than for their Web-based equivalents.²⁰⁷ Procedures that raise response rates include multiple mailings, highly personalized communications, prepaid return envelopes, incentives or gratuities, assurances of confidentiality, first-class outgoing postage, and followup reminders.²⁰⁸

203. Additional disclosure and reporting features applicable to surveys in general are described in Section VII.B, *infra*.

204. See text accompanying note 194, *supra*.

205. See Roger Tourangeau et al., *The Psychology of Survey Response* 289 (2000); Saris, *supra* note 190.

206. See Chase H. Harrison, *Mail Surveys and Paper Questionnaires*, in *Handbook of Survey Research*, *supra* note 1, at 498, 499.

207. See Mick Couper et al., *A Comparison of Mail and E-Mail for a Survey of Employees in Federal Statistical Agencies*, 15 *J. Official Stat.* 39 (1999); Mick Couper, *Web Surveys: A Review of Issues and Approaches* 464, 473 (2001).

208. See, e.g., Richard J. Fox et al., *Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response*, 52 *Pub. Op. Q.* 467, 482 (1988); Kenneth D. Hopkins & Arlen R.

A mail survey will not produce a high rate of return 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 individuals will tend to be underrepresented. 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 respondent's 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 prevent respondents from discussing the questions and answers with others before completing the survey and to control the order in which respondents answer the questions. Although skilled design of questionnaire format, question order, and the appearance of the individual pages of a survey can minimize these problems,²⁰⁹ if it is crucial to have respondents answer questions in a particular order, a mail survey cannot be depended on to provide adequate data.

4. Internet surveys

A more recent innovation in survey technology is the Internet survey in which potential respondents are contacted and their responses are collected over the Internet. Internet surveys in principle can reduce substantially the cost of reaching potential respondents. Moreover, they offer some of the advantages of in-person interviews by enabling the respondent to view pictures, videos, and lists of response choices on the computer screen during the survey. A further advantage is that whenever a respondent answers questions presented on a computer screen, whether over the Internet or in a dedicated facility, the survey can build in a variety of controls. In contrast to a mail survey in which the respondent can examine and/or answer questions out of order and may mistakenly skip questions, a computer-administered survey can control the order in which the questions are displayed so that the respondent does not see a later question before answering an earlier one and so that the respondent cannot go back to change an answer previously given to an earlier question in light of the questions that follow it. The order of the questions or response options can be rotated easily to control for order effects. In addition, the structure permits the survey to remind, or even require, the respondent to answer a question before the next question is presented. One advantage of computer-administered surveys over interviewer-administered

Gullickson, *Response Rates in Survey Research: A Meta-Analysis of the Effects of Monetary Gratuities*, 61 J. Experimental Educ. 52, 54–57, 59 (1992); Eleanor Singer et al., *Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature*, 59 Pub. Op. Q. 66, 71 (1995); see generally Don A. Dillman, *Internet Mail and Mixed-Mode Surveys: The Tailored Design Method* (3d ed. 2009).

209. Dillman, *supra* note 208, at 151–94.

surveys is that they eliminate interviewer error because the computer presents the questions and the respondent records her own answers.

Internet surveys do have limitations, and many questions remain about the extent to which those limitations impair the quality of the data they provide. A key potential limitation is that respondents accessible over the Internet may not fairly represent the relevant population whose responses the survey was designed to measure. Although Internet access has not approached the 95% penetration achieved by the telephone, the proportion of individuals with Internet access has grown at a remarkable rate, as has the proportion of individuals who regularly use a computer. For example, according to one estimate, use of the Internet among adults jumped from 22% in 1997 to 60% in 2003.²¹⁰ Despite this rapid expansion, a digital divide still exists, so that the “have-nots” are less likely to be represented in surveys that depend on Internet access. The effect of this divide on survey results will depend on the population the survey is attempting to capture. For example, if the target population consists of computer users, any bias from systematic underrepresentation is likely to be minimal. In contrast, if the target population consists of owners of television sets, a proportion of whom may not have Internet access, significant bias is more likely. The trend toward greater access to the Internet is likely to continue, and the issue of underrepresentation may disappear in time. At this point, a party presenting the results of a Web-based survey should be prepared to provide evidence on how coverage limitations may have affected the pattern of survey results.

