

# Reference Manual on Scientific Evidence

**Fourth Edition**

---

Committee on Science for Judges—  
Development of the Reference Manual  
on Scientific Evidence, Fourth Edition

Committee on Science, Technology,  
and Law

Policy and Global Affairs

# Reference Guide on Survey Research

SHARI SEIDMAN DIAMOND, MATTHEW KUGLER,  
AND JAMES N. DRUCKMAN

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

*Matthew Kugler, J.D., Ph.D., is Professor of Law at Northwestern University.*

*James N. Druckman, Ph.D., is Martin Brewer Anderson Professor of Political Science at the University of Rochester.*

## CONTENTS

Introduction, 683
Use of Surveys in Court, 685
A Comparison of surveys Evidence and Individual Testimony, 688
Purpose and Design of the Survey, 689
Addressing Relevant Questions, 689
Role of Attorneys in Survey Design and Administration, 690
Skill and Experience of the Experts Who Designed, Conducted, or Analyzed the Survey, 691
Skill and Experience of the Experts Who Will Testify About Surveys Conducted by Others, 691
Population Definition and Sampling, 692
The Survey Universe or Population, 692
The Sampling Frame as an Approximation of the Population, 694
The Sample as a Reflection of the Relevant Characteristics of the Population, 696
Nonresponse as a Potential Source of Bias, 703
Precautions Taken to Ensure That Only Qualified Respondents Were Included in the Survey, 706
Survey Questions and Structure, 707
Clarity, Precision, and Lack of Bias in the Framing of the Survey Questions, 707
What Respondents and Interviewers Knew About the Survey Purpose and Its Sponsorship, 709
Handling Respondents with No Opinions and Reducing Respondent Guessing, 711
Use of Open-Ended and Closed-Ended Questions, 713

Use of Probes to Clarify Ambiguous or Incomplete Answers, 717
Potential Order Effects, 717
Appropriate Inclusion of Control Groups, Control Questions, or Other Comparisons, 718
Benefits and Limitations Associated with the Mode of Data Collection Used in the Survey, 725
In-Person Interviews, 726
Telephone Interviews, 727
Mail Questionnaires, 729
Internet Surveys, 730
Mixed-Format Surveys, 733
Surveys Involving Interviewers, 734
Selection and Training of the Interviewers, 734
Procedures Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias, 735
Accuracy of Data Entry, 736
Disclosure and Reporting, 736
Early Disclosure About the Survey Methodology and Results, 736
Completeness and Accuracy of All Relevant Information in the Survey Report, 738
Concluding Observations, 743
Glossary of Terms, 744
References on Survey Research, 747

# Introduction

Sample surveys gather responses from a subset of a population to draw inferences about the population as a whole. They are used to describe or enumerate beliefs, attitudes, or behaviors of persons or other social units.<sup>1</sup> Surveys typically are offered in legal proceedings to establish or refute claims about the characteristics of those individuals or social units.<sup>2</sup> We focus here primarily on sample surveys with individuals reporting about themselves (e.g., their own beliefs) or their organizations (e.g., the company where they are employed). Such surveys must deal not only with issues of population definition, sampling, and measurement common to all surveys, but also with the specialized issues that arise in obtaining information from human respondents.

In principle, a survey can count or measure every member of the relevant population. A survey that does this full count is sometimes called a census. In practice, however, most surveys typically count or measure only a portion of the individuals or other units that the survey intends to describe. In either case, the goal is to provide information on the relevant population. Sample surveys can be carried out using probability or nonprobability sampling techniques. Although probability sampling offers important advantages over nonprobability sampling,<sup>3</sup> various forms of nonprobability sampling are in wide use. Thus, in this reference guide, we discuss both probability samples and nonprobability samples, including their strengths and weaknesses for achieving various purposes.

As a method of data collection, surveys have several crucial potential advantages over less systematic approaches.<sup>4</sup> When a survey is properly designed,

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

2. See, e.g., Sanderson Farms, Inc. v. Tyson Foods, Inc., 547 F. Supp. 2d 491 (D. Md. 2008); SMS Sys. Maint. Servs. v. Digital Equip. Corp. 188 F.3d 11 22, 22–23 (1st Cir. 1999). For other examples, see *infra* notes 10–26 and accompanying text.

3. See section titled “The Sample as a Reflection of the Relevant Characteristics of the Population” below.

4. This does not mean that surveys can be relied on to address all questions. For example, if survey respondents had been asked in the days before the attacks of 9/11 to predict whether they would volunteer for military service if Washington, D.C., were to be bombed, their answers may not have provided accurate predictions. Although respondents might have willingly answered the question, their assessment of what they would actually do in response to an attack simply may have been inaccurate. Even the option of a “do not know” choice would not have prevented an error in prediction if they believed they could accurately predict what they would do. Thus, although such a survey would have been suitable for assessing the *predictions* of respondents, it might have provided a very inaccurate estimate of what an actual response to the attack would be. If a survey respondent has limited experience (e.g., a child predicting what she will do as an adult) or the hypothetical is far removed from reality or imaginable reality, the survey is unlikely to provide accurate predictions.

executed, and described, it (1) efficiently presents the responses of a group of individuals or other units (e.g., organizations) and (2) permits an assessment of the extent to which the measured responses of the individuals or other units are likely to adequately represent a relevant population of individuals or other units. All questions asked of respondents and all other measuring devices used (e.g., criteria for selecting eligible respondents) can be examined by the court and the opposing party for objectivity, clarity, and relevance, and all answers or other measures obtained can be analyzed for completeness and consistency.

So that the court and the opposing party can closely scrutinize the survey, the party offering the survey as evidence should describe in detail the design, execution, and analysis of the survey, including (1) a description of the population from which the sample was selected, demonstrating that it was a relevant population for the question at hand; (2) a description of how the sample was drawn and an explanation for why that sample design was appropriate; (3) a report on response rate and the ability of the sample to represent the target population; (4) evidence that respondents were attentive and honest in answering the questions on the survey; and (5) an evaluation of any sources of potential bias in respondents' answers or in the ability of the results from the sample to generalize to the relevant population.

The material covered in this reference guide is 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. Questions about a survey 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 to help the judge evaluate what weight, if any, the survey should be given.<sup>5</sup>

All sample surveys should address the issues concerning purpose and design (see section titled "Purpose and Design of the Survey" below), population definition and sampling (see section titled "Population Definition and Sampling" below), and disclosure and reporting (see section titled "Disclosure and Reporting" below). All questionnaire and interview surveys raise methodological issues involving survey questions and structure (see section titled "Survey Questions and Structure" below) and confidentiality (see section titled "Disclosure and Reporting" below). Interview surveys introduce additional issues, such as accuracy of data entry (see section titled "Accuracy of Data Entry" below) and interviewer training and qualifications (see

5. Lanham Act cases involving trademark infringement or deceptive advertising frequently require expedited hearings that request injunctive relief, and 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 Pharmas., Inc.*, 509 U.S. 579, 589 (1993); Fed. R. Evid. 702.

section titled “Surveys Involving Interviewers” below). And online surveys raise special issues and questions (see section titled “Internet Surveys” below).

## Use of Surveys in Court

Sixty years ago, the question of whether surveys were acceptable evidence was unsettled.<sup>6</sup> Early doubts about the admissibility of surveys centered on their use of sampling and their status as hearsay evidence. Federal Rule of Evidence 703 settled both matters for surveys by redirecting attention to the “validity of the techniques employed.”<sup>7</sup> The inquiry under Rule 703 focuses on whether facts or data are “of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject.”<sup>8</sup>

Because the survey method provides an economical and systematic way to gather information and draw inferences about a large number of individuals or other units, surveys are used widely in business, government, administrative settings, and judicial proceedings.<sup>9</sup> Both federal and state courts have accepted survey evidence on a variety of issues. Our review of cases citing survey evidence over the ten-year period between 2012 and 2022 revealed cases involving legal issues in administrative law, copyright, criminal law, deceptive advertising, discrimination, employment, patent, trademark, and unfair competition, among others.

Some of these cases cited surveys conducted specifically for litigation, and some cited surveys not prepared for litigation. The topics of both litigation and nonlitigation surveys are diverse.<sup>10</sup> Surveys appear even in the most routine of

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

7. Fed. R. Evid. 703 advisory committee note. 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*, 509 U.S. 579; *see also* General Elec. Co. v. Joiner, 522 U.S. 136, 147 (1997).

8. Fed. R. Evid. 703 advisory committee note.

9. Some sample surveys are so well accepted that they may not even be recognized as surveys. For example, some U.S. Census Bureau data are based on sample surveys, including the widely relied-upon American Community Survey, <https://perma.cc/XLY8-R4YE>. Similarly, the Standard Table of Mortality, which is accepted as proof of the average life expectancy of an individual of a particular age and gender, is based on survey data. Surveys conducted by federal agencies are generally of high quality. Their demographic statistics are often used as benchmarks for nongovernmental research.

10. *See, e.g.*, Kittle-Aikeley v. Claycomb, 807 F.3d 913, 926 (8th Cir. 2015) (survey to assess the seriousness of the drug abuse problem in schoolchildren to help justify the school’s decision to drug test its community college student body); Miller v. Bonta, 542 F. Supp. 3d 1009, 1021 (S.D. Cal. 2021), *vacated and remanded*, No. 21-55608, 2022 WL 3095986 (9th Cir. Aug. 1, 2022) (in a Second Amendment case, survey used to note that many people in California own firearms); Missouri v. Biden, 576 F. Supp. 3d 622, 634 (E.D. Mo. 2021) (survey used to predict workers’ response to a vaccine mandate for federal contractors); United States v. Cloud, No. 1:19-cr-02032-SMJ-1, 2021 U.S. Dist. LEXIS 235350, at \*13 (E.D. Wash. Dec. 8, 2021) (survey used to determine whether a jury pool was sufficiently biased to warrant a change of venue); Lord & Taylor LLC v. Zim Integrated Shipping Servs., Ltd., 108 F. Supp. 3d 197, 227 (S.D.N.Y. 2015) (survey used to note that 79% of

cases. Vocational experts testifying in Social Security Administration disability hearings regularly rely on a variety of surveys conducted by the Bureau of Labor Statistics, specifically the Occupational Requirements Survey,<sup>11</sup> the Occupational Employment Survey,<sup>12</sup> and the National Compensation Survey,<sup>13</sup> to support their opinions on whether a person would be able to find gainful employment given their documented disabilities. One might think that reliance on these surveys is unexceptional, but federal courts have sometimes struggled to determine whether a given expert's use of survey data is scientifically valid.<sup>14</sup>

Employment and discrimination law cases sometimes incorporate survey evidence as well. These surveys may be conducted as part of a company's normal business operations and may be relevant to the performance of a particular employee or the feelings of a subgroup of employees.<sup>15</sup> Other times the surveys may be conducted by plaintiffs to provide evidence of wage and working conditions.<sup>16</sup>

In cases in which courts set attorneys' fees, judges are charged with determining reasonable fees.<sup>17</sup> Surveys frequently provide courts with evidence of the prevailing market rates.<sup>18</sup>

coastal residents thought the impact of Hurricane Sandy was worse than expected, informing whether it should be treated as an Act of God for contract purposes).

11. Roberta G. v. Kijakazi, No. CV 20-07796-DFM, 2021 U.S. Dist. LEXIS 179074, at \*6-7 (C.D. Cal. Sept. 20, 2021).

12. Sok v. Kijakazi, No. 20-cv-489-wmc, 2021 U.S. Dist. LEXIS 183024, at \*12 (W.D. Wis. Sept. 24, 2021); Dawn L.C. v. Comm'r of Soc. Sec., No. 3:20-cv-00626-GCS, 2021 U.S. Dist. LEXIS 191829, at \*19 (S.D. Ill. Sept. 24, 2021); Stephanie D. v. Comm'r of Soc. Sec., No. 3:20-CV-0768 (ML), 2021 U.S. Dist. LEXIS 168701, at \*8 (N.D.N.Y. Sept. 7, 2021).

13. Missey v. Saul, No. 4:20-CV-701-ERW, 2021 U.S. Dist. LEXIS 120435, at \*14 (E.D. Mo. June 29, 2021).

14. Alaura v. Colvin, 797 F.3d 503, 507-08 (7th Cir. 2015) (describing one apparently common practice in apportioning job availability numbers extracted from the Occupational Employment Survey as "preposterous"). *See also Dawn L.C.*, 2021 U.S. Dist. LEXIS 191829, at \*17 (describing the subsequent debate among the district courts).

15. Spivey v. Mohawk ESV, Inc., No. 7:19-cv-00670, 2021 U.S. Dist. LEXIS 111905, at \*3 (W.D. Va. June 15, 2021) (use of a manager's staff approval ratings in an age discrimination case); Griffin v. Shelby Residential & Vocational Servs., No. 2:18-cv-2665, 2021 U.S. Dist. LEXIS 111866, at \*24 (W.D. Tenn. June 15, 2021) ("The Court finds that the 2016 QA surveys are the most probative and objective evidence of Defendant's rationale for termination."); Milioto-Maruca v. Lauren, No. 3:11-CV-2120, 2012 U.S. Dist. LEXIS 173945, at \*5 (M.D. Pa. Dec. 7, 2012) (work climate survey was conducted at the stores managed by the plaintiff in response to subordinate complaints).

16. Medlock v. Taco Bell Corp., No. 1:07-cv-01314-SAB, 2015 U.S. Dist. LEXIS 165235 (E.D. Cal. Dec. 9, 2015).

17. Courts are directed to calculate a lodestar figure by multiplying the number of hours reasonably expended on the litigation times a reasonable hourly rate and then adjust upward or downward based on particular features of the work involved. *Jones v. George Fox Univ.*, No. 3:19-cv-0005-JR, 2022 U.S. Dist. LEXIS 162869, at \*3 (D. Or. Sept. 9, 2022) (citing *Blum v. Stenson*, 465 U.S. 886, 888 (1984)).

18. *See, e.g., Jones*, 2022 U.S. Dist. LEXIS 162869, at \*8-9 (a survey of attorneys was conducted by the Portland State University Survey Research Lab for the Oregon State Bar; the fee

Surveys also are common in the areas of trademark and false-advertising law. Survey evidence has been used to establish whether a proposed mark is a generic term,<sup>19</sup> assess whether a mark has secondary meaning or achieved sufficient fame for a dilution claim,<sup>20</sup> evaluate whether a defendant's product presents a likelihood of confusion with a plaintiff's mark,<sup>21</sup> and probe how consumers would have interpreted an allegedly misleading advertisement.<sup>22</sup>

Surveys have additionally been conducted in false-advertising cases to assess the value of the deceptive claim to consumers<sup>23</sup> and in patent cases to assess the value of product features.<sup>24</sup> The use of conjoint analysis, a relatively new method of making such value attributions for litigation, is discussed below.<sup>25</sup>

On occasion, courts ruling on the admissibility of scientific claims have examined surveys of scientific experts to assess the extent to which a theory or

award adopted by the court used the 95th percentile rate in light of the attorney's prior experience before admission to the bar).

19. *See, e.g.*, Snyder's Lance, Inc. v. Frito-Lay N. Am., Inc., 542 F. Supp. 3d 371, 397–99, 401–02 (W.D.N.C. 2021) (contrasting the results of two surveys considering "Pretzel Crisps"); Primary Children's Med. Ctr. Found. v. Scentsy, Inc., No. 2:11-cv-1141-TC, 2012 U.S. Dist. LEXIS 86318, at \*12 (D. Utah June 20, 2012) (finding important the results of a Teflon survey analyzing perceptions of "Festival of Trees").

20. Warner Bros. Ent. v. Glob. Asylum, Inc., No. CV 12-9547 PSG (CWx), 2012 U.S. Dist. LEXIS 185695, at \*11 (C.D. Cal. Dec. 10, 2012) ("The survey results show[] that nearly 50 percent of respondents associated the term 'Hobbit' with" a single source); MZ Wallace Inc. v. Fuller, No. 18cv2265(DLC), 2018 U.S. Dist. LEXIS 214754, at \*32–35 (S.D.N.Y. Dec. 20, 2018) (finding helpful a study showing a lack of secondary meaning in the plaintiff's alleged mark); ProFoot, Inc. v. MSD Consumer Care, Inc., No. 11-7079, 2012 U.S. Dist. LEXIS 83427, at \*25–26 (D.N.J. June 14, 2012) (insufficient awareness to support a claim of fame).

21. Under Armour, Inc. v. Battle Fashions, Inc., No. 5T9-CV-297-BO, 2021 U.S. Dist. LEXIS 114292, at \*5–6 (E.D.N.C. June 18, 2021) (admitting a survey evaluating whether the defendant's use of "I can do all things" infringes on the plaintiff's mark); Nat'l Fin. Partners Corp. v. Paycom Software, Inc., No. 14 C 7424, 2015 U.S. Dist. LEXIS 74700, at \*32–33 (N.D. Ill. June 10, 2015) (finding probative one of the two Squirt-style studies conducted to assess likelihood of confusion).

22. Naimi v. Starbucks Corp., 798 F. App'x 67, 69 (9th Cir. 2019) (plaintiff's survey sufficient to establish implied representation at the motion to dismiss stage); Benson v. Newell Brands, Inc., No. 19 C 6836, 2021 U.S. Dist. LEXIS 220986, at \*19–22 (N.D. Ill. Nov. 16, 2021); *In re Elysium Health-ChromaDex Litig.*, No. 17-cv-7394 (LJL), 2022 U.S. Dist. LEXIS 25063, at \*5–36 (S.D.N.Y. Feb. 11, 2022) (contrasting the results of two experts' false advertising studies).

23. *In re Dial Complete Mktg. & Sales Pracs. Litig.*, 320 F.R.D. 326 (D.N.H. 2017) (alleged damages, claiming that the soap did not kill 99.9% of germs, but some smaller percentage).

24. Apple, Inc. v. Samsung Elecs. Co., No. 11-CV-01846-LHK, 2012 U.S. Dist. LEXIS 90877, at \*37–38, \*42–43 (N.D. Cal. June 29, 2012); SimpleAir, Inc. v. Google Inc., No. 2:14-CV-11, 2015 U.S. Dist. LEXIS 135915, at \*6–7 (E.D. Tex. Oct. 5, 2015); TV Interactive Data Corp. v. Sony Corp., 929 F. Supp. 2d 1006, 1020–22 (N.D. Cal. 2013); Microsoft Corp. v. Motorola, Inc., 904 F. Supp. 2d 1109, 1119–20 (W.D. Wash. 2012).

25. *See infra* pp. 722–25.

technique has received widespread acceptance.<sup>26</sup> Such surveys can be valuable in assisting the court, but they must reflect the views of a representative group of recognized experts who have responded to questions about the relevant issue.

In addition, survey methodology has been used creatively to assist federal courts in managing mass torts litigation. For instance, faced with the prospect of conducting discovery concerning 10,000 plaintiffs, the plaintiffs and defendants in *Wilhoite v. Olin Corp.*<sup>27</sup> jointly drafted a discovery survey that was administered in person by neutral third parties, thus replacing interrogatories and depositions. It resulted in substantial savings in both time and cost. Greater use of this approach would be beneficial to all parties.

