

Fall 1-1-2012

What's Good in Theory May Be Flawed in Practice: Potential Legal Consequences of Poor Implementation of a Theoretical Sample

A. Lynn Phillips

G. Michael Phillips

Melanie Stallings Williams

Follow this and additional works at: https://repository.uchastings.edu/hastings_business_law_journal



Part of the [Business Organizations Law Commons](#)

Recommended Citation

A. Lynn Phillips, G. Michael Phillips, and Melanie Stallings Williams, *What's Good in Theory May Be Flawed in Practice: Potential Legal Consequences of Poor Implementation of a Theoretical Sample*, 9 *Hastings Bus. L.J.* 77 (2012).
Available at: https://repository.uchastings.edu/hastings_business_law_journal/vol9/iss1/3

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in *Hastings Business Law Journal* by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.

What's Good in Theory May Be Flawed in Practice: Potential Legal Consequences of Poor Implementation of a Theoretical Sample

A. Lynn Phillips,* G. Michael Phillips** and Melanie Stallings Williams***

The article discusses the problem with the use of statistical sampling in litigation. Sample-based research is increasingly used in cases as diverse as products liability, antitrust, intellectual property, and criminal law. Sample-based research provides objective evidence upon which decisions, damages, and liability may rest. Despite its importance, however, statistical evidence is often misused and misunderstood by attorneys who may be unfamiliar with the underlying form of analysis. The paper explores common errors when using litigative samples, comments upon best practices for the use in law of sample-based research, and demonstrates the importance of sound statistical sampling and data collection in a recent case.

* M.S., (candidate) University of Nebraska, Lincoln; B.A., Grove City College, 2011.

** Professor of Finance, Real Estate and Insurance, California State University, Northridge. Ph.D., University of California, San Diego, 1982; M.A., B.S., Illinois State University, 1980, 1977.

*** Professor & Chair, Department of Business Law, California State University, Northridge; J.D., Boston University, 1981; B.A., University of California, Santa Cruz, 1977.

I. INTRODUCTION

In recent decades, statistical sampling has grown in importance in litigation. Used in cases as diverse as class-action suits regarding discrimination or product defects, collective trials of mass torts, and instances of “antitrust, environment, tax, rate-making, and a variety of other cases in which issues of fact are sought to be resolved,” sample-based research provides solid quantitative evidence on which rulings, damages, and liability can and have been based.¹ Unfortunately, while the theory behind sampling is straightforward, implementing such a theory can lead to significant problems for those unaware of the statistical work that must be done before and after data collection to ensure that any resulting analysis is accurate and useful. This problem is confounded by the lack of experience many attorneys have with statistical theory and practice. This paper will address several common errors made when implementing theoretically sound litigative samples, comment upon best practices for sample-based research, and demonstrate the importance of sound statistical sampling and data collection techniques in a recent case.

II. OVERVIEW OF THE SAMPLING PROCESS

Sampling gives researchers a statistically sound, legally admissible, way to make predictions about a large population.² As opposed to a census, which is a collection of relevant data about all members of a specific population, a sample is a portion of that population from which one can collect data, then extrapolate from the results to the population as a whole.³ However, as David H. Kaye and David A. Freedman noted,

1. The Special Committee on Empirical Data in Legal Decision Making, *Recommendations on Pretrial Proceedings in Cases with Voluminous Data*, in *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS* 245 (Stephen E. Fienberg, ed. 1989). See also Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 *STAN. L. REV.* 815, 824, 835 (1992); Joseph C. Bright, Jr., Joseph B. Kadane & Daniel S. Nagin, *Statistical Sampling in Tax Audits*, 13 *LAW & SOC. INQUIRY* 310, 310 (1988); Laurens Walker & John Monahan, *Sampling Liability*, 85 *VA. L. REV.* 329, 339 (1999).

2. David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 211, 214 (3d ed., 2011); Shari Seidman Diamond, *Reference Guide on Survey Research*, in *REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 359, 363–67 (3d ed. 2011). See also The Special Committee on Empirical Data in Legal Decision Making, *Recommendations on Pretrial Proceedings in Cases with Voluminous Data*, in *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS* 245, 245 (Stephen E. Fienberg ed. 1989); Michael J. Saks & Peter David Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 *STAN. L. REV.* 815, 826–27 (1992); Laurens Walker & John Monahan, *Sampling Evidence at the Crossroads*, 80 *S. CAL. L. REV.* 969, 971 (2007).

3. Kaye & Freedman, *supra* note 2, at 216–17.

“inferences from the part to the whole are justified only when the sample is representative.”⁴ A sample that accurately represents the population from which it is drawn can be a powerful tool.⁵ An unrepresentative sample, on the other hand, can be useless or even harmful to one’s position.⁶ This is important in litigation for two reasons. First, it presents an opportunity to present evidence that can support a compelling argument. Second, there is the danger of unreliable data going unchallenged if practitioners and the court are unfamiliar with the research method.⁷ While it has been said of law students that they “are typically smart people who do not like math,”⁸ it behooves attorneys to learn the effective use of sampling both to promote a client’s position and to block misleading or unsubstantiated evidence from consideration.

III. HISTORY OF SAMPLING AND ITS ADMISSIBILITY

A. USE AS EVIDENCE MATHEMATICALLY AND SCIENTIFICALLY PRECEDED LEGAL ADMISSIBILITY

The origins of sampling are obscure, perhaps beginning with the notion of a food taster to test whether a king’s food had been poisoned.⁹ On a more philosophical level, some commentators view empirical research itself as the first example of sampling. Since “all empirical knowledge is, in a fundamental sense, derived from incomplete or imperfect observations” it is, therefore, “a sampling of experience.”¹⁰ By the eighteenth century, it was well-established that taking a sample of a population (rather than a census) could lead to mathematically valid results.¹¹ Still, it was a flawed process. Prior to the twentieth century, sampling (or “partial investigations”) was conducted by a researcher who

4. Kaye & Freedman, *supra* note 2, at 216–17.

5. Diamond, *supra* note 2, at 362, 372.

6. *Id.* at 377–78.

7. Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116, 126 (1978). Margaret Bull Kovera & Bradley D. McAuliff, *The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?*, 85 J. APPLIED PSYCHOL. 574, 574 (2000). Margaret Bull Kovera, Melissa B. Russano, & Bradley D. McAuliff, *Assessment of the Commonsense Psychology Underlying Daubert: Legal Decision Makers’ Abilities to Evaluate Expert Evidence in Hostile Work Environment Cases*, 8 PSYCHOL. PUB. POL’Y & L. 180, 180 (2002).

