Reference Guide on Survey Research

SHARI SEIDMAN DIAMOND

Shari Seidman Diamond, J.D., Ph.D., is Professor of Law and Psychology, Northwestern University, Evanston, Illinois, and Senior Research Fellow, American Bar Foundation, Chicago, Illinois.

CONTENTS

I. Introduction, 231
- A. Use of Surveys in Court, 233
- B. A Comparison of Survey Evidence and Individual Testimony, 235

II. Purpose and Design of the Survey, 236
- A. Was the Survey Designed to Address Relevant Questions? 236
- B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey? 237
- C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced? 238
- D. Are the Experts Who Will Testify About Surveys Conducted by Others Appropriately Skilled and Experienced? 239

III. Population Definition and Sampling, 239
- A. Was an Appropriate Universe or Population Identified? 239
- B. Did the Sampling Frame Approximate the Population? 240
- C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population? 242
- D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey? 245
- E. What Procedures Were Used to Reduce the Likelihood of a Biased Sample? 246
- F. What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey? 247

IV. Survey Questions and Structure, 248
- A. Were Questions on the Survey Framed to Be Clear, Precise, and Unbiased? 248
- B. Were Filter Questions Provided to Reduce Guessing? 249
- C. Did the Survey Use Open-Ended or Closed-Ended Questions? How Was the Choice in Each Instance Justified? 251
- 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? 253

FINAL

SDARS Ex. 2

E. What Approach Was Used to Avoid or Measure Potential Order or Context Effects? 254
F. If the Survey Was Designed to Test a Causal Proposition, Did the Survey Include an Appropriate Control Group or Question? 256
G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey? 260
1. In-person interviews, 260
2. Telephone surveys, 261
3. Mail surveys, 263
4. Internet surveys, 264

V. Surveys Involving Interviewers, 264
A. Were the Interviewers Appropriately Selected and Trained? 264
B. What Did the Interviewers Know About the Survey and Its Sponsorship? 266
C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias? 267

VI. Data Entry and Grouping of Responses, 268
A. What Was Done to Ensure That the Data Were Recorded Accurately? 268
B. What Was Done to Ensure That the Grouped Data Were Classified Consistently and Accurately? 268

VII. Disclosure and Reporting, 269
A. When Was Information About the Survey Methodology and Results Disclosed? 269
B. Does the Survey Report Include Complete and Detailed Information on All Relevant Characteristics? 270
C. In Surveys of Individuals, What Measures Were Taken to Protect the Identities of Individual Respondents? 271

Glossary of Terms, 273

References on Survey Research, 276

I. Introduction

Surveys are used to describe or enumerate objects or 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 objects, individuals, or social units. Although surveys may count or measure every member of the relevant population (e.g., all plaintiffs eligible to join in a suit, all employees currently working for a corporation, all trees in a forest), sample surveys count or measure only a portion of the objects, individuals, or social units that the survey is intended to describe.[2]

Some statistical and sampling experts apply the phrase "sample survey" only to a survey in which probability sampling techniques are used to select the sample.[3] Although probability sampling offers important advantages over nonprobability sampling,[4] 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.[5] Thus, in this reference guide, both the probability sample and the nonprobability sample are discussed. The strengths of probability sampling and the weaknesses of various types of nonprobability sampling are described so that the trier of fact can consider these features in deciding what weight to give to a particular sample survey.

As a method of data collection, surveys have several crucial potential advantages over less systematic approaches.[6] When properly designed, executed, and

1. Social scientists describe surveys as "conducted for the purpose of collecting data from individuals about themselves, about their households, or about other larger social units." Peter H. Rossi et al., *Sample Surveys: History, Current Practice, and Future Prospects, in* Handbook of Survey Research 1, 2 (Peter H. Rossi et al. eds., 1983). Used in its broader sense, however, the term *survey* applies to any description or enumeration, whether or not a person is the source of this information. Thus, a report on the number of trees destroyed in a forest fire might require a survey of the trees and stumps in the damaged area.

2. 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.

3. *E.g.*, Leslie Kish, Survey Sampling 26 (1965).

4. *See infra* § III.C.

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

6. This does not mean that surveys can be relied on to address all types of questions. For example, some respondents may not be able to predict accurately whether they would volunteer for military service if Washington, D.C., were to be bombed. Their inaccuracy may arise not because they are unwilling to answer the question or to say they don't know, but because they believe they can predict accurately, and they are simply wrong. Thus, the availability of a "don't know" option cannot cure the inaccuracy. Although such a survey is suitable for assessing their predictions, it may not provide a very accurate estimate of what their actual responses would be.

described, surveys (1) economically present the characteristics of a large group of objects or respondents and (2) permit an assessment of the extent to which the measured objects or respondents are likely to adequately represent a relevant group of objects, individuals, or social units.[7] All questions asked of respondents and all other measuring devices used can be examined by the court and the opposing party for objectivity, clarity, and relevance, and all answers or other measures obtained can be analyzed for completeness and consistency. 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 describe in detail the design and execution of the survey.

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.[8] These questions can be (1) raised from the bench during a pretrial proceeding to determine the admissibility of the survey evidence; (2) presented to the contending experts before trial for their joint identification of disputed and undisputed issues; (3) presented to counsel with the expectation that the issues will be addressed during the examination of the experts at trial; or (4) raised in bench trials when a motion for a preliminary injunction is made to help the judge evaluate what weight, if any, the survey should be given.[9] These questions are intended to improve the utility of cross-examination by counsel, where appropriate, not to replace it.

All sample surveys, whether they measure objects, individuals, or other social 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 raise methodological issues involving survey questions and structure (section IV) and confidentiality (section VII.C), and interview surveys introduce additional issues (e.g., interviewer training and qualifications) (section V). The sections of 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.

7. The ability to quantitatively assess the limits of the likely margin of error is unique to probability sample surveys.

8. *See infra* text accompanying note 27.

9. 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).

A. Use of Surveys in Court

Forty years ago the question whether surveys constituted acceptable evidence still was unsettled.[10] Early doubts about the admissibility of surveys centered on their use of sampling techniques[11] and their status as hearsay evidence.[12] Federal Rule of Evidence 703 settled both matters for surveys by redirecting attention to the "validity of the techniques employed."[13] 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."[14] 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

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

11. 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 Sprowls, *The Admissibility of Sample Data into a Court of Law: A Case History*, 4 UCLA L. Rev. 222, 226–29 (1956–1957).

12. 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 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.

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 due to lack of 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 which the proponent can procure through reasonable efforts, and admissibility serves the interests of justice. Fed. R. Evid. 807; *e.g.*, Keith v. Volpe, 618 F. Supp. 1132 (C.D. Cal. 1985); *Schering*, 189 F.3d at 232. Admissibility as an exception to the hearsay exclusion thus depends on the trustworthiness of the survey.

13. Fed. R. Evid. 703 advisory committee's note.

14. Fed. R. Evid. 703.

statistically correct way?"[15] 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.*[16]

Because the survey method provides an economical and systematic way to gather information about a large number of individuals or social units, surveys are used widely in business, government, and, increasingly, administrative settings and judicial proceedings. Both federal and state courts have accepted survey evidence on a variety of issues. In a case involving allegations of discrimination in jury panel composition, the defense team surveyed prospective jurors to obtain age, race, education, ethnicity, and income distribution.[17] Surveys of employees or prospective employees are used to support or refute claims of employment discrimination.[18] 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.[19] Some courts have admitted surveys in obscenity cases to provide evidence about community standards.[20] 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.[21] The plaintiff in an antitrust suit conducted a survey to assess what characteristics, including price, affected consumers' preferences. The sur-

15. 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] and the Manual for Complex Litigation, Third § 21.493 (1995) [hereinafter MCL 3d]. Note, however, that experts who collect survey data, along with the professions that rely on those surveys, may differ in some of their methodological standards and principles. 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.

16. 509 U.S. 579 (1993). *See also* General Elec. Co. v. Joiner, 522 U.S. 136, 147 (1997).