## A Comparison of Survey Evidence and Individual Testimony

To illustrate the value of a survey, it is useful to compare the information that can be obtained from a competently done survey with the information obtained by other means. A survey is presented by a survey expert, who testifies about the responses of a substantial number of individuals who have been selected according to an explicit sampling plan and asked the same set of questions. In contrast, a party using a nonsurvey method generally identifies several witnesses who testify about their own characteristics, experiences, or impressions. Although the party has no obligation to select these witnesses in any particular way or to report on how they were chosen, the party is not likely to select witnesses whose attitudes or beliefs conflict with the party's interests. The witnesses who testify are aware of the parties involved in the case and have discussed the case with the party 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 group of potential witnesses. In some cases, courts have described surveys as the most direct form of

26. For instance, courts determined that the polygraph test has failed to achieve general acceptance in the scientific community based on the inconsistent reactions revealed in several surveys. *See United States v. Scheffer*, 523 U.S. 303, 309–10 (1998); *United States v. Bishop*, 64 F. Supp. 2d 1149 (D. Utah 1999); *United States v. Varoudakis*, No. 97-10158-RGS, 1998 WL 151238 (D. Mass. Mar. 27, 1998); *State v. Shively*, 999 P.2d 952, *aff'd*, 999 P.2d 259 (Kan. 2000); *Lee v. Martinez*, 96 P.3d 291, 304–06 (N.M. 2004). In contrast, an eyewitness identification expert was permitted to testify about scientific studies of factors affecting the perceptual ability and memory of eyewitnesses based in part on survey evidence showing general acceptance of those findings by experts within the field. *See People v. Williams*, 830 N.Y.S.2d 452, 465–66 (N.Y. Sup. Ct. 2006).

27. *Wilhoite v. Olin Corp.*, 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*, 51 Law & Contemp. Probs. 41, 49 (1988).

evidence that can be offered.<sup>28</sup> Indeed, several courts have drawn negative inferences from the absence of a survey, taking the position that failure to undertake a survey may strongly suggest that a properly done survey would not support the plaintiff's position.<sup>29</sup>

## Purpose and Design of the Survey

### Addressing Relevant Questions

Key to evaluating any survey is assessing whether it is relevant to the disputed issues. Surveys conducted in the normal course of business and not in anticipation of, or in response to, litigation may have a lesser risk of bias in favor of the interests of the party conducting the survey, but such surveys may ask irrelevant questions, lack important quality controls, or be conducted on inappropriate populations.<sup>30</sup> In contrast, surveys conducted for litigation are more likely to be designed to address the legally relevant issues in the case (e.g., to estimate damages in an antitrust suit or to assess consumer confusion in a trademark case), but may have a greater risk of bias since they are typically solicited by one of the parties to aid in the case. Thus, the content and execution of a survey must be scrutinized whether or not the survey was explicitly designed to provide relevant data on the issue before the court.<sup>31</sup>

28. See, e.g., *Morrison Ent. Grp. v. Nintendo of Am.*, 56 F. App'x 782, 785 (9th Cir. 2003); *Monster, Inc. v. Dolby Lab's Licensing Corp.*, 920 F. Supp. 2d 1066, 1072 (N.D. Cal. 2013); *Heaven Hill Distilleries, Inc. v. Log Still Distilling, LLC*, No. 3:21-cv-190-BJB-CHL, 2021 U.S. Dist. LEXIS 240373 (W.D. Ky. Dec. 16, 2021) (survey is the “most persuasive evidence” of consumer recognition).

29. *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, 357–58 (7th Cir. 1983); *Medici Classics Prods. LLC v. Medici Grp. LLC*, 590 F. Supp. 2d 548, 556 (S.D.N.Y. 2008); *Citigroup v. City Holding Co.*, No. 99 Civ. 10115 (RWS), 2003 U.S. Dist. LEXIS 1845 (S.D.N.Y. Feb. 10, 2003); *Chum Ltd. v. Lisowski*, 198 F. Supp. 2d 530 (S.D.N.Y. 2002); *Cairns v. Franklin Mint Co.*, 24 F. Supp. 2d 1013, 1041 (C.D. Cal. 1998) (“[A] plaintiff's failure to conduct a survey, assuming it has the financial resources to do so, may lead to an inference that the results of such a survey would be unfavorable.”).

30. 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).

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

## Role of Attorneys in Survey Design and Administration

An early handbook for judges recommended that survey interviews be “conducted independently of the attorneys in the case.”<sup>32</sup> Some courts interpreted this to mean that any evidence of attorney participation is objectionable.<sup>33</sup> A better interpretation is that the attorney should not take part in survey implementation or interact with survey participants.<sup>34</sup> However, some attorney involvement in the survey design is often necessary to ensure that relevant questions are directed to a relevant population,<sup>35</sup> particularly if the survey expert is not an expert in the substantive area of law under dispute. Federal Rule of Civil Procedure 26(4)(b) does not allow an inquiry into the nature of communications between attorneys and experts, and so the role of attorneys in constructing surveys may not always be fully apparent. The key issues for the trier of fact are the design of the survey, the objectivity and relevance of the questions on the survey, the appropriateness of the population used to guide sample selection, and the method of 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 survey administration itself, whether online or in an interview, may not be directly visible. 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 administering questionnaires, conducting interviews, tabulating results, and interpreting the data.<sup>36</sup>

32. Judicial Conference of the United States, Handbook of Recommended Procedures for the Trial of Protracted Cases, 25 F.R.D. 351, 429 (1960).

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

34. Upjohn Co. v. Am. Home Prods. Corp., No. 1-95-CV-237, 1996 U.S. Dist. LEXIS 8049, at \*42 (W.D. Mich. Apr. 5, 1996) (objection that “counsel reviewed the design of the survey carries little force with this Court because [opposing party] has not identified any flaw in the survey that might be attributed to counsel's assistance”). For cases in which attorney participation was linked to significant flaws in the survey design or execution, see *Hurt v. Commerce Energy, Inc.*, No. 1:12-CV-00758, 2015 U.S. Dist. LEXIS 10566 (N.D. Ohio Jan. 29, 2015); *Johnson v. Big Lots Stores, Inc.*, No. 04-321, 2008 U.S. Dist. LEXIS 35316, at \*52–53 (E.D. La. Apr. 29, 2008); *United States v. Southern Indiana Gas & Electric Co.*, 258 F. Supp. 2d 884, 894 (S.D. Ind. 2003); and *Gibson v. County of Riverside*, 181 F. Supp. 2d 1057, 1069 (C.D. Cal. 2002).

35. See 6 J. Thomas McCarthy, McCarthy on Trademarks and Unfair Competition § 32:166 (5th ed. 2021). See also Jerre B. Swann, *A History of the Evolution of Likelihood of Confusion Methodologies*, 113 Trademark Rep. 723 (2023) (generally on the importance of context in determining survey design methodology).

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

## Skill and Experience of the Experts Who Designed, Conducted, or Analyzed the Survey

Experts prepared to design, conduct, and analyze a survey generally should have graduate training in psychology (especially social, cognitive, or consumer psychology), sociology, political science, marketing, communication sciences, statistics, or a related discipline; that training should include courses in survey research methods, sampling, measurement, interviewing, and statistics. In some cases, professional experience in teaching or conducting and publishing survey research may provide the requisite background. In all cases, the expert must demonstrate an understanding of best practices in survey methodology, including sampling,<sup>37</sup> instrument design (questionnaire and interview construction), and statistical analysis.<sup>38</sup> Publication in peer-reviewed journals, authored books, fellowship status in professional organizations, faculty appointments, consulting experience, research grants, and membership on scientific advisory panels for government agencies or private foundations are indications of a professional's area and level of expertise. In addition, some surveys involving highly technical subject matter or specific (sub)populations may require experts to have some further specialized knowledge. Under these conditions, the survey expert also should be able to demonstrate sufficient familiarity with the topic and 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.

## Skill and Experience of the Experts Who Will Testify About Surveys Conducted by Others

Parties often call on an expert to testify about a survey conducted by someone else (e.g., by one of the parties to the suit, or by another entity when the survey was not conducted specifically for the case). The secondary expert's role may be 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

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

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

expertise.<sup>39</sup> All experts who give opinions about the adequacy and interpretation of a survey not only should have general skills and experience with surveys and be familiar with all of the issues addressed in this reference guide, but also should demonstrate familiarity with the following properties of the survey being discussed:

1. Purpose of the survey;
2. Survey methodology,<sup>40</sup> including
  - a. the target population,
  - b. the sampling design used in conducting the survey,
  - c. the survey instrument (questionnaire or interview schedule), and
  - d. (for interview surveys) interviewer training and instruction;
3. Results, including rates and patterns of missing data; and
4. Statistical analyses used to interpret the results.

## Population Definition and Sampling

### The Survey Universe or Population

One of the first steps in designing or evaluating a survey is to identify the target population (or universe).<sup>41</sup> The target population consists of all elements (e.g., individuals) whose characteristics or perceptions the survey is intended to represent. Thus, in trademark litigation, the relevant population in some disputes may include all prospective and past purchasers of the pertinent party's category of

39. For a discussion of the admissibility of expert opinion testimony, see Liesa L. Richter and Daniel J. Capra, *The Admissibility of Expert Testimony*, “Federal Rule of Evidence 702: An Overview” section, in this manual.

40. *See A & M Records, Inc. v. Napster, Inc.*, No. C 99-05183 MHP, 2000 U.S. Dist. LEXIS 20668, at \*22-25 (N.D. Cal. Aug. 10, 2000) (holding that expert could not attest credibly that the surveys upon which he relied conformed to accepted survey principles because of his minimal role in overseeing the administration of the survey and limited expert report); *Hrg Tr.* at 29-30, *Munchkin Inc. v. Playtex Prods., LLC*, No. CV11-503-AHM(RZx) (C.D. Cal. May 1, 2012), ECF No. 285 (excluding survey when the expert “didn’t know how [the survey] was administered[,] . . . didn’t know how the panel was selected [and] didn’t know what the statistical technique was that was used to weigh the survey and provide weight to it”), *cited in Cohen v. Trump*, No. 3:13-cv-2519-GPC-WVG, 2016 U.S. Dist. LEXIS 117059, at \*16-17 (S.D. Cal. Aug. 29, 2016).

41. Identification of the proper target population or universe is recognized uniformly as a key element in the development of a survey. *See, e.g.*, Manual for Complex Litigation, Fourth, § 11.493 (2004), <https://www.uscourts.gov/file/3228/download> [hereinafter MCL 4th]; *see also* 3 McCarthy, *supra* note 35, § 32:166; Council of Am. Survey Rsch. Orgs., Code of Standards and Ethics for Survey Research § III.A.3 (2011), <https://perma.cc/698Z-CF86> [hereinafter CASRO]. (Note that CASRO merged with the Marketing Research Association to form the Insights Association in 2017.)

goods or services. Similarly, the population for a discovery survey may include all potential plaintiffs or all employees who worked for Company A between two specific dates. In a community survey designed to provide evidence for a motion for a change of venue, the relevant population consists of all jury-eligible citizens in the community in which the trial is to take place.<sup>42</sup> 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.<sup>43</sup> For example, in a survey testing whether the defendant made misleading representations about their digital evidence-related software used by police departments, the appropriate universe consisted of potential purchasers of the software in the law enforcement community. Instead, the sample consisted of respondents who merely used photographs in their law enforcement work and had no influence on purchase decisions involving the relevant software or were even necessarily potential users of such software.<sup>44</sup> Defects in the universe may lead to misleading results that should reduce the weight that is given to the survey;<sup>45</sup> a survey should be excluded if it consists of respondents who do not substantially reflect the characteristics of the relevant population.<sup>46</sup>

42. An additional 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 more than one community, 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. Cloud*, No. 1:19-cr-02032-SMJ-1, 2021 U.S. Dist. LEXIS 260839, at \*8–9 (E.D. Wash. Dec. 8, 2021) (order granting a motion for intra-district transfer from Yakima based in part on a telephone survey of jury-eligible respondents in Yakima, Richland, and Spokane); *United States v. Haldeman*, 559 F.2d 31, 140, 151 (MacKinnon, J., dissenting), 176–79 (app. A) (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); *see also People v. Venegas*, 31 Cal. Rptr. 2d 114, 117 (Cal. Ct. App. 1994) (change of venue denied because defendant failed to show that the defendant would face a less hostile jury in a different court).

43. *See Merck Eprova AG v. Brookstone Pharmas.*, 920 F. Supp. 2d 404, 418 (S.D.N.Y. 2013) (“In determining whether a challenged advertisement is likely to confuse or mislead customers, courts must look to the person to whom the advertisement is addressed.”) (citation omitted).

44. *See, e.g., Kwan Software Eng’g, Inc. v. Foray Techs., LLC*, 110 U.S.P.Q.2d 1637, 1641–42 (N.D. Cal. 2014). *See also Warner Bros., Inc. v. Gay Toys, Inc.*, 658 F.2d 76 (2d Cir. 1981) (surveying child users of the product rather than parent purchasers). Children and some other populations create special challenges for researchers. For example, very young children should not be asked about sponsorship or licensing, concepts that are foreign to them. Concepts, as well as wording, should be age appropriate.

45. *See, e.g., Chi. Mercantile Exchange Inc. v. ICE Clear US, Inc.*, No. 18 C 1376, 2020 WL 1905760, at \*12–14 (N.D. Ill. Apr. 12, 2020).

46. *See, e.g., Kwan*, 110 U.S.P.Q.2d at 1642.

## The Sampling Frame as an Approximation of the Population

The target population consists of all the individuals or units whose responses the researcher would like to describe. 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 all members of the target population. The survey expert should identify how the sampling frame was compiled—that is, the source(s) used to obtain the list of members of the population. Ideally, even if a list of every member of the population is not available, the survey expert will still have access to demographic characteristics of the full population (e.g., sex, race/ethnicity, age). In some cases, this is straightforward, such as when the population consists of all American residents (so the Census's American Community Survey can provide the demographic information). In other situations, population descriptions may be less easily obtained (e.g., the consumers of a particular product, whose demographic characteristics may be identifiable only from the marketing research of the company that produces the product). In practice, the target population often includes some members who cannot be contacted or who cannot be identified in advance. As a result, reasonable compromises are sometimes required in developing the sampling frame. The survey report should contain (1) a description of the target population, (2) a description of the sampling frame from which the sample is drawn, (3) a discussion of the difference between the two, and, importantly, (4) an evaluation of the likely consequences of that difference.

A survey that provides information about a wholly irrelevant population is itself irrelevant.<sup>47</sup> Courts are likely to exclude such a survey or accord it minimal

47. A survey aimed at assessing how persons in the trade respond to an advertisement should be conducted on a sample of persons in the trade and not on a sample of consumers. *See Home Box Off. v. Showtime/The Movie Channel*, 665 F. Supp. 1079, 1083–84 (S.D.N.Y. 1987), *aff'd in part and vacated in part*, 832 F.2d 1311 (2d Cir. 1987); *J & J Snack Food Corp. v. Earthgrains Co.*, 220 F. Supp. 2d 358, 371–72 (D.N.J. 2002); *Parks, LLC v. Tyson Foods, Inc.*, 186 F. Supp. 3d 405, 419–20 (E.D. Pa. 2016) (giving little weight to a survey aimed at consumers of the plaintiff's products when it should have been targeted at users of the defendant's). *But see Lon Tai Shing Co., LTD v. Koch & Lowy*, No. 90 Civ. 4464, 1990 U.S. Dist. LEXIS 19123, at \*50–57 (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 reasonably believed that the consuming public would be likely to rely on the salespersons' inaccurate statements about the name of the company that manufactured the lamp they were selling.

weight.<sup>48</sup> More commonly, however, the sampling frame and the target population have some overlap, but the overlap is imperfect. This is called coverage error, and it has two types: the sampling frame may be underinclusive by excluding part of the target population, or it may be overinclusive by including individuals who are not members of the target population. If the coverage is underinclusive, the survey's value depends on the extent to which the excluded population is likely to respond differently from the included population. Thus, a survey of spectators and participants at running events would be sampling a sophisticated subset of those likely to purchase running shoes. Because this subset would probably consist of the consumers most knowledgeable about the trade dress used by companies that sell running shoes, a survey based on this sampling frame would be likely to substantially overrepresent the strength of a particular design as a trademark. Worse still, the extent of that overrepresentation would be unknown and not susceptible to any reasonable estimation.<sup>49</sup>

In some cases, it is difficult to determine whether a sampling frame that omits some members of the population distorts the results of the survey and, if so, the extent and likely direction of the bias. For example, a trademark survey was designed to test the likelihood of confusing an analgesic currently on the market with a new product that was similar in appearance.<sup>50</sup> The plaintiff's survey included only respondents who had used the plaintiff's analgesic, and the court found that the target population should have included users of other analgesics, "so that the full range of potential customers for whom plaintiff and defendants would compete could be studied."<sup>51</sup> In this instance, it is unclear whether users of the plaintiff's product would be more or less likely to be confused than users of the defendants' product or users of a third analgesic, but the omission of users

48. See *Varner v. Dometic Corp.*, No. 16-22482-CIV-SCOLA/OTAZO-RYES, 2022 U.S. Dist. LEXIS 127353, at \*27–30 (S.D. Fla. Feb. 2, 2022) (survey of consumers excluded because gas-absorption refrigerators are not sold directly to consumers, but rather to manufacturers and OEM and RV dealers); *see also In re Fluidmaster, Inc., Water Connector Components Prods. Liab. Litig.*, No. 14-cv-5696, 2017 WL 1196990, at \*29 (N.D. Ill. Mar. 21, 2017); *Wells Fargo & Co. v. WhenU.com, Inc.*, 293 F. Supp. 2d 734 (E.D. Mich. 2003).

49. See *Brooks Shoe Mfg. Co. v. Suave Shoe Corp.*, 533 F. Supp. 75, 80 (S.D. Fla. 1981), *aff'd*, 716 F.2d 854 (11th Cir. 1983); *see also Hodgdon Powder Co. v. Alliant Techsystems, Inc.*, 512 F. Supp. 2d 1178, 1181–82 (D. Kan. 2007) (excluding survey on gunpowder brands distributed at plaintiff's promotional booth at a shooting tournament); *Winning Ways, Inc. v. Holloway Sports-wear, 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).

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

51. *Am. Home Prods.*, 656 F. Supp. at 1070.

of other analgesics made it impossible to assess the effect of that difference.<sup>52</sup> If the sampling frame does not include important groups in the target population, the survey cannot provide information on how the unrepresented members of the target population would have responded.<sup>53</sup>

An overinclusive sampling frame generally presents less of a problem for interpretation than does an underinclusive sampling frame.<sup>54</sup> If the survey expert can demonstrate that a sufficiently large and representative subset of respondents in the survey was drawn from the appropriate sampling frame, the responses obtained from that subset can be examined, and inferences about the relevant population can be drawn based on that subset.<sup>55</sup> If the relevant subset cannot be identified, however, an overbroad sampling frame will reduce (or eliminate) the value of the survey.<sup>56</sup>

## The Sample as a Reflection of the Relevant Characteristics of the Population

Identification of a survey population must be followed by selection of a sample that accurately represents that population.<sup>57</sup> The use of probability sampling techniques maximizes both the representativeness of the survey results and the

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

53. See, e.g., *Amstar Corp. v. Domino's Pizza, Inc.*, 615 F.2d 252, 263–64 (5th Cir. 1980) (court found both plaintiff's and defendant's surveys substantially defective for a systematic failure to include parts of the relevant population); *Scott Fetzer Co. v. House of Vacuums, Inc.*, 381 F.3d 477, 487–88 (5th Cir. 2004) (universe drawn from plaintiff's customer list is underinclusive and customers are likely to differ in their familiarity with plaintiff's marketing and distribution techniques); *Hi Ltd. P'Ship v. Winghouse of Fla., Inc.*, No. 6:03-cv-116-Orl-22JGG, 2004 U.S. Dist. LEXIS 30687, at \*25–26 (M.D. Fla. Oct. 5, 2004) ("[T]he failure to include a single female in the survey, when women comprise nearly a third of Hooters' customer base and perhaps even more of Winghouse's clientele, reflects a patently flawed methodology striking at the very heart of the survey's validity.").