8. Michael J. Saks, *Legal Policy Analysis and Evaluation*, 44 AM. PSYCHOL. 1110, 1115 (1989).

9. Laurens Walker & John Monahan, *Sampling Evidence at the Crossroads*, 80 S. CAL. L. REV. 969, 971 n.11 (2007) (citing Frederick F. Stephan, *History of the Uses of Modern Sampling Procedures*, 434 J. AM. STAT. ASS’N 12, 13 (1948)).

10. *Id.* at n.12.

11. CHARLES COULSTON GILLISPIE, PIERRE-SIMON LAPLACE, 1749-1827: A LIFE IN EXACT SCIENCE 93–94 (1997).

selected those whom he considered to represent the larger whole.¹² The inherent problem in this technique, of course, is obvious: a bias in selection biases the outcome. It was not until the twentieth century that the process of random sampling became an established form of empirical research.¹³ Randomization, properly conducted, eliminates biases that can invalidate the inference.¹⁴

Attempts were made to decrease selection bias by creating a “purposive selection,” i.e., a selection of representative participants.¹⁵ Problematically, this process called upon the designer to decide what characteristics qualified as “representative” of the pool. Further development of “purposive selection” led to the use of a quota system, which required pollsters to obtain responses from groups with quotas imposed for particular characteristics (for example, age, sex, or income).¹⁶ The purpose, again, was to try to obtain representative responses.

One early demonstration of “purposive quota sampling” involved the conflicting predictions of the presidential election of 1936.¹⁷ The Literary Digest, a longtime predictor of election outcomes, used lists obtained from those who owned telephones or automobiles.¹⁸ Based on their survey of more than ten million people, the journal predicted that the winner would be Alfred Landon, beating Franklin D. Roosevelt by a 55-41 margin.¹⁹ George Gallup, using a much smaller sample of 3,000 people, but with quotas over six variables predicted Roosevelt’s win by a 54-46 margin.²⁰ Roosevelt won with 61 percent of the vote.²¹ Gallup’s prediction accuracy was aided by the use of quota sampling, but also by avoiding the selection bias of culling only respondents who owned what were, at the time, expensive and relatively uncommon consumer goods.²² The Literary Digest became defunct two years later,²³ while Gallup, of course, went on

12. Walker & Monahan, *supra* note 9, at 971–72.

13. *Id.* at 971.

14. Joseph B. Kadane, *Probability Sampling in Litigation*, 18 CONN. INS. L.J. 297, 297 (2011-2012).

15. Walker & Monahan, *supra* note 9, at 971–72.

16. Kaye & Freedman, *supra* note 2, at 225.

17. Walker & Monahan, *supra* note 9, at 972.

18. Cynthia Crossen, *Fiasco in 1936 Survey Brought ‘Science’ to Election Polling*, WALL ST. J., Oct. 2, 2006, at B.1.

19. Peeverill Squire, *Why the 1936 Literary Digest Poll Failed*, 52 PUB. OPIN. Q. 125, 126–27 (1988).

20. Walker & Monahan, *supra* note 9, at 973.

21. Squire, *supra* note 19, at 127.

22. Crossen, *supra* note 18. *But cf.* Dominic Lusinchi, “President” Landon and the 1936 Literary Digest Poll: Were Automobile and Telephone Owners to Blame?, 36 SOCIAL SCIENCE HISTORY 23, 32 (2012), available at <http://ssh.dukejournals.org/content/36/1/23.full.pdf+html> (noting that nonresponse bias was a more significant factor in the error); Squire, *supra* note 19, at 131 (low response rate and nonresponse bias were significant factors).

23. Frank Rich, *The Moose Stops Here*, N.Y. TIMES, Nov. 16, 2008, at 12.

to be a pioneer in the field.²⁴

B. PROBLEMS WITH EXPERT TESTIMONY GENERALLY: THE TREND FROM *FRYE TO DAUBERT*

Concurrent with the gradual acceptance of sampling results was a changing legal acceptance of expert evidence generally.²⁵ Common law originally made no distinction between expert testimony and any other witness testimony.²⁶ The first apparent record of expert testimony introduced by a party was in *Folkes v. Chadd*.²⁷ Early American cases permitted the use of expert testimony if the evidence was beyond the understanding of a typical juror and an expert of relevant expertise could testify in ways that would assist the jurors' understanding.²⁸

As scientific methods became more refined and more accepted generally, they became more accepted in the courtroom.²⁹ Courts initially performed little review of scientific or expert evidence, simply weighing probative value against prejudicial effect.³⁰ The weight, if any, that such testimony was to be accorded rested largely with triers of fact.³¹ This changed with *Frye v. United States*.³² Alphonse Frye, accused of murder, sought to introduce evidence of his innocence based on a test created by psychologist and Harvard professor William Moulton Marston.³³ Professor Marston had developed a forerunner of a lie detector test and was prepared to testify that the test had demonstrated Frye's innocence.³⁴ The trial court excluded the evidence as too speculative.³⁵ In affirming the decision, the U.S. Court of Appeals noted that "when a scientific principle or discovery

24. See generally Steven H. Chaffee, *George Gallup and Ralph Nafziger: Pioneers of Audience Research*, 3 MASS. COMM. & SOC'Y 317 (2000).

25. See D.H. Kaye, *The Dynamics of Daubert: Methodology, Conclusions, and Fit in Statistical and Econometric Studies*, 87 VA. L. R. 1933, 1941 (2001).

26. *Id.* at 1938 (citing SIMON GREENLEAF, A TREATISE ON THE LAW OF EVIDENCE, 483 (1958)).

27. John Basten, *The Court Expert in Civil Trials – A Comparative Appraisal*, 40 MOD. L. REV. 174, 175–76 (1977) (citing 99 *Folkes v. Chadd*, Eng. Rep. 589, 589–90 (1972)) (noting that before the 16th century, what we would today call experts were members of the jury. Until *Folkes v. Chadd*, it is unclear whether experts were called by the court or by parties). Cf. TAL GOLAN, LAWS OF MEN AND LAWS OF NATURE: THE HISTORY OF SCIENTIFIC EXPERT TESTIMONY IN ENGLAND AND AMERICA 20–22 (2004) (By the sixteenth century, the process of having the parties call expert witnesses was well established in England. This is distinct from the practice of permitting court-appointed experts. The earliest reference for the latter is 1299, when physicians and surgeons were called to testify in London concerning the medical value of wolf flesh.).

28. See, e.g., MCCORMICK ON EVIDENCE 23 (John W. Strong ed., West Group 6th ed. 2006).

29. Kaye, *supra* note 25, at 1938–39.

30. *Id.*

31. *Id.*

32. 293 F. 1013 (D.C. Cir. 1923).