Even if noncoverage error is not a significant concern, courts evaluating a Web-based survey must still determine whether the sampling approach is adequate. That evaluation will depend on the type of Internet survey involved, because Web-based surveys vary in fundamental ways.

At one extreme is the list-based Web survey. This Web survey is sent to a closed set of potential respondents drawn from a list that consists of the e-mail addresses of the target individuals (e.g., all students at a university or employees at a company where each student or employee has a known e-mail address).

At the other extreme is the self-selected Web survey in which Web users in general, or those who happen to visit a particular Web site, are invited to express their views on a topic and they participate simply by volunteering. Whereas the list-based survey enables the researcher to evaluate response rates and often to assess the representativeness of respondents on a variety of characteristics, the self-selected Web survey provides no information on who actually participates or how representative the participants are. Thus, it is impossible to evaluate nonresponse error or even participation rates. Moreover, participants are very likely to self-select on the basis of the nature of the topic. These self-selected pseudosurveys resemble reader polls published in magazines and do not meet standard criteria for legitimate surveys

210. Jennifer C. Day et al., *Computer and Internet Use in the United States: 2003*, 8–9 (U.S. Census Bureau 2005).

admissible in court.²¹¹ Occasionally, proponents of such polls tout the large number of respondents as evidence of the weight the results should be given, but the size of the sample cannot cure the likely participation bias in such voluntary polls.²¹²

Between these two extremes is a large category of Web-based survey approaches that researchers have developed to address concerns about sampling bias and nonresponse error. For example, some approaches create a large database of potential participants by soliciting volunteers through appeals on well-traveled sites.²¹³ Based on the demographic data collected from those who respond to the appeals, a sample of these panel members are asked to participate in a particular survey by invitation only. Responses are weighted to reduce selection bias.²¹⁴ An expert presenting the results from such a survey should be prepared to explain why the particular weighting approach can be relied upon to achieve that purpose.²¹⁵

Another approach that is more costly uses probability sampling from the initial contact with a potential respondent. Potential participants are initially contacted by telephone using random-digit dialing procedures. Those who lack Internet access are provided with the technology to participate. Members from the panel are then invited to participate in a particular survey, and the researchers know the characteristics of participants and nonparticipants from the initial telephone contact.²¹⁶ For all surveys that rely on preselected panels, whether nonrandomly or randomly selected, questions have been raised about panel conditioning (i.e., the effect of having participants in earlier surveys respond to later surveys) and the relatively low rate of response to survey invitations. An expert presenting results from a Web-based survey should be prepared to address these issues and to discuss how they may have affected the results.

Finally, the recent proliferation of Internet surveys has stimulated a growing body of research on the influence of formatting choices in Web surveys. Evidence from this research indicates that formatting decisions can significantly affect the quality of survey responses.²¹⁷

211. See, e.g., *Merisant Co. v. McNeil Nutritionals, LLC*, 242 F.R.D. 315 (E.D. Pa. 2007) (report on results from AOL “instant poll” excluded).

212. See, e.g., Couper (2001), *supra* note 207, at 480–81 (a self-selected Web survey conducted by the National Geographic Society through its Web site attracted 50,000 responses; a comparison of the Canadian respondents with data from the Canadian General Social Survey telephone survey conducted using random-digit dialing showed marked differences on a variety of response measures).

213. See, e.g., *Ecce Panis, Inc. v. Maple Leaf Bakery, Inc.* 2007 U.S. Dist. LEXIS 85780 (D. Ariz. Nov. 7, 2007).

214. See, e.g., *Philip Morris USA, Inc. v. Otamedia Limited*, 2005 U.S. Dist. LEXIS 1259 (S.D.N.Y. Jan. 28, 2005).

215. See, e.g., *A&M Records, Inc. v. Napster, Inc.* 2000 WL 1170106 (N.D. Cal. Aug. 10, 2000) (court refused to rely on results from Internet panel survey when expert presenting the results showed lack of familiarity with panel construction and weighting methods).