54. See *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992, 1135 (E.D.N.Y. 2006) ("Studies evaluating broadly the beliefs of low tar smokers generally are relevant to the beliefs of 'light' smokers more specifically."); *Jacobs v. Fareportal, Inc.*, No. 8:17CV362, 2020 U.S. Dist. LEXIS 211840, at \*25–26 (D. Neb. May 29, 2020) (critique about the overinclusiveness of sampling past as well as future purchasers goes to weight rather than admissibility).

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

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

57. MCL 4th, *supra* note 41, § 11.493.

ability to assess the precision of estimates obtained from the survey. Probability samples range from simple random samples to complex multistage sampling designs that use stratification, clustering of population elements into various groupings, or both. In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample.<sup>58</sup> These sample selection probabilities do not need to be the same for all population elements (i.e., some members may have a higher or lower chance of being selected than others). If the probabilities are unequal, however, compensatory adjustments should be made in the analysis. Failure to adjust by weighting to reflect the known distribution in the population on relevant characteristics may undermine representativeness and warrant exclusion of the survey.<sup>59</sup>

Probability sampling offers two important advantages over nonprobability sampling. First, a probability sample can provide an unbiased estimate that summarizes the responses of all persons in the population from which the sample was drawn; that is, the expected value of the sample estimate is the population value being estimated. For instance, if a probability sample leads to an estimate that 80% of respondents were confused as to whether an analgesic was an existing or new option, the researcher could be confident that approximately 80% of those in the population would have that belief (within some margin of error). A probability sample can provide an unbiased estimate of the population even when the size of the sample is relatively small. The advantage of larger samples is that they provide more precision (smaller margins of error). In general, the absolute size of the sample, rather than its size relative to the population, determines the precision of the estimate.<sup>60</sup>

Second, probability sampling allows the researcher to calculate a confidence interval that explicitly provides information on the reliability of the sample estimate of the population value (i.e., the value one would obtain if everyone in the population responded). The difference between the estimate and the exact value

58. See Thomas Piazza, *Fundamentals of Applied Sampling*, in *Handbook of Survey Research*, *supra* note 1, at 139, 145.

59. *Fish v. Kobach*, 309 F. Supp. 3d 1048, 1059–61 (D. Kan. 2018), *aff'd sub nom. Fish v. Schwab*, No. 18–3133, 2020 U.S. App. LEXIS 13723 (10th Cir. Apr. 29, 2020) (expert's survey intended to represent the eligible voting population of Kansas was excluded, in part, for failure to weight responses to reflect the distribution of educational attainment and household income level in the population).

60. An exception to this rule applies when a population is (identifiably) finite and the sample size is large relative to the population size, typically taken to occur when the sample size is greater than 5% of the total population (e.g., the population strictly includes 1,000 people, and the sample includes more than 50 people). In this case, it is appropriate to adjust the standard error/margin of error by the finite population correction (FPC) factor. The FPC factor incorporates both the size of the sample and its share of the population in determining the degree of precision. See William G. Cochran, *Sampling Techniques* 24–25 (3d ed. 1977).

is called the sampling error.<sup>61</sup> Thus, suppose a survey collected responses from a simple random sample of 400 dentists selected from the population of all dentists licensed to practice in the United States and found that 20% of them mistakenly believed that a new toothpaste, Goldgate, was manufactured by the makers of Colgate. A survey expert could properly compute a confidence interval around the 20% estimate obtained from this sample. If the survey were repeated a large number of times, and a 95% confidence interval was computed each time, 95% of the confidence intervals would include the actual percentage of dentists in the entire population who would believe that Goldgate was manufactured by the makers of Colgate.<sup>62</sup> In this example, the margin of error is  $\pm 4\%$ , and so the confidence interval is the range between 16% and 24%—that is, the estimate (20%) plus or minus 4%.

All sample surveys (i.e., surveys that do not measure responses from every member of the population) produce estimates of population values, not exact measures of those values. Assuming a probability sample, a confidence interval describes how stable the mean response in the sample is likely to be. The width of the confidence interval depends on three primary characteristics:

1. Size of the sample (larger samples produce narrower intervals);
2. Variability of the response being measured (more variability leads to wider intervals); and
3. Confidence level the researcher wants to have.<sup>63</sup>

Traditionally, scientists adopt the 95% level of confidence, which means that if one hundred samples of the same size were drawn, the confidence interval expected for ninety-five of the samples would be expected to include the true population value.<sup>64</sup>

Stratified probability sampling uses what is known about the population characteristics to correct for imbalances produced by random sampling. Consider the dentist example presented earlier: if it is known that 60% of dentists have at least

61. See the glossary of this chapter, and David H. Kaye & Hal S. Stern, *Reference Guide on Statistics and Research Methods*, in this manual, for a more detailed definition of sampling error.

62. Actually, because survey interviewers would be unable to locate some dentists, and some dentists would be unwilling to participate in the survey, technically the population to which this sample would be projectable would be all dentists with current addresses who would be willing to participate in the survey if they were asked. The expert should be prepared to discuss possible sources of bias due to, for example, an address list that is not current.

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

64. To increase the likelihood that the confidence interval contains the actual population value (e.g., from 95% to 99%) without increasing the sample size, the width of the confidence interval can be expanded. An increase in the confidence interval brings an increase in the confidence level. For further discussion of confidence intervals, see the glossary to the current chapter, and David H. Kaye & Hal S. Stern, *Reference Guide on Statistics and Research Methods*, in this manual.

fifteen years of experience and 40% have fewer years, the researcher could randomly sample dentists within each of those separate categories to obtain a sample that reflected those proportions. This would ensure the sample was unbiased with respect to years of experience—that proportion would be set by the researcher—and would therefore reduce sampling error.<sup>65</sup> Disproportionate sampling from subgroups may be used to enable the survey to provide separate estimates for particular subgroups but should be weighted (as discussed below) to reflect the population as a whole when conducting an overall analysis.

The last decade has seen a dramatic rise in the availability of nonprobability samples. Most of these samples come in one of three general varieties. The first variety is drawn from large nonprobability internet panel samples overseen by a vendor. For these panels, individuals opt in (e.g., via advertisements) to receive compensation for taking surveys. The companies that oversee these panels can produce a set of respondents that consists of members of the target population (e.g., the population of elementary school teachers) and matches the distribution of that population on specified characteristics (e.g., percentage of women and minorities). These samples vary in their ability to hit benchmarks, but some perform quite well.<sup>66</sup> Regardless, these samples differ from probability samples because they are not drawn from a list of the entire population. Instead, there is an attempt to identify relevant respondents and to balance the sample to resemble the population on certain observable features (e.g., the sample and population have the same distribution of gender, age, income).

The second source of nonprobability samples is crowdsourcing labor market platforms, the best-known of which is Amazon's Mechanical Turk (MTurk). Researchers can directly use MTurk or analogous platforms to hire individuals to complete tasks, including taking surveys, for direct compensation. Here, researchers can try to draw samples that match populations on observable variables, but it is often difficult since these platforms do not easily allow for the kind of selective participant invitations that permit managed panels to build custom samples.

The third source of nonprobability samples are purposive samples. These use nonprobability methods designed to target hard-to-reach populations (for whom probability sampling is extremely difficult or impossible), such as low-income people, young people, Indigenous people, and people in poor health.<sup>67</sup> These sampling methods are sometimes the subject of judicial concerns about researcher bias in selecting respondents, as these methods require the researcher to engage

65. See *Pharmacia Corp. v. Alcon Lab'y's, Inc.*, 201 F. Supp. 2d 335, 365 (D.N.J. 2002).

66. Lynn Vavreck & Douglas Rivers, *The 2006 Cooperative Congressional Election Study*, 18 J. Elections, Pub. Op., & Parties, 355–66 (2008), <https://doi.org/10.1080/17457280802305177>.

67. Hard-to-Survey Populations (Roger Tourangeau et al. eds., 2014). Abdolreza Shaghaghi et al., *Approaches to Recruiting 'Hard-To-Reach' Populations into Research: A Review of the Literature*, 1 Health Promotion Persps. 86, 94 (2011), <https://doi.org/10.5681/hpp.2011.009>.

in active and targeted outreach.<sup>68</sup> They also require particular attention to language and survey mode.<sup>69</sup>

Generally speaking, researchers employ nonprobability samples because they are substantially less expensive, often allow for larger samples (due to the lower cost), may include a higher number of individuals from less prevalent subgroups (e.g., members of racial or ethnic minority groups) that could be of interest, and/or offer the only option for hard-to-reach populations. Indeed, the use of nonprobability samples has been common in litigation for some time. For instance, an overwhelming majority of the consumer surveys conducted for Lanham Act litigation present results from nonprobability samples,<sup>70</sup> and the standard modern practice is for these surveys to be conducted online using professionally managed nonprobability internet panels (i.e., the first type discussed above).<sup>71</sup> A common rationale for admitting such surveys into evidence is that they are used widely in marketing research and that “results of these studies are used by major American companies in making decisions of considerable consequence.”<sup>72</sup>

Nonprobability samples are appropriate for some types of litigation surveys but still carry important limitations. In many cases, researchers cannot compute response rates for nonprobability samples because they are unaware of how many potential respondents viewed the invitation to participate in the survey and decided not to do so (e.g., researchers often just post participation invitations to those in a labor market without knowing how many people viewed it). In contrast, when trying to assess the base rate of some variable in the national population, probability samples offer superior accuracy relative to nonprobability samples

68. See, e.g., *Branson v. All. Coal, LLC*, No. 4:19-CV-00155-JHM-HBB, 2022 U.S. Dist. LEXIS 123731, at \*23–24 (W.D. Ky. July 13, 2022) (rejecting the use of purposive sampling in a class certification survey: “Put simply, the inclusion of purposefully selected data questions whether the data accurately represents the population as a whole. To preserve the reliability of the discovery data, no party shall select certain opt-in plaintiffs from whom to collect information or depose.”).

69. Robin Bayes et al., *Studying Science Inequities: How to Use Surveys to Study Diverse Populations*, 700 Annals Am. Acad. Pol. & Soc. Sci., 220, 233 (2022), <https://doi.org/10.1177/00027162221093970>.

70. Jacob Jacoby & Amy H. Handlin, *Non-Probability Sampling Designs for Litigation Surveys*, 81 Trademark Rep. 169, 173 (1991). For probability surveys conducted in trademark cases, see *James Burrough, Ltd. v. Sign of Beefeater, Inc.*, 540 F.2d 266 (7th Cir. 1976); *Nightlight Systems v. Nitelites Franchise Systems*, No. 1:04-CV-2112-CAP, 2007 U.S. Dist. LEXIS 95565 (N.D. Ga. July 17, 2007); *National Football League Properties, Inc. v. Wichita Falls Sportswear, Inc.*, 532 F. Supp. 651 (W.D. Wash. 1982).

71. Matthew B. Kugler & R. Charles Henn, *Internet Surveys in Trademark Cases: Benefits, Challenges, and Solutions*, in *Trademark and Deceptive Advertising Surveys* 291, 293 (Shari Diamond & Jerre Swann eds., 2d ed. 2022).

72. *Nat'l Football League Props., Inc. v. N.J. 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 70, at 172–73, 176.

on validated demographic measures such as sex, age, race, and ethnicity,<sup>73</sup> as well as validated behavioral measures such as vaccine uptake.<sup>74</sup> This is because probability samples are more representative of the underlying population. In evaluating the representativeness of a nonprobability sample relative to the population, the expert should consider both observable and nonobservable sample demographics. On critical variables, it may be necessary to include the joint distributions of the measured variables (e.g., the percentage of minority women).<sup>75</sup> More generally, there is always the possibility of unobserved relevant factors even in a well-conducted nonprobability survey.<sup>76</sup> For instance, a nonprobability sample with appropriate quotas may still not include typical members from each quota group. This can lead to inaccuracies in point estimates. Nonprobability samples can provide unbiased estimates if the sample is representative on all variables that correlate with the outcome of interest, or has been weighted to be representative, but this can be challenging when it comes to unobserved factors.<sup>77</sup>

Ultimately, the importance of using a probability sample depends on the survey's goal. If the survey is designed to make a causal inference based on differences between randomly assigned experimental conditions (40% of people in

73. Bo MacInnis et al., *The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension*, 82 Pub. Op. Q. 707 (2018), <https://doi.org/10.1093/poq/nfy038>. The authors compare the accuracy of probability surveys (RDD telephone and internet), internet surveys that combine nonprobability and probability samples, and internet surveys that are fully nonprobability (but with different types of compensation). This study is the most extensive of its kind, using a set of fifty measures and forty benchmark variables from federal face-to-face surveys with high response rates. The analyses showed that probability samples provide more accurate estimates of population quantities than nonprobability samples as well as samples that combined methods.

74. Valerie C. Bradley et al., *Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake*, 600 Nature 695 (2021), <https://doi.org/10.1038/s41586-021-04198-4>.

75. Connor Huff & Dustin Tingley, *"Who Are These People?" Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents*, 2 Rsch. & Politics 1 (2015), <https://doi.org/10.1177/2053168015604648>.

76. The court in *Kinetic Concepts, Inc. v. BlueSky Medical Corp.*, No. SA-03-CA-0832, 2006 U.S. Dist. LEXIS 60187, at \*14–17 (W.D. Tex. Aug. 11, 2006), found the plaintiff's survey using a nonprobability sample to be admissible and permitted the plaintiff's expert to present results from a survey using a convenience sample. The court then assisted the jury by providing an instruction on the differences between probability and convenience samples and the estimates obtained from each.

77. See, e.g., Robert M. Groves et al., *Survey Methodology* (2d ed. 2009). Groves and colleagues point out that using statistical significance tests and confidence intervals with nonprobability samples is technically inappropriate since it becomes impossible to obtain an unbiased estimate of sampling variance. That said, they accept that nonprobability samples can approximate probability samples when the researcher knows what population attributes correlate with the key statistics and ensures that the sample is balanced on those attributes (i.e., using quotas). *Id.* at 409–10. Vasia Vehovar and colleagues agree: "We thus recommend to be more openly accepting of the reality of using a standard statistical inference approach as an approximation in non-probability settings" as long as the nature of how the sample is drawn is clear. Vasia Vehovar et al., *Non-Probability Sampling*, in *The SAGE Handbook of Survey Methodology* (2016), <https://doi.org/10.4135/9781473957893>.

the experimental condition were confused by defendant's advertising, but only 20% by control advertising), a survey-experiment using a nonprobability sample of relevant respondents will generally suffice. If the aim is to obtain a precise measure of a particular feature of a population, however, a nonprobability sample faces challenges.<sup>78</sup> An expert should consider the potential biases in the sample and whether they impact factors correlated with the variable being estimated, to assess the likely magnitude of the biases (i.e., how much they are likely to affect the estimates). The expert should be prepared to explain how observable and unobservable sample characteristics might impact their estimate.

When using a nonprobability sample in an experimental context (i.e., a survey-experiment), the primary concern about the characteristics of the sample is whether they modify the causal relationship being tested in the survey.<sup>79</sup> For instance, if more and less experienced dentists react the same way to the name Goldgate for toothpaste—if experience does not moderate (that is, change) the effect of the name—a survey-experiment that does not include more experienced dentists will still produce results that generalize across experience level. In contrast, if reaction to the name does vary with experience—if experience *does* moderate the reaction—a sample that includes only less experienced dentists will produce a biased result. The survey expert should identify any potential sample characteristics that may moderate the effects being assessed and demonstrate that the sampling approach takes them into account.

In sum, the historic gold standard was a probability sample because it could provide unbiased estimates of the population. Yet, increased costs of probability samples, more ways to obtain nonprobability samples, challenges with hard-to-reach populations, and nonresponse bias (discussed below) have all led to greater use of nonprobability samples. When such samples are used, it is vital to clarify how the sample compares to known-probability benchmarks (e.g., demographic

78. However, even if the initial sampling approach uses a probability sample, systematic nonresponse can undermine the representativeness of the sample and hence the accuracy of a precise estimate. This effect has been demonstrated by studies of the effect of a two-stage process to recruit survey participants in which the initial invitation used probability sampling through U.S. Postal Service mailings, phone contact, and modest incentives, and a follow-up effort involving FedEx mailings, in-person recruitment by field interviewers, and enhanced incentives attempted to recruit initial nonresponders. Comparisons of the two groups indicated small but statistically significant differences in reported political views and income level. The importance of such differences will depend on the claim being made for the accuracy of the estimate. Ipek Bilgen et al., *The Undercounted: Measuring the Impact of 'Nonresponse Follow-up' on Research Data* (2019), <https://perma.cc/P5RS-268Q>.

79. See James N. Druckman & Cindy D. Kam, *Students as Experimental Participants: A Defense of the 'Narrow Data Base'*, in Cambridge Handbook of Experimental Political Science (2011); James N. Druckman, *Experimental Thinking: A Primer on Social Science Experiments* (2022).

characteristics and other features that could influence the outcome being studied), what adjustments may have been made (e.g., weights; see below) and how biases may influence inferences—particularly when the goal is to arrive at a precise estimate of a population value. Nonprobability samples tend to be on stronger footing when used in an experimental context, but even here it is vital for the researcher to discuss potential variables that could moderate the experimental effect and whether the sampling approach influenced the impact of those moderators.

## Nonresponse as a Potential Source of Bias

Even when a sample is drawn randomly from a complete list of elements in the target population (i.e., a probability sample), responses or measures are typically obtained from only part of the sample. If this lack of response is distributed randomly, valid inferences about the population can be drawn with assurance using the measures obtained from the available sample. But nonresponse often is not random. For example, persons who are single typically have three times the “not at home” rate in U.S. Census Bureau surveys as do people living with family.<sup>80</sup> Nonresponse also occurs when contact with the sampled individual is made, but that person declines to participate. This problem arose recently in political polls where Republicans seemed less likely to participate than Democrats.<sup>81</sup> Efforts to increase response rates include making several attempts to contact potential respondents, sending advance letters,<sup>82</sup> and providing financial or nonmonetary incentives for participating in the survey.<sup>83</sup>

The key to evaluating the effect of nonresponse in a survey is to determine the extent to which nonrespondents differ from respondents in a way that would impact the results of the survey. If nonresponse has biased the pattern of responses, the size and direction of that bias need to be assessed. On some occasions, it may be possible to anticipate systematic patterns of nonresponse. For example, high-volume medical professionals may be less willing to respond to a survey than those with lower-volume practices. If the survey includes questions about volume of

80. 2 Gastwirth, *supra* note 30, at 501.

81. Courtney Kennedy et al., Pew Research Center, *Confronting 2016 and 2020 Polling Limitations* (2021), <https://perma.cc/P5CC-HJH2>; see also Joshua D. Clinton et al., *Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-election Telephone Polls*, 86 Pub. Op. Q. 247 (2022), <https://doi.org/10.1093/poq/nfac011>.

82. Edith De Leeuw et al., *The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis*, 71 Pub. Op. Q. 413 (2007), <https://doi.org/10.1093/poq/nfm014> (advance letters effective in increasing response rates in telephone as well as mail and face-to-face surveys).