17. People v. Harris, 679 P.2d 433 (Cal.), *cert. denied*, 469 U.S. 965 (1984).

18. EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1308 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988); Stender v. Lucky Stores, Inc., 803 F. Supp. 259, 326 (N.D. Cal. 1992); Richardson v. Quik Trip Corp., 591 F. Supp. 1151, 1153 (S.D. Iowa 1984).

19. United States v. Scheffer, 523 U.S. 303, 309 (1998); Meyers v. Arcudi, 947 F. Supp. 581, 588 (D. Conn. 1996); United States v. Varoudakis, No. 97-10158, 1998 WL 151238 (D. Mass. Mar. 27, 1998); United States v. Bishop, 64 F. Supp. 2d 1149 (D. Utah 1999); United States v. Orians, 9 F. Supp. 2d 1168, 1174 (D. Ariz. 1998) (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).

20. *E.g.*, People v. Page Books, Inc., 601 N.E.2d 273, 279–80 (Ill. App. Ct. 1992); People v. Nelson, 410 N.E.2d 476, 477–79 (Ill. App. Ct. 1980); State v. Williams, 598 N.E.2d 1250, 1256–58 (Ohio Ct. App. 1991).

21. *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); Powell v. Superior Court, 283 Cal. Rptr. 777, 783 (Ct. App. 1991).

vey was offered as one way to estimate damages.[22] A routine use of surveys in federal courts occurs in Lanham Act[23] cases, where the plaintiff alleges trademark infringement[24] or claims that false advertising[25] 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 an inaccurate message?).[26] 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.*[27] jointly drafted a discovery survey that was administered in person by neutral third parties, thus replacing interrogatories and depositions. It resulted in substantial savings in both time and cost.

B. A Comparison of Survey Evidence and Individual Testimony

To illustrate the value of a survey, it is useful to compare the information that can be obtained from a competently done survey with the information obtained

22. Dolphin Tours, Inc. v. Pacifico Creative Servs., Inc., 773 F.2d 1506, 1508 (9th Cir. 1985). *See also* SMS Sys. Maintenance 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., 1986). Surveys also are used in litigation to help define relevant markets. In *United States v. E.I. DuPont de Nemours & Co.*, 118 F. Supp. 41, 60 (D. Del. 1953), *aff'd*, 351 U.S. 377 (1956), a survey was used to develop the "market setting" for the sale of cellophane. 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.

23. Lanham Act § 43(a), 15 U.S.C. § 1125(a) (1946) (amended 1992).

24. *E.g.*, Union Carbide Corp. v. Ever-Ready, Inc., 531 F.2d 366 (7th Cir.), *cert. denied*, 429 U.S. 830 (1976); 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 on other grounds*, 13 F.3d 1297 (9th. Cir. 1994), *rev'd on other grounds*, 514 U.S. 159 (1995). 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.

25. *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).

26. Courts have observed that "the court's reaction is at best not determinative and at worst irrelevant. The question in such cases is, what does the person to whom the advertisement is addressed find to be the message?" American Brands, Inc. v. R.J. Reynolds Tobacco Co., 413 F. Supp. 1352, 1357 (S.D.N.Y. 1976). The wide use of surveys in recent years was foreshadowed in *Triangle Publications, Inc. v. Rohrlich*, 167 F.2d 969, 974 (2d Cir. 1948) (Frank, J., dissenting). Called on to determine whether a manufacturer of girdles labeled "Miss Seventeen" infringed the trademark of the magazine *Seventeen*, Judge Frank suggested 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.

27. 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.

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,[28] the court can and should scrutinize the method of respondent selection for any survey that is presented.

A party using a nonsurvey method generally identifies several witnesses who testify about their own characteristics, experiences, or impressions. While the party has no obligation to select these witnesses in any particular way or to report on how they were chosen, the party is not likely to select witnesses whose attributes conflict with the party's interests. The witnesses who testify are aware of the parties involved in the case and have discussed the case before testifying.

Although surveys are not the only means 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.[29] 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.[30]

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 con-

28. Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 205 (S.D.N.Y. 1981) (distinguishing between surveys conducted in anticipation of litigation and surveys conducted for nonlitigation purposes which cannot be reproduced because of the passage of time, concluding that parties should not be compelled to introduce the former at trial, but may be required to provide the latter).

29. *E.g.*, Charles Jacquin et Cie, Inc. v. Destileria Serralles, Inc., 921 F.2d 467, 475 (3d Cir. 1990). *See also* Brunswick Corp. v. Spinit Reel Co., 832 F 2d 513, 522 (10th Cir. 1987).

30. E.S. Originals, Inc. v. Stride Rite Corp., 656 F. Supp. 484, 490 (S.D.N.Y. 1987); *see also* 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); Information Clearing House, Inc. v. Find Magazine, 492 F. Supp. 147, 160 (S.D.N.Y. 1980).

sumer confusion in a trademark case). Surveys not conducted specifically in preparation for, or in response to, litigation may provide important information,[31] but they frequently ask irrelevant questions[32] or select inappropriate samples of respondents for study.[33] Nonetheless, surveys do not always achieve their stated goals. Thus, the content and execution of a survey must be scrutinized even if the survey was designed to provide relevant data on the issue before the court. 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 prove its side of the issue in controversy.

B. Was Participation in the Design, Administration, and Interpretation of the Survey Appropriately Controlled to Ensure the Objectivity of the Survey?

An early handbook for judges recommended that survey interviews be "conducted independently of the attorneys in the case."[34] Some courts have interpreted this to mean that any evidence of attorney participation is objectionable.[35] A better interpretation is that the attorney should have no part in carrying out the survey.[36] However, some attorney involvement in the survey design is

31. *See, e.g.,* Wright v. Jeep Corp., 547 F. Supp. 871, 874 (E.D. Mich. 1982). Indeed, as courts increasingly have been faced with scientific issues, parties have requested in a number of recent cases that the courts compel production of research data and testimony by unretained experts. The circumstances under which an unretained expert can be compelled to testify or to disclose research data and opinions, as well as the extent of disclosure that can be required when the research conducted by the expert has a bearing on the issues in the case, are the subject of considerable current debate. *See, e.g.,* Richard L. Marcus, *Discovery Along the Litigation/Science Interface,* 57 Brook. L. Rev. 381, 393–428 (1991); Joe S. Cecil, *Judicially Compelled Disclosure of Research Data,* 1 Cts. Health Sci. & L. 434 (1991); *see also* Symposium, *Court-Ordered Disclosure of Academic Research: A Clash of Values of Science and Law,* Law & Contemp. Probs., Summer 1996, at 1.

32. Loctite Corp. v. National Starch & Chem. Corp., 516 F. Supp. 190, 206 (S.D.N.Y. 1981) (marketing surveys conducted before litigation were designed to test for brand awareness, whereas the "single issue at hand . . . [was] whether consumers understood the term 'Super Glue' to designate glue from a single source").

33. 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).

34. Judicial Conference of the U.S., Handbook of Recommended Procedures for the Trial of Protracted Cases 75 (1960).

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

36. 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

necessary to ensure that relevant questions are directed to a relevant population.[37] The trier of fact evaluates the objectivity and relevance of the questions on the survey and the appropriateness of the definition of the population used to guide sample selection. These aspects of the survey are visible to the trier of fact and can be judged on their quality, irrespective of who suggested them. In contrast, the interviews themselves are not directly visible, and any potential bias is minimized by having interviewers and respondents blind to the purpose and sponsorship of the survey and by excluding attorneys from any part in conducting interviews and tabulating results.

C. Are the Experts Who Designed, Conducted, or Analyzed the Survey Appropriately Skilled and Experienced?

Experts prepared to design, conduct, and analyze a survey generally should have graduate training in psychology (especially social, cognitive, or consumer psychology), sociology, marketing, communication sciences, statistics, or a related discipline; that training should include courses in survey research methods, sampling, measurement, interviewing, and statistics. In some cases, professional experience in conducting and publishing survey research may provide the requisite background. In all cases, the expert must demonstrate an understanding of survey methodology, including sampling,[38] instrument design (questionnaire and interview construction), and statistical analysis.[39] Publication in peer-reviewed journals, authored books, membership 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, if the survey involves 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 involves a special population (e.g., developmentally disabled adults with limited cognitive skills), the survey expert also should be able to demonstrate sufficient familiarity with the topic or population (or assistance from an individual on the research team with suitable expertise) to design a survey instrument that will communicate clearly with relevant respondents.