216. See, e.g., *Price v. Philip Morris, Inc.*, 219 Ill. 2d 182, 848 N.E.2d 1 (2005).

217. See, e.g., Mick P. Couper et al., *What They See Is What We Get: Response Options for Web Surveys*, 22 Soc. Sci. Computer Rev. 111 (2004) (comparing order effects with radio button and

A final approach to data collection does not depend on a single mode, but instead involves a mixed-mode approach. By combining modes, the survey design may increase the likelihood that all sampling members of the target population will be contacted. For example, a person without a landline may be reached by mail or e-mail. Similarly, response rates may be increased if members of the target population are more likely to respond to one mode of contact versus another. For example, a person unwilling to be interviewed by phone may respond to a written or e-mail contact. If a mixed-mode approach is used, the questions and structure of the questionnaires are likely to differ across modes, and the expert should be prepared to address the potential impact of mode on the answers obtained.²¹⁸

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.”²¹⁹ Properly trained interviewers receive detailed written 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. Moreover, interviewers 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 practice in resisting invitations to express the interviewer’s beliefs or opinions). Although procedures vary, there is evidence that interviewer performance suffers with less than a day of training in general interviewing skills and techniques for new interviewers.²²⁰

drop-box formats); Andy Peytchev et al., *Web Survey Design: Paging Versus Scrolling*, 70 Pub. Op. Q. 212 (2006) (comparing the effects of presenting survey questions in a multitude of short pages or in long scrollable pages).

218. Don A. Dillman & Benjamin L. Messer, *Mixed-Mode Surveys*, in Wright & Marsden, *supra* note 1, at 550, 553.

219. Toys “R” Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983).

220. Fowler & Mangione, *supra* note 158, at 117; Nora Cate Schaeffer et al., *Interviewers and Interviewing*, in *Handbook of Survey Research*, *supra* note 1, at 437, 460.

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. Note, however, that in surveys conducted using CAPI or CATI procedures, the interviewer will be guided by the computer used to administer the questionnaire.

If the questions require specific probes to clarify 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 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 outside 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 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.²²¹

Some compromises may be accepted when surveys must be conducted swiftly. In trademark and deceptive advertising cases, 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, as well as monitoring and validation to ensure quality control,²²² and complete disclosure of the methods used for all of the procedures followed 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 in surveys conducted for litigation to do double-blind

221. *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).

222. See Section V.C., *infra*.

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 or implicit clues about the sponsorship of the survey or the expected responses. Explicit clues could include a sponsor's letterhead appearing on the survey; implicit clues could include 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 *no* will be checked.²²³

Nonetheless, in some surveys (e.g., some government surveys), disclosure of the survey's sponsor to respondents (and thus to interviewers) is required. Such surveys call for an evaluation of the likely biases introduced by interviewer or respondent awareness of the survey's sponsorship. 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 kinds of responses the sponsor expects or would find acceptable.²²⁴ When interviewers are well trained, their awareness of sponsorship may be a less serious threat than respondents' awareness. The empirical evidence for the effects of interviewers' prior expectations on respondents' answers generally reveals modest effects when the interviewers are well trained.²²⁵

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. Some

223. See *Centaur Communications, Ltd. v. A/S/M Communications, Inc.*, 652 F. Supp. 1105, 1111 n.3 (S.D.N.Y. 1987) (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).

224. See, e.g., Stanley Presser et al., *Survey Sponsorship, Response Rates, and Response Effects*, 73 Soc. Sci. Q. 699, 701 (1992) (different responses to a university-sponsored telephone survey and a newspaper-sponsored survey for questions concerning attitudes toward the mayoral primary, an issue on which the newspaper had taken a position).

225. See, e.g., Seymour Sudman et al., *Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses*, 6 Soc. Methods & Res. 171, 181 (1977).

monitoring systems, both telephone and field, now use recordings, procedures that may require permission from respondents.

