
SAMPLING EVIDENCE AT THE CROSSROADS

LAURENS WALKER* AND JOHN MONAHAN†

I. INTRODUCTION

McLaughlin v. Phillip Morris USA, Inc., has been certified as a nationwide class action on behalf of an estimated 50 million “light” cigarette smokers.¹ Plaintiffs seek more than \$280 billion in damages, to be trebled to over \$800 billion.² In certifying this mass tort, District Judge Jack B. Weinstein announced his plan to completely abandon individualized adjudication in favor of aggregate factual determinations based on evidence from statistical samples.³ Prior to *McLaughlin*, at least two federal trial judges had permitted the use of sampled evidence in major consolidated or class action trials,⁴ but both included some adjudication of individual claims.⁵ In *McLaughlin*, Judge Weinstein’s plan would entirely eliminate proof of individual class member claims in the face of the

* T. Munford Boyd Professor of Law and Caddell and Chapman Research Professor, University of Virginia. We wish to thank David Barnes, Shari Diamond, and participants at a University of Virginia School of Law workshop for comments and suggestions. We also thank Allyson McKenzie, Kelly DeMarchis, Chris Roche, Gregory Walters, and Meredith Wilson for able research assistance.

† John S. Shannon Distinguished Professor of Law, University of Virginia. This article was completed while Professor Monahan was in residence at the New York University School of Law.

1. See *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992 (E.D.N.Y. 2006), *appeal docketed sub nom.* *McLaughlin v. Philip Morris USA, Inc.*, No. 06-4666 (2d Cir. Nov. 17, 2006). The plaintiffs’ case relies on the combined federal “RICO” and Mail and Wire Fraud Acts and alleges fraud in the advertising and marketing of “light” cigarettes. Racketeer Influenced and Corrupt Organizations Act, 18 U.S.C. §§ 1961–1968 (2000); Mail and Wire Fraud Acts, 18 U.S.C. §§ 1341–1343 (2000 & Supp. IV 2004).

2. See *Schwab*, 449 F. Supp. 2d 992.

3. *Id.* at 1022.

4. See, e.g., *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460 (D. Haw. 1995), *aff’d sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649 (E.D. Tex. 1990), *rev’d*, 151 F.3d 297 (5th Cir. 1998).

5. See *In re Estate of Marcos*, 910 F. Supp. at 1462, 1464–66; *Cimino*, 751 F. Supp. at 653, 665–66.

overwhelming cost of gathering such evidence from tens of millions of plaintiffs. The central issue in the interlocutory appeal now before the Second Circuit⁶ is the legality of Judge Weinstein's plan to use sampled evidence to determine whether the plaintiff class members relied on representations by the defendants that "light" cigarettes were less harmful than regular cigarettes, and, if so, to determine the aggregate amount of damages.⁷

In this Article, we address and defend Judge Weinstein's controversial proposal to statistically sample evidence, rather than to obtain evidence on an individualized, case-by-case basis. We endorse his view that statistical sampling combined with other evidence "is a necessary and pragmatic evidentiary approach that reflects full due process in this and many other mass tort cases."⁸ In Part I, we describe the history of sampling both in science, where it originated, and in law, where in recent decades it has taken root in many areas of litigation. In Part II, we present the proposed plan to sample evidence on reliance and on damages in *McLaughlin*. In Part III, we argue that a new common law rule of evidence is emerging that not only permits but also *requires* judges to consider sampling in the face of exorbitant litigation costs. Finally, in Part IV, we present our own proposal for integrating the results of statistical sampling with traditional legal techniques for evaluating the sufficiency of evidence. We conclude that the use of sampled evidence now before the Second Circuit is both acceptable in *McLaughlin* and should be required in any consolidated case or class action where group-based determinations are more cost-efficient than individually-based ones.

II. A BRIEF HISTORY OF SAMPLING

The history of sampling—"selecting some part of a population to observe so that one may estimate something about the whole

6. *Schwab*, 449 F. Supp. 2d 992, *appeal docketed sub nom. McLaughlin v. Philip Morris USA, Inc.*, No. 06-4666 (2d Cir. Nov. 17, 2006).

7. Defendants maintain that reliance and damages must be proven on an individual basis, stating:

Under the court's aggregate approach, the predominance requirement of Rule 23(b)(3) would be a dead letter. All individual issues could be reduced to statistical analysis, resulting in trials that are little more than a battle of statistics, without any actual, individual issue being "tried" in a judicial sense. . . . The statistical approach proposed here also would violate defendants' due process and trial by jury rights.

Petition for Review pursuant to Federal Rule of Civil Procedure 23(f) at 18-19, *Schwab v. Philip Morris USA, Inc.*, No. 06-4666 (2d Cir. Oct. 6, 2006).

8. *Schwab*, 449 F. Supp. 2d at 1240.

population”⁹—is remarkably similar in science and in law. In each case, several decades of early controversy over the use of samples were followed by a decisive acceptance of sampling methods. In science, that acceptance has been absolute. At a current meeting of members of any scientific discipline, one would be no more likely to hear sampling called into question than to hear a claim that the earth is flat. In many areas of law, the use of sampling has also become routine. Whether sampling will achieve the same unquestioned acceptance in law as it has in science may depend, for the foreseeable future, on the Second Circuit’s decision in *McLaughlin*.

A. SAMPLING IN SCIENCE

“Difficulties beset anyone who attempts to trace the developing uses of sampling,” Frederick Stephan wrote in 1948, “because they are scattered throughout many branches of science and technology and are described, if at all, in subordinate portions of reports and articles whose titles provide no hint of what they may contain on the subject of sampling.”¹⁰ Some find rudimentary notions of sampling in the ancient practice of having a food-taster eat a small portion of each dish in the King’s meal to assure that the food was not poisoned.¹¹ More philosophically, others believe that “[a]ll empirical knowledge is, in a fundamental sense, derived from incomplete or imperfect observation and is, therefore, a sampling of experience.”¹²

What is known beyond conjecture about the history of sampling, is that while the mathematical theory that justifies the practice of taking a sample of some members of a population rather than conducting a census of all members of a population was well established by the eighteenth century, it was not until the beginning of the twentieth century that sampling began to be seriously pursued by scientists. It was not until well into the twentieth century that *random* sampling became a defining characteristic of empirical research.

Prior to the early twentieth century, “[p]artial investigations,” as sampling was then referred to, “were considered imprecise and unscientific.”¹³ When a “complete investigation”—a census—could not be taken, the researcher set out to select a sample of people or objects who

9. STEVEN K. THOMPSON, *SAMPLING* 1 (2d ed. 2002).

10. Frederick F. Stephan, *History of the Uses of Modern Sampling Procedures*, 43 J. AM. STAT. ASS’N 12, 12 (1948).

11. *See id.* at 13.

12. *Id.*

13. GERD GIGERENZER ET AL., *THE EMPIRE OF CHANCE: HOW PROBABILITY CHANGED SCIENCE AND EVERYDAY LIFE* 116 (1989).

were believed to be representative of the population of interest.¹⁴ This was referred to as “purposive selection:” cases were selected “on purpose,” because they were believed to mirror the average characteristics of the population.¹⁵ The difficulty with purposive selection, however, soon became apparent: how was one to know what the “average” characteristics of the population were, without first doing a census to find them out?¹⁶ Further, if a census had to be conducted before a sample could be taken, what was the purpose of sampling?

One of the most widespread attempts to reduce the human judgment involved in purposive sampling’s search for “average” people was known as “quota sampling.”¹⁷ In this procedure, which was developed by early commercial pollsters, interviewers were given specific percentages or “quotas” of cases to be selected from each of a given number of categories (for example, gender, age group, or income level).¹⁸ The final selection of cases was still purposive—“average” men and “average” women were still sought, for example—but it was purposive within a specified category.