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").

37. 3 J. Thomas McCarthy, McCarthy on Trademarks and Unfair Competition § 32:166 (4th ed. 1996).

38. The one exception is that sampling expertise is unnecessary if the survey is administered to all members of the relevant population. *See, e.g.*, McGovern & Lind, *supra* note 27.

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

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.[40] The secondary expert who gives an opinion about the adequacy and interpretation of a survey not only should have general skills and experience with surveys and be familiar with all of the issues addressed in this reference guide, but also should demonstrate familiarity with the following properties of the survey being discussed:

1. the purpose of the survey;
2. the survey methodology, including
   a. the target population,
   b. the sampling design used in conducting the survey,
   c. the survey instrument (questionnaire or interview schedule), and
   d. (for interview surveys) interviewer training and instruction;
3. the results, including rates and patterns of missing data; and
4. the statistical analyses used to interpret the results.

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).[41] The target population consists of all elements (i.e., objects, individuals, or other social 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 actual purchasers of the plaintiff's goods or services and all prospective and actual purchasers of the defendant's goods or services. Similarly, the population for a discovery survey may include all potential plaintiffs or all em-

40. Margaret A. Berger, The Supreme Court's Triology on the Admissibility of Expert Testimony § IV.C, in this manual.

41. Identification of the proper universe is recognized uniformly as a key element in the development of a survey. *See, e.g.*, Judicial Conference of the U.S., *supra* note 34; MCL 3d, *supra* note 15, § 21.493. *See also* 3 McCarthy, *supra* note 37, § 32:166; Council of Am. Survey Res. Orgs., Code of Standards and Ethics for Survey Research § III.B.4 (1997).

ployees 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.[42] The definition of the relevant population is crucial because there may be systematic differences in the responses of members of the population and nonmembers. (For example, consumers who are prospective purchasers may know more about the product category than consumers who are not considering making a purchase.)

The universe must be defined carefully. For example, a commercial for a toy or breakfast cereal may be aimed at children, who in turn influence their parents' purchases. If a survey assessing the commercial's tendency to mislead were conducted based on the universe of prospective and actual adult purchasers, it would exclude a crucial group of eligible respondents. Thus, the appropriate population in this instance would include children as well as parents.[43]

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, compromises are sometimes required in developing the sampling frame. The survey report should contain a description of the target population, a description of the survey population actually sampled, a discussion of the difference between the two populations, and an evaluation of the likely consequences of that difference.

42. 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 trial, that the defendants were guilty), *cert. denied*, 431 U.S. 933 (1977); *see also* People v. Venegas, 31 Cal. Rptr. 2d 114, 117 (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).

43. 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.

A survey that provides information about a wholly irrelevant universe of respondents is itself irrelevant.[44] Courts are likely to exclude the survey or accord it little weight. 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.[45] More commonly, however, the sampling frame is either underinclusive or overinclusive relative to the target population. If it is underinclusive, the survey's value depends on the extent to which the excluded population is likely to react differently from the included population. Thus, a survey of spectators and participants at running events would be sampling a sophisticated subset of those likely to purchase running shoes. Because this subset probably would consist of the consumers most knowledgeable about the trade dress used by companies that sell running shoes, a survey based on this population would be likely to substantially overrepresent the strength of a particular design as a trademark, and the extent of that overrepresentation would be unknown and not susceptible to any reasonable estimation.[46]

Similarly, in a survey designed to project demand for cellular phones, the assumption that businesses would be the primary users of cellular service led surveyors to exclude potential nonbusiness users from the survey. The Federal Communications Commission (FCC) found the assumption unwarranted and concluded that the research was flawed, in part because of this underinclusive universe.[47]

44. 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. Home Box Office v. Showtime/The Movie Channel, 665 F. Supp. 1079, 1083 (S.D.N.Y.), *aff'd in part & vacated in part*, 832 F.2d 1311 (2d Cir. 1987). *But see* Lon Tai Shing Co. v. Koch + Lowy, No. 90-C4464, 1990 U.S. Dist. LEXIS 19123, at *50 (S.D.N.Y. Dec. 14, 1990), in which the judge was willing to find likelihood of consumer confusion from a survey of lighting store salespersons questioned by a survey researcher posing as a customer. The court was persuaded that the salespersons who were misstating the source of the lamp, whether consciously or not, must have believed reasonably that the consuming public would be misled by the salespersons' inaccurate statements about the name of the company that manufactured the lamp they were selling.

45. R.L. Winston Rod Co. v. Sage Mfg. Co., 838 F. Supp. 1396, 1401-02 (D. Mont. 1993).

46. 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* 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).

47. 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).

In some cases, it is difficult to determine whether an underinclusive universe distorts the results of the survey and, if so, the extent and likely direction of the bias. For example, a trademark survey was designed to test the likelihood of confusing an analgesic currently on the market with a new product that was similar in appearance.[48] The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the universe should have included users of other analgesics, "so that the full range of potential customers for whom plaintiff and defendants would compete could be studied."[49] In this instance, it is unclear whether users of the plaintiff's product would be more or less likely to be confused than users of the defendant's product or users of a third analgesic.[50]

An overinclusive universe generally presents less of a problem in interpretation than does an underinclusive universe. If the survey expert can demonstrate that a sufficiently large (and representative) subset of respondents in the survey was drawn from the appropriate universe, the responses obtained from that subset can be examined, and inferences about the relevant universe can be drawn based on that subset.[51] If the relevant subset cannot be identified, however, an overbroad universe will reduce the value of the survey.[52] If the sample is drawn from an underinclusive universe, there is generally no way to know how the unrepresented members would have responded.[53]

C. How Was the Sample Selected to Approximate the Relevant Characteristics of the Population?

Identification of a survey population must be followed by selection of a sample that accurately represents that population.[54] 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 simple random sampling, the most basic type

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

49. *Id.* at 1070.

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

51. This occurred in *National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651, 657–58 (W.D. Wash. 1982).

52. Schieffelin & Co. v. Jack Co. of Boca, 850 F. Supp. 232, 246 (S.D.N.Y. 1994).

53. *See, e.g.*, Amstar Corp. v. Domino's Pizza, Inc., 615 F.2d 252, 263–64 (5th Cir.) (court found both plaintiff's and defendant's surveys substantially defective for a systematic failure to include parts of the relevant population), *cert. denied*, 449 U.S. 899 (1980).

54. MCL 3d, *supra* note 15, § 21.493. *See also* David H. Kaye & David A. Freedman, Reference Guide on Statistics § II.B, in this manual.

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.[55] In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample.[56]

Probability sampling offers two important advantages over other types of sampling. First, the sample can provide an unbiased estimate of the responses of all persons in the population from which the sample was drawn; that is, the 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. Thus, suppose a survey tested a sample of 400 dentists randomly selected from the population of all dentists licensed to practice in the United States and found that 80, or 20%, of them mistakenly believed that a new toothpaste, Goldgate, was manufactured by the makers of Colgate. A survey expert could properly compute a confidence interval around the 20% estimate obtained from this sample. If the survey was 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.[57] In this example, the confidence interval, or margin of error, is the estimate (20%) plus or minus 4%, or the distance between 16% and 24%.

All sample surveys produce estimates of population values, not exact measures of those values. Strictly speaking, the margin of sampling error associated with the sample estimate assumes probability sampling. Assuming a probability sample, a confidence interval describes how stable the mean response in the sample is likely to be. The width of the confidence interval depends on three characteristics:

55. Systematic sampling, in which every *n*th unit in the population is sampled and the starting point is selected randomly, fulfills the first of these conditions. It does not fulfill the second, because no systematic sample can include elements adjacent to one another on the list of population members from which the sample is drawn. Except in very unusual situations when periodicities occur, systematic samples and simple random samples generally produce the same results. Seymour Sudman, *Applied Sampling, in* Handbook of Survey Research, *supra* note 1, at 145, 169.