Second, validation of interviews occurs when respondents in a sample are recontacted to ask whether the initial interviews took place and to determine whether the respondents were qualified to participate in the survey. Validation callbacks may also collect data on a few key variables to confirm that the correct respondent has been interviewed. The standard procedure for validation of in-person interviews is to telephone a random sample of about 10% to 15% of the respondents.²²⁶ 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 typically 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. In computer-assisted interviews, further validation information can be obtained from the timings that can be automatically recorded when an interview occurs.

A third way to verify that the interviews were conducted properly is to examine the work done by each individual interviewer. By reviewing the interviews and individual responses recorded by each interviewer and comparing patterns of response across interviewers, researchers can identify any response patterns or inconsistencies that warrant further investigation.

When a survey is conducted at the request of a party for litigation rather than in the normal course of business, a heightened standard for validation checks may be appropriate. Thus, independent validation of a random sample of interviews by a third party rather than by the field service that conducted the interviews increases the trustworthiness of the survey results.²²⁷

VI. Data Entry and Grouping of Responses

A. What Was Done to Ensure That the Data Were Recorded Accurately?

Analyzing the results of a survey requires that the data obtained on each sampled element be recorded, edited, and often coded before the results can be tabulated

226. See, e.g., *Davis v. Southern Bell Tel. & Tel. Co.*, No. 89-2839, 1994 U.S. Dist. LEXIS 13257, at *16 (S.D. Fla. Feb. 1, 1994); *National Football League Properties, Inc. v. New Jersey Giants, Inc.*, 637 F. Supp. 507, 515 (D.N.J. 1986).

227. In *Rust Environment & Infrastructure, Inc. v. Teunissen*, 131 F.3d 1210, 1218 (7th Cir. 1997), the court criticized a survey in part because it “did not comport with accepted practice for independent validation of the results.”

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 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 all cases, the verbatim responses should be available so that they can be recoded using alternative criteria.²²⁸

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 often too late to correct methodological deficiencies that could have been addressed in the planning stages of the survey. The plaintiff in a trademark case²²⁹ submitted a set of proposed survey questions to the trial judge, who ruled that the survey results

228. See, e.g., *Revlon Consumer Prods. Corp. v. Jennifer Leather Broadway, Inc.*, 858 F. Supp. 1268, 1276 (S.D.N.Y. 1994) (inconsistent scoring and subjective coding led court to find survey so unreliable that it was entitled to no weight), *aff'd*, 57 F.3d 1062 (2d Cir. 1995); *Rock v. Zimmerman*, 959 F.2d 1237, 1253 n.9 (3d Cir. 1992) (court found that responses on a change-of-venue survey incorrectly categorized respondents who believed the defendant was insane as believing he was guilty); *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 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); see also *Cumberland Packing Corp. v. Monsanto Co.*, 140 F. Supp. 2d 241 (E.D.N.Y. 2001) (court examined verbatim responses that respondents gave to arrive at a confusion level substantially lower than the level reported by the survey expert).

229. *Union Carbide Corp. v. Ever-Ready, Inc.*, 392 F. Supp. 280 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366 (7th Cir. 1976).

would be admissible at trial while reserving the question of the weight the evidence would be given.²³⁰ The Seventh Circuit 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.”²³¹

The *Manual for Complex Litigation, Second*, recommended 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.”²³² The parties then were encouraged to attempt to resolve any methodological disagreements before the survey was conducted.²³³ Although this passage in the second edition of the Manual has been cited with apparent approval,²³⁴ the prior agreement that the Manual recommends has occurred rarely, and the *Manual for Complex Litigation, Fourth*, recommends, but does not advocate requiring, prior disclosure and discussion of survey plans.²³⁵ As the Manual suggests, however, early disclosure can enable the parties to raise prompt objections that may permit corrective measures to be taken before a survey is completed.²³⁶

Rule 26 of the Federal Rules of Civil Procedure requires extensive disclosure of the basis of opinions offered by testifying experts. However, Rule 26 does not produce disclosure of all 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,

230. Before trial, the presiding judge was appointed to the court of appeals, and so the case was tried by another district court judge

231. *Union Carbide*, 531 F.2d at 386. More recently, the Seventh Circuit recommended filing 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). On one recent occasion, the parties jointly developed a survey administered by a neutral third-party survey firm. *Scott v. City of New York*, 591 F. Supp. 2d 554, 560 (S.D.N.Y. 2008) (survey design, including multiple pretests, negotiated with the help of the magistrate judge).