33. See *id.* See also William G. Iacono & Christopher J. Patrick, *Polygraph ("Lie Detector") Testing: The State of the Art*, in THE HANDBOOK OF FORENSIC PSYCHOLOGY (Allen K. Hess & Irving B. Weiner eds., 1999).

34. See *Frye*, 293 F. at 1013; see also Iacono & Patrick, *supra* note 33.

35. *Id.*

crosses the line between the experimental and the demonstrable stages is difficult to determine,”³⁶ that expert evidence could be admitted only when it had gained general acceptance in the field of expertise.³⁷

The *Frye* “general acceptance” standard survived until *Daubert v. Merrell Dow Pharmaceuticals, Inc.*³⁸ In *Daubert*, the U.S. Supreme Court held that trial judges had a “gatekeeping role” that required them to admit only evidence that was reliable.³⁹ This required courts to establish both the type of evidence warranting increased scrutiny and then to determine whether it was sufficiently reliable to warrant admission.⁴⁰ There were two obvious problems: first, determining what kinds of evidence triggered this heightened judicial scrutiny and second, reducing the risk that a court would wrongfully exclude evidence from a jury.⁴¹ The *Frye* standard was modified when the U.S. Supreme Court extended the reliability standard to all expert testimony.⁴² While *Kumho* therefore solved the “boundary problem” of determining when the court’s gatekeeping function was triggered,⁴³ it remained a risk that courts would improperly usurp a jury’s role in considering the weight to be given expert testimony (the “usurpation problem”⁴⁴).⁴⁵ The Federal Rules of Evidence, however, “admit a somewhat broader range of scientific testimony than would have been admissible under *Frye*” while leaving in place a trial court’s “gatekeeping” role in screening the reliability of such evidence.⁴⁶ The judicial role was expanded in *General Electric Co. v. Joiner*, which encouraged trial courts to examine the validity not only of the scientific principles and methods but to review, as well, the expert’s conclusions. Encouraging an integrated and comprehensive approach to admissibility, the court noted that “conclusions and methodology are not entirely distinct from one another.”⁴⁷ A court may properly exclude evidence, the *Joiner* court held, when “there is simply too great an analytical gap between the data and the opinion proffered.”⁴⁸ This calls for a much-expanded role for the judiciary in evaluating the admission of expert testimony.

36. *Frye*, 293 F. at 1014.

37. *Id.*

38. 509 U.S. 579, 587 (1993).

39. *Id.* at 597.

40. *Id.* at 592–93.

41. See Kaye, *supra* note 25, at 1941–47.

42. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 147 (1999).

43. Kaye, *supra* note 25, at 1935.

44. *Id.*

45. *Id.* For an example of a rebuke of an appellate court’s “overly stringent” exclusion of expert testimony, see *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 143 (1997).

46. *Joiner*, 522 U.S. at 142.

47. *Joiner*, 522 U.S. at 142.

48. *Id.*

C. PROBLEMS WITH LEGAL USE OF STATISTICAL EVIDENCE

While courts are asked to increase their scrutiny of not only an expert's methodology and principles but the expert's conclusions, attorneys and judges get little exposure to the type of training needed for meaningful analysis of statistical evidence.⁴⁹ "Law students," after all, "are typically smart people who do not like math."⁵⁰ Historically, this usually resulted in skepticism of such evidence.⁵¹ We therefore examine the early resistance to the use of statistical evidence and its gradual judicial acceptance.

1. *Early Resistance to Use of Statistical Evidence in the Courtroom*

The first reported introduction of statistical analysis may be in *Robinson v. Mandell*.⁵² In a will contest, Professors Oliver Wendell Holmes and Louis Agassiz countered the testimony of fellow Harvard Professor Benjamin Peirce on whether codicil signatures were traced.⁵³ Peirce testified that, as a professor of mathematics, the chances that the disputed signature's characteristics matched a known original was "once in 2,666 millions of millions of millions"—a number, he testified, that "far transcends human experience."⁵⁴ As late as 1953, the Supreme Court of Florida refused to admit a public opinion survey of community sentiment, noting that not only was the survey hearsay,⁵⁵ but that "its competency was suspect."⁵⁶ Subsequently, the California Supreme Court reversed a criminal conviction after a college mathematics instructor testified that a combination of characteristics of the criminal suspects was one in twelve million.⁵⁷ While challenging the mathematical basis for the expert's conclusions, the decision rested not on the validity of the statistical calculations, but instead on the court's finding that the study's prejudicial impact outweighed its probative value.⁵⁸

49. Michael I. Meyerson & William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771, 772 (2010).

50. Michael J. Saks, *Legal and Policy Analysis and Evaluation*, 44 AM. PSYCHOL. 1110, 1115 (1989).

51. *Id.* See also Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1479 (1999).

52. 20 F. Cas. 1027 (C.C.D. Mass. 1868) (No. 11,959).

53. *Kaye*, *supra* note 25, at 1940–41.

54. Paul Meier & Sandy Zabell, *Benjamin Peirce and the Howland Will*, 75 J. AM. STAT. ASS'N 497, 499 (1980). For discussion, see *Kaye*, *supra* note 25, at 1941.

55. Regarding the observation that public survey data is hearsay, see *Zippo Manufacturing Co. v. Rogers Imports, Inc.*, 216 F. Supp. 670, 682–83 (S.D.N.Y. 1963) (providing alternative bases for admitting opinion surveys: first, that such surveys are not hearsay because they are not offered to prove the truth of the matters asserted and that, even if hearsay, they fall under the exception of "present sense impression").

56. *Irvin v. State*, 66 So. 2d 288, 291–92 (Fla. 1953) (discussing a survey intended to show that defendant could not get a fair jury pool based on the community's widespread beliefs in racial stereotypes).

57. *People v. Collins*, 438 P.2d 33, 37 (Cal. 1968).