56. 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. Martin Frankel, *Sampling Theory, in* Handbook of Survey Research, *supra* note 1, at 21, 37, 47.

57. Actually, since survey interviewers would be unable to locate some dentists and some dentists would be unwilling to participate in the survey, technically the population to which this sample would be projectable would be all dentists with current addresses who would be willing to participate in the survey if they were asked.

1. the size of the sample (the larger the sample, the narrower the interval);
2. the variability of the response being measured; and
3. the confidence level the researcher wants to have.

Traditionally, scientists adopt the 95% level of confidence, which means that if 100 samples of the same size were drawn, the confidence interval expected for at least 95 of the samples would be expected to include the true population value.[58]

Although probability sample surveys often are conducted in organizational settings and are the recommended sampling approach in academic and government publications on surveys, probability sample surveys can be expensive when in-person interviews are required, the target population is dispersed widely, or qualified respondents are scarce. A majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability convenience samples.[59] 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."[60] 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.[61] In addition, quantitative values computed from such samples (e.g., percentage of respondents indicating confusion) should be viewed as rough indicators rather than as precise quantitative estimates. Confidence intervals should not be computed.

58. To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%), the width of the confidence interval can be expanded. An increase in the confidence interval brings an increase in the confidence level. For further discussion of confidence intervals, see David H. Kaye & David A. Freedman, Reference Guide on Statistics § IV.A, in this manual.

59. 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 *National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651 (W.D. Wash. 1982); James Burrough, Ltd. v. Sign of Beefeater, Inc., 540 F.2d 266 (7th Cir. 1976).

60. National Football League Properties, 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 59, at 172–73, 176.

D. Was the Level of Nonresponse Sufficient to Raise Questions About the Representativeness of the Sample? If So, What Is the Evidence That Nonresponse Did Not Bias the Results of the Survey?

Even when a sample is drawn randomly from a complete list of elements in the target population, responses or measures may be obtained on only part of the selected sample. If this lack of response were distributed randomly, valid inferences about the population could be drawn from the characteristics of the available elements in the sample. The difficulty is that nonresponse often is not random, so that, for example, persons who are single typically have three times the "not at home" rate in U.S. Census Bureau surveys as do family members.[62] Efforts to increase response rates include making several attempts to contact potential respondents and providing financial incentives for participating in the survey.

One suggested formula for quantifying a tolerable level of nonresponse in a probability sample is based on the guidelines for statistical surveys issued by the former U.S. Office of Statistical Standards.[63] According to these guidelines, response rates of 90% or more are reliable and generally can be treated as random samples of the overall population. Response rates between 75% and 90% usually yield reliable results, but the researcher should conduct some check on the representativeness of the sample. Potential bias should receive greater scrutiny when the response rate drops below 75%. If the response rate drops below 50%, the survey should be regarded with significant caution as a basis for precise quantitative statements about the population from which the sample was drawn.[64]

Determining whether the level of nonresponse in a survey is critical generally requires an analysis of the determinants of nonresponse. For example, even a survey with a high response rate may seriously underrepresent some portions of the population, such as the unemployed or the poor. If a general population sample was used to chart changes in the proportion of the population that knows someone with HIV, the survey would underestimate the population value if some groups more likely to know someone with HIV (e.g., intravenous drug users) were underrepresented in the sample. The survey expert should be prepared to provide evidence on the potential impact of nonresponse on the survey results.

61. *See infra* § III.E.

62. 2 Gastwirth, *supra* note 33, at 501. This volume contains a useful discussion of sampling, along with a set of examples. *Id.* at 467.

63. This standard is cited with approval by Gastwirth. *Id.* at 502.

64. For thoughtful examples of judges closely scrutinizing potential sample bias when response rates were below 75%, see *Vuyanich v. Republic National Bank*, 505 F. Supp. 224 (N.D. Tex. 1980); Rosado v. Wyman, 322 F. Supp. 1173 (E.D.N.Y.), *aff'd*, 437 F.2d 619 (2d Cir. 1970), *aff'd*, 402 U.S. 991 (1971).

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. 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.[65]

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 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 malls, the expert can test for and report on any differences observed across sites. To the extent that similar results are obtained in different locations using different on-site interview operations, it is less likely that idiosyncrasies of sample selection or administration can account for the results.[66] Similarly, since the characteristics of persons visiting a shopping center vary by day of the week and time of day, bias in sampling can be reduced if the survey design calls for sampling time segments as well as mall locations.[67]

In mall intercept surveys, the organization that manages the on-site interview facility generally employs recruiters who approach potential survey respondents in the mall and ascertain if they are qualified and willing to participate in the survey. If a potential respondent agrees to answer the questions and meets the specified criteria, he or she is escorted to the facility where the survey interview 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-

65. Andy B. Anderson et al., *Missing Data: A Review of the Literature, in* Handbook of Survey Research, *supra* note 1, at 415.

66. Note, however, that differences in results across sites may be due to genuine differences in respondents across geographic locations or to a failure to administer the survey consistently across sites.

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

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.[68]

F. *What Precautions Were Taken to Ensure That Only Qualified Respondents Were Included in the Survey?*

In a carefully executed survey, each potential respondent is questioned or measured on the attributes that determine his or her eligibility to participate in the survey. Thus, the initial questions screen potential respondents to determine if they are within the target population of the survey (e.g., Is she at least fourteen years old? Does she own a dog? Does she live within ten miles?). The screening questions must be drafted so that they do not convey information that will influence the respondent's answers on the main survey. For example, if respondents must be prospective and recent purchasers of Sunshine orange juice in a trademark survey designed to assess consumer confusion with Sun Time orange juice, potential respondents might be asked to name the brands of orange juice they have purchased recently or expect to purchase in the next six months. They should not be asked specifically if they recently have purchased, or expect to purchase, Sunshine orange juice, because this may affect their responses on the survey either by implying who is conducting the survey or by supplying them with a brand name that otherwise would not occur to them.

The 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.*,[69] physicians were asked a screening question to determine whether they prescribed particular drugs. The court found that the screener conditioned the physicians to respond with the name of a drug rather than a condition.[70]

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 avail-

68. 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.

69. 858 F. Supp. 1305, 1321 & n.13 (S.D.N.Y. 1994).

70. *Id.* at 1321.

able 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.[71]

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.[72] 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.[73] 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.[74] The court noted that the phrase "compensatory damages," although well-defined for attorneys, was unlikely to be meaningful for laypersons.[75]

Texts on survey research generally recommend pretests as a way to increase the likelihood that questions are clear and unambiguous,[76] and some courts have

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

72. *Id.* at 219.

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

74. *Id.* at 102, 103.

75. *Id.* at 103. When there is any question about whether some respondent will understand a particular term or phrase, the term or phrase should be defined explicitly.

76. For a thorough treatment of pretesting methods, see Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire 51 (1986). *See also* Fred W. Morgan, *Judicial Standards for Survey Research: An Update and Guidelines*, 54 J. Marketing 59, 64 (1990).

recognized the value of pretests.[77] In a pretest or pilot test,[78] the proposed survey is administered to a small sample (usually between twenty-five and seventy-five)[79] of the 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. 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 can improve the quality of a survey and to anticipate that it often results in word changes that increase clarity and correct misunderstandings. Thus, changes may indicate informed survey construction rather than flawed survey design.[80]

B. Were Filter Questions Provided to Reduce Guessing?

Some survey respondents may have no opinion on an issue under investigation, either because they have never thought about it before or because the question mistakenly assumes a familiarity with the issue. For example, survey respondents may not have noticed that the commercial they are being questioned about guaranteed the quality of the product being advertised and thus 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 one-year guarantee, a sixty-day guarantee, or a thirty-day guarantee?"). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an answer even if (in this example) the respondent did not notice the guarantee (or is unfamiliar with the parental leave policy). Such answers will

77. *E.g.*, Zippo Mfg. Co. v. Rogers Imports, Inc., 216 F. Supp. 670 (S.D.N.Y. 1963).