The advantages of quota sampling over more unstructured purposive sampling methods were dramatically demonstrated when pollster George Gallup challenged the *Literary Digest*, one of the largest-selling magazines of the day, to see who could more accurately predict the results of the 1936 presidential election.¹⁹ The *Literary Digest*, long famous for its election

14. *Id.*

15. Alain Desrosières, *Three Studies on the History of Sampling Surveys: Norway, Russia-USSR, United States*, 15 *SCI. IN CONTEXT* 377 (2002).

16. GIGERENZER ET AL., *supra* note 13, at 116. Early efforts at sampling consisted of “a systematic search for a sample that agreed in important characteristics with the population at large. These characteristics had to be learned from a complete investigation, a census.” GIGERENZER ET AL., *supra* note 13, at 116.

17. JEAN M. CONVERSE, *SURVEY RESEARCH IN THE UNITED STATES: ROOTS AND EMERGENCE 1890–1960* 92 (1987).

18. In *Survey Research in the United States: Roots and Emergence 1890–1960*, Jean Converse describes the quota system used by the J. Walter Thompson advertising agency in 1924, called the “ABCD system.” Interviewers were instructed to select respondents in certain percentages from the following groups:

Class A. Homes of substantial wealth above the average in culture that have at least one servant. The essential point, however, in this class is that the persons interviewed shall be people of intelligence and discrimination.

Class B. Comfortable middle class homes, personally directed by intelligent women.

Class C. Industrial homes of skilled mechanics, mill operators, or petty trades people (no servants).

Class D. Homes of unskilled laborers or in foreign districts where it is difficult for American ways to penetrate.

Id. at 93 (citations omitted). Converse notes that “[e]ven this weak reed introduced some system in the field procedure which was an improvement over ‘man-on-the-street’ interviews or talks with customers in any retail store and which could provide some record of where interviewers had been.” *Id.*

19. See CONVERSE, *supra* note 17, at 117.

polling, sampled by mail (without using quotas) from lists of people who owned telephones and automobiles.²⁰ Over two million “straw ballot” surveys were returned.²¹ The results indicated that Alfred Landon would defeat Franklin Roosevelt by a margin of 55–41 percent.²² Gallup, then a newcomer, used a much smaller national sample—3,000 people—but with quotas for six different variables.²³ Gallup predicted Roosevelt over Landon by 54–46 percent.²⁴ Roosevelt won the election (with 61 percent of the vote), the *Literary Digest* was “hooted out of the business,”²⁵ and Gallup’s quota sampling techniques enjoyed a two-decade claim to respectability in the mind of the general public.²⁶ In the scientific community, however, purposive sampling—even with quotas—was already doomed.

Two years before the *Literary Digest* fiasco, Jerzy Neyman, a Polish statistician, published a landmark article directly comparing purposive selection—with and without quotas—and purely random selection.²⁷ So powerful was his mathematical demonstration of the superiority of random sampling that purposive selection soon became an object of ridicule and random selection quickly achieved general acceptance among statisticians throughout the world.²⁸ Among the broader scientific community, however, it took more than mathematical proof for random selection to triumph as *the* approach to sampling. It took World War II, with the federal government’s voracious need for quick and accurate empirical information for military and industrial purposes, to institutionalize random sampling as

20. Peverill Squire, *Why the 1936 Literary Digest Poll Failed*, PUB. OPINION Q., Spring 1988, at 125, 126–27.

21. *Id.* at 127.

22. *Id.*

23. *See id.* at 126 n.3. The variables for which Gallup constructed quotas were region of the country, urban/rural setting, gender, age, race, and socioeconomic status. *See* Claude E. Robinson, *Recent Developments in the Straw-Poll Field*, 1 PUB. OPINION Q., July 1937, at 45, 47.

24. CONVERSE, *supra* note 17, at 119.

25. W. Phillips Shively, *A Reinterpretation of the New Deal Realignment*, PUB. OPINION Q. 621, Winter 1971–1972, at 621, 621–22. *See also* S.E. Fienberg & J.M. Tanur, *History of Sample Surveys*, 20 INT’L ENCYCLOPEDIA OF THE SOC. & BEHAV. SCI. 13453, 13456 (Neil J. Smelser & Paul B. Baltes eds., 2001) (“Then, as now, election forecasting was taken as the acid test of survey validity. A reputation for accuracy in ‘calling’ elections was thought to spill over to a presumption of accuracy in other, less verifiable areas.”).

26. Gallup himself abandoned quota sampling in favor of random sampling for the 1956 Presidential election. *See generally* SARAH E. IGO, *THE AVERAGED AMERICAN: SURVEYS, CITIZENS, AND THE MAKING OF A MASS PUBLIC* (2007).

27. Jerzy Neyman, *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*, 97 J. ROYAL STAT. SOC’Y 558 (1934).

28. *See* Martin R. Frankel & Lester R. Frankel, *Fifty Years of Survey Sampling in the United States*, PUB. OPINION Q. 50th Anniversary Issue 1987, at S127, S127–29.

a sine qua non of scientific research.²⁹ With its mathematical superiority demonstrated by Neyman and its practical advantages established by the War, random sampling has, for the past sixty years, been a hallmark of the scientific method.

B. SAMPLING IN LAW

Sampling came to law later than to science. The earliest sustained development of the use of sampled evidence occurred in the arena of trademark litigation.³⁰ In the 1928 case *Elgin National Watch Co. v. Elgin Clock Co.*, the plaintiff Elgin National Watch Company sought an injunction prohibiting the defendant from using the name “Elgin.”³¹ The plaintiff submitted an affidavit from an expert witness who offered the opinion that the general public understood the name “Elgin” to mean only time pieces manufactured by the Elgin National Watch Company.³² The expert witness based his opinion on a mail survey that sampled approximately 2000 retail dealers of clocks and watches.³³ The dealers had been asked, “[i]f you saw the name Elgin Clock Company on an automobile clock or watch, who would you think made it?” and “[w]ho would your customers think made such a clock or watch?”³⁴ The court found the survey responses to be inadmissible hearsay.³⁵

Sampled evidence fared better in a 1940 decision in *Oneida, Ltd. v. National Silver Co.*³⁶ In that case, the plaintiff Oneida sought trademark protection for a silverware pattern called Coronation, against the defendant’s use of a pattern called Princess Royal.³⁷ Oneida supported its

29. JOSEPH W. DUNCAN & WILLIAM C. SHELTON, U.S. DEP’T OF COMMERCE, *REVOLUTION IN UNITED STATES GOVERNMENT STATISTICS 1926–1976* 35, 50 (1978) (“There were few examples of sampling in the Federal Government before 1933, . . . By about the time the United States entered World War II, probability sampling was no longer an esoteric subject in the Federal Government; it was well recognized as a method, and agencies were beginning to depend on it.”). As Converse notes, however, “purposive sampling of the ‘quota control’ variety continued to thrive in commercial research well into the 1940s and 1950s.” CONVERSE, *supra* note 17, at 44 (citation omitted).

30. See Trademark Act of 1948 (Lanham Act), 15 U.S.C. § 1051, *amended by* Trademark Law Revision Act of 1988, 15 U.S.C. § 1127 (2000); JOHN MONAHAN & LAURENS WALKER, *SOCIAL SCIENCE IN LAW* 95–126 (6th ed. 2006) (discussing the use of surveys and sampling in trademark litigation).

31. *Elgin Nat’l Watch Co. v. Elgin Clock Co.*, 26 F.2d 376, 376 (D. Del. 1928).

32. *Id.*

33. *Id.* at 377.

34. *Id.*

35. *Id.* at 377–78.

36. *Oneida, Ltd. v. Nat’l Silver Co.*, 25 N.Y.S.2d 271 (N.Y. Sup. Ct. 1940).