83. Erica Ryu et al., *Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error*, 18 Int'l J. Pub. Op. Rsch. 89 (2006), <https://doi.org/10.1093/ijpor/edh089>.

practice, the expert can assess how experience level may have affected the pattern of results.<sup>84</sup>

Although high response rates are desirable because they reduce the potential impact of nonresponse bias, they are increasingly difficult to achieve. Survey nonresponse rates have risen substantially in the twenty-first century, along with the costs of obtaining responses, and so the issue of nonresponse has attracted considerable attention from survey researchers.<sup>85</sup>

Researchers have developed a variety of approaches to adjust for nonresponse, including weighting obtained responses in proportion to known demographic characteristics of the target population (which we discuss below), comparing the pattern of responses from early and late responders to mail surveys or the pattern of responses from easy-to-reach and hard-to-reach responders in telephone surveys, and imputing estimated responses to nonrespondents based on known characteristics of those who have responded. All of these techniques can only approximate the response patterns that would have been obtained if nonrespondents had responded. Nonetheless, they are useful for testing the robustness of the estimates obtained from responders.

To assess the general impact of the lower response rates, researchers have compared results obtained from surveys with varying response rates.<sup>86</sup> Surprisingly comparable results have been obtained in many surveys with varying response rates, suggesting that surveys may achieve reasonable estimates even with relatively low response rates. The key is whether nonresponse is associated with systematic differences in response that cannot be adequately modeled or assessed. Generally, the representativeness of a sample (regardless of response rate) matters much more for accurate inference than the response rate.<sup>87</sup> Determining whether

84. In *People v. Williams*, 830 N.Y.S.2d 452 (N.Y. Sup. Ct. 2006), a published survey of experts in eyewitness research was used to show general acceptance of various eyewitness phenomena. See Saul Kassin et al., *On the “General Acceptance” of Eyewitness Testimony Research: A New Survey of the Experts*, 56 Am. Psychologist 405 (2001), <https://doi.org/10.1037/0003-066X.56.5.405>. The survey included questions on the publication activity of respondents and compared the responses of those with high and low research productivity. Productivity levels in the respondent sample suggested that respondents constituted a blue-ribbon group of leading researchers. *Williams*, 830 N.Y.S.2d at 457, 459 n.26. See also *Pharmacia Corp. v. Alcon Lab’ys, Inc.*, 201 F. Supp. 2d 335 (D.N.J. 2002).

85. E.g., Richard Curtin et al., *Changes in Telephone Survey Nonresponse over the Past Quarter Century*, 69 Pub. Op. Q. 87 (2005), <https://doi.org/10.1093/poq/nfi002>. See also Robert M. Groves & Emilia Peytcheva, *The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis*, 72 Pub. Op. Q. 167 (2008), <https://doi.org/10.1093/poq/nfn011>; Peter Miller et al., *Federal Committee on Statistical Methodology, A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys* (2020), <https://perma.cc/6WCQ-AFNK>.

86. E.g., Daniel M. Merkle & Murray Edelman, *Nonresponse in Exit Polls: A Comprehensive Analysis*, in *Survey Nonresponse* 243–57 (Robert M. Groves et al. eds., 2002) (finding minimal nonresponse error associated with refusals to participate in in-person exit polls); see also Jon A. Krosnick, *Survey Research*, 50 Ann. Rev. Psych. 537 (1999), <https://doi.org/10.1146/annurev.psych.50.1.537>.

87. Bo MacInnis et al., *The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension*, 82 Pub. Op. Q. 707 (2018), <https://doi.org/10.1093/poq/nfy038>;

the level of nonresponse in a survey seriously impairs inferences drawn from the results 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. The survey expert should be prepared to provide evidence on the potential impact of nonresponse on the survey results.

In surveys that include sensitive or difficult questions, some respondents may refuse to provide answers or may provide incomplete answers.<sup>88</sup> 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 a respondent's other answers to predict the missing response (i.e., imputation).<sup>89</sup>

Survey researchers commonly use weights, even with many probability samples, to address coverage gaps and control for differential nonresponse among subgroups. Weighting adjustments can sometimes improve the accuracy of survey results.<sup>90</sup> Responses are weighted to reflect distributions in the target population, typically demographics.<sup>91</sup> When the population includes all Americans, the U.S. Census or American Community Survey can provide demographic population figures that can guide how to weight the survey responses. For instance, if the population consists of 50% men but the sample contains only 40% men, then male sample respondents will be weighted to count more in computations from the sample (and women will be counted less) to better approximate the target

Scott Keeter, *Evidence About the Accuracy of Surveys in the Face of Declining Response Rates*, in The Palgrave Handbook of Survey Research 19–22 (David L. Vannette & Jon A. Krosnick eds., 2018), [https://doi.org/10.1007/978-3-319-54395-6\\_4](https://doi.org/10.1007/978-3-319-54395-6_4).

88. See Roger Tourangeau et al., *The Psychology of Survey Response* (2000); California v. Ross, 358 F. Supp. 3d 965 (N.D. Cal. 2019) (surveys were used to show that if a question on citizenship was included in the 2020 census, some individuals would be less likely to participate or answer the citizenship question, and that this refusal was likely to be higher in the Latinx community).

89. See Paul D. Allison, *Missing Data*, in *Handbook of Survey Research*, *supra* note 1, at 630; see also Groves & Peytcheva, *supra* note 85.

90. Heidi Jensen et al., *The Impact of Non-Response Weighting in Health Surveys for Estimates on Primary Health Care Utilization*, 32 Eur. J. Public Health 450 (2022), <https://doi.org/10.1093/eurpub/ckac032>. But see Kenneth Bollen et al., *Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis*, 3 Ann. Rev. Stat. Application 375 (2016), <https://doi.org/10.1146/annurev-statistics-011516-012958> (suggesting that weighting is sometimes unhelpful and inefficient).

91. Jelke Bethlehem & Mario Callegaro, *Part IV Weighting Adjustments*, in *Online Panel Research: A Data Quality Perspective* 264–72 (Mario Callegaro et al. eds., 2014). See also Fish v. Kobach, 309 F. Supp. 3d 1048, 1089 (D. Kan. 2018) (criticizing an expert for not assigning weights to their results given the differences between their sample's demographics and the expected population demographics); California v. Ross, 358 F. Supp. 3d 965, 984 (N.D. Cal. 2019) (commenting that the expert used weighting to balance the demographics of the sample).

population.<sup>92</sup> When responses are weighted, it is crucial to be transparent about how the data were weighted, and to acknowledge that weighting reduces statistical power and, hence, the precision of the estimates.<sup>93</sup>

There are three main challenges with survey weights. First, the survey expert needs to know the true proportions in the population (how many are men, aged 18–35, college educated, etc.), or the sample weights will be largely guesses. Weighting cannot compensate for lack of surveyor knowledge about the underlying distributions in the population. Second, weighting does not actually increase sample size. If a sample of forty African American respondents is weighted at 1.5 (meaning each counts as one and a half people for purposes of analysis), the size of their subsample (and consequently the number used to calculate the margin of error of that subsample) is still forty, not sixty. In fact, weighting decreases the precision of estimates, leading to larger confidence intervals. Finally, weighting cannot correct for unobserved factors. If the survey has recruited unrepresentative members of a subpopulation (for example, atypical Republicans), weighting cannot somehow make them representative. This may be a particular problem when weighting is used to substantially amplify the responses of a small number of people.<sup>94</sup>

## Precautions Taken to Ensure That Only Qualified Respondents Were Included in the Survey

In a carefully executed survey, each potential respondent is questioned on the attributes that determine their eligibility to participate in the survey. Thus, the initial questions screen potential respondents to determine if they are members of the target population of the survey (e.g., Does she own a dog? Does he live within ten miles of where he works?). The screening questions must be drafted so that they do not appeal to or deter specific groups within the target population or convey information that will influence a respondent's answers on the main survey (i.e., create a context effect). For example, if respondents must be prospective or recent purchasers of Sunshine orange juice in a trademark survey

92. The process often involves iteratively adjusting for multiple demographics. The idea is to assign an adjustment weight to each respondent such that those from underrepresented groups receive weights greater than 1 and those from overrepresented groups receive weights less than 1. There are a host of ways to weight, and caps should be placed on how much a given observation can be weighted so that one respondent does not have a grossly disproportionate effect on the outcomes. See Matthew DeBell, *Computation of Survey Weights*, in *The Palgrave Handbook of Survey Research* 519–27 (2018).

93. Thomas Piazza, *Fundamentals of Applied Sampling*, in *Handbook of Survey Research*, *supra* note 1. See also *Texas v. Holder*, 888 F. Supp. 2d 113, 131 (D.D.C. 2012) (criticizing an expert for inconsistent use of weighting).

94. Nate Cohn, *How One 19-Year-Old Illinois Man Is Distorting National Polling Averages*, N.Y. Time, Oct. 12, 2016, <https://perma.cc/76R7-J363>.

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.*,<sup>95</sup> physicians were asked a screening question to determine whether they prescribed particular drugs. The survey question that followed the screener asked, “Thinking of the practice of cardiovascular medicine, what first comes to mind when you hear the letters XL?” The court found that the screener conditioned the physicians to respond with the name of a product (a drug) rather than a function (long acting).<sup>96</sup>

The criteria for determining whether to include a potential respondent in the survey should be objective and clearly conveyed, preferably using written instructions addressed to those who administer the screening questions. These instructions and the completed screening questionnaire should be made available to the court and the opposing party along with the interview or survey form for each respondent. Computerized administration, described below, has allowed for this to be automated.

## **Survey Questions and Structure**

### **Clarity, Precision, and Lack of Bias in the Framing of the Survey Questions**

Although it seems obvious that questions on a survey should be clear and precise, phrasing questions to reach that goal is often difficult. Even questions that appear clear can convey unexpected meanings and ambiguities to potential respondents. For example, the question “What is the average number of days each week you have butter?” appears to be straightforward. Yet some respondents wondered whether margarine counted as butter, and when the question was revised to include the introductory phrase “not including margarine,” the reported frequency of butter use dropped dramatically.<sup>97</sup>

95. 858 F. Supp. 1305, 1321 & n.13 (S.D.N.Y. 1994).

96. *Id.* at 1321.

97. Floyd J. Fowler, Jr., *How Unclear Terms Affect Survey Data*, 56 Pub. Op. Q. 218, 225–26 (1992), <https://doi.org/10.1086/269312>.

When unclear questions are included in a survey, they may threaten the validity of the survey by systematically distorting responses if respondents are misled in a particular direction, or by inflating random error if respondents guess because they do not understand the question.<sup>98</sup> 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.<sup>99</sup> 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.<sup>100</sup> The court noted that the phrase “compensatory damages,” although well-defined for attorneys, was unlikely to be meaningful for laypersons.<sup>101</sup>

Pilot work is a standard and valuable way to improve the quality of a survey.<sup>102</sup> When there is any doubt whether a term or phrase will be clear to respondents, researchers should pretest to assess understanding, which may include focus groups or cognitive interviewing.<sup>103</sup>

The value of pilot work is greatest when the issues and questions in the survey differ from the issues and questions addressed in previous surveys, or when the target population differs significantly from previously surveyed populations.<sup>104</sup>

98. *See id.* at 219.

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

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

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

102. *See* Jon A. Krosnick & Stanley Presser, *Questions and Questionnaire Design*, in *Handbook of Survey Research*, *supra* note 1, at 294 (“No matter how closely a questionnaire follows recommendations based on best practices, it is likely to benefit from pretesting. . . .”). *See also* Jean M. Converse & Stanley Presser, *Survey Questions: Handcrafting the Standardized Questionnaire* 51 (1986); Fred W. Morgan, *Judicial Standards for Survey Research: An Update and Guidelines*, 54 J. Mktg. 59, 64 (1990), <https://doi.org/10.1177/002224299005400104>; OMB Standards and Guidelines for Statistical Surveys, Standard 1.4, Pretesting Survey Systems (2006), <https://perma.cc/D8XZ-4UX7> (specifying that to ensure that all components of a survey function as intended, pretests of survey components should be conducted unless those components have previously been successfully fielded); American Association for Public Opinion Research, Best Practices (2022), <https://perma.cc/Y3HU-XCHK> (“Before fielding a survey, it is important to pretest the questionnaire.”).

103. Cognitive interviewing includes a combination of think-aloud and verbal probing techniques. Gordon B. Willis et al., *Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques*, in *Cognition and Survey Research* 136 (Monroe G. Sirken et al. eds., 1999). *See also* Gordon B. Willis, *Cognitive Interviewing in Survey Design: State of the Science and Future Directions*, in *The Palgrave Handbook of Survey Research* 103–07 (2018). *See also* Tourangeau et al., *supra* note 88, at 326–27.

104. Ivan R. Ross, *The Use of Pilot Tests and Pretests in Consumer Surveys*, in *Trademark and Deceptive Advertising Surveys* 13, 26 (Shari Diamond & Jerre Swann eds., 2d ed. 2022).

In many pretests or pilot tests,<sup>105</sup> the proposed survey is administered to a small sample (usually between twenty-five and seventy-five)<sup>106</sup> of the same type of respondents who would be eligible to participate in the full-scale survey. Some courts have explicitly recognized the value of pretests<sup>107</sup> and that lack of pretesting may suggest a weakness in the survey.<sup>108</sup>

Litigants would be more likely to conduct pilot work and disclose it in expert reports if courts recognized that pilot work can maximize the likelihood that respondents understand the questions they are being asked. Moreover, the Federal Rules of Civil Procedure may require that a testifying expert disclose pilot work that serves as a basis for the expert's opinion. The situation is more complicated when a nontestifying expert conducts the pilot work and the testifying expert learns about the pilot testing only indirectly through the attorney's advice about the relevant issues in the case. Some commentators suggest that attorneys are obligated to disclose such pilot work.<sup>109</sup>

## What Respondents and Interviewers Knew About the Survey Purpose and Its Sponsorship

One way to protect the objectivity of survey administration is to avoid telling respondents or interviewers who is sponsoring the survey. Respondents who know the identity of the sponsor of the survey may adjust their responses, and interviewers who know the identity of the survey's sponsor may affect results inadvertently by communicating to respondents their expectations or what they believe are the preferred responses of the survey's sponsor. To ensure objectivity in the administration of the survey, it is standard interview practice in surveys conducted for litigation to do double-blind research whenever possible: Both the interviewer and the respondent are blind to the sponsor of the survey and its

105. 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.

106. Converse & Presser, *supra* note 102, at 69. Converse and Presser suggest that a pretest with twenty-five respondents is appropriate when the survey uses professional interviewers.

107. See, e.g., Zippo Mfg. Co. v. Rogers Imports, Inc., 216 F. Supp. 670 (S.D.N.Y. 1963); Scott v. City of New York, 591 F. Supp. 2d 554, 560 (S.D.N.Y. 2008) ("[T]he survey went through multiple pretests in order to insure its usefulness and statistical validity."); Estes Park Taffy Co. v. Original Taffy Shop, Inc., No. 15-cv-01697-CBS, 2017 U.S. Dist. LEXIS 88113, at \*9 (D. Colo. June 8, 2017).

108. GOLO, LLC. v. Goli Nutrition, Inc., No.20-667-RGA, 2020 U.S. Dist. LEXIS 158508, at \*28 (D. Del. Sept. 1, 2020) (including lack of pretesting in a list of potential weaknesses in an expert's method).

109. See Yvonne C. Schroeder, *Pretesting Survey Questions: The Procedural and Ethical Ramifications*, 11 Am. J. Trial Advoc. 195, 197–201 (1987).

purpose. Thus, the survey instrument provides no explicit or implicit clues about the sponsorship of the survey (e.g., a sponsor's letterhead) or the expected responses (e.g., reversing the usual order of the yes and no response boxes on a key question, potentially increasing the likelihood that *no* will be checked).<sup>110</sup>

Nonetheless, in some situations (e.g., on some government surveys), disclosure of the survey's sponsor to respondents (and thus to interviewers) is required. Such surveys call for an evaluation of the likely biases introduced by interviewer or respondent awareness of the survey's sponsorship. In evaluating the consequences of sponsorship awareness, it is important to consider (1) whether the sponsor has views and expectations that are apparent and (2) whether awareness is confined to the interviewers or involves the respondents. For example, if a survey concerning attitudes toward gun control is sponsored by the National Rifle Association, it is clear that responses opposing gun control are likely to be preferred. In contrast, if the survey on gun control attitudes is sponsored by the Department of Justice, the identity of the sponsor may not suggest the kinds of responses the sponsor expects or would find acceptable.<sup>111</sup> When a survey involves interviewers who are well-trained, their awareness of sponsorship may be a less serious threat than respondents' awareness.<sup>112</sup>

In most situations, the survey expert can follow the traditional CASRO Code of Standards and Ethics<sup>113</sup> and promise the respondent anonymity in the interest of obtaining the most accurate and candid responses. Moreover, in most situations, there is no need to disclose the identity of the survey sponsor. In some cases, however, respondents to a survey are involved in litigation (e.g., class-action wage and hour cases) and are likely to recognize that a survey asking them questions about relevant issues (such as unpaid overtime) is being sponsored by a party to the litigation. They may even be unwilling to participate in the survey unless they are told that the attorney representing them is sponsoring it. This creates a difficult problem for survey researchers. To get an acceptable response rate, it may be necessary to reveal who is sponsoring the survey. But this revelation may incentivize respondents to adjust their answers for potential financial gain, and promising anonymity may exacerbate this issue. If the survey is conducted by a friendly party, one approach that can be used to bolster accuracy is to tell respondents

110. See *Centaur Commc'n's, Ltd. v. A/S/M Commc'n's, 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).

111. 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).

112. See, e.g., Seymour Sudman et al., *Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses*, 6 Soc. Methods & Rsch. 171, 181 (1977), <https://doi.org/10.1177/004912417700600203>.

113. CASRO, *supra* note 41, § I.A.

that their responses may not be anonymous and thus that exaggeration or other inaccuracies may be detected. It is unclear what the best practice is in this case, and the survey expert should be prepared to justify their choices.

## Handling Respondents with No Opinions and Reducing Respondent Guessing

Some survey respondents may have no opinion about 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, some respondents in a consumer survey 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. The following three alternative question structures will affect how those respondents answer and how their responses are counted.

First, the survey can ask all respondents to answer the question (e.g., “Did you understand the guarantee offered by Clover to be a 1-year guarantee, a 60-day guarantee, or a 30-day guarantee?”). Faced with a direct question, particularly one that provides response alternatives, the respondent obligingly may supply an answer even if (in this example) the respondent did not notice the guarantee. Such answers will reflect only what the respondent can glean from the question, or they may reflect pure guessing.