78. 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.

79. Converse & Presser, *supra* note 76, at 69. Converse and Presser suggest that a pretest with twenty-five respondents is appropriate when the survey uses professional interviewers.

80. *See infra* § VII.B for a discussion of obligations to disclose pilot work.

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?").[81] 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%–25% increase in the proportion of respondents selecting that response.[82]

Finally, the survey can include full-filter questions, that is, questions that lay the groundwork for the substantive question by first asking the respondent if he or she has an opinion about the issue or happened to notice the feature that the interviewer is preparing to ask about (e.g., "Based on the commercial you just saw, do you have an opinion about how long Clover stated or implied that its guarantee lasts?"). The interviewer then asks the substantive question only 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.[83] Respondents are more likely to say they do not have an opinion on an issue if a full filter is used than if a quasi-filter is used.[84] 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 follow-up questions by saying that they have no opinion.[85]

81. 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).

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

83. Considerable research has been conducted on the effects of filters. For a review, see George F. Bishop et al., *Effects of Filter Questions in Public Opinion Surveys*, 47 Pub. Opinion Q. 528 (1983).

84. Schwarz & Hippler, *supra* note 81, at 45–46.

85. *Id.* at 46.

In general, then, a survey that uses full filters tends to provide a conservative estimate of the number of respondents holding an opinion, whereas a survey that uses neither full filters nor quasi-filters tends to overestimate the number of respondents with opinions, because some respondents offering opinions are guessing. The strategy of including 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?"[86] 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?"

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 in these waters?"[87]). 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?"[88]) 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."[89]).

Open-ended and closed-ended questions may elicit very different responses.[90]

86. 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).

87. A relevant example from *Wilhoite v. Olin Corp.* is described in McGovern & Lind, *supra* note 27, at 76.

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

89. 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?"

90. Howard Schuman & Stanley Presser, *Question Wording as an Independent Variable in Survey Analysis*,

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.[91]

The advantage of open-ended questions is that they give the respondent fewer hints about the answer that is expected or preferred. Precoded responses on a closed-ended question, in addition to reminding respondents of options that they might not otherwise consider,[92] 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 other brands.[93] 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?"[94] 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.[95] Note, however, that the open-ended question also may suggest that the answer is more than one. By asking "how many different brands," the question suggests (1) that the viewer should have received some message from the commercial about the number of brands each woman tried and (2) that different brands were tried. Thus, the wording of a question, open-ended or closed-ended, can be leading, and the degree of suggestiveness of each question must be considered in evaluating the objectivity of a survey.

6 Soc. Methods & Res. 151 (1977); Schuman & Presser, *supra* note 82, at 79–112; Converse & Presser, *supra* note 76, at 33.

91. 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 82, 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.

92. Schwarz & Hippler, *supra* note 81, at 43.

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

94. This was the wording of the stem of the closed-ended question in the survey discussed in *Vidal Sassoon*, 661 F.2d at 275–76.

95. 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.

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.[96] Moreover, if respondents are told explicitly that they are not limited to the choices presented, most respondents nevertheless will select an answer from among the listed ones.[97]

Although many courts prefer open-ended questions on the grounds that they tend to be less leading, the value of any open-ended or closed-ended question depends on the information it is intended to elicit. Open-ended questions are more appropriate when the survey is attempting to gauge what comes first to a respondent's mind, but closed-ended questions are 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. 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. In either situation, interviewers should record verbatim both what the respondent says and what the interviewer says in the attempt to get 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.

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. Opinion Q. 275, 292 (1996).

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

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

If the survey is designed to allow for probes, interviewers must be given explicit instructions on when they should probe and what they should say in probing. Standard probes used to draw out all that the respondent has to say (e.g., "Any further thoughts?" "Anything else?" "Can you explain that a little more?") are relatively innocuous and noncontroversial in content, 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.[98] 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.[99] Thus, 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

98. *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).

99. *See* Schuman & Presser, *supra* note 82, at 23, 56–74; Norman M. Bradburn, *Response Effects*, *in* Handbook of Survey Research, *supra* note 1, at 289, 302. In *R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc.*, 511 F. Supp. 867, 875 (S.D.N.Y. 1980), the court recognized the biased structure of a survey which 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).

the questions go from the general (e.g., "What do you recall being discussed 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?").[100]

The mode of questioning can influence the form that an order effect takes. In mail surveys, respondents are more likely to select the first choice offered (a primacy effect); in telephone surveys, respondents are more likely to choose the last choice offered (a recency effect). Although these effects are typically small, no general formula is available that can adjust values to correct for order effects, because the size and even the direction of the order effects may 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,[101] which order of presentation will produce the more accurate response?[102]

To control for order effects, the order of the questions and the order of the response choices in a survey should be rotated,[103] so that, for example, one-third of the respondents have Product A listed first, one-third of the respondents have Product B listed first, and one-third of the respondents have Product C listed first. If the three different orders[104] 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.[105]

100. This question was accepted by the court in *U-Haul International, Inc. v. Jartran, Inc.*, 522 F. Supp. 1238, 1249 (D. Ariz. 1981), *aff'd*, 681 F.2d 1159 (9th Cir. 1982).

101. Similarly, candidates in the first position on the ballot tend to attract extra votes when the candidates are not well known. Henry M. Bain & Donald S. Hecock, Ballot Position and Voter's Choice: The Arrangement of Names on the Ballot and Its Effect on the Voter (1973).

102. *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).

103. *See, e.g.*, Stouffer Foods Corp., 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *24–25 (Sept. 26, 1994); *cf.* 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).

104. 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.

105. Although rotation is desirable, many surveys are conducted with no attention to this potential bias. Since it is impossible to know in the abstract whether a 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?

Most surveys that are designed to provide evidence of trademark infringement or deceptive advertising are not conducted to describe consumer beliefs. Instead, they are intended to show how a trademark or the content of a commercial influences respondents' perceptions or understanding of a product or commercial. Thus, the question is whether the commercial misleads the consumer into thinking that Product A is a superior pain reliever, not whether consumers hold inaccurate beliefs about the product. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that records consumers' impressions after they view the commercial may reflect those preexisting beliefs rather than impressions produced by the commercial.

Surveys that record consumer impressions have a limited ability to answer questions about the origins of those impressions. The difficulty is that the consumer's response to any question on the survey may be the result of information or misinformation from sources other than the trademark the respondent is being shown or the commercial he or she has just watched. In a trademark survey attempting to show secondary meaning, for example, respondents were shown a picture of the stripes used on Mennen stick deodorant and asked, "[W]hich [brand] would you say uses these stripes on their package?"[106] 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"[107]; 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.[108]

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 source of their impressions.[109] The more routine the idea being examined in the survey (e.g., that the advertised pain reliever is more effective than others on the

106. 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.*

107. *Id.*

108. *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).

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

market; that the mark belongs to the brand with the largest market share), the more likely it is that the respondent's answer is influenced by preexisting impressions, by expectations about what commercials generally say (e.g., the product being advertised is better than its competitors), or by guessing, rather than by the actual content of the commercial message or trademark being evaluated.

It is possible to adjust many survey designs so that causal inferences about the effect of a trademark or an allegedly deceptive commercial become clear and unambiguous. By adding an appropriate control group, the survey expert can test directly the influence of the stimulus.[110] In the simplest version of a survey experiment, respondents are assigned randomly to one of two conditions.[111] 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.[112] Respondents in both the experimental and control groups answer the same set of questions. The effect of the 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 with 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 error[113]) can be attributed only to the allegedly deceptive commercial. Without the control group, it is not possible to determine how much of the 40% is due to respondents' preexisting beliefs 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

110. *See* Shari S. Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 Law & Hum. Behav. 239, 244–46 (1989); Shari S. Diamond & Linda Dimitropoulos, Deception and Puffery in Advertising: Behavioral Science Implications for Regulation (American Bar Found. Working Paper Series No. 9105, 1994); Jacob Jacoby & Constance Small, *Applied Marketing: The FDA Approach to Defining Misleading Advertising*, 39 J. Marketing 65, 68 (1975). For a more general discussion of the role of control groups, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, § II.A, in this manual.