37. *Id.* at 275.

case with a survey that sampled 1000 women in their homes.³⁸ The women were shown a piece of Princess Royal silverware and asked, “[w]ho do you think puts out this silverware?” and “[i]f you wanted to buy a set of this silverware, how would you ask for it?”³⁹ A “considerable portion” of the women responded “Oneida” to the first question and “Coronation” to the second.⁴⁰ The court found the survey evidence to be “competent” and admitted it into evidence, but felt it necessary to note that individualized evidence was also introduced:

The plaintiff did not rest alone upon this survey. From the housewives who had in the questionnaire identified the pattern as one of plaintiff or its manufacture, twenty-four women were called as witnesses who gave their reasons for being confused or deceived in identifying the Princess Royal pattern shown them.⁴¹

In was not until 1963 that sampled evidence gained full acceptance in trademark law, in *Zippo Manufacturing Co. v. Rogers Imports, Inc.*⁴² In that case, Zippo alleged trademark infringement on the part of the defendant’s sale of cigarette lighters said to closely resemble Zippo’s.⁴³ Zippo supported its claim with three separate surveys sampling the national “smoking population.”⁴⁴ In one of the surveys, for example, 500 randomly sampled smokers were handed a Rogers’ lighter and asked to identify its brand.⁴⁵ Almost three times as many people said “Zippo” as said “Rogers.”⁴⁶ The defendant, as in *Elgin*, argued that surveys constituted hearsay and therefore were inadmissible as evidence.⁴⁷ This time, however, the court held that the surveys were admissible, and found that Zippo had

38. *Id.* at 286.

39. *Id.*

40. *Id.*

41. *Id.*

42. *Zippo Mfg. Co. v. Rogers Imps., Inc.*, 216 F. Supp. 670 (S.D.N.Y. 1963).

43. *Id.* at 671.

44. *Id.* at 681.

45. *Id.*

46. *Id.* at 690.

47. *Id.* at 682.

established the likelihood of confusion among potential consumers.⁴⁸ Today the use of surveys involving sampling is commonplace and expected not only in trademark cases, but also in misleading advertising cases, change of venue motions, obscenity prosecution, racial and gender discrimination cases, and many other areas of litigation.⁴⁹

Despite a three-decade history of admissibility in law, sampled evidence was not used in a mass tort case until 1990. The case *Cimino v. Raymark Industries, Inc.* involved the consolidated compensatory damage claims of 2298 plaintiffs who allegedly acquired one of five diseases through exposure to asbestos.⁵⁰ District Judge Robert Parker divided the plaintiffs into five categories according to the type of disease at issue, and then randomly selected from each category a number of cases for trial.⁵¹ Several judges, dozens of lawyers, and two juries tried a total of 160 randomly sampled cases.⁵² The verdicts rendered in each of these 160 cases were awarded to the individual plaintiffs whose cases had been tried, and the average verdict for cases in each of the five categories was awarded to the non-sampled members of that category.⁵³ The total amount of damages awarded by this procedure to all 2298 plaintiffs was in excess of \$1 billion.⁵⁴ The defendants appealed, and a divided panel of the Fifth Circuit held that the use of sampling violated Texas tort law as well as the Seventh Amendment.⁵⁵

48. *Id.* at 683. The court pointed out that:

Regardless of whether the surveys in this case could be admitted under the non-hearsay approach, they are admissible because the answers of respondents are expressions of presently existing state of mind, attitude, or belief. There is a recognized exception to the hearsay rule for such statements, and under it the statements are admissible to prove the truth of the matter contained therein.

Id. The court further pointed out that even if the surveys did not fit within this exception, they would be admissible based on "well-reasoned authority" requiring the court to examine the "need for the statement at trial" and the "circumstantial guaranty of trustworthiness surrounding the making of the statement." *Id.* Following the *Zippo* decision, the hearsay objection to evidence from surveys was virtually eradicated.

49. See Jack P. Lipton, *A New Look at the Use of Social Science Evidence in Trademark Litigation*, 78 TRADEMARK REP. 32, 63-64 (1988) (noting the growing acceptance of surveys among members of both the bench and the bar and that the failure of a trademark owner to present survey evidence may now give rise to an adverse inference by the court); Shari Seidman Diamond, *Survey Research*, in 1 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 393, 393-448 (David L. Faigman et al. eds., 2006-2007) (noting the areas of law in which the use of surveys has become common).

50. *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652-53 (E.D. Tex. 1990), *rev'd*, 151 F.3d 297 (5th Cir. 1998).

51. *Id.* at 653.

52. *Id.*

53. *Id.*

54. Robert M. Parker, *Streamlining Complex Cases*, 10 REV. LITIG. 547, 553 (1991).

55. *Cimino v. Raymark Indus., Inc.*, 151 F.3d 297, 297 (5th Cir. 1998).

Only one other mass tort case to reach trial has involved sampled evidence. *In re Estate of Marcos Human Rights Litigation* was a class action involving approximately 10,000 plaintiffs.⁵⁶ The case was certified to adjudicate claims for compensatory damages alleged to have resulted from human rights violations carried out by agents of Ferdinand Marcos during his tenure as president of the Philippines.⁵⁷ District Judge Manuel Real divided claimants into three categories based on the nature of the human rights violation alleged, and ordered depositions taken from a random sample of claimants in each category.⁵⁸ The results of these randomly sampled depositions from a total of 137 claimants were reported to a jury in a single trial.⁵⁹ The jury was asked to return individual verdicts for each of the claimants whose cases had been tried, and the average verdict for cases in each of the three categories was awarded to the non-sampled members of that category.⁶⁰ The total amount of damages awarded by this procedure to all 10,000 plaintiffs was in excess of \$766 million.⁶¹ The defendant appealed and a divided panel of the Ninth Circuit approved the District Court's reliance upon sampling.⁶²

We applaud the trial judges' pioneering use of sampling in *Cimino* and *Marcos*. We argue here, however, that the time taken by the jury to determine individual damage claims in each case was both unnecessary, from a legal standpoint, and inefficient as a matter of cost. It is precisely the elimination of determining individual claims that makes *McLaughlin* such a potentially paradigm-shifting event in mass tort litigation.⁶³

III. THE MCLAUGHLIN SAMPLING PLAN

An action for damages under the RICO statute requires proof of reliance on the representations of the defendant in order to establish causation.⁶⁴ Rule 23(b)(3) of the Federal Rules of Civil Procedure, in turn, requires that a proposed class action must be "superior to other available

56. *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1462 (D. Haw. 1995), *aff'd sub nom.* Hilao v. Estate of Marcos, 103 F.3d 767 (9th Cir. 1996).

57. *Id.*

58. *Id.* at 1462, 1464.

59. *Id.* at 1464.

60. *Id.* at 1467.

61. *Id.* at 1464.

62. Hilao v. Estate of Marcos, 103 F.3d 767, 767-68 (9th Cir. 1996).

63. See Laurens Walker & John Monahan, *Sampling Damages*, 83 IOWA L. REV. 545, 554-55 (1998).

64. Schwab v. Philip Morris USA, Inc., 449 F. Supp. 2d 992, 1272 (E.D.N.Y. 2006).

methods for the fair and efficient adjudication of the controversy.”⁶⁵ The Rule lists several matters “pertinent” to the determination of whether this superiority requirement has been met, including “the difficulties likely to be encountered in the management of a class action.”⁶⁶ Clearly, if proof of reliance were required by each of tens of millions of class members, a class action would be “unmanageable,” and therefore not “superior” to the alternative of individual damage actions. On the other hand, proving reliance with evidence from statistical samples would be relatively easy to manage. Thus the method of proof in *McLaughlin*—individual versus aggregate—is central to determining the threshold issue of whether a class can be certified.

Judge Weinstein’s response to the management issue was a strong endorsement of the use of sampling. “Extrapolation that comports with due process can be carried out using very small samples; the burden on the parties and the court of generating and evaluating the evidence would not be over-great.”⁶⁷ Judge Weinstein’s plan is responsive to an earlier proposal by plaintiffs. According to that proposal, plaintiffs will present “statistical evidence indicating that almost all members of the class value health in the selection of a ‘light’ cigarette, and would, therefore, have relied on Defendants’ ‘health reassurance’ message in purchasing ‘lights.’”⁶⁸ Proving damages with sampled evidence receives the same endorsement as proving reliance.⁶⁹ Judge Weinstein’s plan to rely so strongly on statistical sampling and to omit entirely any individual adjudication⁷⁰ for class members in *McLaughlin*, if upheld, will mark the beginning of a new era in sampling evidence. Although we understand the cautious behavior of the trial courts in *Cimino* and in *Marcos*, in our view these aspects of individualized adjudication were a waste of time and money.