Second, the survey can use a quasi-filter question to reduce guessing by providing “don’t know” or “no opinion” options as part of the question (e.g., “Did you understand the guarantee offered by Clover to be for more than a year, a year, or less than a year, or don’t you have an opinion?”).<sup>114</sup> By signaling to the respondent that it is acceptable 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 an interviewer than if it is merely accepted when the respondent spontaneously offers it as a response. Similarly, in an online survey, explicitly providing “don’t know” as a potential response rather than accepting it only if the respondent writes it in as an “other” response can increase the use of that option. The consequence of this change in format can be 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

114. 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); Spangler Candy Co. v. Tootsie Roll Indus., LLC, 372 F. Supp. 3d 588, 599 (N.D. Ohio 2019) (“Because [the expert] offered a ‘Don’t Know/No Opinion’ option for each closed-ended question, I conclude the questions were not unduly suggestive or guess-inducing.”).

know” or “no opinion” alternative commonly leads to a 20% to 25% increase in the proportion of respondents selecting that response.<sup>115</sup>

Finally, the survey can include full-filter questions—that is, questions that lay the groundwork for the substantive question by first asking respondents if they have 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?”).<sup>116</sup> The survey 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.<sup>117</sup> Respondents are more likely to say that they do not have an opinion on an issue if a full filter is used than if a quasi-filter is used.<sup>118</sup> However, in maximizing respondent expressions of “no opinion,” full filters may produce an underreporting of opinions. Some evidence indicates 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.<sup>119</sup>

In general, then, a survey that uses full filters provides a conservative estimate of the number of respondents holding an opinion, while a survey that uses neither full filters nor quasi-filters may overestimate the number of respondents with opinions if some respondents offering opinions are guessing. The strategy of including a “no opinion” or “don’t know” response as a quasi-filter avoids both of these extremes, although some research suggests that even a quasi-filter may discourage a substantive answer from a respondent who would be able to provide one.<sup>120</sup>

One solution that some survey researchers use is to provide respondents with a general instruction not to guess at the beginning of an interview, rather than supplying a “don’t know” or “no opinion” option as part of the options attached to each question.<sup>121</sup> Another approach is to eliminate the “don’t know” option and to add follow-up questions that measure the strength of the respondent’s

115. Howard Schuman & Stanley Presser, *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context* 113–46 (1996).

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

117. Considerable research has been conducted on the effects of filters. For a review, see George F. Bishop et al., *Effects of Filter Questions in Public Opinion Surveys*, 47 *Pub. Op. Q.* 528 (1983), <https://doi.org/10.1086/268810>.

118. Schwarz & Hippler, *supra* note 114, at 45–46.

119. *Id.* at 46.

120. Jon A. Krosnick et al., *The Impact of “No Opinion” Response Options on Data Quality: Non-Attitude Reduction or Invitation to Satisfice?*, 66 *Pub. Op. Q.* 371 (2002), <https://doi.org/10.1086/341394>.

121. *Anheuser-Busch, Inc. v. VIP Prods., LLC*, No. 4:08cv0358, 2008 U.S. Dist. LEXIS 82258, at \*6 (E.D. Mo. Oct. 16, 2008).

opinion.<sup>122</sup> More generally, survey experts can conduct pilot tests or cognitive interviews to assess whether many in the population in fact lack any opinion or the relevant knowledge to arrive at an opinion. If a substantial majority of people do have an opinion, it may be appropriate to exclude a “don’t know” option and thus avoid reducing data quality by encouraging less motivated respondents to use that response. In contrast, if many people do not have formed opinions or the knowledge required to arrive at them, the option should be included.<sup>123</sup> The survey expert should be prepared to explain why the “don’t know” option was included or excluded.

## Use of Open-Ended and Closed-Ended Questions

The questions that make up a survey instrument may be open-ended, closed-ended, or a combination of both. Both are valid choices in appropriate situations. Open-ended questions require the respondent to formulate and express an answer in their own words (e.g., “What was the main point of the commercial?”). 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?”<sup>124</sup>) 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.”<sup>125</sup>). When a survey involves in-person interviews, the interviewer may show the respondent these choices on a showcard that lists them.

Open-ended and closed-ended questions may elicit very different responses.<sup>126</sup> Most responses are less likely to be volunteered by respondents who are asked an

122. Krosnick & Presser, *supra* note 102, at 285.

123. Dragana Bolcic-Jankovic et al., *Using “Don’t Know” Responses in a Survey of Oncologists Regarding Medicinal Cannabis*, 14 Surv. Prac. (2021), <https://doi.org/10.29115/SP-2020-0016>; see also Erika A. Waters et al., *Dismissing “Don’t Know” Responses to Perceived Risk Survey Items Threatens the Validity of Theoretical and Empirical Behavior-Change Research*, 17 Persps. Psych. Sci. 841 (2022), <https://doi.org/10.1177/17456916211017860>.

124. President & Trs. of Colby College v. Colby College—N.H., 508 F.2d 804, 809 (1st Cir. 1975).

125. 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 likelihoods 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?”

126. Howard Schuman & Stanley Presser, *Question Wording as an Independent Variable in Survey Analysis*, 6 Socio. Methods & Rsch. 151 (1977), <https://doi.org/10.1177/004912417700600202>; Schuman & Presser, *supra* note 115, at 79–112; Converse & Presser, *supra* note 102, at 33.

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.<sup>127</sup>

The advantage of open-ended questions is that they give the respondent fewer hints about expected or preferred answers. Response choices in a closed-ended question, in addition to reminding respondents of options that they might not otherwise consider, 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.<sup>128</sup> According to a competitor, the commercial deceptively implied that each woman in the test rated more than one shampoo, when in fact each woman rated only one. To test consumer impressions, the survey showed the commercial and asked: "How many different brands mentioned in the commercial did each of the 900 women try?"<sup>129</sup> 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.<sup>130</sup>

An open-ended question may also suggest that the answer is more than one, however. 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. Instead, a nonleading version of the question would have simply asked: "How many brands mentioned in the commercial did each of the 900 women try?"

127. 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 115, 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.

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

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

130. Ninety-five percent of the respondents who answered the closed-ended question in the plaintiff's survey said that each woman had tried two or more brands. The open-ended question was never asked. *Vidal Sassoon*, 661 F.2d at 276. Norbert Schwarz, *Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Questionnaire Construction*, in *Research Methods in Personality and Social Psychology* 98 (Clyde Hendrick & Margaret S. Clark eds., 1990), suggests that respondents often rely on the range of response alternatives as a frame of reference when they are asked for frequency judgments. See, e.g., Roger Tourangeau & Tom W. Smith, *Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context*, 60 *Pub. Op. Q.* 275, 292 (1996), <https://doi.org/10.1086/297751>.

rather than “How many *different* brands” did each try? Similarly, an open-ended question that asks, “[W]hich company or store do you think puts out this shirt?” indicates to the respondent that the appropriate answer is the name of a company or store. The question would be leading if the respondent would have considered other possibilities (e.g., an individual or website if the question had not provided the frame of a company or store).<sup>131</sup> 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.

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 or more responses from among several choices, the response or responses 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 their opinion.<sup>132</sup> The omission of a relevant choice option can only be partially attenuated by telling respondents explicitly that they are not limited to the choices presented, because most respondents nevertheless will select an answer from among the listed ones.<sup>133</sup>

One form of closed-ended question format that can produce some distortion is the popular agree/disagree, true/false, or yes/no question. Although this format is appealing because it is easy to write and score these questions and their responses, the format is also seriously problematic. With its simplicity comes acquiescence: “[T]he tendency to endorse any assertion made in a question, regardless of its content,” is a systematic source of bias that has produced an inflation effect of 10% across a number of studies.<sup>134</sup> Only when control groups or control questions are added to the survey design, or when multiple items are counterbalanced, can this question format provide reasonable response estimates.<sup>135</sup>

One challenge with open-ended responses is translating them into meaningful numeric metrics.<sup>136</sup> Imagine that a researcher asks respondents what they think about self-driving cars. If the researcher were interested in the prevalence and nature of safety concerns about self-driving cars, they would need a mechanism for translating the varied open-ended responses into a set of binary values: each possible safety concern was listed or not. This requires developing a list of different types of concerns such that each concern is unique (e.g., can malfunction, inability of driver to have control, etc.), and each has a text label. The list

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

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

133. See Howard Schuman, *Ordinary Questions, Survey Questions, and Policy Questions*, 50 Pub. Op. Q. 432, 435–36 (1986), <https://doi.org/10.1093/pubopq/50.3.432>.

134. Krosnick, *supra* note 86, at 552–53.

135. See section titled “Potential Order Effects” below.

136. Groves et al., *supra* note 77, at 332–44.

should also be exhaustive (cover all written answers) and include both a code for responses that do not refer to safety issues (e.g., reflects technological evolution) and also a code for nonresponses (e.g., left blank). In some cases, a respondent may give more than one answer, and that response should be coded in more than one category (e.g., “can malfunction and has no human control”). The researcher should be transparent on the construction of the coding system, which may be based on a priori categories or constructed based on the answers the respondents have given. Moreover, unless there are substantial data privacy concerns, the full open-ended answers should be made available to the other party for analysis.<sup>137</sup>

The coding system should be clear enough that different coders would assign the same responses to the same categories in the vast majority of cases. In most academic research, this is accomplished by having multiple coders who are blind to the hypotheses and purpose of the project, giving them detailed instructions and training, and measuring inter coder reliability. Although this approach constitutes the best way to ensure reliable coding, because the raw data underlying the coding and the coding itself are made available to the court and the opposing party in litigation, the court and opposing party can directly examine and evaluate the appropriateness of the coding. It is therefore not uncommon for coding of simple measures in litigation to dispense with having multiple blind coders (though the resultant coding should be given close scrutiny).<sup>138</sup>

Although many courts prefer open-ended questions on the ground that they are likely to be less leading, the value of any open-ended or closed-ended question depends on the information it conveys in the question and, in the case of closed-ended questions, in the choices provided. Open-ended questions are more appropriate when the survey is attempting to gauge what comes first to a respondent’s mind, but closed-ended questions are more suitable for assessing choices between well-identified options or obtaining ratings on a clear set of alternatives.

137. Mary B. Vardigan & Peter Granda, *Archiving, Documentation, and Dissemination*, in *Handbook of Survey Research* 707 (Peter V. Marsden & James D. Wright eds., 2d ed. 2010); *see also* section titled “Disclosure and Reporting” *infra*. Partial redaction of a response might be appropriate if the response was individually identifiable and the content sensitive. For example, a complaint about a named supervisor in an employment survey would fall in this category.

138. *See, e.g.*, Revlon Consumer Prods. Corp. v. Jennifer Leather Broadway, Inc., 858 F. Supp. 1268, 1276 (S.D.N.Y. 1994) (inconsistent scoring and subjective coding led court to find survey so unreliable that it was entitled to no weight), *aff’d*, 57 F.3d 1062 (2d Cir. 1995); Rock v. Zimmerman, 959 F.2d 1237, 1253 n.9 (3d Cir. 1992) (court found that responses on a change-of-venue survey incorrectly categorized respondents who believed the defendant was insane as believing he was guilty); Coca-Cola Co. v. Tropicana Prods., Inc., 538 F. Supp. 1091, 1094–96 (S.D.N.Y.) (plaintiff’s expert stated that respondents’ answers to the open-ended questions revealed that 43% of respondents thought Tropicana was portrayed as fresh-squeezed; the court’s own tabulation found no more than 15% believed this was true), *rev’d on other grounds*, 690 F.2d 312 (2d Cir. 1982); *see also* Cumberland Packing Corp. v. Monsanto Co., 140 F. Supp. 2d 241 (E.D.N.Y. 2001) (court examined verbatim responses that respondents gave to arrive at a confusion level substantially lower than the level reported by the survey expert).

## Use of Probes to Clarify Ambiguous or Incomplete Answers

When questions allow respondents to express their opinions in their own words, some of the respondents may give ambiguous or incomplete answers. If the survey is administered online, some probes (e.g., “Why did you choose that answer?”) can be programmed into the survey instrument and automatically administered as part of the survey. If interviewers are asking the questions, they may be instructed to probe to obtain a more complete response or clarify the meaning of the ambiguous response. Interviewers should be instructed what clarification, if any, they can provide. The record should reflect both what the respondent said initially, what the interviewer said in the attempt to get or provide clarification, and how the respondent answered the probe; this information will allow the court and the opposing party to evaluate whether the probe affected the views expressed by the respondent.

If the survey involves interviewers who are permitted to administer follow-up questions, they must be given explicit instructions on when they should probe and what they should say in probing.<sup>139</sup> Standard probes used to draw out all that the respondent has to say (e.g., “Any further thoughts?”; “Anything else?”; “Can you explain that a little more?”; or “Could you say that another way?”) are relatively noncontroversial and can be programmed into an online survey instrument. But probing should be limited. Persistent continued requests for further responses to the same or nearly identical questions may convey the idea to the respondent that they have not yet produced the “right” answer, particularly if the probes are administered by a live interviewer.<sup>140</sup>

## Potential Order Effects

The order in which questions are asked on a survey and the order in which response alternatives are provided in a closed-ended question can influence the answers.<sup>141</sup> For example, although asking a general question before a more specific

139. Floyd J. Fowler, Jr. & Thomas W. Mangione, Standardized Survey Interviewing: Minimizing Interviewer-Related Error 41–42 (1990), <https://doi.org/10.4135/9781412985925>.

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

141. See Schuman & Presser, *supra* note 115, at 23, 56–74; Krosnick & Presser, *supra* note 102, at 278–81. In *R.J. Reynolds Tobacco Co. v. Loew's Theatres, Inc.*, 511 F. Supp. 867, 875 (S.D.N.Y. 1980), the court recognized the biased structure of a survey that disclosed the tar content of the cigarettes being compared before questioning respondents about their cigarette preferences. Not surprisingly, respondents expressed a preference for the lower tar product. See also E. & J. Gallo Winery v. Pasatiempos Gallo, S.A., 905 F. Supp. 1403, 1409–10 (E.D. Cal. 1994) (court recognized

question on the same topic is unlikely to affect the response to the specific question, reversing the order of the questions may influence responses to the general question. As a rule, then, surveys are less likely to be subject to order effects if the questions move from the general (e.g., “What do you recall being discussed 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?”).<sup>142</sup>

The mode of questioning can influence the form that an order effect takes. When respondents are shown response alternatives visually, as in mail surveys and self-administered online surveys or in face-to-face interviews when respondents are shown a card containing response alternatives, they are more likely to select the first choice offered (a primacy effect).<sup>143</sup> In contrast, when response alternatives are presented orally, as in telephone surveys, respondents are more likely to choose the last choice offered (a recency effect).<sup>144</sup> 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. To control for order effects, the order of the response choices in a survey should be rotated, randomized, or counterbalanced,<sup>145</sup> so that no response alternative will have an inflated chance of being selected because of its position.

## Appropriate Inclusion of Control Groups, Control Questions, or Other Comparisons

Many surveys are designed not simply to describe attitudes, beliefs, or reported behaviors, but to determine the source of those attitudes, beliefs, or behaviors. That is, the purpose of the survey is to test a causal proposition. For example, how does a trademark or the content of a commercial affect respondents' perceptions or understanding of a product or commercial? Thus, the question is not merely whether consumers hold inaccurate beliefs about Product A, but whether exposure to the commercial misleads the consumer into thinking that Product A

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).

142. 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).

143. Krosnick & Presser, *supra* note 102, at 280.

144. *Id.*

145. See, e.g., *Winning Ways, Inc. v. Holloway Sportswear, Inc.*, 913 F. Supp. 1454, 1465–67 (D. Kan. 1996) (failure to rotate the order in which the jackets were shown to the consumers led to reduced weight for the survey); *Procter & Gamble Pharms., Inc. v. Hoffmann-La Roche, Inc.*, No. 06 Civ. 0034 (PAC), 2006 U.S. Dist. LEXIS 64363 (S.D.N.Y. Sept. 6, 2006).

is a superior pain reliever. Yet if consumers already believe, before viewing the commercial, that Product A is a superior pain reliever, a survey that simply records consumers' impressions after they view the commercial may reflect those preexisting beliefs rather than impressions produced by the commercial.

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.<sup>146</sup> The more routine the idea being examined in the survey, the more likely it is that the respondent's answer is influenced by (1) preexisting impressions; (2) general expectations about what commercials typically say (e.g., the product being advertised is better than its competitors); or (3) guessing, rather than by the actual content of the commercial message or the trademark being evaluated. A similar limitation occurs when respondents are asked to explain why they chose a particular response: they can justify their choice, but may not be able to report accurately what actually led them to make that choice.

With respondents randomly assigned to one or more appropriate comparison conditions, the survey expert can draw clear causal inferences about the influence of the stimulus.<sup>147</sup> In the simplest version of such a survey-experiment (as discussed above), respondents are assigned randomly to one of two conditions.<sup>148</sup> For example, respondents assigned to the experimental condition view an allegedly deceptive commercial, and respondents assigned to the control condition view the same commercial with the allegedly deceptive material removed or an alternative commercial that does not contain the allegedly deceptive material.<sup>149</sup>

146. See Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 Psych. Rev. 231 (1977), <https://doi.org/10.1037/0033-295X.84.3.231>.

147. See Shari S. Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 L. & Hum. Behavior 239, 244–46 (1989), <https://doi.org/10.1007/BF01067028>; Jacob Jacoby & Constance Small, *Applied Marketing: The FDA Approach to Defining Misleading Advertising*, 39 J. Mktg. 65, 68 (1975), <https://doi.org/10.1177/002224297503900413>. See also R. Charles Henn, *Why Ask Why? A Critical Assessment of an Historical Survey Artifact*, 113 Trademark Rep. 772, 773–76 (2023).

148. 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, on average, 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 section titled "The Sample as a Reflection of the Relevant Characteristics of the Population" above.

149. 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.

Respondents in both the experimental and control groups answer the same set of questions about the allegedly deceptive message. The effect of the commercial's allegedly deceptive message is evaluated by comparing the responses made by the experimental group members with those of the control group members. If 40% of the respondents in the experimental group responded indicating a belief in the deceptive claim, 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<sup>150</sup>) can be attributed to the allegedly deceptive commercial.

A survey-experimental design with an appropriate control group can account for more than preexisting beliefs.<sup>151</sup> Other sources of systematic and random error can influence responses to survey questions. Thus, if a respondent is asked "Have you heard of a security company named Titan Alarm?" some individuals will say they have, even if they have not. A control group can help address this type of error as well; this and other background noise should have produced similar response levels in the experimental and control groups, so comparing average scores in those groups should control for those sources of error. In addition, a leading question may cause participants to be more likely to endorse or reject a particular statement. But if respondents who viewed the allegedly deceptive commercial respond differently than respondents who viewed the control commercial, the difference cannot be merely the result of a leading question, because both groups answered the same question. The ability to evaluate the effect of the wording of a particular question makes the control group design particularly useful in assessing responses to closed-ended questions,<sup>152</sup> which may encourage guessing or particular responses. Thus, the focus is not on the absolute response level in the experimental group, but on the difference between the responses from the experimental group and those from the control group.<sup>153</sup>

In designing a survey-experiment, the expert should select a stimulus for the control group that shares as many characteristics with the experimental stimulus as possible, with the key exception of the characteristic whose influence is being assessed.<sup>154</sup> Although a survey with an imperfect control group may provide

150. For a discussion of sampling error, see the glossary to the current chapter and David H. Kaye and Hal S. Stern, *Reference Guide on Statistics and Research Methods*, in this manual.

151. For a more extensive discussion of controls, see Shari Seidman Diamond, *Control Foundations: Rationale and Approaches*, in Trademark and Deceptive Advertising Surveys 239 (Shari Diamond & Jerre Swann eds., 2d ed. 2022).

152. The Federal Trade Commission has long recognized the need for some kind of control for closed-ended questions, although it has not specified the type of control that is necessary. See *In re Stouffer Foods Corp.*, 118 F.T.C. 746, 808–09 (Sept. 26, 1994).