58. *Id.* at 39 (In *Collins*, a victim was robbed by a woman with a blond ponytail who escaped,

One much-noted case in which statistical sampling was excluded was *Sears, Roebuck and Co. v. City of Inglewood*.⁵⁹ The issue was whether Sears had correctly reported the proportion of sales made to non-residents of the city, since such sales were not subject to sales tax.⁶⁰ A statistician testified that a random survey of 33 of 826 working days revealed that such sales amounted to \$28,250 with a standard deviation of \$2,100 or a 95 percent confidence interval of \$24,000 to \$32,400 per quarter and then projected those figures for the entire period.⁶¹ After the survey was not admitted into evidence, Sears then undertook a complete audit of 950,000 transactions, finding the figure to be \$26,750 per quarter—in other words, just as predicted by the random sample.⁶²

2. Use of Statistical Evidence in Litigation Increasing

Despite the early resistance to the judicial admissibility of statistical evidence, its use in the courtroom is increasing.⁶³ Survey sampling was permitted in a case deciding trademark law, where consumer surveys helped the petitioner support a claim of consumer confusion.⁶⁴ Survey

according to witnesses, with a bearded, mustachioed black man driving a yellow car. The expert calculated the probability multiplying the individual probabilities of the various factors (yellow car, interracial couple, black man with beard, black man with mustache, woman with blond hair, and woman with ponytail) to arrive at the estimate. Among other things, the court found, the analysis was faulty because (1) the lack of evidence supporting the probabilities assumed by the expert, (2) inadequate proof of the “statistical independence of the six factors” and therefore the appropriateness of applying the multiplication rule, and (3) the risk that jurors were confused by the unchallenged admission of mathematical “proof” of guilt).

59. R. Clay Sprows, *The Admissibility of Sample Data Into a Court of Law*, 4 UCLA L. REV. 222–232 (1957) (discussing Los Angeles Superior Court case from 1955).

60. *Symposium: Probability Sampling in Litigation*, 18 CONN. INS. L.J. 297, 299 (2011–2012).

61. Diamond, *supra* note 2, at 228 n.12; *Symposium: Probability Sampling in Litigation*, *supra* note 60, at 299.

62. R. Clay Sprows, *The Admissibility of Sample Data into a Court of Law: a Case History*, 4 UCLA L. REV. 222, 226–29 (1956–1957). For discussion see Diamond, *supra* note 2, at 228 n.12; *Symposium: Probability Sampling in Litigation*, *supra* note 60, at 299; Gregory Todd Jones & Reidar Hagtvedt, *Sample Data as Evidence: Meeting the Requirements of Daubert and the Recently Amended Federal Rules of Evidence*, 18 GA. ST. U.L. REV. 721, 727–28 (2002).

63. STEPHEN E. FEINBERG, *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS 7–9* (1989) (Feinberg attributes the steep incline of the use of statistical evidence to the following factors: (1) expanding government regulation raise factual issues calling for statistical analysis, (2) the ease with which large masses of data can be managed and increased number of experts able to use and interpret such data, (3) increased access of computer programs to process data, (4) younger lawyers who are more trained and receptive to statistical information, (5) reliance on data at the U.S. Supreme Court in high profile cases has increased its acceptability, (6) higher financial stakes in litigation that increases the likelihood that attorneys and clients will invest in such data generation and analysis, and (7) enactment of the Federal Rules of Evidence in the mid-70s that eased restrictions on the use of expert witnesses, including those who specialized in statistical analysis).

64. *Zippo Mfg. Co. v. Rogers Impts., Inc.*, 216 F. Supp. 670, 680–82 (S.D.N.Y. 1963). For a study concluding that surveys purporting to show consumer confusion are excluded more often than those that show little or no consumer confusion, see MICHAEL O. FINKELSTEIN & BRUCE LEVIN, *STATISTICS FOR LAWYERS*, 269–70 (2d ed. 2001).

sampling has been used in misleading advertising cases, venue challenges, obscenity prosecutions and employment discrimination cases, among other examples.⁶⁵ This increase has brought with it concerns of a jury's ability to interpret data, particularly when experts differ on the evidence and its significance.⁶⁶ The increasing complexity of evidence has led to some calls for courts to appoint neutral managerial experts:⁶⁷ "[A] judge could better fulfill this gatekeeper function if he or she had help from scientists [and thus], [j]udges should be strongly encouraged to make greater use of their inherent authority . . . to appoint experts."⁶⁸

IV. THE PROBLEM: SAMPLING AND NON-SAMPLING ERRORS HAVE A SIGNIFICANT IMPACT IN LITIGATION

The use of sampling is routine in litigation today, and responsible use thereof was defined in *United States v. United Shoe Machinery Corp.*⁶⁹ Responsible use includes following sampling methods that "conform to generally recognized statistical standards," as listed by the Federal Judicial Center.⁷⁰ These federal standards "include whether:

- the population was properly chosen and defined;
- the sample chosen was representative of that population;
- the data gathered were accurately reported; and
- the data were analyzed in accordance with accepted statistical principles."⁷¹

Simply meeting these statistical standards, however, is not enough. While it can be difficult to determine an appropriate population for sampling, sampling frame, or list of elements in this population, and data collection methods to meet the above standards, these difficulties are well-

65. Laurens Walker & John Monahan, *Sampling Evidence at the Crossroads*, 80 S. CAL. L. REV. 969, 975-77 (2007).

66. Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1124-25 (2001).

67. Ellen E. Deason, *Managing the Managerial Expert*, U. ILL. L. REV. 341 (1998); Ellen E. Deason, *Court-Appointed Expert Witnesses: Scientific Positivism Meets Bias and Deference*, 77 OR. L. REV. 59 (1998).

68. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 150-51 (1997) (Breyer, J., concurring) (quoting amicus brief of the *New England Journal of Medicine*) (noting that the Federal Rules of Evidence permitted the appointment of independent experts).

69. *U.S. v. United Shoe Mach. Corp.*, 110 F. Supp. 295, 305-06 (D. Mass. 1953) *aff'd*, 347 U.S. 521 (1954). See also HANS ZEISEL & DAVID KAYE, *PROVE IT WITH FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION* 102 (1997).

70. FEDERAL JUDICIAL CENTER, *MANUAL FOR COMPLEX LITIGATION*, § 11.493, at 103 (Stanley Marcus et al., eds., 4th ed. 2004).