65. FED. R. CIV. P. 23(b)(3) .

66. *Id.*

67. *Schwab*, 449 F. Supp. 2d at 1120.

68. Class Plaintiffs’ Proposed Litigation Plan at 14 (Aug. 22, 2005) (Ex. 8 to Class Plaintiffs’ Response to the Court’s Memorandum of June 6, 2005, *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992 (E.D.N.Y. 2006) (No. 04 Civ. 1945)) [hereinafter *Proposed Litigation Plan*].

69. *Schwab*, 449 F. Supp. 2d at 1244.

70. It is important to note that some particularized evidence will be introduced at trial in *McLaughlin*. The plaintiffs’ litigation plan is explicit that the focus of this particularized evidence is on the RICO elements of “conduct of a racketeering enterprise” and “conspiracy.” For example, the Proposed Litigation Plan notes that plaintiffs intend to introduce as evidence documentary evidence as well as fact witnesses to prove that the defendants knew that low tar cigarettes were not less harmful than other cigarettes. *Proposed Litigation Plan*, *supra* note 68, at 8–14.

IV. TOWARD A NEW REQUIREMENT FOR COST-EFFICIENT EVIDENCE

Judge Weinstein's trial plan in *McLaughlin* may reflect the emergence of a far-reaching new approach to the judicial management of the often-prohibitive costs of litigation. In this Part, we begin by considering how science has long handled the issue of cost in research involving extensive data-gathering. We then examine how several areas of law have more recently come to deal with the cost of large-scale information gathering that occurs during litigation, and focus on the role of cost in mass tort litigation. Finally, we identify a new rule of evidence that is emerging from these concerted judicial efforts at cost-containment.

A. THE PROBLEM OF COST IN SCIENCE AND IN LAW

Although the origin and development of science and law are plainly different, these two enterprises share significant commonalities.⁷¹ For example, both science and law deal with facts, and both are concerned not only with the accuracy of the facts with which they deal, but also with the costs of obtaining those facts. Given these common concerns, legal practice in containing the costs of gathering evidence in litigation profitably might be informed by comparing it to longstanding practices of cost-containment in science.

1. Containing Costs in Science

The historical justification for sampling in science was that it cost much less to take a sample than to conduct a census—that is, it was cheaper to measure some of the cases than to measure all of the cases in a population. This is still the primary justification for sampling in science. As stated in a leading contemporary text, “[t]he purpose of sample survey design is to maximize the amount of information for a given cost.”⁷²

71. See generally John Monahan & Laurens Walker, *Empirical Questions Without Empirical Answers*, 1991 WIS. L. REV. 569 (1991); John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating and Establishing Social Science in Law*, 134 U. PA. L. REV. 477 (1986); Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 CAL. L. REV. 877 (1988); Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987).

72. RICHARD L. SCHEAFFER ET AL., *ELEMENTARY SURVEY SAMPLING* 59 (2d ed. 1979). See also Gary T. Henry, *Practical Sampling*, in *HANDBOOK OF APPLIED SOCIAL RESEARCH METHODS* 101, 101 (Leonard Bickman & Debra J. Rog eds., 1998) (“Researchers would like to be able to speak about entire populations of interest, such as adult residents of the United States or children enrolled in public preschool programs. However, time and costs permit them to collect data from only a limited number of

From the beginning, however, a secondary justification for sampling as a scientific enterprise was also asserted: *more accurate* results could be obtained by taking a sample than by conducting a census.⁷³ The reason for this counterintuitive conclusion is not difficult to grasp: “estimates based on sample surveys are often more accurate than those based on a census because investigators can be more careful when collecting data.”⁷⁴ For example, to determine the prevalence of a given disease in the American population, conducting a census would necessitate the use of only the crudest screening tests, for the simple reason that there do not exist enough physicians and enough laboratories to carefully examine a population of 300,000,000 Americans. In contrast, taking a random sample of one or two thousand people would allow for much more thorough medical examinations, and hence provide much more accurate scientific estimates of the prevalence of the disease.

The primary justification for sampling in science, however, remains one of economy. It simply costs less time and money to measure fewer things than it does to measure more things. Depending on the difference between the number of cases in the sample and the number of cases in the population from which the sample is taken, and on the cost of measuring each case, it can cost *much* less to take a sample than to conduct a census. To take a perhaps extreme example, the great majority of Americans receive the “short form” of the U.S. Census, consisting of seven questions.⁷⁵ The cost of having a commercial survey firm take a random sample of 1,500 Americans and ask these seven questions—producing results accurate to the true population values plus-or-minus three percent⁷⁶—is approximately \$50,000.⁷⁷ The actual cost of conducting the 2000 Census, which involved questioning all Americans, was \$6.5 billion.⁷⁸

population members. The researcher needs a bridge to connect the goals of the study with the practical considerations of conducting the research. Sampling methods, or the methods by which members of a population are selected for a study, provide that bridge.”)

73. See SHARON L. LOHR, SAMPLING: DESIGN AND ANALYSIS 16 (1999).

74. *Id.*

75. These questions include: Age; Hispanic or Latino origin; Household relationship; Race; Sex; Number of people in the home; and Home. See Letter from the Director, U.S. Census Bureau, available at http://www.census.gov/acs/www/SBasics/congress_toolkit/ACS%20Tool%20Kit.pdf (last visited July 10, 2007) (describing how most households received “short forms,” consisting of seven questions).

76. See *infra* Part V (discussing the 95% confidence interval).

77. E-mail from John McNee, Vice President, Gallup Organization, to John Monahan (June 14, 2003) (on file with author).

78. U.S. GEN. ACCT. OFFICE, 2000 CENSUS: SIGNIFICANT INCREASE IN COST PER HOUSING UNIT COMPARED TO 1990 CENSUS 2 (2001), available at <http://www.gao.gov/new.items/d0231.pdf>.

Since the cost of sampling will depend on the number of cases that have to be sampled, sample size is a crucial issue. Perhaps counterintuitively, the size of the *population* being sampled does not determine the size of the *sample* that needs to be taken. It is the absolute, rather than the relative, size of the sample that matters. As stated in one leading sampling text:

The size of the population from which a sample of a particular size is drawn has virtually no impact on how well that sample is likely to describe the population. A sample of 150 people will describe a population of 15,000 or 15 million with virtually the same degree of accuracy, assuming that all other aspects of the sample design and sampling procedures are the same.

....

... Specifying a fraction of the population to be included in the sample is never the right way to decide on a sample size.⁷⁹

The size that a sample needs to be depends on several things, among them the width of the *confidence interval* that the researcher wishes to apply to the results of the sampling, and the stringency of the *confidence level* that the researcher desires to characterize this interval.⁸⁰

The *confidence interval* describes “how narrowly the researcher needs to zero in”⁸¹ on the correct answer, that is, on the actual value that the researcher would find if the researcher conducted a census of the population rather than took a sample. Is being within ten percent of the population mean “good enough” for one’s purposes, or is it necessary to be within one percent of the population mean? The narrower the confidence interval required—the more precise the sample estimate needs to be—the larger the sample size has to be.

The *confidence level* describes “how confident” the researcher needs

79. FLOYD J. FOWLER, JR., *SURVEY RESEARCH METHODS* 35–36 (3d ed. 2002). *See also* HANS ZEISEL & DAVID KAYE, *PROVE IT WITH FIGURES: EMPIRICAL METHODS IN LAW AND LITIGATION* 109 (1997) (“[A]n occasional difficulty of accepting a small sample derives from the popular belief that the sampling error is related to the size of the population. Most people find it hard to believe that the same sample size will produce almost equally accurate estimates for the proportion of Republicans in California as the proportion, say, in Rhode Island. In sampling from very large populations, however, the size of the sample in relation to the whole population does not matter. What counts is the total number of units in the sample—the absolute, not the relative size of the sample.”).