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

154. See, e.g., *Skechers USA, Inc. v. Vans, Inc.*, No. CV-07-01703 DSF (PLAx), 2007 WL 4181677, at \*8–9 (C.D. Cal. Nov. 20, 2007) (in trade dress infringement case, control stimulus should have retained design elements not at issue); *Procter & Gamble Pharms., Inc. v. Hoffman-LaRoche*,

better information than a survey with no control group at all, the choice of an appropriate control group requires care and should influence the admissibility of the survey and the weight that the survey receives. For example, a control stimulus should not be less attractive than the experimental stimulus if the survey is designed to measure how familiar the experimental stimulus is to respondents, because attractiveness may affect perceived familiarity.<sup>155</sup> Nor should the control stimulus share with the experimental stimulus the feature whose impact is being assessed. If the control stimulus in a case of alleged trademark infringement is itself a likely source of consumer confusion, for example, reactions to the experimental and control stimuli may not differ because both cause respondents to express a similar level of confusion.<sup>156</sup> In an extreme case, an inappropriate control may do nothing more than control for the effect of the nature or wording of the survey questions (e.g., acquiescence).<sup>157</sup> That generally will not be enough to rule out other explanations for different or similar responses to the experimental and control stimuli. Finally, it may sometimes be appropriate to have more than one control group to assess precisely what is causing the response to the experimental stimulus (e.g., in the case of an allegedly deceptive ad, whether it is a misleading graph or a misleading claim by the announcer, or in the case of allegedly infringing trade dress, whether it is the style of the font used or the coloring of the packaging).<sup>158</sup>

Courts have increasingly come to recognize the central role that control groups play in evaluating causal claims.<sup>159</sup> Litigants have taken the cue, and most

Inc., No. 06 Civ 0034 (PAC), 2006 U.S. Dist. LEXIS 64363, at \*87–88 (S.D.N.Y. Sept. 6, 2006) (in false advertising action, disclaimer was inadequate substitute for appropriate control group).

155. See, e.g., Indianapolis Colts, Inc. v. Metro. Balt. Football Club L.P., 34 F.3d 410, 415–16 (7th Cir. 1994) (court recognized that the name “Baltimore Horses” was less attractive for a sports team than the name “Baltimore Colts”); *see also* Reed-Union Corp. v. Turtle Wax, Inc., 77 F.3d 909, 912 (7th Cir. 1996) (court noted that one expert’s choice of a control brand with a well-known corporate source was less appropriate than the opposing expert’s choice of a control brand whose name did not indicate a specific corporate source).

156. See, e.g., W. Publ’g Co. v. Publ’ns Int’l, Ltd., No. 94-C-6803, 1995 U.S. Dist. LEXIS 5917, at \*45 (N.D. Ill. May 2, 1995) (court noted that the control product was “arguably more infringing than” the defendant’s product) (emphasis omitted). *See also* Classic Foods Int’l Corp. v. Kettle Foods, Inc., No. SACV 04–725 CJC (Ex), 2006 U.S. Dist. LEXIS 97200 (C.D. Cal. Mar. 2, 2006); McNeil-PPC, Inc. v. Merisant Co., No. 04–1090 (JAG), 2004 U.S. Dist. LEXIS 27733 (D.P.R. July 29, 2004).

157. *See supra* text accompanying note 134.

158. See, e.g., Masterfoods USA v. Arcor USA, Inc., 230 F. Supp. 2d 302 (W.D.N.Y. 2002).

159. See, e.g., Colangelo v. Champion Petfoods USA, Inc., No. 6:18-CV-1228, 2022 U.S. Dist. LEXIS 60489, at \*32 (N.D.N.Y. Mar. 31, 2022) (“[W]ithout a control group it is impossible to establish cause and effect.”); Longoria v. Million Dollar Corp., No. 18-CV-02266-PAB-NYW, 2021 U.S. Dist. LEXIS 38478, at \*27 (D. Colo. Mar. 2, 2021) (survey excluded because “a causal study . . . requires a control group”); SmithKline Beecham Consumer Healthcare, L.P. v. Johnson & Johnson-Merck, No. 01 Civ. 2775 (DAB), 2001 U.S. Dist. LEXIS 7061, at \*37–39 (S.D.N.Y. June 1, 2001) (survey to assess implied falsity of a commercial not probative in the absence of a control group);

experts recognize the need to submit a survey with a control group (i.e., a survey-experiment) if they wish to make a causal claim and rule out other explanations for the responses to the survey questions.<sup>160</sup>

A less common control methodology is a control question. Rather than administering a control stimulus to a separate group of respondents, the survey asks all respondents one or more control questions along with the question about the product or service at issue. This technique is used to evaluate whether a brand name is generic. The genericness survey presents survey respondents with a series of product or service names and asks 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 demonstrated that their respondents understood the difference between brands and product categories (so-called common names) and that the Teflon mark at issue was perceived as a brand name; they retained their trademark.<sup>161</sup> It is crucial to control for order effects in such designs.

The issue of appropriate comparisons is especially salient in the context of conjoint analysis, which has been used in some patent and false-advertising cases. In cases of patent infringement, the plaintiff must submit an estimate of the damages produced by the defendant's infringement. When the infringement involves a multicomponent product, the plaintiff must apportion damages to determine the value of the patented component at issue.<sup>162</sup> Similarly, in class actions involving false advertising, damages depend on the value to consumers that can be attributed to the deceptive portion of the advertisement.<sup>163</sup> To estimate these damages, survey experts have increasingly turned to conjoint analysis. When done

Am. 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).

160. William, R. Shadish et al., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (2002); James N. Druckman, *Experimental Thinking: A Primer on Social Science Experiments* (2022).

161. E.I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502, 526-27 & n.54 (E.D.N.Y. 1975); *see also* Donchez v. Coors Brewing Co., 392 F.3d 1211, 1218 (10th Cir. 2004) (respondents evaluated eight brand and generic names in addition to the disputed name); Reinalt-Thomas Corp. v. Mavis Tire Supply, LLC, 391 F. Supp. 3d 1261, 1271-73 (N.D. Ga. 2019). A similar approach is used in assessing secondary meaning. *See, e.g.*, T-Mobile US, Inc. v. AIO Wireless LLC, 991 F. Supp. 2d 888 (S.D. Tex. 2014).

162. For example, in *Apple, Inc. v. Samsung Electronics Co., Ltd.*, No. 11-CV-01846-LHK, 2014 U.S. Dist. LEXIS 29721 (N.D. Cal. Mar. 6, 2014), Apple used a conjoint survey to isolate the costs associated with Samsung's alleged patent infringement regarding iPhone and iPad features.

163. *E.g.*, Price v. L'Oréal U.S., Inc., No. 17-Civ-614 (LGS), 2020 Dist. LEXIS 153255 (S.D.N.Y. Aug. 24, 2020).

well, conjoint analysis can provide useful information, but there are many decisions made in designing a conjoint analysis that can undermine reliability and bias the resulting estimates.<sup>164</sup>

In contrast to a survey that asks respondents directly how much they would be willing to pay (WTP) for a given feature or attribute of a product, a task that may be difficult for consumers, a conjoint survey-experiment typically presents respondents with choices between products with different randomly assigned profiles consisting of combinations of features.<sup>165</sup> Suppose the survey is designed to determine the value of a patented component of a washing machine that shuts off the machine automatically in response to overflow (automatic shutoff). The respondent will be presented with a series of choices between washing machines that may, for example, vary on brand (LG, General Electric, Whirlpool), number of settings (eight, ten, twelve), color (white, beige, stainless steel), capacity (3, 4, 4.5 cu. ft.), and price (\$400, \$598, \$648). Each respondent makes their choice in response to several sets (profiles) of products where the attributes (e.g., brand, number of settings) have randomly assigned levels (e.g., LG, General Electric, Whirlpool; eight, ten, twelve) for each profile. A regression is then used to compute part worths, the value that each feature adds to the total product. As in any survey, it is crucial to identify the appropriate universe of individuals likely to purchase the product. For instance, it would be inappropriate to conduct the washing machine conjoint survey on college students who can be assumed to have no intention of purchasing a washing machine in the near future.

Although conjoint analysis offers a potentially useful method for assessing WTP for a single feature of a multifeature product as a result of its experimental nature, the design of the profiles can lead to misleading results. The first issue is what features to include in the product profiles. If important features of the product are omitted (e.g., in the washing machine example, whether the machine is front-loading or top-loading) and the profiles include predominantly less important features e.g., door shape as round versus square, maximum spin speed), the estimated values for the component of interest are likely to be inflated.<sup>166</sup> Thus, an important preliminary step by the expert designing a conjoint analysis is to assess the importance of the various features of the product.<sup>167</sup>

164. Bernard Chao & Sydney Donovan, *Does Conjoint Analysis Reliably Value Patents?*, 58 Am. Bus. L.J. 225 (2021), <https://doi.org/10.1111/ablj.12182>. See also Suneal Bedi & David Reibstein, *Damaged Damages: Errors in Patent and False Advertising Litigation*, 73 Ala. L. Rev. 385 (2021); David Franklyn & Adam Kuhn, *The Problem of Mop Heads in the Era of Apps: Toward More Rigorous Standards of Value Apportionment in Contemporary Patent Law*, 98 J. Pat. & Trademark Off. Soc'y 182 (2016).

165. Some conjoint analyses ask respondents for rankings or ratings, but asking them to make choices is the most generally accepted approach because it is more reflective of what they would do in the marketplace. Bedi & Reibstein, *supra* note 164, at 399 nn.76 & 77.

166. Bedi & Reibstein, *supra* note 164.

167. In *Apple, Inc. v. Samsung Electronics Co.*, No. 11-CV-01846-LHK, 2014 U.S. Dist. LEXIS 29721 (N.D. Cal. Mar. 6, 2014), the conjoint survey used by Apple included seven attributes;

An additional challenge in designing a conjoint survey is to ensure that the features are described clearly and accurately, while simultaneously not drawing substantially more attention to features than consumers might naturally devote to them in the real world. This allows respondents to understand what each feature is and ensures that their understanding matches the meanings and definitions used in the case.<sup>168</sup> If a feature is unclear (e.g., electronic spinner), the respondent cannot make an informed decision that takes that feature into account. It also is essential, as with any survey, that the population be carefully chosen.<sup>169</sup>

Further, if the conjoint analysis is being used to identify prices, there is some controversy on how to capture supply-side aspects of the pricing that reflect the market. Given that market conditions may have changed due to the infringement, this can present a challenge.<sup>170</sup>

Samsung critiqued it for excluding other important attributes such as brand name and battery life. In *MacDougall v. American Honda Motor Co.*, No. 20-56060, 2021 U.S. App. LEXIS 37780 (9th Cir. Dec. 21, 2021), the court of appeals reversed after the district court had rejected the use of a conjoint survey for “the reduction of the amount of vehicle attributes in the final survey from thirty-three to four . . . The more limited a consumer’s choice of vehicle features, the more artificially inflated the importance of the remaining features. . . .” *MacDougall v. Am. Honda Motor Co.*, No. SACV 17-1079 JGB, 2020 U.S. Dist. LEXIS 166786, at \*22 (C.D. Cal., Sept. 11, 2020). Including thirty-three attributes may be too many, but at issue here was the lack of justification for the four chosen (it was unclear how the pretest led to choosing those four). On appeal, the court found that the attributes chosen go to weight and not admissibility; however, the lack of a reasonable basis for attribute choice can seriously undermine the value of the analysis. There often is an inevitable tradeoff between overwhelming respondents with too many attributes and capturing those that are most important. The expert should explicitly provide reasons for the specific inclusion and exclusion of attributes.

168. For instance, in *Price*, the court rejected an analysis that used the term “Kerantindose Pro-Keratin + Silk” to isolate the impact of “Kerantindose” and “Pro-Keratin” since it includes “+ Silk.” 2020 Dist. LEXIS 153255. In *Allegra v. Luxottica Retail North America*, 341 F.R.D. 373 (E.D.N.Y. Dec. 13, 2021), the court rejected the use of a conjoint survey that it said failed to properly describe the allegedly fraudulent omission in the advertising.

169. In *Cardenas v. Toyota Motor Corp.*, 418 F. Supp. 3d 1090 (S.D. Fla. 2019), the case involved a defect in a nonhybrid Toyota. The defendants argued that hybrid purchasers should not have been included. The crucial question was whether purchasers of hybrid cars would differ in their assessments from nonhybrid purchasers, which the court determined was a question of weight rather than admissibility. The analysis of subgroup effects is tricky because respondent subgroups may differ in the evaluations of comparison points and thus analyses needed to account for those distinctions; see Thomas Leeper et al., *Measuring Subgroup Preferences in Conjoint Experiments*, 28 Pol. Analysis, 207-21 (2020), <https://doi.org/10.1017/pan.2019.30>.

170. J. Gregory Sidak & Jeremy O. Skog, *Using Conjoint Analysis to Apportion Patent Damages*, 25 Fed. Cir. Bar J. 581 (2016). More generally, how to estimate supply side considerations is a matter of some debate. In *Cardenas v. Toyota Motor Corp.*, the court acknowledged differing opinions on whether the appropriate pricing information should emulate actual market prices and sales during the class period or the probable prices and sales in the situation where the relevant attribute took on the value that was under contention (the relevant attribute in this case was an odor emitted from a car’s HVAC system). The court ruled that the approach to incorporating supply-side information does not go to admissibility, but could affect weight. See also *Colangelo v. Champion*

Some aspects of conjoint analysis are controversial. These include whether it is always necessary to give respondents the choice of declining to select any of the products offered in the profiles presented (i.e., “none of the above”). If the relevant population includes people who may prefer not to make the purchase, inclusion is warranted, but note there is a danger that respondents may choose the “no purchase” option to avoid assessing the products.<sup>171</sup> Similarly, there is some dispute about how many features of a product can be listed (e.g., can respondents process and respond to more than seven features?).

As with any survey, the expert should be prepared to explain the choices made in constructing the survey design. In light of the complexity of conjoint survey experimental designs and the controversial nature of some of the decisions the expert needs to make, a detailed explanation of those decisions is warranted to guide the court in evaluating whether the survey is admissible (or what weight to give it).

## Benefits and Limitations Associated with the Mode of Data Collection Used in the Survey

Four primary methods have traditionally been used to collect survey data: (1) in-person interviews, (2) telephone interviews, (3) mail questionnaires, and (4) internet surveys. The choice of any data collection method for a survey should be justified by its strengths and weaknesses.

Common across in-person, telephone, and internet surveys is the use of computer-assisted techniques.<sup>172</sup> The interviewer conducting a computer-assisted interview (CAI), whether by telephone (CATI) or face-to-face (computer-assisted personal interviewing, or CAPI), follows the computer-generated script for the interview and enters the respondent’s answers as the interview proceeds. A primary advantage of CATI and other CAI procedures is that skip patterns can be built into the program. If, for example, the respondent answers “yes” when asked whether she has ever been the victim of a burglary, the computer will generate further questions about the burglary; if she answers

Petfoods USA, Inc., No. 6:18-CV-122 (LEK/ML), 2022 U.S. Dist. LEXIS 60489, at \*25 (N.D.N.Y. Mar. 31, 2022) (“[T]here is a legitimate difference of opinions, both among judges and experts, about the significance of supply side information in calculating loss of value.”) (citing *In re Fisher-Price Rock ‘n’ Play Sleeper Mktg.*, 567 F. Supp. 3d 406, 415 (W.D.N.Y. 2021)).

171. Daniel McFadden, *Stated Preference Methods and Their Applicability to Environmental Use and Non-use Valuations, in Contingent Valuation of Environmental Goods: A Comprehensive Critique*, 153–87 (Daniel McFadden & Kenneth Train eds., 2017). See also Moshe Ben-Akiva et al., *Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-Based Conjoint Analysis*, 10 Founds. & Trends in Econometrics 1 (2019), <http://dx.doi.org/10.1561/0800000036>.

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

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.<sup>173</sup> CAI procedures also can be used to control for order effects by having the program rotate the order in which the questions or choices are presented and facilitate the implementation of complex experiments with many conditions by randomly generating many potential factors at once.<sup>174</sup>

CAI procedures also include audio computer-assisted self-interviewing (ACASI) in which the respondent listens to recorded questions over the telephone or reads questions from a computer screen while listening to recorded versions of them through headphones. The respondent then answers verbally or directly on a keypad. ACASI procedures are particularly useful for collecting sensitive information (e.g., illegal drug use and other HIV risk behavior).<sup>175</sup> This is also useful in face-to-face interviews where the respondent can avoid having to reveal their (potentially sensitive) answer to the interviewer since they enter it themselves.

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

## *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. When the questions are complex and the interviewers are skilled, in-person interviewing provides the maximum opportunity to clarify or probe. Unlike a mail survey, in-person, telephone, and web-based surveys 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. In-person

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

174. 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).

175. See, e.g., P.C. Cooley et al., *Automating Telephone Surveys: Using T-ACASI to Obtain Data on Sensitive Topics*, 16 *Comput. in Hum. Behav.* 1 (2000), <https://perma.cc/L94R-AKKY>.

interviewers also can directly verify who is completing the survey, a check that is unavailable in mail and web-based surveys. As described below in the “Selection and Training of the Interviewers” section, appropriate interviewer training, as well as monitoring of the implementation of interviewing, is necessary if these potential benefits are to be realized. Objections to the use of in-person interviews arise primarily from their high cost or, on occasion, from evidence of inept or biased interviewers. The latter concern is somewhat mitigated by computer assistance, described above.

## Telephone Interviews

Telephone surveys offer a comparatively fast and lower-cost alternative to in-person surveys and can be particularly useful when the population is large and geographically dispersed. Telephone interviews (unless supplemented with mailed materials) should be used only when it is unnecessary to show the respondent any visual materials. Thus, an attorney may present the results of a telephone survey of jury-eligible citizens in a motion for a change of venue in order to provide evidence that community prejudice raises a reasonable suspicion of potential jury bias.<sup>176</sup> Similarly, potential confusion between a restaurant called McBagel’s and the McDonald’s fast-food chain was established in a telephone survey. Over objections from defendant McBagel’s that the survey did not show respondents the defendant’s print advertisements, the court found likelihood of confusion based on the survey, noting that “by soliciting audio responses [the telephone survey] was closely related to the radio advertising involved in the case.”<sup>177</sup> In contrast, when words are not sufficient because, for example, the survey is assessing reactions to the trade dress or packaging of a product that is alleged to promote confusion, a telephone survey alone does not offer a suitable vehicle for questioning respondents.<sup>178</sup>

176. See, e.g., *State v. Baumruk*, 85 S.W.3d 644 (Mo. 2002) (overturning the trial court’s decision to ignore a survey that found about 70% of county residents remembered the shooting that led to the trial, and that of those who had heard about the shooting, 98% believed that the defendant was either definitely guilty or probably guilty); *State v. Erickstad*, 620 N.W.2d 136, 140 (N.D. 2000) (denying change-of-venue motion based on media coverage, concluding that “defendants [need to] submit qualified public opinion surveys, other opinion testimony, or any other evidence demonstrating community bias caused by the media coverage”). For a discussion of surveys used in motions for change of venue, see Neal Miller, *Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation, Part II*, 40 Rutgers L. Rev. 467, 470–74 (1988); National Jury Project, *Jurywork: Systematic Techniques* (2d ed. 2008).

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

178. *See Thompson Med. Co. v. Pfizer Inc.*, 753 F.2d 208 (2d Cir. 1985); *Inc. Publ’g Corp. v. Manhattan Mag., Inc.*, 616 F. Supp. 370 (S.D.N.Y. 1985), *aff’d without op.*, 788 F.2d 3 (2d Cir. 1986).