232. MCL 2d, *supra* note 16, § 21.484.

233. *See id.*

234. *See, e.g., National Football League Props., Inc. v. New Jersey Giants, Inc.*, 637 F. Supp. 507, 514 n.3 (D.N.J. 1986).

235. MCL 4th, *supra* note 16, § 11.493 (“including the specific questions that will be asked, the introductory statements or instructions that will be given, and other controls to be used in the interrogation process.”).

236. *See id.*

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. In *McNeilab, Inc. v. American Home Products Corp.*,²³⁷ Judge William C. Conner encouraged the parties to submit their survey plans for court approval to ensure their evidentiary value; the plaintiff did so and altered its research plan based on Judge Conner's recommendations. 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.²³⁸

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 provide in detail:

1. The purpose of the survey;
2. A definition of the target population and a description of the sampling frame;
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 method, and other pertinent information;
4. A description of the results of sample implementation, including the number of
 - a. potential respondents contacted,
 - b. potential respondents not reached,
 - c. noneligibles,
 - d. refusals,
 - e. incomplete interviews or terminations, and
 - f. completed interviews;
5. The exact wording of the questions used, including a copy of each version of the actual questionnaire, interviewer instructions, and visual exhibits;²³⁹

237. 848 F.2d 34, 36 (2d Cir. 1988) (discussing with approval the actions of the district court). See also *Hubbard v. Midland Credit Mgmt*, 2009 U.S. Dist. LEXIS 13938 (S.D. Ind. Feb. 23, 2009) (court responded to plaintiff's motions to approve survey methodology with a critique of the proposed methodology).

238. Larry C. Jones, *Developing and Using Survey Evidence in Trademark Litigation*, 19 Memphis St. U. L. Rev. 471, 481 (1989).

239. The questionnaire itself can often reveal important sources of bias. See *Marria v. Broadus*, 200 F. Supp. 2d 280, 289 (S.D.N.Y. 2002) (court excluded survey sent to prison administrators based

6. A description of any special scoring (e.g., grouping of verbatim responses into broader categories);
7. A description of any weighting or estimating procedures used;
8. Estimates of the sampling error, where appropriate (i.e., in probability samples);
9. Statistical tables clearly labeled and identified regarding the source of the data, including the number of raw cases forming the base for each table, row, or column; and
10. Copies of interviewer instructions, validation results, and code books.²⁴⁰

Additional information to include in the survey report may depend on the nature of sampling design. For example, reported response rates along with the time each interview occurred may assist in evaluating the likelihood that nonresponse biased the results. In a survey designed to assess the duration of employee preshift activities, workers were approached as they entered the workplace; records were not kept on refusal rates or the timing of participation in the study. Thus, it was impossible to rule out the plausible hypothesis that individuals who arrived early for their shift with more time to spend on preshift activities were more likely to participate in the study.²⁴¹

Survey professionals generally do not describe pilot testing in their survey reports. They would be more likely to do so if courts recognized that surveys are improved by pilot work that maximizes the likelihood that respondents understand the questions they are being asked. Moreover, the Federal Rules of Civil Procedure 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 non-testifying 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

on questionnaire that began, "We need your help. We are helping to defend the NYS Department of Correctional Service in a case that involves their policy on intercepting Five-Percenter literature. Your answers to the following questions will be helpful in preparing a defense.")

240. These criteria were adapted from the Council of American Survey Research Organizations, *supra* note 76, § III.B. Failure to supply this information substantially impairs a court's ability to evaluate a survey. *In re Prudential Ins. Co. of Am. Sales Practices Litig.*, 962 F. Supp. 450, 532 (D.N.J. 1997) (citing the first edition of this manual). *But see Florida Bar v. Went for It, Inc.*, 515 U.S. 618, 626–28 (1995), in which a majority of the Supreme Court relied on a summary of results prepared by the Florida Bar from a consumer survey purporting to show consumer objections to attorney solicitation by mail. In a strong dissent, Justice Kennedy, joined by three other Justices, found the survey inadequate based on the document available to the court, pointing out that the summary included "no actual surveys, few indications of sample size or selection procedures, no explanations of methodology, and no discussion of excluded results . . . no description of the statistical universe or scientific framework that permits any productive use of the information the so-called Summary of Record contains." *Id.* at 640.