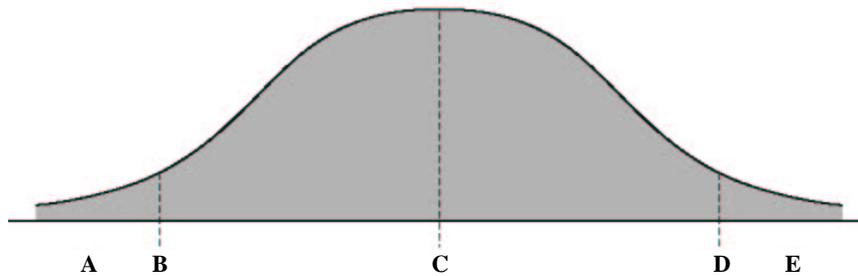
80. TERANCE D. MIETHE, *SIMPLE STATISTICS: APPLICATIONS IN CRIMINOLOGY AND CRIMINAL JUSTICE* 131 (2007).

81. DAVID L. FAIGMAN ET AL., *Scientific Method: The Logic of Drawing Inferences from Empirical Evidence*, in 1 *MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY* 179, 196 (2006–2007).

to be that the obtained confidence interval contains the actual population value in which the researcher is interested.⁸² Traditionally, scientists adopt the 95% level of confidence, which means that if 100 samples of the same size were drawn, the confidence interval expected for at least ninety-five of the samples would include the true population value.⁸³ An illustration may be useful: Figure 1 describes the data that are obtained in a hypothetical sample, with a sample mean of C. If 100 samples of the same size as this one were taken, in ninety-five out of those 100 samples the actual mean of the population would be between B and D—the lower and upper bounds of the “95% confidence interval,” respectively. In the remaining 5 samples, the actual mean of the population would be outside the 95% confidence interval—that is, in area A or E of the distribution.

FIGURE 1

- A Outside lower bound of 95% confidence interval
- B Lower bound of the 95% confidence interval
- C Sample Mean
- D Upper bound of 95% confidence interval
- E Outside upper bound of 95% confidence interval



Increasing the size of the sample will narrow the confidence interval, so that one can “zero in” more precisely on the true population value of interest. The key point in terms of cost, however, is that the size of the

82. *Id.* at 274–76.

83. See Diamond, *supra* note 49. As said elsewhere in the same volume, “a high confidence level alone means very little, but a high confidence level for a small [confidence] interval is impressive, indicating that the random error in the sample estimate is low.” David H. Kaye & David A. Freedman, *Statistical Proof*, in 1 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 219, 276 (David L. Faigman et al. eds., 2006–2007) (citations omitted).

sample and the narrowness of the confidence interval do not bear a linear, one-to-one, relationship: the gain in precision is much greater when a small sample is made larger than when an already-sizeable sample is made even larger. As stated in *Modern Scientific Evidence*, with “n” referring to the number of cases in the sample, “while a real gain in accuracy can be achieved when going from n=10 to n=25 or to n=100, as larger numbers are added to the sample, the marginal gain in accuracy shrinks quite considerably.”⁸⁴ Precision increases up to sample sizes of 150 to 200.⁸⁵ “After that point, there is a much more modest gain from increasing the sample size.”⁸⁶ In other words, after a certain point, it simply fails to be cost-efficient to increase the sample size, since the (marginal) gain in accuracy will not offset the (substantial) increased costs of enlarging the sample.

2. Containing Costs in Law

Judicial concern that the cost of litigation might overwhelm the adjudication process is long-standing. One prominent early example is Judge Charles Wyzanski’s exclusion of thousands of patents and related evidence in *United States v. United Shoe Machinery Corp.*⁸⁷ According to Judge Wyzanski, “[i]t is the function of counsel to aid the Court by selecting from the mass of evidence made available through . . . such portions as the Court can study within a reasonable time.”⁸⁸ He decried the “dump truck” approach to trial management:⁸⁹ “[c]ounsel cannot dump into

84. FAIGMAN ET AL., *supra* note 81, at 196.

85. FOWLER, *supra* note 79, at 36.

86. *Id.* For an illustration, see also Norbert Schwarz et al., *Survey Methods*, in 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY 143, 144 (Gilbert et al. eds., 4th ed. 1998), which states that:

Suppose that a survey based on simple random sampling . . . indicates that 70 percent of the sample would vote for candidate A. With a sample size of $N = 500$ and a desired confidence interval of 95 percent, sampling error would be around 4 percent. Hence, one may conclude that between 66 percent and 74 percent of the population would vote for candidate A. Sampling error decreases with increasing sample size, but the decrease is nonlinear. For the above example, sampling error declines from 9 percent for a sample of 100 to 4 percent for a sample of 500. But doubling the sample size from 500 to 1,000 further decreases sampling error only modestly to 3 percent, and to reduce sampling error to 1 percent one would need a sample of approximately 10,500.

Finally, it is important to note in this context that the sample size required to achieve a given 95% confidence interval usually cannot be specified in advance of conducting the sample. This is so because the 95% confidence interval will depend in part on how much variability there is in the data, and the data have to be collected before the degree of variability is known. FOWLER, *supra* note 79, at 35. (“[I]t is unusual for a researcher to be able to specify a desired level of precision in more than the most general way. It is only the exception, rather than the common situation, when a specific acceptable margin for error can be specified in advance.”).

87. *United States v. United Shoe Mach. Corp.*, 93 F. Supp. 190 (D. Mass. 1950).

88. *Id.* at 191.

89. *See, e.g., SEC v. Glass Marine Indus., Inc.*, 194 F. Supp. 879, 881 n.2 (D. Del. 1961).

the lap of the Court an undigested mass of documents comprising hundreds of thousands of pages and then expect the Court to read all of them, even if they were all to some degree both relevant and persuasive.”⁹⁰ Rather than examining each item in the “mass of evidence” individually, Judge Wyzanski proposed that the government sample the evidence.⁹¹

More specifically, he suggested that the government conduct by deposition a sample of forty-five shoe manufacturers operating fifty-five factories:⁹²

The Court arbitrarily selected from a standard directory of shoe manufacturers, the first 15 names that began with the first letter of the alphabet, the first 15 names that began with the eleventh letter of the alphabet, all 8 of the names that began with the twenty-first letter of the alphabet, and the first seven of the names that began with the twenty-second letter of the alphabet.⁹³

Today, concern about cost is reflected in at least three Federal Rules of Evidence. The earliest to occur is Rule 102, which provides that the Rules shall be construed to secure “elimination of unjustifiable expense and delay”⁹⁴ Specific clarification is provided by Rule 403, which states that relevant evidence may be excluded if the probative value of that evidence is outweighed “by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.”⁹⁵ The court’s obligation to control cost is confirmed by Rule 611, which provides that the court shall control the presentation of evidence to avoid “needless consumption of time.”⁹⁶

Since these three rules were adopted in 1975, they have received widespread judicial discussion and application, chiefly to justify trial time limitations—that is, costs measured in time—and occasionally to support the admission of evidence that provides an alternative to other, more

90. *United Shoe Mach. Corp.*, 93 F. Supp. at 191.

91. *See id.* Judge Wyzanski presented several options for how the evidence might be sampled. One possibility was calling a witness familiar with the shoe industry to testify in detail about a “half dozen [of the] commercially most important patents in the field of clicking machines and the half dozen commercially most important patents in the field of lasting machines” and the extent of defendant’s monopoly of those patents. Another suggested possibility was using an “analysis of the inventive history, ownership, commercial exploitation and the like of all shoe machinery patents issued say in the first three months of the year 1940” *Id.*

92. *United States v. United Shoe Mach. Corp.*, 110 F. Supp. 295, 305 (D. Mass. 1953).