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

1. Whether (when prospective respondents are not business personnel) some form of random-digit dialing<sup>179</sup> was used instead of or to supplement telephone numbers obtained from telephone directories, because a high percentage of all residential telephone numbers in some areas may be unlisted;<sup>180</sup>
2. What types of phones were included—that is, landlines, cell phones, or both. Over ninety-six percent of American adults live in households with cell phones, leading to a shift in sampling procedures that sometimes rely only on cell phones.<sup>181</sup>
3. Whether the sampling procedures required the interviewer to sample within the household or business, instead of allowing the interviewer to administer the survey to any qualified individual who answered the telephone;<sup>182</sup> and
4. Whether interviewers were required to call back multiple times at several different times of the day and on different days to increase the likelihood of contacting individuals or businesses with different schedules and, as a result, to reduce nonresponse levels.<sup>183</sup>

Telephone surveys that do not include these procedures may not provide precise measures of the characteristics of a representative sample of respondents, but may be adequate for providing rough approximations. The vulnerability of the survey depends on the information being gathered. More elaborate procedures are advisable for achieving a representative sample of respondents if the survey instrument requests information that is likely to differ for individuals with listed

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

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

181. Between 2% and 3% of the U.S. population has only landline access. Stephen J. Blumberg & Julian V. Luke, *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January–June 2020* (U.S. Dep’t of Health & Hum. Servs., Feb. 2021), at 4, <https://perma.cc/5TSM-V88J>; see also Courtney Kennedy et al., *Implications of Moving Public Opinion Surveys to a Single-Frame Cell-Phone Random-Digit-Dial Design*, 82 Pub. Op. Q. 279 (2018), <https://doi.org/10.1093/poq/nfy016> (“Analysis of more than 250 survey questions show[ed] that when landlines [were] excluded, estimates change[d] by less than one percentage point, on average.”).

182. 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.

183. This applies equally to in-person interviews.

telephone numbers versus individuals with unlisted telephone numbers, individuals rarely at home versus those usually at home, or groups who are more versus less likely to rely on landlines or cell phones.

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

- The procedures that were used to identify potential respondents, including both the procedures used to select the telephone numbers that were called and the procedures used to identify the qualified individual to question;
- The number of telephone numbers for which no contact was made; and
- The number of contacted potential respondents who refused to participate in the survey (and how many follow-up attempts were made).<sup>184</sup>

Like computer-assisted personal interviewing (CAPI),<sup>185</sup> computer-assisted telephone interviewing (CATI) facilitates the administration and data entry of large-scale surveys.<sup>186</sup> A computer protocol may be used to generate and dial telephone numbers as well as to guide the interviewer.

## *Mail Questionnaires*

In general, mail surveys tend to be substantially less costly than both in-person and telephone surveys.<sup>187</sup> Procedures that raise response rates for mail surveys include multiple mailings, highly personalized communications, prepaid return envelopes, incentives or gratuities, assurances of confidentiality, first-class outgoing postage, and follow-up reminders.<sup>188</sup>

A mail survey will not produce a high rate of return unless the recruitment process begins with an accurate and up-to-date list of names and addresses for

184. Additional disclosure and reporting features applicable to surveys in general are described in the section titled “Completeness and Accuracy of All Relevant Information in the Survey Report” below.

185. See *infra* text accompanying note 205.

186. See Tourangeau et al., *supra* note 88, at 289; Saris, *supra* note 173.

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

188. See, e.g., Richard J. Fox et al., *Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response*, 52 Pub. Op. Q. 467, 482–84 (1988), <https://doi.org/10.1086/269125>; 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), <https://doi.org/10.1080/002220973.1992.9943849>; Eleanor Singer et al., *Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature*, 59 Pub. Op. Q. 66, 71 (1995), <https://doi.org/10.1086/269458>; see generally Don A. Dillman et al., *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (3d ed. 2009).

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 (i.e., differential nonresponse). For example, if a survey targets a population that includes individuals with literacy problems, these individuals will tend to be underrepresented. Open-ended questions are generally of limited value on a mail survey because they depend entirely on the respondent to answer fully and do not provide the opportunity to probe or clarify unclear answers. Similarly, if eligibility to answer some questions depends on the respondent's answers to previous questions, such skip sequences may be difficult for some respondents to follow. Finally, because respondents complete mail surveys without supervision, survey personnel are unable to prevent respondents from discussing the questions and answers with others before completing the survey or to control the order in which respondents answer the questions. Although skilled design of questionnaire format, question order, and the appearance of the individual pages of a survey can minimize these problems, if it is crucial to have respondents answer questions in a particular order, a mail survey cannot be depended on to provide adequate data.

## Internet Surveys

Over the past two decades there has been a massive increase in the use of internet surveys. One recent analysis found that the overwhelming majority of expert reports reviewed in the trademark and false advertising space were based on online surveys.<sup>189</sup> As of 2024, “an estimated” 96% of U.S. adults have access to the internet, including 99% of those between the ages of eighteen and forty-nine.<sup>190</sup> Although concerns remain about internet survey administration, and especially internet administration through online panels, many of those concerns can be mitigated through quality-control measures.

The benefits of internet administration are clear. The cost of conducting surveys online is a small fraction of the cost of hiring human interviewers to speak with potential respondents; the speed with which a researcher can find, screen, and survey hundreds of respondents online is far greater than in non-internet-based environments (e.g., via the mail); and there is less risk of interviewer bias and error when questions are presented by computer on a screen and respondents

189. Kugler & Henn, *supra* note 71, at 293–94.

190. Pew Rsch. Ctr., Internet, Broadband Fact Sheet (Nov. 2024), <https://www.pewinternet.org/fact-sheet/internet-broadband>. Of those age fifty to sixty-four, 98% had internet access. The sixty-five-plus group was lower at 90%. The same research shows that 79% of the U.S. population has broadband internet at home, while an additional 15% do not have broadband but do have internet access via smartphone. In 2011, only 79% of Americans had internet access. *Id.*

type their own responses.<sup>191</sup> Further, internet surveys may be particularly appropriate when the goal of the researcher is to simulate everyday commercial conduct that now frequently occurs online.

Computer administration also allows careful randomization of survey items, branching survey logic, and the display of complicated visual and auditory stimuli. In addition, the structure permits the survey to remind, or even require, the respondent to answer a question before the next question is presented. A further advantage of computer-administered surveys over interviewer-administered surveys is that they eliminate interviewer error because the computer presents the questions and the respondent records their own answers.

One major question for internet surveys is the adequacy of the sampling approach. The evaluation of this will depend on the type of internet survey involved, because the samples used in web-based surveys vary in fundamental ways. At one extreme is the list-based web survey. This web-based survey is sent to a closed set of potential respondents drawn from a list that consists of the email addresses of the target individuals (e.g., all employees at a company where each employee has a known email address). Here there is no reason not to use the tools of probability sampling, described above.

At the other extreme is the self-selected web pseudosurvey in which web users in general, or those who happen to visit a particular website, are all invited to express their views on a topic and they participate simply by volunteering. Participants are very likely to self-select on the basis of the nature of the topic, substantially distorting the results. These self-selected pseudosurveys resemble reader polls published in magazines and do not meet standard criteria for legitimate surveys admissible in court.<sup>192</sup> Occasionally, proponents of such polls tout the large number of respondents as evidence of the weight the results should be given, but the size of the sample cannot cure the likely participation bias in such voluntary polls.<sup>193</sup>

Between these two extremes is a large category of web-based survey approaches that researchers have developed to address concerns about sampling bias and nonresponse error. Many of these use the professionally managed non-probability panels described above. These companies recruit diverse panels of respondents. Some respondents come from other well-traveled sites; others are

191. Reg Baker et al., *Research Synthesis: AAPOR Report on Online Panels*, 74 Pub. Op. Q. 711, 739 (2010), <https://doi.org/10.1093/poq/nfq048> (“Overall, the research reported here generally suggests higher data quality for computer administration than for oral administration.”).

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

193. See Mick P. Couper, *Web Surveys: A Review of Issues and Approaches*, 64 Pub. Op. Q. 464, 480–81 (2000), <https://doi.org/10.1086/318641> (a self-selected web survey conducted by the National Geographic Society through its website attracted 50,000 responses; a comparison of the Canadian respondents with data from the Canadian General Social Survey telephone survey conducted using random digit dialing showed marked differences on a variety of response measures).

pursued because their qualifications make them valuable for marketing studies. When a researcher contracts with a panel provider, the company can sift through its database to invite an appropriate mix of people to participate in a given survey. Those invited do not know the topic of the survey in advance (i.e., they are not opting in based on the topic). The final sample can be balanced to match desired demographics either through recruitment quotas or through response weighting.<sup>194</sup> The researcher using such data should obtain the procedures used by the vendor to ensure the quality of the data and make it available to the opposing party and the trier of fact. For example, what steps did the vendor take to minimize fraudulent respondents<sup>195</sup> and minimize duplicative respondents? Did the vendor themselves outsource the sampling?<sup>196</sup>

Generally, a researcher conducting an internet survey must take active steps to ensure an honest and attentive sample. On the most technical level, use of CAPTCHA<sup>197</sup> questions can weed bots out from the sample; panel providers can audit their panels to verify participant demographics, or at least check that self-reported demographics are consistent over time,<sup>198</sup> and survey platforms, which host the actual survey questions, can use cookies and other means to prevent multiple submissions from the same person or computer.<sup>199</sup> As with many other issues in the online space, there is constant evolution in the fraud-detection domain. A researcher should ensure that they, and their panel provider, are using current best practices to detect fraudulent responses and be prepared to explain how these responses work and their known impact.

Moving beyond the technology and into the survey itself, the researcher should take further steps to ensure an attentive sample. In the trademark space, a

194. See, e.g., Philip Morris USA, Inc. v. Otamedia Ltd., No. 02 Civ 7575 (GEL) (KNF), 2005 U.S. Dist. LEXIS 1259, at \*10–11 (S.D.N.Y. Jan. 28, 2005).

195. Andrew M. Bell & Thomas Gif, *Fraud in Online Surveys: Evidence from a Nonprobability, Subpopulation Sample*, 10 J. Experimental Pol. Sci. 148–53 (2023), <https://doi.org/10.1017/XPS.2022.8>.

196. Peter K. Enns & Jake Rothschild, *Do You Know Where Your Survey Data Come From?*, Medium, May 2, 2022, <https://perma.cc/BD9T-SK25>.

197. CAPTCHA is an acronym for “Completely Automated Public Turing test to tell Computers and Humans Apart.” The respondent may be asked to decipher a set of distorted letters or complete a categorization task that would be difficult to automate.

198. Courtney Kennedy et al., Pew Rsch. Ctr., Evaluating Online Nonprobability Surveys 36 (May 2, 2016), <https://perma.cc/VJ4L-BTJF>:

Most panels are double opt-in, meaning potential panelists first enter an email address and then respond to an email sent from the panel provider in order to confirm the email account. Depending on the vendor, other quality control features include IP address validation and digital fingerprinting, which guards against a single person having multiple accounts in a given panel.

199. “The most common technique for identifying duplicate respondents is digital fingerprinting. Specific applications of this technique vary, but they all involve the capture of technical information about a respondent’s IP address, browser, software settings, and hardware configuration to construct a unique ID for that computer.” Baker et al., *supra* note 191, at 756 (emphasis omitted).

few checks are particularly common.<sup>200</sup> One method of measuring attention in a traditional likelihood-of-confusion survey is to include a free-response question prompting participants to explain a prior answer. If the participant responds with gibberish or irrelevant responses, they are either not paying attention or not taking the survey seriously.<sup>201</sup> A similar check is possible in a Teflon survey testing whether a mark is generic.<sup>202</sup> There, a participant is asked whether each in a list of terms is a “brand name” or a “common name.” A participant who responds that all terms are brand names or all are common names, that is, who gives a straight-line response, should be excluded. They either do not understand the task or are not trying to do it. For surveys aimed at purchasers of a particular product, it is common to ask which, of a list of products, they have bought in the past X months. A researcher can include nonexistent products on those lists, screening out people who implausibly claim to have purchased them.

The science behind these efforts to ensure honest responding is ever-evolving. A researcher should employ some of the above mechanisms in any online survey, and should be prepared to justify their choices. A court should not treat the methods described here as a comprehensive checklist, however. No particular check is independently necessary or sufficient, but use of appropriate checks is required to ensure quality when the respondent is completing the survey online.

## *Mixed-Format Surveys*

A final approach to data collection does not depend on a single mode, but instead involves a mixed-mode approach. By combining modes, the survey design may increase the likelihood that all sampled members of the target population will be contacted. For example, a person without a landline may be reached by mail or email. Similarly, response rates may be increased if members of the target population are more likely to respond to one mode of contact versus another. For example, a person unwilling to be interviewed by phone may respond to a written or email contact. If a mixed-mode approach is used, the questions and structure of the questionnaires are likely to differ across modes, and the expert should be prepared to address the potential impact of mode on the answers obtained.<sup>203</sup>

200. These checks are described in greater length in Kugler & Henn, *supra* note 71, at 300–06.

201. Nat'l Fin. Partners Corp. v. Paycom Software, Inc., No. 14 C 7424, 2015 U.S. Dist. LEXIS 74700, at \*25–26 (N.D. Ill. June 10, 2015) (questioning the results of a survey in part because the expert did not appear to have even read the free response answers, which included responses like “cool” and “LOL,” when the expert should have excluded nonresponsive answers).

202. See E. Deborah Jay, *Generics Surveys in Trademark Disputes: Under the Gavel*, in *Trademark and Deceptive Advertising Surveys* 107, 113–16, 120–25 (Shari Diamond & Jerre Swann eds., 2d ed. 2022).

203. Don A Dillman & Benjamin L. Messer, *Mixed-Mode Surveys*, in *Handbook of Survey Research*, *supra* note 1, at 550, 553.

# Surveys Involving Interviewers

## Selection and Training of the Interviewers

When interviews are used to question survey respondents, the results are trustworthy only if “sound interview procedures were followed by competent interviewers.”<sup>204</sup> 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. Interviewers should be instructed to record verbatim the respondent’s answers, to indicate explicitly whenever they repeat a question to the respondent, and to record any statements they make to the respondent or supplementary questions they ask.

Interviewers require training to ensure that they can follow directions in administering the survey questions and employ optimal interviewing techniques (e.g., pausing to give the respondent enough time to answer, resisting invitations to express their own beliefs or opinions, knowing when and how to use probes). Although procedures vary, there is evidence that interviewer performance suffers with less than a day of training in general interviewing skills and techniques for new interviewers.<sup>205</sup> Additional training is needed when the interviewer is responsible for last-stage sampling (i.e., selecting the particular respondents to be interviewed in an unbiased fashion). Further, in-person interviews also should be conducted in situations without distractions and where others cannot overhear the answers. Failure to follow these dictates was one ground used by a court to reject as inadmissible a survey that purported to demonstrate consumer confusion.<sup>206</sup>

Some compromises may be accepted when surveys must be conducted swiftly. In trademark and deceptive advertising cases, the plaintiff’s usual request is for a preliminary injunction, because a delay means irreparable harm. Nonetheless, careful instruction and training of interviewers who administer the survey, as well as monitoring and validation to ensure quality control<sup>207</sup> and complete disclosure of the methods used for all of the procedures followed, are crucial

204. Toys “R” Us, Inc. v. Canarsie Kiddie Shop, Inc., 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983). *See also* Wisconsin v. Indivior IncNo. 16–5073, 2020 U.S. Dist. LEXIS 219949, at \*58 (E.D. Pa. Nov. 24, 2020) (praising a survey expert for “training interviewers carefully on interviewing techniques and the subject matter of the survey”).

205. Fowler & Mangione, *supra* note 139, at 117; Nora Cate Schaeffer et al., *Interviewers and Interviewing*, in *Handbook of Survey Research*, *supra* note 1, at 437, 460.

206. *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).

207. See section titled “Procedures Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias” below.

elements that, if compromised, seriously undermine the trustworthiness of any survey.

## Procedures 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, but can occur in the field via recordings (if respondents consent to it). 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.<sup>208</sup> This validation procedure does not determine whether the initial interview as a whole was conducted properly, but it warns interviewers that their work is being checked and can detect gross failures in the administration of the survey. In computer-assisted interviews, further validation information can be obtained from the timings that can be automatically recorded when an interview occurs.

A third way to verify that the interviews were conducted properly is to examine the work done by each individual interviewer. By reviewing the interviews and individual responses recorded by each interviewer and comparing patterns of response across interviewers, researchers can identify any response patterns or inconsistencies that warrant further investigation. When interviewers conduct the survey, the identity of the interviewer (or a unique code) should be included in the data provided to the opposing party and trier of fact.

When a survey is conducted at the request of a party for litigation rather than in the normal course of business, a heightened standard for validation checks may be appropriate. Thus, independent validation of a random sample of interviews by a third party rather than by the field service that conducted the interviews increases the trustworthiness of the survey results.<sup>209</sup>

208. See, e.g., *Davis v. S. Bell Tel. & Tel. Co.*, No. 89-2839-CIV-NESBITT, 1994 U.S. Dist. LEXIS 13257, at \*16 (S.D. Fla. Feb. 1, 1994); *Nat'l Football League Props., Inc. v. N.J. Giants, Inc.*, 637 F. Supp. 507, 515 (D.N.J. 1986).

209. 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."

## Accuracy of Data Entry

Analyzing the results of a survey requires that the data obtained on each sampled element be recorded 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.

The need for these checks and rules to control mistakes in data entry is particularly great when data collected from survey respondents are combined with data on those same respondents obtained from institutional databases or records (e.g., a survey of jurors is combined with court data on their jury service).<sup>210</sup> In such cases, both data entries and matches need to be closely scrutinized.

## Disclosure and Reporting

### Early Disclosure About the Survey Methodology and Results

Objections to the definition of the relevant population, the method of selecting the sample, and the wording of questions generally are raised for the first time when the results of the survey are presented. By that time, it is often too late to correct methodological deficiencies that could have been addressed in the planning stages of the survey. The plaintiff in a trademark case<sup>211</sup> submitted a set of proposed survey questions to the trial judge, who ruled that the survey results would be admissible at trial while reserving the question of the weight the evidence would be given.<sup>212</sup> The Seventh Circuit called this approach a commendable procedure and suggested that it would have been even more desirable if the parties had “attempt[ed] in good faith to agree upon the questions to be in such a survey.”<sup>213</sup>

210. See generally Jan Van den Broeck et al., *Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities*, 2 PLoS Med 2(10): e267, <https://doi.org/10.1371/journal.pmed.0020267>. See also Mary R. Rose & Marc A. Musick, *How Can You Tell if There Is a Crisis? Data and Measurement Challenges in Assessing Jury Representation*, 98 Chicago-Kent L. Rev. 35, 42 (2023).

211. *Union Carbide Corp. v. Ever-Ready, Inc.*, 392 F. Supp. 280, 291 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366 (7th Cir. 1976).