111. Random assignment should not be confused with random selection. When respondents are assigned randomly to different treatment groups (e.g., respondents in each group watch a different commercial), the procedure ensures that within the limits of sampling error the two groups of respondents will be equivalent except for the different treatments they receive. Respondents selected for a mall intercept study, and not from a probability sample, may be assigned randomly to different treatment groups. Random selection, in contrast, describes the method of selecting a sample of respondents in a probability sample. *See supra* § III.C.

112. 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.

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

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 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,[114] 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.

In designing a control group study, 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. A survey with an imperfect control group generally provides better information than a survey with no control group at all, but the choice of the specific 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, since attractiveness may affect perceived familiarity.[115] 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 differ because both cause respondents to express the same level of confusion.[116]

Explicit attention to the value of control groups in trademark and deceptive-advertising litigation is a recent phenomenon, but it is becoming more common.[117] A LEXIS search using *Lanham Act* and *control group* revealed fourteen

114. 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. Stouffer Foods Corp., 118 F.T.C. 746, No. 9250, 1994 FTC LEXIS 196, at *31 (Sept. 26, 1994).

115. *See, e.g.*, Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Ltd. Partnership, 34 F.3d 410, 415–16 (7th Cir. 1994) (The 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).

116. *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).

117. *See, e.g.*, 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).

district court cases in the six years since the first edition of this manual in 1994,[118] five district court cases in the seven years from 1987 to 1993,[119] and only one case before 1987[120] in which surveys with control groups were discussed. Other cases, however, have described or considered surveys using control group designs without labeling the comparison group a control group.[121] 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 less ambiguous findings, which may lead to a resolution before a preliminary injunction hearing or trial occurs.[122]

Another more common use of control methodology is a control question. Rather than administering a control stimulus to a separate group of respondents,

118. National Football League Properties, Inc. v. Prostyle, Inc., 57 F. Supp. 2d 665 (E.D. Wis. 1999); Nabisco, Inc. v. PF Brands, Inc., 50 F. Supp. 2d 188 (S.D.N.Y. 1999); Proctor & Gamble Co. v. Colgate-Palmolive Co., No. 96 Civ. 9123, 1998 U.S. Dist. LEXIS 17773 (S.D.N.Y. Nov. 5, 1998); Mattel, Inc. v. MCA Records, Inc., 28 F. Supp. 2d 1120 (C.D. Cal. 1998); Westchester Media Co. v. PRL USA Holdings, No. H-97-3278, 1998 U.S. Dist. LEXIS 11737 (S.D. Tex. July 2, 1998); Time Inc. v. Petersen Publ'g Co., 976 F. Supp. 263 (S.D.N.Y. 1997), *aff'd*, 173 F.3d 113 (2d Cir. 1999); Adjusters Int'l, Inc. v. Public Adjusters Int'l, Inc., No. 92-CV-1426, 1996 U.S. Dist. LEXIS 12604 (N.D.N.Y. Aug. 27, 1996); Upjohn Co. v. American Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049 (W.D. Mich. Apr. 5, 1996); Copy Cop, Inc. v. Task Printing, Inc., 908 F. Supp. 37 (D. Mass. 1995); Volkswagen Aktiengesellschaft v. Uptown Motors, No. 91-CIV-3447, 1995 U.S. Dist. LEXIS 13869 (S.D.N.Y. July 13, 1995); Western Publ'g Co. v. Publications Int'l, Ltd., No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917 (N.D. Ill. May 2, 1995); Dogloo, Inc. v. Doskocil Mfg. Co., 893 F. Supp. 911 (C.D. Cal. 1995); Reed-Union Corp. v. Turtle Wax, Inc., 869 F. Supp. 1304 (N.D. Ill. 1994), *aff'd*, 77 F.3d 909 (7th Cir. 1996); Pfizer, Inc. v. Miles, Inc., 868 F. Supp. 437 (D. Conn. 1994).

119. ConAgra, Inc. v. Geo. A. Hormel & Co., 784 F. Supp. 700 (D. Neb. 1992), *aff'd*, 990 F.2d 368 (8th Cir. 1993); Johnson & Johnson-Merck Consumer Pharms. Co. v. Smithkline Beecham Corp., No. 91 Civ. 0960, 1991 U.S. Dist. LEXIS 13689 (S.D.N.Y. Sept. 30, 1991), *aff'd*, 960 F.2d 294 (2d Cir. 1992); Goya Foods, Inc. v. Condal Distribs., Inc., 732 F. Supp. 453 (S.D.N.Y. 1990); Sturm, Ruger & Co. v. Arcadia Mach. & Tool, Inc., No. 85-8459, 1988 U.S. Dist. LEXIS 16451 (C.D. Cal. Nov. 7, 1988); Frisch's Restaurant, Inc. v. Elby's Big Boy, Inc., 661 F. Supp. 971 (S.D. Ohio 1987), *aff'd*, 849 F.2d 1012 (6th Cir. 1988).

120. American Basketball Ass'n v. AMF Voit, Inc., 358 F. Supp. 981 (S.D.N.Y.), *aff'd*, 487 F.2d 1393 (2d Cir. 1973).

121. Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Ltd. Partnership, No. 94-727-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).

122. The relatively infrequent mention of control groups in surveys discussed in federal cases is not confined to Lanham Act litigation. A LEXIS search using *survey* and *control group* revealed thirty district court cases in the six years from 1994 in which *control group* was used to refer to a methodological feature: the fourteen Lanham Act cases cited *supra* note 118; nine that referred to medical, physiological, or pharmacological experiments; and seven others.

the survey asks all respondents one or more control questions along with the question about the product or service. In a trademark dispute, for example, a survey indicated that 7.2% of respondents believed that "The Mart" and "K-Mart" 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.[123]

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.[124]

Every measure of opinion or belief in a survey reflects some degree of error. Control groups and control questions are the most reliable means for assessing response levels against the baseline level of error associated with a particular question.

G. What Limitations Are Associated with the Mode of Data Collection Used in the Survey?

Three primary methods are used to collect survey data: (1) in-person interviews, (2) telephone surveys, and (3) mail surveys.[125] The choice of a data collection method for a survey should be justified by its strengths and weaknesses.

1. In-person interviews

Although costly, in-person interviews generally are the preferred method of data collection, especially when visual materials must be shown to the respondent under controlled conditions.[126] When the questions are complex and the interviewers are skilled, in-person interviewing provides the maximum oppor-

123. 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).

124. E.I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502, 526–27 & n.54 (E.D.N.Y. 1975).

125. 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.

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

tunity to 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. As described in section V.A, appropriate interviewer training is necessary if these potential benefits are to be realized. Objections to the use of in-person interviews arise primarily from their high cost or, on occasion, from evidence of inept or biased interviewers.

2. *Telephone surveys*

Telephone surveys offer a comparatively fast and low-cost alternative to in-person surveys and are particularly useful when the population is large and geographically dispersed. Telephone interviews (unless supplemented with mailed materials) can be used only when it is unnecessary to show the respondent any visual materials. Thus, an attorney may present the results of a telephone survey of jury-eligible citizens in a motion for a change of venue in order to provide evidence that community prejudice raises a reasonable suspicion of potential jury bias.[127] 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."[128] In contrast, when words are not sufficient because, for example, the survey is assessing reactions to the trade dress or packaging of a product that is alleged to promote confusion, a telephone survey alone does not offer a suitable vehicle for questioning respondents.[129]

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

- (when prospective respondents are not business personnel) whether some form of random-digit dialing[130] was used instead of or to supplement tele-

127. United States v. Partin, 320 F. Supp. 275, 279-80 (E.D. La. 1970). For a discussion of surveys used in motions for change of venue, see Neal Miller, *Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation, Part II*, 40 Rutgers L. Rev. 467, 470–74 (1988); National Jury Project, Jurywork: Systematic Techniques (Elissa Krauss & Beth Bonora eds., 2d ed. 1983).

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

129. 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).