93. *Id.* The defendant objected and Judge Wyzanski responded, “[i]f antitrust trials are to be kept manageable, samples must be used” *Id.*

94. FED. R. EVID. 102.

95. FED. R. EVID. 403.

96. FED. R. EVID. 611.

expensive, evidence—that is, costs measured in money. As an example of the use of the rules to justify trial time limitations, the court in *MCI Communications v. AT&T Corp.* directed the parties to submit lists of their witnesses and a summary of the testimony of each, together with an estimate of the time required for trial.⁹⁷ MCI's list named seventeen witnesses and predicted that it would require twenty-six days to present its case.⁹⁸ AT&T predicted that trial would take eight to nine months and that more than 160 witnesses would be called.⁹⁹ After reviewing the materials of both sides, the court decided to impose a twenty-six day limit for each side's case.¹⁰⁰

As an example of the use of the rules to justify the admission of evidence that provides an alternative to other, more expensive, evidence, the trial court in *Holbrook v. Lykes Brothers Steamship Co.* refused to permit two of the plaintiff's experts, including the treating physician, to testify, because the two experts lacked sufficient specialization to do so.¹⁰¹ The court of appeals reversed this decision on the ground that

[t]he district court's approach . . . would unjustly increase litigation costs by requiring litigants in countless cases to hire a host of experts out of fear that their treating physicians, in whom they entrusted their health and lives, would not "make the grade" when it came time to testify in court.¹⁰²

3. Containing Costs in Mass Torts

The judges in mass tort sampling cases have echoed Judge Wyzanski's fifty-year-old concern about cost. In *Cimino*, Judge Parker wrote, "[i]f the Court could somehow close thirty cases a month, it would take six and one-half years to try these cases and there would be pending over 5,000 untouched cases at the present rate of filing. Transaction costs would be astronomical."¹⁰³ He concluded bluntly, "[d]amages must be determined in the aggregate. Whether it is by the mechanism of the Court's

97. *MCI Commc'ns Corp. v. AT&T Corp.*, 708 F.2d 1081, 1170 (7th Cir. 1983).

98. *Id.*

99. *Id.*

100. *Id.* at 1170–71. *See also* *Enright v. Auto-Owners Ins. Co.*, 2 F. Supp. 2d 1072, 1074 (N.D. Ind. 1998) (imposing a thirty-hour time limit on the litigants to further the goal of "testimony through well-planned and thoughtful direct and cross-examination," leaving the decisions to exclude particular witnesses up to the attorneys); *Tabas v. Tabas*, 166 F.R.D. 10, 12–13 (E.D. Pa. 1996) (imposing a thirty-hour time limit on presentation of evidence).

101. *Holbrook v. Lykes Bros. S.S. Co.*, 80 F.3d 777 (3d Cir. 1996).

102. *Id.* at 782 n.1.

103. *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652 (E.D. Tex. 1990), *rev'd*, 151 F.3d 297 (5th Cir. 1998).

plan or by some other procedure approved or suggested by the Court of Appeals, without the ability to determine damages in the aggregate, the Court cannot try these cases.”¹⁰⁴ Exactly as Judge Parker had predicted, since the Fifth Circuit disallowed sampling,¹⁰⁵ the *Cimino* plaintiffs’ claims have yet to be adjudicated.¹⁰⁶

Likewise, in *Marcos*, Judge Real said “[p]ragmatically, the jury could not hear testimony of nearly 10,000 plaintiffs in this action within any practicable and reasonable time, to do justice to the class members.”¹⁰⁷ He continued, “[h]ere, individual trials for each of the 9,541 plaintiffs would take decades. Most of that time would be wasted since the nature of the injuries would be similar, if not identical, and the testimony would be largely duplicative.”¹⁰⁸

Finally, in *Schwab*, Judge Weinstein wrote:

If . . . an individualized process were undertaken, it would have to continue beyond all lives in being. Assuming tobacco companies were willing to expend the resources and monies necessary both in discovery and at trial to mount such an undertaking, the litigation costs in doing so would far exceed any monies saved by avoiding erroneous payments. . . . The interest of plaintiffs in avoiding the additional litigation costs that would arise if defendants were permitted to confront each possible plaintiff at trial is weighty. The necessary additional litigation costs to plaintiffs would exceed any recovery possible from defendants, making continued pursuit of the litigation fruitless.¹⁰⁹

Judges Parker, Real, and Weinstein, just as Judge Wyzanski, considered alternatives to individualized proof, and each, largely for reasons of cost, chose to randomly sample relevant evidence. Judge Parker explained, “[t]he reasons the courts have come to rely on statistics are the same reasons that society embraces the science. It has been proved to provide information with an acceptable degree of accuracy and

104. *Id.* at 667.

105. *See Cimino v. Raymark Indus., Inc.*, 151 F.3d 297 (5th Cir. 1998).

106. Following remand, the Joint Panel on Multidistrict Litigation (“MDL”), over the *Cimino* plaintiffs’ objections, transferred those claims to the Eastern District of Pennsylvania. The claims apparently have been merged with the ongoing effort to settle the many asbestos cases pending in the Pennsylvania MDL.

107. *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1462 (D. Haw. 1995), *aff’d sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996).

108. *Id.* at 1467.

109. *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992, 1247 (E.D.N.Y. 2006). Judge Weinstein also noted that “[t]ransaction costs would be enormous. Most of these costs would be borne by the public through financing of a court system that would require expansion.” *Id.*

economy.”¹¹⁰ Likewise, Judge Real stated that the court “must weigh defendant’s claim to the right to trial in each individual case against judicial economy and manageability by use of a valid statistical procedure.”¹¹¹ Finally, Judge Weinstein explained his approval of sampled evidence in similar terms: “[s]ampling and survey techniques are well-accepted alternatives for the trial judge facing crippling discovery and evidentiary costs. . . . In some cases sampling techniques may ‘provide the only practicable means to collect and present relevant data.’”¹¹²

As we have suggested elsewhere, Judges Parker and Real ultimately chose an inefficient version of random sampling that incorporated elements of individualized case adjudication.¹¹³ Judge Weinstein adopted sampling “without apology”¹¹⁴ and without individualization. The common question that these three cases raise is whether now, after more than a half century in development, a new rule of sampling evidence is emerging.

B. A NEW RULE OF COST-EFFICIENT EVIDENCE GATHERING

Beginning, at least, in *United States v. United Shoe Machinery Corp.*,¹¹⁵ continuing in scattered opinions, and gaining momentum in recent mass tort cases, a new rule of evidence is emerging. This rule can be thought of as a judicial construction of Federal Rules of Evidence 102, 403, and 611, a modification *requiring* consideration of alternative proof in the face of potentially overwhelming trial costs. We view the opinions of Judges Parker, Real, and Weinstein as strong common law precedent holding that federal trial judges are now obliged to consider alternative methods of proof in order to serve the goal of efficient trial announced in Federal Rule of Evidence 102 and clarified in Rules 403 and 611. Since the problem of overwhelming cost can be a threat to litigating any kind of case and can arise from any kind of evidence, this rule may have a broad scope of application. In any particular case several alternatives may be available, but in most cases involving potentially prohibitive costs, random sampling of relevant evidence will likely be the best form of alternative proof. As described above, scientists confronting overwhelming costs, if they were to

110. *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 663 (E.D. Tex. 1990), *rev'd*, 151 F.3d 297 (5th Cir. 1998).

111. *In re Estate of Marcos*, 910 F. Supp. at 1467.

112. *Schwab*, 449 F. Supp. 2d at 1244 (quoting the MANUAL FOR COMPLEX LITIGATION (FOURTH) § 11.493 (2004)).

113. *See Walker & Monahan*, *supra* note 63, at 554–55.

114. *Id.* at 546.

115. *United States v. United Shoe Mach. Corp.*, 93 F. Supp. 190 (D. Mass. 1950).

examine an entire population, have pioneered the strategy of using a random sample to estimate values for the whole population.¹¹⁶ The remarkable virtue of the random sample, they demonstrated, is its potential not only for enormous cost savings, but also for providing a *more* accurate picture of the facts at issue than the study of each case one-by-one. In many areas of the law, judges have seen the clear relevance of sampling to the management of large amounts of evidence in litigation.