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

213. *Union Carbide*, 531 F.2d at 386. On one occasion, the Seventh Circuit recommended filing a motion in limine, asking the district court to determine the admissibility of a survey based on an examination of the survey questions and the results of a preliminary survey before the party undertakes the expense of conducting the actual survey. *Piper Aircraft Corp. v. Wag-Aero, Inc.*,

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.”<sup>214</sup> The parties then were encouraged to attempt to resolve any methodological disagreements before the survey was conducted.<sup>215</sup> Although this passage in the second edition of the *Manual* has been cited with apparent approval,<sup>216</sup> the prior agreement that the *Manual* recommends has occurred rarely, and the *Manual for Complex Litigation, Fourth*, recommends, but does not advocate requiring, prior disclosure and discussion of survey plans.<sup>217</sup> As the *Manual* suggests, however, early disclosure can enable the parties to raise prompt objections that may permit corrective measures to be taken before a survey is completed.<sup>218</sup>

Rule 26 of the Federal Rules of Civil Procedure requires extensive disclosure of the basis of opinions offered by testifying experts. However, Rule 26 does not produce disclosure of all survey materials, because parties are not obligated to disclose information about nontestifying experts. Parties considering whether to commission or use a survey for litigation are not obligated to present a survey that produces unfavorable results. Prior disclosure of a proposed survey instrument places the party that ultimately would prefer not to present the survey in the position of presenting damaging results or leaving the impression that the results are not being presented because they were unfavorable. Anticipating such a situation, parties do not decide whether an expert will testify until after the results of the survey are available.

Nonetheless, courts can encourage early disclosure and discussion even if they do not lead to agreement between the parties. In *McNeilab, Inc. v. American Home Products Corp.*,<sup>219</sup> Judge William C. Conner encouraged the parties to submit their survey plans for court approval to ensure their evidentiary value; the plaintiff did

741 F.2d 925, 929 (7th Cir. 1984). On another occasion, the parties jointly developed a survey administered by a neutral third-party survey firm. *Scott v. City of New York*, 591 F. Supp. 2d 554, 560 (S.D.N.Y. 2008) (survey design, including multiple pretests, negotiated with the help of the magistrate judge).

214. *Manual for Complex Litigation, Second* § 21.484 (1985).

215. *See id.*

216. *See, e.g., Nat'l Football League Props., Inc. v. N.J. Giants, Inc.*, 637 F. Supp. 507, 514 n.3 (D.N.J. 1986).

217. MCL 4th, *supra* note 41, § 11.493 (“including the specific questions that will be asked, the introductory statements or instructions that will be given, and other controls to be used in the interrogation process”).

218. *See id.*

219. 848 F.2d 34, 36 (2d Cir. 1988) (discussing with approval the actions of the district court). *See also Hubbard v. Midland Credit Mgmt.*, No. 1:05-cv-0216, 2009 U.S. Dist. LEXIS 13938 (S.D. Ind. Feb. 23, 2009) (court responded to plaintiff’s motions to approve survey methodology with a critique of the proposed methodology).

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.<sup>220</sup>

## Completeness and Accuracy of All Relevant Information in the Survey Report

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. The American Association for Public Opinion Research (AAPOR), a professional organization that brings together producers and users of survey data, offers standards for disclosure to be reported with any survey results.<sup>221</sup> The following list, which draws on the AAPOR guidelines, describes the elements that a survey report generally should include:

1. The purpose of the survey, its research sponsor, and those who conducted the research;
2. A definition of the target population;
3. A description of how the sample was recruited, including the method of selection (and how much of the population is covered); population quotas used, if any; respondent compensation, if any; the method (mode) of interview; the number of callbacks; respondent eligibility or screening criteria and method; and other pertinent information (e.g., the name of the vendor, if used);
4. A description of the results of sample implementation, including the number of
  - a. potential respondents contacted,
  - b. potential respondents not reached,
  - c. noneligibles,
  - d. refusals,
  - e. incomplete interviews or terminations, and
  - f. completed interviews and final sample size;
5. The dates of data collection;
6. The exact wording of the questions used, including a copy of the actual questionnaire (and screenshots of the questionnaire as viewed by

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

221. AAPOR also provides professional ethics guidelines as well as a transparency initiative where survey organizations can be certified as publicly disclosing their basic research methods and making them public, <https://perma.cc/4GPB-7QKK>.

respondents if administered electronically), interviewer instructions, and visual exhibits, with any cross-condition variations or randomizations marked;<sup>222</sup>

7. A description of any weighting or estimating procedures used;
8. A description of data-processing procedures (e.g., validity checks such as measures of the time respondents took to complete the survey, any data-imputation procedures, criteria for removing any responses);
9. Estimates of the sampling error, where appropriate (i.e., in probability samples);
10. Statistical tables clearly labeled and identified regarding the source of the data, including the number of raw cases forming the base for each table, row, or column;
11. Copies of interviewer instructions, validation results, and code books, including coding instructions for free-response data;<sup>223</sup> and
12. Acknowledgment of limitations to the survey.

As a general rule, any research should provide such information or explain why doing so is not possible. Additional information to include in the survey report may depend on the nature of sampling design. For example, reported response rates along with the exact time each interview occurred may assist in evaluating the likelihood that nonresponses biased the results. In a survey designed to assess the duration of employee pre-shift activities, workers were approached as they entered the workplace; records were not kept on refusal rates or the timing of participation in the study. Thus, it was impossible to rule out the plausible hypothesis that individuals who arrived early for their shift with more time to spend on pre-shift activities were more likely to participate in the study.<sup>224</sup>

222. The questionnaire itself can often reveal important sources of bias. *See Marria v. Broad-dus*, 200 F. Supp. 2d 280, 289 (S.D.N.Y. 2002) (court excluded survey sent to prison administrators based on questionnaire that began, “We need your help. We are helping to defend the NYS Department of Correctional Service in a case that involves their policy on intercepting Five-Percenter literature. Your answers to the following questions will be helpful in preparing a defense.”).

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

224. *See Chavez v. IBP, Inc.*, No. CT-01-5093-RHW, 2004 U.S. Dist. LEXIS 28837 (E.D. Wash. Aug. 18, 2004).

Researchers should also ask, and disclose, whether respondents have previously participated in similar prior surveys. Many nonprobability samples come from vendors with online panels where respondents participate in multiple surveys over time. There also are vendors who have established analogous online panels with probability samples.<sup>225</sup> Some evidence suggests that repeat respondents from these panels differ from fresh sample draws.<sup>226</sup> Vendors often do not provide information on prior participation—something about which a researcher can inquire prior to using a given vendor. Even if not, researchers can, on their own, attempt to identify “professional respondents” and consider whether they might bias inferences (e.g., by measuring prior participation and evaluating its effect on outcomes).<sup>227</sup> If the expert asks participants whether they have participated in previous studies on the same topic, or obtains that information from other sources, the results obtained from the experienced and naive participants can be compared. Currently, many experts take the sensible approach of simply excluding participants who report having completed similar prior surveys. Regardless, researchers should always disclose prior participation and its possible effects or be explicit that such information does not exist and explain why it was not obtained.<sup>228</sup>

When the presenting party has access to the raw data from a survey, that data should generally be made available to opposing counsel upon request. This dataset should include responses from all participants whose data were ultimately used in the original expert’s report and also all participants whose data were excluded by the original expert for reasons of quality after the completion of the survey, with an indication of why they were excluded (i.e., data-processing procedures).<sup>229</sup> If a researcher wishes to exclude unusually fast and slow responses, they should disclose this in the report and provide the data from those respondents as well.<sup>230</sup> This permits an opposing expert to rerun statistical analyses and determine if any debatable analytic choices made by the original expert had substantial effects on the report’s conclusions.

225. Online Panel Research: A Data Quality Perspective (Mario Callegaro et al, eds., 2014).

226. Andrew Halpern-Manners & John Robert Warren, *Panel Conditioning in Longitudinal Studies: Evidence From Labor Force Items in the Current Population Survey*, 49 Demography 1499 (2012), <https://doi.org/10.1007/s13524-012-0124-x>.

227. D. Sunshine Hillygus et al., *Professional Respondents in Nonprobability Online Panels*, in Online Panel Research: A Data Quality Perspective 219–37 (2014).

228. K.H. Jamieson et al., *Protecting the Integrity of Survey Research*, 2 PNAS Nexus 1 (2023), <https://doi.org/10.1093/pnasnexus/pgad049>. More generally, these authors supplement AAPOR’s code by suggesting researchers should provide details on question order, details on respondent attrition (i.e., dropping out of the survey), and so on.

229. For example, an expert may exclude respondents who responded with gibberish to free-response items or who completed the survey much more quickly than did other participants.

230. The issue of completion time-based exclusions is explored at greater length in Kugler & Henn, *supra* note 71, at 305–06. At present, there is no agreed-upon standard for identifying “too fast” responses; most researchers appear to be using rules of thumb (for example, excluding those who took less than one third the median time). *Id.*

The public (and, presumably, also judges) sometimes express concern that surveys are not sufficiently reliable, generally citing polling accuracy in recent elections.<sup>231</sup> Election pollsters have the unenviable task of modeling a changing electorate's intended behavior in a circumstance where it matters a great deal whether the correct answer is 49% support or 51% support. Other surveys do not require this level of precision relative to a specified standard (i.e., greater than 50%). Rather, the relevant estimate the survey is seeking to reflect is whether approximately 20% of people (vs. 40%, 60%, or even 5%) in the target market are confused by a given advertisement. Thus, the accuracy of a survey estimate should be assessed in light of the degree of precision needed for the context.

Though transparency is very important, some information must be removed from data files before they are shared beyond the survey expert to protect participant confidentiality. All identifying information, such as the respondent's name, address, and telephone number, should be omitted. In the case of internet surveys, it is appropriate to also remove the participant's panel ID number, IP address, and precise geolocation information; these might all be linked to the participant's identity.<sup>232</sup> Keeping the survey duration and survey start time in the file may aid analysis, however, and will not compromise participant anonymity in most cases.

Greater efforts to promote participant anonymity are appropriate in cases where the population surveyed is small or otherwise susceptible to easy reidentification. If a survey is conducted of a workplace, for example, it may be that the total population numbers only in the hundreds. The identity of a respondent of any given age, ethnicity, and gender may therefore be narrowed down to one of only a few possibilities.<sup>233</sup> Much greater care should be taken to protect participant anonymity in such cases. Possibilities include omitting data fields that the opposing expert does not expect to use in their own analyses, or restricting access

231. See, e.g., Scott Keeter et al., Pew Rsch. Ctr., *What 2020's Election Poll Errors Tell Us About the Accuracy of Issue Polling* (Mar. 2, 2021), <https://perma.cc/LDF9-KUWK> (commenting on this concern and evaluating it). For a discussion of the accuracy of 2020 polling, see American Association of Public Opinion Research, *Task Force on 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls*, <https://perma.cc/64K9-GVDJ> (not reaching firm conclusions, but suggesting that this may have been due to greater nonresponse among Trump voters and difficulty in accounting for new voters in screening).

232. IP addresses are sometimes used to detect whether individuals took the survey more than once. If so, shared data should indicate which respondents shared an IP address but should generally not include the IP address itself.

233. This problem also arises in the case of privacy statutes such as HIPAA and FERPA, where deidentified data can be shared but identifiable data is tightly restricted. The Department of Education reviews a number of questions that may be relevant to a school or workplace survey in their FERPA guidance, <https://perma.cc/CCD3-47Y8> (updated May 2013).

to the data to ensure that it is not shared beyond the opposing expert—and particularly not with the ultimate client.<sup>234</sup>

The respondents questioned in a survey generally do not testify in legal proceedings and are unavailable for cross-examination. Indeed, one of the advantages of a survey is that it avoids a repetitious and unrepresentative parade of witnesses. To verify that interviews occurred with qualified respondents, standard survey practice includes validation procedures,<sup>235</sup> the results of which should be included in the survey report.

Conflicts may arise when an opposing party asks for survey respondents' names and addresses so that they can re-interview some respondents. The party introducing the survey or the survey organization that conducted the research generally resists supplying such information.<sup>236</sup> Professional surveyors as a rule promise confidentiality to increase participation rates and encourage candid responses although, to the extent that identifying information is collected, such promises may not effectively prevent a lawful inquiry. Because failure to extend confidentiality may bias both the willingness of potential respondents to participate in a survey and their responses, the professional standards for survey researchers generally prohibit disclosure of respondents' identities. "The use of survey results in a legal proceeding does not relieve the Survey Research Organization of its ethical obligation to maintain in confidence all Respondent-identifiable information or lessen the importance of Respondent anonymity."<sup>237</sup> Although no surveyor–respondent privilege currently is recognized, the need for surveys and the availability of other means to examine and ensure their trustworthiness argue for deference to legitimate claims for confidentiality in order to avoid seriously compromising the ability of surveys to produce accurate information.<sup>238</sup> In

234. In the case of a workplace survey, the client might well be the current employer of the survey participant.

235. See section titled "Procedures Used to Ensure and Determine That the Survey Was Administered to Minimize Error and Bias" above.

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

237. CASRO, *supra* note 41, § I.A.3f; Am. Ass'n for Pub. Op. Rsch., AAPOR Code of Professional Ethics and Practices (revised Apr. 2021), <https://perma.cc/DBC8-LTWJ>.

238. *United States v. Dentsply Int'l, Inc.*, No. 99-5 MMS, 2000 U.S. Dist. LEXIS 6994, at \*23 (D. Del. May 10, 2000) (Fed. R. Civ. P. 26(a)(1) does not require party to produce the identities of individual survey respondents); *In re Litton Indus., Inc.*, No. 9123, 1979 FTC LEXIS 311, at \*13 & n.12 (June 19, 1979) (Order Concerning the Identification of Individual Survey-Respondents with Their Questionnaires) (citing Frederick H. Boness & John F. Cordes, *The Researcher-Subject Relationship: The Need for Protection and a Model Statute*, 62 Geo. L.J. 243, 253 (1973)); see also *Applera Corp. v. MJ Rsch., Inc.*, 389 F. Supp. 2d 344, 350 (D. Conn. 2005) (denying access to names of survey respondents); *Lamphire 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

general, the better approach is for the opposing party to conduct their own survey rather than re-interrogate the participants in the previous one.

## Concluding Observations

As this chapter has shown, judges have a variety of factors to consider when determining whether a survey should be admitted and how much weight it should be given. The *Manual for Complex Litigation, Fourth (MCL4)*, published in 2004, suggested a set of relevant factors for judges to evaluate.<sup>239</sup> In the past twenty years since the *MCL4* was published, survey methods have evolved (e.g., with the growth of internet and other computer technology, and recognition of the importance of survey-experimental methodology for causal inference), but these factors remain important. Thus, we close by presenting an updated list that clarifies and builds on the list of factors presented in the *MCL4*.<sup>240</sup>

Relevant factors include whether:

- the population was properly chosen and defined;
- the sampling frame and the sample itself were representative of that population;
- the data and details on data collection were fully and accurately reported;
- the survey questions asked were clear and not leading;
- the survey was conducted by qualified persons following proper procedures;
- the data were analyzed following accepted statistical principles;
- statements about causal relationships were supported by sufficient evidence about causality; and
- the results were presented in accordance with statistical standards.<sup>241</sup>

“any order which justice requires to protect a party or persons from annoyance, embarrassment, oppression, or undue burden or expense”) (citation omitted).

239. MCL 4th, *supra* note 41, § 11.493.

240. The *Manual for Complex Litigation* distinguished between factors to be considered regarding admissibility and factors to be considered in assigning weight. As all of the factors can affect either admissibility or weight, depending on their quality, we have combined them in one set.

241. These include, where appropriate, information on sampling error and confidence intervals.

## Glossary of Terms

The following terms and definitions were adapted from a variety of sources, including *Handbook of Survey Research* (Peter H. Rossi et al. eds., 1st ed. 1983; Peter V. Marsden & James D. Wright eds., 2d ed. 2010); *Measurement Errors in Surveys* (Paul P. Biemer et al. eds., 1991); Willem E. Saris, *Computer-Assisted Interviewing* (1991); Seymour Sudman, *Applied Sampling* (1976).

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

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

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

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

**closed-ended question.** A question that provides the respondent with a list of choices and asks the respondent to choose from among them.

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

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

**conjoint survey.** Survey-experiment designed to identify and estimate the causal effects of many treatment components simultaneously by using an orthogonal, fractional factorial design.

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

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

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

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

**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.

**margin of error.** An indication of the likely precision of an estimate from a probability sample; used to compute a confidence interval.

**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 their own response.

**order effect.** A tendency of respondents to choose an item based in part on the order of response alternatives on the questionnaire (see primacy effect and recency effect).

**parameter.** See population value.

**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. 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 probability sample.

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

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

**sampling error.** The estimated size of the difference between the result obtained from a sample study and the result that would be obtained by attempting a complete study of all units in the sampling frame from which the sample was selected in the same manner and with the same care.

**sampling frame.** The source or sources from which the 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 in which the researcher subdivides the population into mutually exclusive and exhaustive subpopulations, or strata; within these strata, separate samples are selected. Results can be combined to form overall population estimates or used to report separate within-stratum estimates.

**survey-experiment.** A survey with randomly assigned control and treatment groups, enabling the researcher to test a causal proposition.

**survey 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).

**universe.** See population.

## References on Survey Research

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

Mick P. Couper, *Designing Effective Web Surveys* (2008).

Matthew DeBell, *Computation of Survey Weights*, in *The Palgrave Handbook of Survey Research*, 519 (David L. Vannette & Jon A. Krosnick eds., 2018).

Shari S. Diamond, *Control Foundations: Rationale and Approaches*, in *Trademark and Deceptive Advertising Surveys: Law, Science and Design* 239 (Shari S. Diamond and & Jerre Swann, eds., 2d ed. 2022).

Don A. Dillman, Jolene Smyth & Leah M. Christian, *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method* (3d ed. 2009).

James N. Druckman, *Experimental Thinking: A Primer on Social Science Experiments* (2022).

*Experimental Methods in Survey Research: Techniques that Combine Random Sampling and Random Assignment* (Paul J. Lavrakas, Michael W. Traugott, Courtney Kennedy, Allyson L. Holbrook, Edith D. de Leeuw & Brady T. West eds., 2019).

Arlene Fink, *How to Conduct Surveys: A Step-By-Step Guide* (4th ed. 2009).

Robert M. Groves, Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer & Roger Tourangeau, *Survey Methodology* (2d ed. 2009).

*Handbook of Survey Research* (Peter V. Marsden & James D. Wright eds., 2d ed. 2010).

Matthew B. Kugler & R. Charles Henn, *Internet Surveys in Trademark Cases: Benefits, Challenges, and Solutions*, in *Trademark and Deceptive Advertising Surveys: Law, Science and Design*, 293 (Shari S. Diamond and Jerre B. Swann, eds., 2d ed. 2022).

Sharon Lohr, *Sampling: Design and Analysis* (2d ed. 2010).

*Measurement Errors in Surveys* (Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz & Seymour Sudman eds., 2004).

*Online Panel Research: A Data Quality Perspective* (Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick & Paul J. Lavrakas eds., 2014).

*The Palgrave Handbook of Survey Research* (David L. Vannette & Jon A. Krosnick eds., 2018).

*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).

William Shadish, Thomas D. Cook & Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inferences* (2002).

Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur & Roger Tourangeau, *Cognition and Survey Research* (1999).

Seymour Sudman, *Applied Sampling* (1976).

*Survey Nonresponse* (Robert M. Groves, Don A. Dillman, John L. Eltinge & Roderick J. A. Little eds., 2002).

*Telephone Survey Methodology* (Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey & William L. Nicholls eds., 1988).

Roger Tourangeau, Lance J. Rips & Kenneth Rasinski, *The Psychology of Survey Response* (2000).

*Trademark and Deceptive Advertising Surveys: Law, Science and Design* (Shari S. Diamond & Jerre B. Swann eds., 2d ed. 2021).