130. Random digit dialing provides coverage of households with both listed and unlisted telephone numbers by generating numbers at random from the frame of all possible telephone numbers. James M. Lepkowski, *Telephone Sampling Methods in the United States*, *in* Telephone Survey Methodology 81–91 (Robert M. Groves et al. eds., 1988).

phone numbers obtained from telephone directories, because up to 65% of all residential telephone numbers in some areas may be unlisted;[131]

- 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;[132] and
- whether interviewers were required to call back at several different times of the day and on different days to increase the likelihood of contacting individuals or businesses with different schedules.

Telephone surveys that do not include these procedures may, like other nonprobability sampling approaches, be adequate for providing rough approximations. The vulnerability of the survey depends on the information being gathered. More elaborate procedures for achieving a representative sample of respondents are advisable if the survey instrument requests information that is likely to differ for individuals with listed telephone numbers and individuals with unlisted telephone numbers, or individuals rarely at home and those usually at home.

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

1. the procedures that were used to identify potential respondents;
2. the number of telephone numbers for which no contact was made; and
3. the number of contacted potential respondents who refused to participate in the survey.

Computer-assisted telephone interviewing, or CATI, is increasingly used in the administration and data entry of large-scale surveys.[133] A computer protocol may be used to generate telephone numbers and dial them as well as to guide the interviewer. The interviewer conducting a computer-assisted interview (CAI), whether by telephone or in a face-to-face setting, follows the script for the interview generated by the computer program and types in 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 so that, for example, if the respondent is asked whether she has ever been the victim of a burglary and she says yes, the computer will generate further questions about

131. In 1992, the percentage of households with unlisted numbers reached 65% in Las Vegas and 62% in Los Angeles. Survey Sampling, Inc., The Frame 2 (March 1993). Studies comparing listed and unlisted household characteristics show some important differences. Lepkowski, *supra* note 130, at 76.

132. 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.

133. William L. Nicholls II & R.M. Groves, *The Status of Computer-Assisted Telephone Interviewing*, 2 J. Official Stat. 93 (1986); Mary A. Spaeth, *CATI Facilities at Academic Research Organizations*, 21 Surv. Res. 11 (1990); William E. Saris, Computer-Assisted Interviewing (1991).

the burglary, but if she says no, the program will automatically skip the follow-up 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.[134] CAI procedures can also be used to control for order effects by having the program rotate the order in which questions or choices are presented.[135] CAI procedures, however, 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 answers that the respondent is giving to those questions.

3. Mail surveys

In general, mail surveys tend to be substantially less costly than both in-person and telephone surveys.[136] Although response rates for mail surveys are often low, researchers have obtained 70% response rates in some general public surveys and response rates of over 90% with certain specialized populations.[137] Procedures that encourage high response rates include multiple mailings, highly personalized communications, prepaid return envelopes and incentives or gratuities, assurances of confidentiality, and first-class outgoing postage.[138]

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

134. Saris, *supra* note 133, at 20, 27.

135. *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).

136. Don A. Dillman, *Mail and Other Self-Administered Questionnaires, in* Handbook of Survey Research, *supra* note 1, at 359, 373.

137. *Id.* at 360.

138. *See, e.g.,* Richard J. Fox et al., *Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response,* 52 Pub. Opinion Q. 467, 482 (1988); Eleanor Singer et al., *Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature,* 59 Pub. Opinion Q. 66, 71 (1995); Kenneth D. Hopkins & Arlen R. Gullickson, *Response Rates in Survey Research: A Meta-Analysis of the Effects of Monetary Gratuities,* 61 J. Experimental Educ. 52, 54–57, 59 (1992).

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. If it is crucial to have respondents answer questions in a particular order, a mail survey cannot be depended on to provide adequate data.[139]

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 can substantially reduce the cost of reaching potential respondents and offer some of the advantages of in-person interviews by allowing the computer to show the respondent pictures or lists of response choices in the course of asking the respondent questions. The key limitation is that the respondents accessible over the Internet must fairly represent the relevant population whose responses the survey was designed to measure. Thus, a litigant presenting the results of a web-based survey should be prepared to provide evidence on the potential bias in sampling that the web-based survey is likely to introduce. If the target population consists of computer users, the bias may be minimal. If the target population consists of owners of television sets, significant bias is likely.

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."[140] 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. Interviewers also should be instructed to record verbatim the respondent's

139. Dillman, *supra* note 136, at 368–70.
140. Toys "R" Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983).

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, one treatise recommends at least five hours of training in general interviewing skills and techniques for new interviewers.[141]

The more complicated the survey instrument is, the more training and experience the interviewers require. Thus, if the interview includes a skip pattern (where, e.g., Questions 4–6 are asked only if the respondent says yes to Question 3, and Questions 8–10 are asked only if the respondent says no to Question 3), interviewers must be trained to follow the pattern. Similarly, if the questions require specific probes to 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 out of the hearing of any other respondent). The trustworthiness of a survey is questionable if there is evidence that some interviews were conducted in a setting in which respondents were likely to have been distracted or in which others were present and could overhear. Such evidence of careless administration of the survey was one ground used by a court to reject as inadmissible a survey that purported to demonstrate consumer confusion.[142]

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 and complete disclosure of the methods used for instruction and training are crucial elements that, if compromised, seriously undermine the trustworthiness of any survey.

141. Eve Weinberg, *Data Collection: Planning and Management*, *in* Handbook of Survey Research, *supra* note 1, at 329, 332.

142. *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).

B. What Did the Interviewers Know About the Survey and Its Sponsorship?

One way to protect the objectivity of survey administration is to avoid telling interviewers who is sponsoring the survey. Interviewers who know the identity of the survey's sponsor may affect results inadvertently by communicating to respondents their expectations or what they believe are the preferred responses of the survey's sponsor. To ensure objectivity in the administration of the survey, it is standard interview practice to conduct double-blind research whenever possible: both the interviewer and the respondent are blind to the sponsor of the survey and its purpose. Thus, the survey instrument should provide no explicit clues (e.g., a sponsor's letterhead appearing on the survey) and no implicit clues (e.g., reversing the usual order of the yes and no response boxes on the interviewer's form next to a crucial question, thereby potentially increasing the likelihood that *no* will be checked[143]) about the sponsorship of the survey or the expected responses.

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 kind of responses the sponsor expects or would find acceptable.[144] 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.[145]

143. Centaur Communications, Ltd. v. A/S/M Communications, Inc., 652 F. Supp. 1105, 1111 n.3 (S.D.N.Y.) (pointing out that reversing the usual order of response choices, yes or no, to no or yes may confuse interviewers as well as introduce bias), *aff'd*, 830 F.2d 1217 (2d Cir. 1987).

144. *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).

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

C. What Procedures Were Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias?

Three methods are used to ensure that the survey instrument was implemented in an unbiased fashion and according to instructions. The first, monitoring the interviews as they occur, is done most easily when telephone surveys are used. A supervisor listens to a sample of interviews for each interviewer. Field settings make monitoring more difficult, but evidence that monitoring has occurred provides an additional indication that the survey has been reliably implemented.

Second, validation of interviews occurs when 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. The standard procedure for validation of in-person interviews is to telephone a random sample of about 10% to 15% of the respondents.[146] Some attempts to reach the respondent will be unsuccessful, and occasionally a respondent will deny that the interview took place even though it did. Because the information checked is limited to whether the interview took place and whether the respondent was qualified, this validation procedure does not determine whether the initial interview as a whole was conducted properly. Nonetheless, this standard validation technique warns interviewers that their work is being checked and can detect gross failures in the administration of the survey.

A third way to verify that the interviews were conducted properly is to compare the work done by each individual interviewer. By reviewing the interviews and individual responses recorded by each interviewer, researchers can identify any response patterns or inconsistencies for 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 at least 50% of interviews by a third party rather than by the field service that conducted the interviews increases the trustworthiness of the survey results.[147]

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

147. 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."