When is this emerging judicial requirement to consider alternative forms of proof to be triggered? Sometimes, of course, it will be evident that no trial can be held unless randomly sampled evidence is permitted. This was clearly the case in *Cimino* and *Marcos* and is the case in *McLaughlin*. Given the remarkable potential of randomly sampled evidence to provide greater accuracy at lower cost, however, the emerging rule may have a broader reach in the future. The individual cases in *McLaughlin* number in the millions; *Cimino* and *Marcos* each concerned thousands of people.¹¹⁷ Accuracy and efficiency would be as well served if the requirement to consider alternative forms of proof were triggered when the number of cases to be litigated was in the hundreds.

V. A PROPOSAL FOR THE USE OF SAMPLED EVIDENCE

We have to this point argued that a requirement is emerging that judges consider alternative forms of proof when individualized adjudication is infeasible, and that in most cases the use of statistical sampling will be the least expensive and most accurate of those alternatives. But *how* should courts employ statistical sampling? In this Part, we integrate statistical theory with traditional legal doctrine to offer a detailed proposal for how courts and juries should treat sampled evidence at trial.

Importantly, in making this proposal, we assume that the sampled evidence being proffered has been reviewed by the court and found admissible pursuant to Federal Rule of Evidence 702,¹¹⁸ the key precedent

116. See *supra* Part IV.A.1.

117. Schwab v. Philip Morris USA, Inc., 449 F. Supp. 2d 992 (E.D.N.Y. 2006), *appeal docketed sub nom.* McLaughlin v. Philip Morris USA, Inc., No. 06-4666 (2d Cir. Nov. 17, 2006) (estimated 50 million plaintiffs); *In re* Estate of Marcos Human Rights Litig., 910 F. Supp. 1460, 1462 (D. Haw. 1995), *aff'd sub nom.* Hilao v. Estate of Marcos, 103 F.3d 767 (9th Cir. 1996) (approximately 10,000 plaintiffs); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 652-53 (E.D. Tex. 1990), *rev'd*, 151 F.3d 297 (5th Cir. 1998) (2,298 plaintiffs).

118. FED. R. EVID. 702.

underlying that rule, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹¹⁹ and related Supreme Court decisions concerning expert testimony.¹²⁰ That is, we assume that Rule 702 and the *Daubert* vetting process has already excluded sampled evidence with substantial “non-sampling error.” Non-sampling error, or systematic error, results from aspects of the research design that lead to misleading or biased results. Examples of non-sampling error include choosing the wrong population to sample, failing to obtain responses from many of the cases randomly selected for the sample, or simply asking the wrong question.¹²¹ Being found admissible under Rule 702 and *Daubert*, however, does not necessarily mean that the research is completely free of non-sampling error. As Judge Posner stated, “[t]rials would be very short if only perfect evidence were admissible.”¹²² The role of the jury in assessing any residual bias in research that a court has already found to possess fundamental “scientific validity”¹²³ is addressed below.¹²⁴

Our proposal pivots on the concept of the “95% confidence interval,” introduced above.¹²⁵ Recall that if 100 samples of a similar size were taken to estimate the mean of a given population, in ninety-five out of those 100 samples the actual mean of the population would be between the lower and upper bounds of the 95% confidence interval. In the remaining 5 samples, the actual mean of the population would be outside—either lower or higher than—the 95% confidence interval. In the illustration given in Figure 1, the 95% confidence interval would be between the points labeled B and D, and areas A or E would be outside the 95% confidence interval. Our proposal,

119. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993).

120. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999) (holding that a court’s “gatekeeping” obligation under *Daubert* extends to all expert testimony, not just scientific testimony); *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997) (holding that the trial judge’s decision to admit or exclude scientific evidence should be reviewed for abuse of discretion and exclusion based on a perceived “analytical gap” between data and opinion is not an abuse of discretion).

121. These three forms of non-sampling error are called “frame errors,” “nonresponse errors,” and “measurement errors,” respectively. JUDITH T. LESSLER & WILLIAM D. KALSBECK, NONSAMPLING ERROR IN SURVEYS 10 (1992). *See also* 1 MCCORMICK ON EVIDENCE § 208 (Strong ed., 5th ed. 1999).

122. *Indianapolis Colts, Inc. v. Metro. Balt. Football Club Ltd.*, 34 F.3d 410, 416 (7th Cir. 1994).

123. *Daubert*, 509 U.S. at 590 n.9.

124. To simplify our proposal, we limit our discussion to the case where only one party offers into evidence the results of a single study that relied on sampled data. This was the situation in *Cimino*. *See Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 653 (E.D. Tex. 1990) (describing survey in the form of damages trial for randomly selected sample class members), *rev’d*, 151 F.3d 297 (5th Cir. 1998). It was also the case in *Marcos*. *See In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1464 (D. Haw. 1995) (describing survey in the form of depositions of and testimony from random sample of class members in compensatory damages phase of trial), *aff’d sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996). Our proposal, however, is based on broad principals that might readily be applied in more complex situations.

125. *See supra* Part IV.A.1.

in essence, is that the jury should be permitted to make any factual finding based on sampled evidence that is within the sample's 95% confidence interval, and that the court should be required to intervene whenever a factual finding based on sampled evidence exceeds the sample's 95% confidence interval.

A. THE JURY: FACT-FINDING WITHIN THE 95% CONFIDENCE INTERVAL

The jury's task is to resolve uncertainty regarding factual propositions at issue in litigation. In *McLaughlin*, those factual propositions include the extent to which plaintiffs relied on defendants' assertions that "light" cigarettes were less harmful than regular cigarettes, and—should reliance be proven—the amount of injury suffered by the plaintiffs. The method that Judge Weinstein proposes for resolving uncertainty regarding these propositions is the use of sampled evidence.

In a perfectly designed survey that involves sampling, the "best estimate" of the mean of the population of interest is the mean of the sample. Perfect design is not likely, however, and hence there is important work for the jury: the search for any systematic non-sampling error ("bias") in the design of the survey.¹²⁶ For example, in a survey to determine smoker reliance on "light" cigarette advertising, subjects might all have been asked a question which suggested that they were influenced more by one advertisement than by another, like "which of these four cigarette advertisements are most persuasive to you?," without being given the option of responding that the four advertisements were equally persuasive. A jury determination of non-sampling error or bias can properly be made based on the testimony of the expert introducing the survey results and the experts for the opposing side critiquing those results, as well as by the jury's own consideration of the methodology of the study. With respect to the methodology of the hypothetical survey of smokers' reliance on "light" cigarette advertising, for example, the jury could properly be shown questionnaires used in the research, listen to descriptions about the setting and manner of data collection, and even view illustrative videotaped subject responses.

The key point, however, is that the jury would be free to make a determination based on sampled evidence *only within the sample's 95% confidence interval*. We make this proposal because, before the sampled evidence has been presented to the jury, the trial court would have made

126. But the results could not contain an excessive amount of systematic error. Otherwise, the results would not have been admissible as scientifically valid in light of *Daubert* and Rule 702.

the general determination that the use of sampled evidence was “superior” to the use of individualized evidence in litigating the case, as well as the specific determination that the sampled evidence at issue survives the Rule 702 and *Daubert* vetting process and is “scientifically valid”—that is, without disqualifying non-sampling error. In essence, the court would have told the jury that sampling was the best way to determine a fact in the case, and that the particular sample before them was a sound one. Jury consideration of non-sampling error or bias in the design of a sample survey can and often should result in jury findings that differ from the sample mean, but not that exceed the 95% confidence interval around that mean.¹²⁷

Indeed, it would follow from this analysis that a jury finding that exceeded the sample’s 95% confidence interval should result in a new trial before a second jury. Federal Rule of Civil Procedure 59(a) provides that:

A new trial may be granted to all or any of the parties and on all or part of the issues . . . in an action in which there has been a trial by jury, for any of the reasons for which new trials have heretofore been granted in actions at law in the courts of the United States¹²⁸

One of the permissible reasons for a new trial is a jury verdict “against the weight of the evidence.”¹²⁹ This ground for turning to another jury emanates from the proposition that a jury must “choose between plausible versions of the evidence.”¹³⁰ Any jury finding based on a sample determined by Rule 702 and *Daubert* to be scientifically valid that is outside that sample’s 95% confidence interval is against the weight of the evidence and therefore not plausible. Any course of action other than ordering a new trial before a second jury would be a miscarriage of justice.¹³¹

127. The specific task of the jury will vary from trial to trial and may take the form of determining an outcome, answering questions, or a combination of both functions. *See, e.g.*, FED R. CIV. P. 49.