241. *See Chavez v. IBP, Inc.*, 2004 U.S. Dist. LEXIS 28838 (E.D. Wash. Aug. 18, 2004).

in the case. Some commentators suggest that attorneys are obligated to disclose such pilot work.²⁴²

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,²⁴³ the results of which should be included in the survey report.

Conflicts may arise when an opposing party asks for survey respondents' names and addresses so that they can re-interview some respondents. The party introducing the survey or the survey organization that conducted the research generally resists supplying such information.²⁴⁴ Professional surveyors as a rule promise confidentiality in an effort to increase participation rates and to encourage candid responses, although to the extent that identifying information is collected, such promises may not effectively prevent a lawful inquiry. 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 respondents' identities. "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 importance of Respondent anonymity."²⁴⁵ Although no surveyor–respondent privilege currently is recognized, the need for surveys and the availability of other means to examine and ensure their trustworthiness argue for deference to legitimate claims for confidentiality in order to avoid seriously compromising the ability of surveys to produce accurate information.²⁴⁶

242. See Yvonne C. Schroeder, *Pretesting Survey Questions*, 11 Am. J. Trial Advoc. 195, 197–201 (1987).

243. See *supra* Section V.C.

244. See, e.g., *Alpo Petfoods, Inc. v. Ralston Purina Co.*, 720 F. Supp. 194 (D.D.C. 1989), *aff'd in part and vacated in part*, 913 F.2d 958 (D.C. Cir. 1990).

245. Council of Am. Survey Res. Orgs., *supra* note 76, § I.A.3.f. Similar provisions are contained in the By-Laws of the American Association for Public Opinion Research.

246. *United States v. Dentsply Int'l, Inc.*, 2000 U.S. Dist. LEXIS 6994, at *23 (D. Del. May 10, 2000) (Fed. R. Civ. P. 26(a)(1) does not require party to produce the identities of individual survey respondents); *Litton Indus., 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, *The Researcher–Subject Relationship: The Need for Protection and a Model Statute*, 62 Geo. L.J. 243, 253 (1973)); see also *Applera Corp. v. MJ Research, Inc.*, 389 F. Supp. 2d 344, 350 (D. Conn. 2005) (denying access to names of survey respondents); *Lampshire*

Copies of all questionnaires should be made available upon request so that the opposing party has 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.

VIII. Acknowledgment

Thanks are due to Jon Krosnick for his research on surveys and his always sage advice.

v. Procter & Gamble Co., 94 F.R.D. 58, 60 (N.D. Ga. 1982) (defendant denied access to personal identifying information about women involved in studies by the Centers for Disease Control based on Fed. R. Civ. P. 26(c) giving court the authority to enter "any order which justice requires to protect a party or persons from annoyance, embarrassment, oppression, or undue burden or expense.") (citation omitted).

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., 1st ed. 1983; Peter V. Marsden & James D. Wright eds., 2d ed. 2010); Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991); Willem E. Saris, Computer-Assisted Interviewing (1991); Seymour Sudman, Applied Sampling (1976).

branching. A questionnaire structure that uses the answers to earlier questions to determine which set of additional questions should be asked (e.g., citizens who report having served as jurors on a criminal case are asked different questions about their experiences than citizens who report having served as jurors on a civil case).

CAI (computer-assisted interviewing). A method of conducting interviews in which an interviewer asks questions and records the respondent's answers by following a computer-generated protocol.

CAPI (computer-assisted personal interviewing). A method of conducting face-to-face interviews in which an interviewer asks questions and records the respondent's answers by following a computer-generated protocol.

CATI (computer-assisted telephone interviewing). A method of conducting telephone interviews in which an interviewer asks questions and records the respondent's answers by following a computer-generated protocol.