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 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.[148]

148. *See, e.g.*, Coca-Cola Co. v. Tropicana Prods., Inc., 538 F. Supp. 1091, 1094–96 (S.D.N.Y.) (plaintiff's expert stated that respondents' answers to the several open-ended questions revealed that 43% of respondents thought Tropicana was portrayed as fresh squeezed; the court's own tabulation found no more than 15% believed this was true), *rev'd on other grounds*, 690 F.2d 312 (2d Cir. 1982). *See also* McNeilab, Inc. v. American Home Prods. Corp., 501 F. Supp. 517 (S.D.N.Y. 1980); 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); 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).

VII. Disclosure and Reporting

A. When Was Information About the Survey Methodology and Results Disclosed?

Objections to the definition of the relevant population, the method of selecting the sample, and the wording of questions generally are raised for the first time when the results of the survey are presented. By that time it is too late to correct methodological deficiencies that could have been addressed in the planning stages of the survey. The plaintiff in a trademark case[149] submitted a set of proposed survey questions to the trial judge, who ruled that the survey results would be admissible at trial while reserving the question of the weight the evidence would be given.[150] The court of appeals called this approach a commendable procedure and suggested that it would have been even more desirable if the parties had "attempt[ed] in good faith to agree upon the questions to be in such a survey."[151]

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."[152] The parties then were encouraged to attempt to resolve any methodological disagreements before the survey was conducted.[153] Although this passage in the second edition of the manual has been cited with apparent approval,[154] the prior agreement the manual recommends has occurred rarely and the *Manual for Complex Litigation, Third,* recommends, but does not advocate requiring, prior disclosure and discussion of survey plans.[155]

Rule 26 of the Federal Rules of Civil Procedure requires extensive disclosure of the basis of opinions offered by testifying experts. However, these provisions may not produce disclosure of all survey materials, because parties are not obli-

149. Union Carbide Corp. v. Ever-Ready, Inc., 392 F. Supp. 280 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366 (7th Cir.), *cert. denied*, 429 U.S. 830 (1976).

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

151. *Union Carbide*, 531 F.2d at 386. More recently, the Seventh Circuit recommended the filing of a motion *in limine*, asking the district court to determine the admissibility of a survey based on an examination of the survey questions and the results of a preliminary survey before the party undertakes the expense of conducting the actual survey. Piper Aircraft Corp. v. Wag-Aero, Inc., 741 F.2d 925, 929 (7th Cir. 1984).

152. MCL 2d, *supra* note 15, § 21.484.

153. *Id.*

154. *E.g.*, National Football League Properties, Inc. v. New Jersey Giants, Inc., 637 F. Supp. 507, 514 n.3 (D.N.J. 1986).

155. MCL 3d, *supra* note 15, § 21.493.

gated to disclose information about nontestifying experts. Parties considering whether to commission or use a survey for litigation are not obligated to present a survey that produces unfavorable results. Prior disclosure of a proposed survey instrument places the party that ultimately would prefer not to present the survey in the position of presenting damaging results or leaving the impression that the results are not being presented because they were unfavorable. Anticipating such a situation, parties do not decide whether an expert will testify until after the results of the survey are available.

Nonetheless, courts are in a position to encourage early disclosure and discussion even if they do not lead to agreement between the parties. In *McNeilab, Inc. v. American Home Products Corp.*,[156] 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.[157]

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 population that was actually sampled;
3. a description of the sample design, including the method of selecting respondents, the method of interview, the number of callbacks, respondent eligibility or screening criteria, and other pertinent information;
4. a description of the results of sample implementation, including (a) the number of potential respondents contacted, (b) the number not reached, (c) the number of refusals, (d) the number of incomplete interviews or terminations, (e) the number of noneligibles, and (f) the number of completed interviews;
5. the exact wording of the questions used, including a copy of each version of the actual questionnaire, interviewer instructions, and visual exhibits;
6. a description of any special scoring (e.g., grouping of verbatim responses into broader categories);

156. 848 F.2d 34, 36 (2d Cir. 1988) (discussing with approval the actions of the district court).

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

7. estimates of the sampling error, where appropriate (i.e., in probability samples);
8. statistical tables clearly labeled and identified as to source of data, including the number of raw cases forming the base for each table, row, or column; and
9. copies of interviewer instructions, validation results, and code books.[158]

A description of the procedures and results of pilot testing is not included on this list. Survey professionals generally do not describe pilot testing in their reports. The Federal Rules of Civil Procedure, however, may require that a testifying expert disclose pilot work that serves as a basis for the expert's opinion. The situation is more complicated when a nontestifying expert conducts the pilot work and the testifying expert learns about the pilot testing only indirectly through the attorney's advice about the relevant issues in the case. Some commentators suggest that attorneys are obligated to disclose such pilot work.[159]

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,[160] 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 in order to reinterview some respondents. The party introducing the survey or the survey organization that conducted the research generally resists supplying such information.[161] Professional surveyors as a rule guarantee

158. These criteria were adapted from the Council of Am. Survey Res. Orgs., *supra* note 41, § 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 of his colleagues, 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.

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

160. *See supra* § V.C.

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

confidentiality in an effort to increase participation rates and to encourage candid responses. Because failure to extend confidentiality may bias both the willingness of potential respondents to participate in a survey and their responses, the professional standards for survey researchers generally prohibit disclosure of 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."[162] 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.[163]

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.

162. Council of Am. Survey Res. Orgs., *supra* note 41, § I.A.3.f. Similar provisions are contained in the By-Laws of the American Association for Public Opinion Research.

163. 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, Note, *The Researcher–Subject Relationship: The Need for Protection and a Model Statute*, 62 Geo. L.J. 243, 253 (1973)). *See also* Lampshire 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., 1983); 1 Environmental Protection Agency, Survey Management Handbook (1983); Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991); William 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 answer 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 answer 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. Also, margin of error.

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

double-blind research. Research in which the respondent and the interviewer are not given information that will alert them to the anticipated or preferred 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).

nonprobability sample. Any sample that does not qualify as a probability sample.

open-ended question. A question that requires the respondent to formulate his or her own response.

order effect. A tendency of respondents to choose an item based in part on the order in which it appears in the question, questionnaire, or interview (see primacy effect and recency effect); also referred to as a context effect because the context of the question influences the way the respondent perceives and answers it.

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 (objects, individuals, or other social units) that have some common property of interest; the target population is the collection of elements that the researcher would like to study; the survey population is the population that is actually sampled and for which data may be obtained. Also, universe.

population value, population parameter. The actual value of some characteristic in the population (e.g., the average age); 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 follow-up 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 simple random sample.

recency effect. A tendency of respondents to choose later items from a list of choices; the opposite of a primacy effect.

sample. A subset of a population or universe selected so as to yield information about the population as a whole.

sampling error. The 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 objects, individuals, or other social 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 that permits the researcher to subdivide the population into mutually exclusive and exhaustive subpopulations, or strata; within these strata, separate samples are selected; results can be combined to form overall population estimates or used to report separate within-stratum estimates.

survey population. See population.

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

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

William G. Cochran, Sampling Techniques (3d ed. 1977).

Jean M. Converse & Stanley Presser, Survey Questions: Handcrafting the Standardized Questionnaire (1986).

Thomas D. Cook & Donald T. Campbell, Quasi-Experimentation: Design and Analysis Issues for Field Settings (1979).

Shari S. Diamond, *Methods for the Empirical Study of Law, in* Law and the Social Sciences (Leon Lipson & Stanton Wheeler eds., 1986).

Floyd J. Fowler, Survey Research Methods (2d ed. 1984).

Robert M. Groves & Robert L. Kahn, Surveys by Telephone: A National Comparison with Personal Interviews (1979).

Handbook of Survey Research (Peter H. Rossi et al. eds., 1983).

Leslie Kish, Survey Sampling (1965).

Measurement Errors in Surveys (Paul P. Biemer et al. eds., 1991).

Questions About Questions: Inquiries into the Cognitive Bases of Surveys (Judith M. Tanur ed., 1992).

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

Seymour Sudman, Applied Sampling (1976).

Seymour Sudman & Norman M. Bradburn, Response Effects in Surveys: A Review and Synthesis (1974).

Telephone Survey Methodology (Robert M. Groves et al. eds., 1988).