71. *Id.*

known to sampling experts and are routinely addressed.⁷² More problematic in many cases are the discrepancies between a valid theoretical sample and the data that are actually collected. There are innumerable ways in which a statistically-sound sample may be incorrectly implemented, and if even one of them occurs, the resulting data must be correctly reweighted, analyzed, and interpreted before meaningful projections concerning the population of interest may be made; and even then, results may not be accurate.⁷³

To avoid these costly and often less-than-perfect solutions, Robert Groves, Director of the U.S. Census Bureau, proposed an effective method of collecting data that addresses, and seeks to minimize “Total Survey Error.”⁷⁴ In short, Groves notes that sampling error is only one of seven sources of potential error that can render data misleading or even useless.⁷⁵ These sources of error include problems of accurately identifying information about population elements (specification error, measurement error, and processing error) and problems of data collection as a whole (coverage error, nonresponse error, sampling error, and adjustment error).⁷⁶ While all seven are important, coverage and nonresponse error are especially pernicious sources of error in today’s litigation.⁷⁷

Fortunately, these two sources of error can be minimized if steps are taken to guard against them before data collection, and check and adjust for them afterwards.⁷⁸ As was recommended in Fienberg’s *The Evolving Role of Statistical Assessments as Evidence in the Courts*, after having been collected, data should be statistically analyzed to control for non-response error and non-sampling error (e.g., problems of coverage), and might also be audited to “assess the extent to which there were deviations from stated procedures, and what the effect on study results might be from those deviations.”⁷⁹

72. See, e.g., Joseph C. Bright, Jr., Joseph B. Kadane & Daniel S. Nagin, *Statistical Sampling in Tax Audits*, 13 LAW & SOC. INQUIRY 305, 305 (1988); ROBERT M. GROVES, SURVEY ERRORS AND SURVEY COSTS 81–132, 239–94 (1989); ROBERT M. GROVES, ET AL., SURVEY METHODOLOGY 39–66 (2004); and Robert M. Groves, *Nonresponse Rates and Nonresponse Bias in Household Surveys*, 70 PUB. OPINION Q. 646, 646–75 (2006) [hereinafter Groves, *Nonresponse Rates*].

73. Groves, *Nonresponse Rates* supra note 72, at 646–75.

74. See ROBERT M. GROVES, ET AL., SURVEY METHODOLOGY 49 (2004).

75. *Id.*

76. *Id.*

77. Cf. MICHAEL O. FINKELSTEIN, BASIC CONCEPTS OF PROBABILITY AND STATISTICS IN THE LAW 98-101 (2009). Finkelstein uses the term “selection bias” for what Groves calls “coverage error” and also names “response bias” (“measurement error” in Groves) as a third “principle variety” of bias at in litigation surveys. However, addressing the reduction of measurement error is beyond the scope of this paper except to note that measurement error may occur when respondent memory is faulty or when surveys address sensitive issues.

78. ROBERT M. GROVES, SURVEY ERRORS AND SURVEY COSTS 81-183 (1989)

79. The Special Committee on Empirical Data in Legal Decision Making, *Recommendations on Pretrial Proceedings in Cases with Voluminous Data*, in THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS app. F, at 267 (Stephen E. Fienberg ed., 1988).

Although these recommendations were made in 1989, and best practices for conducting sample-based research have been well established since then, too many of today's litigators still unwittingly use faulty data to support their cases. To avoid being one of those who misuse sample data, one must be aware of and work to counteract the traps and pitfalls that have ensnared many in the profession. Before one can look for problems in procedure, however, one must first understand the proper theory of sampling.

A. PROBABILITY AND NON-PROBABILITY SAMPLING COMPARED

There are two broad types of sampling that researchers may employ: probability sampling and nonprobability sampling.⁸⁰ Each will be defined, evaluated, and discussed with regard to its application to litigation in turn.

Probability sampling is recommended for use in court, for reasons given in the Federal Reference Manual on Scientific Evidence: "In all forms of probability sampling, each element in the relevant population has a known, nonzero probability of being included in the sample."⁸¹ Probability sampling offers two important advantages over other types of sampling:

First, the sample can provide an unbiased estimate that summarizes the responses of all persons in the population from which the sample was drawn; that is, the expected value of the sample estimate is the population value being estimated. Second, the researcher can calculate a confidence interval that describes explicitly how reliable the sample estimate of the population is likely to be.⁸²

In other words, probability sampling gives the researcher the ability to predict a characteristic of a population with statistical accuracy.

To draw a probability sample, one must first know the general characteristics of the population in which one is interested,⁸³ such as "all consumers of XYZ food products in the past twelve months," "all female employees of Alpha Corp.," or "all houses on Main Street." Second, one must identify an appropriate sampling frame—the list of population elements from which one will draw the sample.⁸⁴ For instance, tax auditing cases could use as their sampling frame "the universe of documented transactions available for audit."⁸⁵ Other sampling frames might be a telephone book, a company's employee directory, or a list of all public schools in the country.⁸⁶ Third, one must select the elements of the

80. Shari Seidman Diamond, *Reference Guide on Survey Research*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 359, 361 (Fed. Judicial Ctr. ed., 3d ed. 2011).

81. *Id.* at 380.

82. Diamond, *supra* note 80, at 380.

83. *Id.* at 376.

84. *Id.* at 377.

85. Bright, Kadane & Nagin, *supra* note 72, at 305–33.

86. Arne L. Kalleberg et al., *Comparing Organizational Sampling Frames*, 35 ADMIN. SCI. Q. 658,

population to be included in the sample.⁸⁷ As already noted, for “adequate statistical sampling,” the sample drawn must be probabilistic in nature.⁸⁸ Many strategies for drawing probability samples exist, from simple random sampling to stratified sampling to cluster sampling, each has its own unique advantages.⁸⁹

One common form of probability sampling is simple random sampling, where every element in a population has an equal chance of being chosen for inclusion in the sample.⁹⁰ With this type of statistical sample, one should theoretically obtain a sample that is proportionally similar to the population in every dimension.⁹¹ Even if different elements of the population have different chances of being selected, though—for instance, if one wanted to “oversample” a minority group to ensure statistical significance of the results—one can still use statistical methods to ensure a representative outcome. If one oversamples, for instance, one can appropriately weight the groups before data analysis, as long as one knew the probability of selecting a member of each group ahead of time. For instance, a team of researchers interested in predicting violent tendencies in emergency room patients in 1996 purposefully oversampled black men and undersampled white women so that “patients likely to be involved in violence were more prevalent in the sample” than in the population as a whole, then reweighted the sample based on patients’ “ages, races, genders, and clinicians’ judgments about their dangerousness” before conducting their analysis.⁹²

As a more sedate example, the American Time Use Survey⁹³ (often used in litigation) collects data on each weekday from 10 percent of its respondents, and data on time spent on Saturday or Sunday from 50 percent of its respondents.⁹⁴ From this overweighting of weekend data, it can

658 (1990); Peter M. Smith & Cameron A. Mustard, *How Many Employees Receive Safety Training During Their First Year of a New Job?*, 13 *INJ. PREVENTION* 37, 37 (2007); Anne Cattagni & Elizabeth Farris, *Internet Access in U.S. Public Schools and Classrooms: 1994-2000*, *STATISTICS IN BRIEF* (National Center for Education Statistics, Washington D.C.), May 2001.