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. A sampling technique allowing for the selection of sample elements in groups or clusters, rather than on an individual basis; it may significantly reduce field costs and may increase sampling error if elements in the same cluster are more similar to one another than are elements in different clusters.

confidence interval. An indication of the probable range of error associated with a sample value obtained from a probability sample.

context effect. A previous question influences the way the respondent perceives and answers a later question.

convenience sample. A sample of elements selected because they were readily available.

coverage error. Any inconsistencies between the sampling frame and the target population.

double-blind research. Research in which the respondent and the interviewer are not given information that will alert them to the anticipated or preferred pattern 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.** A sampling design in which sampling takes place in several stages, beginning with larger units (e.g., cities) and then proceeding with smaller units (e.g., households or individuals within these units).
- noncoverage error.** The omission of eligible population units from the sampling frame.
- 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 of response alternatives on the questionnaire (see primacy effect and recency effect).
- parameter.** A summary measure of a characteristic of a population (e.g., average age, proportion of households in an area owning a computer). Statistics are estimates of parameters.
- pilot test.** A small field test replicating the field procedures planned for the full-scale survey; although the terms *pilot test* and *pretest* are sometimes used interchangeably, a pretest tests the questionnaire, whereas a pilot test generally tests proposed collection procedures as well.
- population.** The totality of elements (individuals or other units) that have some common property of interest; the target population is the collection of elements that the researcher would like to study. Also, universe.
- population value, population parameter.** The actual value of some characteristic in the population (e.g., the average age); the population value is estimated by taking a random sample from the population and computing the corresponding sample value.
- pretest.** A small preliminary test of a survey questionnaire. See pilot test.
- primacy effect.** A tendency of respondents to choose early items from a list of choices; the opposite of a recency effect.
- probability sample.** A type of 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 followup question that an interviewer asks to obtain a more complete answer from a respondent (e.g., “Anything else?” “What kind of medical problem do you mean?”).

quasi-filter question. A question that offers a “don’t know” or “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 probability 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 estimated size of 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 source or sources from which the individuals or other units in a sample are drawn.

secondary meaning. A descriptive term that becomes protectable as a trademark if it signifies to the purchasing public that the product comes from a single producer or source.

simple random sample. The most basic type of probability sample; 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 questions in which some should not be asked (should be skipped) based on the respondent’s answer to a previous question (e.g., if the respondent indicates that he does not own a car, he should not be asked what brand of car he owns).

stratified sampling. A sampling technique in which the researcher subdivides 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-experiment. A survey with one or more control groups, enabling the researcher to test a causal proposition.

survey population. See population.

systematic sampling. A sampling technique that consists of a random starting point and the selection of every *n*th member of the population; it is generally analyzed as if it were a simple random sample and generally produces the same results..

target population. See population.

trade dress. A distinctive and nonfunctional design of a package or product protected under state unfair competition law and the federal Lanham Act § 43(a), 15 U.S.C. § 1125(a) (1946) (amended 1992).

true score. The underlying true value, 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

- Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, & Seymour Sudman (eds.), *Measurement Errors in Surveys* (2004).
- Jean M. Converse & Stanley Presser, *Survey Questions: Handcrafting the Standardized Questionnaire* (1986).
- Mick P. Couper, *Designing Effective Web Surveys* (2008).
- Don A. Dillman, Jolene Smyth, & Leah M. Christian, *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method* (3d ed. 2009).
- Robert M. Groves, Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, & Roger Tourangeau, *Survey Methodology* (2004).
- Sharon Lohr, *Sampling: Design and Analysis* (2d ed. 2010).
- 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).
- Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, & Roger Tourangeau, *Cognition and Survey Research* (1999).
- Seymour Sudman, *Applied Sampling* (1976).
- Survey Nonresponse* (Robert M. Groves, Don A. Dillman, John L. Eltinge, & Roderick J. A. Little eds., 2002).
- Telephone Survey Methodology* (Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, & William L. Nicholls eds., 1988).
- Roger Tourangeau, Lance J. Rips, & Kenneth Rasinski, *The Psychology of Survey Response* (2000).