87. Diamond, *supra* note 80, at 380.

88. *Id.* As previously mentioned, probabilistic samples are samples based on known probabilities; that is, every element in a population that might be selected for a sample has a known, nonzero chance of inclusion. In contrast, nonprobabilistic samples do not meet this requirement. As an example, a study where a researcher interviews carefully selected class members who share a certain opinion is nonprobabilistic. Similarly, a mall-intercept study that approaches people who look like they might be willing to answer a survey is not probabilistic; neither is a survey advertised in a publication that asks all interested parties to volunteer and participate.

89. *Id.*; see generally LESLIE KISH, *SURVEY SAMPLING* (1965).

90. Diamond, *supra* note 80, at 380.

91. *Id.*

92. William Gardner et al., *A Comparison of Actuarial Methods for Identifying Repetitively Violent Patients with Mental Illnesses*, 20 *LAW AND HUMAN BEHAVIOR* 35, 38 (1996).

93. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, *AMERICAN TIME USE SURVEY – 2010 RESULTS 4* (2011), <http://www.bls.gov/tus/>.

94. *Id.*

measure far more accurately how time spent on weekends compares to time spent on weekdays on various activities.⁹⁵

In opposition to probability samples, nonprobability samples tend to be biased and are not generalizable, and so are not recommended for statistical analysis.⁹⁶ Sometimes there are good reasons for doing a nonprobability sample—if a researcher does not know the characteristics of the target population or it would be prohibitively expensive to do a probability sample, he might still learn potentially valuable information from a nonprobability sample.⁹⁷ For instance, nonprobability consumer surveys have been admitted as evidence in Lanham Act litigation based on the fact that major companies use such surveys when making major decisions,⁹⁸ and self-report surveys have proved useful in criminological research.⁹⁹ In these instances, however, the Reference Manual on Scientific Evidence directs experts to be prepared to justify their methods, to view results as “rough indicators rather than as precise quantitative estimates,” and to not calculate confidence intervals for the sample.¹⁰⁰ Because of the statistical limitations of non-probability samples, severe problems may arise if one believes that the sample one has drawn is a probability-based sample, when in fact it is not.¹⁰¹ If a study is based on a nonprobability selection, it should so indicate, including an explanation of the methods used in selecting the study sample.¹⁰²

Basic warnings about nonprobability samples abound in academic literature, from general advice against using them,¹⁰³ to specific warnings pertaining to a specific field.¹⁰⁴ In the world of tax auditing, for instance, Bright, Kadane, and Nagin note that if auditors “have preconceived ideas” regarding where problems might be when fact-checking records, and “exploit either their ideas or their initial findings in deciding where to look

95. BUREAU OF LABOR STATISTICS, *supra* note 93, at 8.

96. David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 211, 223-25 (Fed. Jud. Center, 3d ed. 2011). *See also*, The Special Committee on Empirical Data in Legal Decision Making, *Recommendations on Pretrial Proceedings in Cases with Voluminous Data*, in THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS app. F, at 261 (Stephen E. Fienberg ed., 1988) (“Nonprobability sampling is normally a highly subjective operation and relies either on expert judgment, or other criteria, to select particular units for the sample.”).

97. Diamond, *supra* note 80, at 381–82.

98. *Id.*

99. Josine Junger-Tas & Ineke Haen Marshall, *The Self-Report Methodology in Crime Research*, 25 CRIME & JUST. 291, 291 (1999).

100. Diamond, *supra* note 80, at 381–83.

101. *Id.*

102. The Special Committee on Empirical Data in Legal Decision Making, *Recommendations on Pretrial Proceedings in Cases with Voluminous Data*, in THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS app. F, at 261 (Stephen E. Fienberg ed., 1988).

103. Frederick W. H. Ho, L. Biggeri & Graham Kalton, *Survey as a Source of Statistics and Factors Affecting the Quality of Survey Statistics*, 73 INTERNATIONAL STATISTICAL REVIEW 245, 245 (2005).

104. Bright, Kadane & Nagin, *supra* note 72, at 307.

further . . . they are not entitled to treat the cases they examined as a statistical sample, since they chose them purposely. Proper statistical sampling is a skill distinct from auditing.”¹⁰⁵

B. NONRESPONSE AND COVERAGE ERROR

While the danger of error due to nonprobability samples has been made clear to the courts over the years, other potential non-sampling sources of error in surveys have not been as well addressed in the world of litigation. Nonresponse and problems of coverage are two of these subtler sources of error in sample-based research. It is true that academia warns researchers that “deviations from the sampling plan should be avoided,”¹⁰⁶ that one should be cautious about drawing conclusions from samples with too low a response rate,¹⁰⁷ and that nonresponse bias is a significant problem that must be addressed.¹⁰⁸ Even so, many samples used in litigation today are biased due to a failure to implement a properly designed sampling plan.¹⁰⁹ Both error sources can be minimized through careful data collection techniques, post-survey research of the characteristics of refusals, and weighting of the data,¹¹⁰ but only if the possibility of their existence is recognized and identified through careful pre- and post-collection analysis.¹¹¹ Several examples of common, but flawed, sampling practices are given below as illustrations.

1. Example¹¹²

First is an example of simple nonresponse bias. A hypothetical litigator wants to give a survey to all employees of XYZ Corp. to ascertain whether discrimination has taken place. To do this, the litigator, or a statistical consultant, first creates a randomized list of the population of interest.¹¹³ Next, the litigator determines how many responses to the survey (X) are necessary for statistical significance.¹¹⁴ Once that is known, the

105. Bright, Kadane & Nagin, *supra* note 72, at 307.

106. *Id.*

107. Diamond, *supra* note 80, at 382.

108. Groves, *Nonresponse Rates*, *supra* note 72, at 670.

109. Author’s Interview with Sean Chasworth, Statistical Expert, www.rule26.com, in Pasadena, Cal. (Aug. 1, 2011).

110. See generally ROBERT M. GROVES, *SURVEY ERRORS AND SURVEY COSTS* (1989); ROBERT M. GROVES ET AL., *SURVEY METHODOLOGY* (2004).

111. Diamond, *supra* note 80, at 377-379, 383-385.

112. These are original examples created by the authors of the paper, loosely based on general work experience of said authors.

113. Putting lists in a random order is a staple of survey methodology, and can easily be done using many computer programs, including Microsoft Excel.

114. The number of responses necessary for any given research question depends on the dispersion of the data in the population, the desired precision of the estimate, and the type of statistical testing required. The details of determining these figures are beyond the scope of this paper; for more information, see LESLIE KISH, *SURVEY SAMPLING* (1965).

firm's agents are instructed to begin at the top of the randomized list and collect survey data from the first X people on the list. If someone is unavailable or chooses not to participate in the survey, however, the agents are instructed to go on down the list until they reach X total participants.

On its surface, this survey seems methodologically sound. The participants were randomly selected, with an equal chance of each participant being chosen, so the results would seem to be a valid probability sample. There is, however, a potential problem that the litigator has overlooked: the characteristics of those who do not participate in the survey. If those people who chose not to answer the survey are fundamentally different in some way from those who did participate, and if that difference is related to a variable that is pertinent to the litigation, then the survey will not accurately reflect the population as a whole on those variables.¹¹⁵ As the Federal Reference Manual on Scientific Evidence notes,

Even when a sample is drawn randomly from a complete list of elements in the target population, responses or measures may be obtained on only part of the selected sample. If this lack of response is distributed randomly, valid inferences about the population can be drawn with assurance using the measures obtained from the available elements in the sample. The difficulty is that nonresponse often is not random, so that, for example, persons who are single typically have three times the "not at home" rate in U.S. Census Bureau surveys as do family members [citation omitted] . . . Determining whether the level of nonresponse in a survey seriously impairs inferences drawn from the results of a survey 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.¹¹⁶

It should also be noted that a low nonresponse rate does not ensure that one's sample is free from nonresponse bias. Variables such as socioeconomic status, gender, race, and household size all impact nonresponse rates, and so increasing response rates through measures that do not increase minority participation rates can actually increase nonresponse bias.¹¹⁷ For instance, offering a pen as an incentive to answer an exit-polling survey increased Democrat responses more than Republican responses, leading to a survey with higher response rates, but more biased conclusions.¹¹⁸

From a litigation standpoint, it is even more important to note that those who perceive themselves as having an economic self-interest in

115. Groves, *supra* note 78, at 133.

116. Diamond, *supra* note 80, at 383–85.

117. *Nonresponse Rates* *supra* note 72, at 664 .

118. *Id.* at 666.

answering a survey are more likely to do so than those who do not consider it worth their time to respond, as per the social exchange theory of survey response.¹¹⁹ In the case of a discrimination suit, for instance, it is more in the interest of those who believe that they have been discriminated against—and perhaps have even filed lawsuits on their own—to respond than it is for those who do not believe there are grounds for a discrimination case. This means that people who do not believe that discrimination has occurred are more likely to refuse to answer the survey than are those who do believe that discrimination has occurred. Clearly, if this discrepancy between respondents is not identified, the resulting survey would be biased to overestimate the amount of discrimination taking place in XYZ Corp. Before analysis, then, any potential nonresponse bias in the survey data must be identified and minimized.

Coverage error also stems from a lack of responses from relevant population members. In this case, however, it happens when elements are left off a sampling frame on which they belong, thereby taking away their proper chance of being included in a sample.¹²⁰ This could occur, for example, on a case that requires analysis of employee timesheets. If certain groups of data are difficult to access—for instance, if they are stored in boxes in the back of a warehouse or on damaged computer files—then it is conceivable that the actual data collected would be significantly different from the original list of which data should be collected. This could be problematic if the distribution of data is not random, but is rather grouped by date or by department.

Similarly, consider a class action suit where affected members live in three neighborhoods, two of which are middle-class neighborhoods and one of which is a wealthy gated community. If an interviewer is unable to get immediate access to that community, he or she might decide to skip it and only interview members of the middle-class neighborhoods. This, however, would significantly underestimate the actual income levels of the class.

As a final example of coverage error, it is fairly common in class-action suits for several different classes to be legally combined into one larger class, and treated as a single group. However, legal commonality does not always translate to statistical commonality. If a sample is drawn only from one of several subclasses—perhaps because it is the easiest to access, or the largest, or preferable from some other demographic characteristic—the results of the survey may not be generalizable to other subclasses of the overall class. For example, if a group of actors and tech crew all file suits against a particular Hollywood studio, the cases might be combined into a single class-action suit. However, it would be

119. DON A. DILLMAN, JOLENE D. SMYTH, AND LEAH MELANI CHRISTIAN, INTERNET, MAIL, AND MIXED-MODE SURVEYS: THE TAILORED DESIGN METHOD 16–23 (3rd ed. 2009).

120. WILLIAM G. COCHRAN, SAMPLING TECHNIQUES 396 (3rd ed. 1977).

inappropriate to survey only members of the lighting crew, or only actors, to identify the specific grievances held, and the damages sought, by the entire class. Furthermore, even taking a random sample of the entire population would likely lead to significant bias if employment data from actors and technical crew members are combined without regard for their specific professions.

C. RECOMMENDATIONS

To avoid performing biased analysis, researchers (and those who hire them) should take care to follow best practices to minimize Total Survey Error, both while collecting data, and after the data-collection process. First, samples should be probability-based. This “maximizes both the representativeness of the survey results and the ability to assess the accuracy of estimates obtained from the survey,” as the Federal Reference Manual on Scientific Evidence makes clear.¹²¹

Second, samples should be drawn from a proper sampling frame that gives every relevant member of the population a known chance of being selected for inclusion in the sample. After the sample has been selected, accurate records should be kept of refusals, failures to contact, ineligibility, completions, etc., as well as relevant demographic information (race, sex, age, neighborhood, income, etc.) that can help distinguish those who participate in the study from those who do not. Recording these different types of nonresponses during the data collection process, while not commonly done, can make one’s later analysis far easier, more accurate, and more precise.

Third, after sample data has been collected, statistical analysis should be performed on the sample itself to determine any unrepresentative qualities of the sample and correct for them. This can only be accurately done if the sample is a probability sample, and if relevant demographic data has been recorded both for the sample and for the population, but it is a crucial step. If the sample is representative, then analysis can proceed with confidence. If the sample is not representative, the following steps could be taken to statistically reweight or otherwise improve the generalizability of the sample.

To correctly reweight a sample, collected data can be statistically compared to theoretical outcomes and population values based on demographic information. For litigation, relevant categories might include not only traditional demographic measures, but such variables as whether or not the respondent is in a protected class and whether or not the respondent had previously filed a lawsuit against the company. If there were sub-classes of respondents previously identified (such as actors and technical crew members in the example above), these subclass distinctions

121. Diamond, *supra* note 80, at 380.

could also be used to help identify the representativeness of the sample.

If the data are not representative, then the statistician might reweight them by subclasses to approximate the theoretical population. For instance, if the population was composed of 50 percent males and 50 percent females, but the sample drawn had 30 percent males and 70 percent females, then male values would have to be weighted by 1.67 and female values would have to be weighted by 0.71 to accurately calculate a population average. This type of reweighting is both routine and recommended, and can easily be performed by statisticians given accurate population parameters.

Even if information about the population is unknown, however, it is still vital to perform a statistical analysis of the sample data to identify differences between subgroups. There is a difference, for instance, between knowing that 10 percent of the sample suffered wrongful injury, and knowing that 90 percent of a certain subgroup suffered the wrongful injuries, while nearly no people from other subgroups were affected. These differences should also be easy for statisticians to calculate using cross-tabulations, regression, Fisher's Exact Tests, resampling, or other appropriate methodologies.

Whatever may be discovered, it is also vital that the sample's restrictions be taken into account, so that it is not used inappropriately or to make claims that it cannot support.¹²² While sampling is a very powerful technique, bad samples can be severely flawed in their capacity to provide meaning, and can be misleading if used incorrectly. For ethical reasons, then, it is important to verify that one's sample has been collected as was directed, and then to correct or at least acknowledge any weaknesses remaining in the sample before proceeding to analysis.

V. CASE STUDY

The following is a case study drawn from recent work done by the authors that illustrates the importance of following proper survey methodology in issues of litigation. While the case presented settled confidentially in 2011 and thus allows for neither precedent nor appropriate citations, the methodology used in the litigation—and what followed from it—provides a clear example of how the problems raised in this paper can affect legal decisions.

Seventeen individuals living in the same subdivision, governed by a homeowners' association, filed separate construction defect suits due to various flaws in their homes, such as leaky windows and improperly

122. See also AAPOR Code of Professional Ethics and Practices, AAPOR (last modified May 2010), http://www.aapor.org/AAPOR_Code_of_Ethics/4249.htm (for more specific best practices).

installed plumbing. Shortly after this, the homeowners' association itself filed a class action construction defects suit on behalf of all its members. The court, rather than having to try eighteen similar cases, consolidated all of these filings into one class action lawsuit.

During the discovery phase, it was necessary to identify the exact nature and extent of the flaws with the homes in the subdivision. Clearly, seventeen of the homes had already been self-identified as having defects. The defendant, however, asserted that these problems did not extend to the rest of the subdivision—over one hundred other homes. Determining which houses had defects, and the type and extent of defects that existed, was neither easy nor inexpensive, as walls had to be removed, floors torn up, and similar “destructive tests” had to be done to determine whether or not various installations were defective. To conserve time and resources, the plaintiffs' experts decided to take a simple random sample of the homes in the subdivision, then extrapolate the results of this sample to the population of homeowners.

While the decision to take a sample was appropriate, the implementation of this decision was problematic. On its surface, the sampling process was uncontroversial: the experts randomized the entire list of homeowners from which they drew their sample, and then decided to draw the first thirteen units for inclusion in the sample. (Thus, there was no coverage error.) However, if a house selected for inclusion in the sample chose not to participate, field workers were instructed to go on to the next house on the randomized list, and so to continue until they obtained thirteen participants.

Perhaps unsurprisingly, many homes contacted from the randomized list chose not to participate in the sample and undergo the destructive tests necessary to determine whether or not their home had been improperly constructed. Field workers neither recorded the reasons for refusals nor distinguished between failures to contact and refusals after contact. Similarly, they did not distinguish between previous litigants and nonlitigants when collecting data, but treated all potential respondents as though they were from a homogeneous class.

After the sample was collected, the plaintiffs' statisticians apparently made no attempt to determine its generalizability by, for instance, comparing its demographic characteristics to those of the known population (the subdivision). Instead, they assumed that the average number of defects per house found in the sample would reasonably approximate the average value for the subdivision as a whole. With this assumption, they simply multiplied the sample average by an appropriate dollar amount per defect by the total number of houses in the subdivision to calculate damages for the class.

The plaintiffs' experts' lack of post-data collection analysis led to an embarrassing situation when statisticians retained by the defense examined

their sample and cross-referenced the results with other data. First, the sample's composition proved to be strongly biased towards previous litigants due to differential rates of response across sub-classes. While only 14.2 percent of the total homeowners had previously filed construction defect suits, 53.8 percent of the sample had previously filed these suits. This was a significant difference, and one that could have easily been identified if demographic information had been collected and used in sample analysis, or if the subclass of previous litigants had been sampled separately from the subclass of non-litigants.

Second, the defense's statisticians analyzed the differences between the numbers of defects found in the homes of those who had previously filed suits and those who had not. Significant nonresponse bias was found: Individual plaintiffs had far more defects per home than did others in the homeowners' association. On average, the sample overestimated the number of defects per home by more than 25 percent.

These discrepancies are understandable when one takes into account economic self-interest, as discussed above. As opposed to the plaintiffs' assumption that all nonresponses would be randomly assigned (an assumption all too common in litigation and not easily testable), from an economic model it makes sense that respondents with more to gain or who have a vested interest in the results of a sample would be willing to undergo higher costs to participate in the sample. In this case, since all sample participants had to undergo very high costs to participate (agreeing to destructive testing of their homes), those who already believed that their houses had defects were far more likely to participate than were those who had no such beliefs.

If proper data collection and post-data analysis procedures had been followed, even this flawed sample could have been weighted to properly reflect the relevant characteristics of the population and provide an estimate of damages that could have stood up to statistical scrutiny. It was not, however, and after the discovery of the aforementioned data problems the case quickly settled.

VI. CONCLUSION

As evidenced above, it is vital to plan for and implement proper statistical analysis that can reduce Total Survey Error before, during, and after the sampling process. While not difficult for trained statisticians to perform, this analysis needs to be included in the list of budgeted tasks that they are assigned. Before data collection, relevant demographic variables that might bias the resulting sample should be identified and an appropriate method of sampling decided upon. It should be noted that these variables are often case-specific, and may be better known to the legal team than to

third party experts who are unfamiliar with the case. During data collection, any deviations from the theoretical sampling process should be noted and information collected to help determine the extent and type of nonresponse bias present in the sample.¹²³ Demographic data pertaining to the sample should also be collected for post-sample analysis of differences between subgroups. Finally, after data have been collected, statisticians must analyze the composition of the sample and, if necessary, reweight it to minimize the impact of bias. As illustrated by the case study, making simplistic assumptions can be inaccurate, can lead to unwarranted conclusions, and can also be potentially dangerous to one's arguments and positions. Instead, by following proper statistical procedure before, during, and after sampling, one can lean with confidence on the results and extrapolations of a sample that is worthy of being used in court.

123. *See* THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH, STANDARD DEFINITIONS: FINAL DISPOSITIONS OF CASE CODES AND OUTCOME RATES FOR SURVEYS 7–12 (7th ed. 2011).