128. FED. R. CIV. P. 59(a).

129. 11 CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 2805 (2d ed. 1995). The other traditional grounds for a new trial are that damages are excessive, that for other reasons the trial was not fair or because of substantial errors in the admission or rejection of evidence or the giving or refusal of jury instructions. *Id.* at 54.

130. *Jacobs Mfg. Co. v. Sam Brown Co.*, 19 F.3d 1259, 1267 (8th Cir. 1994) (citing *White v. Pence*, 961 F.2d 776, 781 (8th Cir. 1992)). *See also* *Waitek v. Dalkon Shield Claimants Trust*, 934 F. Supp. 1068, 1094 (N.D. Iowa 1996) (“[I]t is the jury’s province to choose between plausible versions of the evidence” (citing *Jacobs Mfg. Co.*, 19 F.3d at 1267)).

131. *See, e.g.*, *Lama v. Borrás*, 16 F.3d 473, 477 (1st Cir. 1994) (noting that there is no abuse of discretion in denying a motion for a new trial unless “the verdict was so clearly against the weight of the evidence as to amount to a manifest miscarriage of justice” (quoting *PH Group Ltd. v. Birch*, 985 F.2d 649, 653 (1st Cir. 1993))); *Song v. Ives Labs., Inc.*, 957 F.2d 1041, 1047 (2d Cir. 1992); *Williamson v. Consol. Rail Corp.*, 926 F.2d 1344, 1348 (3d Cir. 1991) (reversing grant of new trial but

B. THE COURT: JUDGMENT OUTSIDE THE 95% CONFIDENCE INTERVAL

If the jury should be permitted to make any factual finding based on sampled evidence that is within the sample's 95% confidence interval, what is the role of the court? We propose that the court should be required to make determinations whenever a factual finding based on sampled evidence exceeds the sample's 95% confidence interval. We believe that the role of judicial responsibility with respect to sampled evidence is at the extremes. In terms of the distribution of sampled evidence depicted in Figure 1, the role of the court lies in the area designated A and E. When sampled evidence that has passed the Rule 702 and *Daubert* vetting process indicates that the fact at issue lies outside the sample's 95% confidence interval, there is nothing for the jury to determine. The court can either make a summary judgment, or can make a judgment as a matter of law.

1. Summary Judgment

Federal Rule of Civil Procedure 56(c) states that summary judgment "shall be rendered forthwith if the pleadings, depositions, answers to interrogatories, and admissions on file, together with the affidavits, if any, show that there is no genuine issue as to any material fact . . ." ¹³² Any study involving sampling ruled admissible under Rule 702 and *Daubert* would already have been found to be "material" to the litigation. The issue for granting summary judgment is whether the sampled evidence would leave any "genuine issues" to be resolved by the jury.

Although Wright, Miller, and Kane note, regarding summary judgment, that "[d]oubts as to the credibility of the movant's affiants or witnesses may lead the court to conclude that a genuine issue exists," ¹³³

noting that review of district court's granting motion is grounded in whether trial court could reasonably have concluded that a miscarriage of justice would occur if the jury's verdict were left to stand); *Katara v. D.E. Jones Commodities, Inc.*, 835 F.2d 966, 970 (2d Cir. 1987) (noting that a trial court should grant a motion for a new trial "when convinced that the . . . verdict is a miscarriage of justice"); *Boehringer Ingelheim Vetmedica, Inc. v. Schering-Plough Corp.*, 166 F. Supp. 2d 19, 50 (D.N.J. 2001) (noting that, because the verdict is against the weight of the evidence, a new trial should be granted if the jury's verdict results in a miscarriage of justice, but finding no miscarriage of justice based on the record), *aff'd*, 320 F.3d 1339 (Fed. Cir. 2003); *Waitek v. Dalkon Shield Claimants Trust*, 934 F. Supp. 1068, 1092 (N.D. Iowa 1996) (finding the jury's verdict to be based on credible evidence and thus no new trial warranted but noting that a district court should grant a new trial if, after weighing the evidence, the jury's verdict amounts to a miscarriage of justice), *aff'd*, 114 F.3d 117 (8th Cir. 1997); *Henry v. Hess Oil Virgin Islands Corp.*, 163 F.R.D. 237, 242 (D.V.I. 1995) (explaining that on motion for a new trial for verdict against the weight of the evidence, the judge's "essential function" is to prevent a "miscarriage of justice").

132. FED. R. CIV. P. 56(c).

133. 10A CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE

they point to *Trigo Hermanos, Inc. v. Premium Wholesale Groceries, Inc.*¹³⁴ for a limiting proposition. There, the court stated that “[w]hen at the hearing on a motion for summary judgment there is contradictory evidence, or the movant’s evidence is impeached, an issue of credibility is present, provided the contradicting or impeaching evidence is not too incredible to be believed by reasonable minds.”¹³⁵ Thus, a conflict in evidence will prevent summary judgment, but the conflict must occur within limits of credibility. For example, unsupported assertions that a document might be forged or that a witness might be lying will not suffice to prevent summary judgment.

Applying this standard for summary judgment to the context of sampled evidence, we argue that outside the 95% confidence interval there is no “credible” contradicting evidence that may be relied on by the non-moving party. For example, in *McLaughlin*, an issue is what percentage of smokers relied upon inaccurate health information contained in advertisements for “light” cigarettes. If a study that sampled smokers survived the Rule 702 and *Daubert* testing process and concluded that the mean percentage of smokers who relied on the advertisements was Point C in Figure 1, then although the true percentage of smokers who relied on the advertisements *might* fall outside the lower or higher bounds of the 95% confidence interval (into areas A or E of Figure 1), this possibility lacks sufficient statistical “credibility” (that is, it will happen 5 or fewer times in 100 samples) to preclude summary judgment. If the party moving for summary judgment is the party favored by a “high” mean on some set of sampled data, then no “genuine issue” of fact exists at the lower bound of the 95% confidence interval (Point D in Figure 1), because any lower result is outside the accepted range of credibility, given that the sample mean is Point C. In this example, since the party claiming that smokers were misled by their reliance on “light” cigarette advertising would benefit by a finding of a high rate of reliance on misleading information, that party might move for and obtain summary judgment granted at the lower bound of the 95% confidence interval.¹³⁶

AND PROCEDURE § 2726 (3d ed. 1998).

134. *Trigo Hnos., Inc. v. Premium Wholesale Groceries, Inc.*, 424 F. Supp. 1125 (S.D.N.Y. 1976).

135. *Id.* at 1129 (quoting 6 JAMES WM. MOORE ET AL., MOORE’S FEDERAL PRACTICE ¶ 56.15(4) (2d ed. 1976)).

136. On the other hand, and for similar reasons, in this situation the non-moving party may claim that no genuine issue of fact exists at the high bound. Of course, in this situation it seems unlikely that the non-moving party would wish to cross-move for summary judgment (or make an initial motion), but the formal option does exist under our analysis.

2. Judgment as a Matter of Law

Federal Rule of Civil Procedure 50(a) permits judgment as a matter of law “if a party has been fully heard on an issue during a jury trial and the court finds that a reasonable jury would not have a legally sufficient evidentiary basis to find for that party on that issue”¹³⁷ Rule 50(b) provides in the case of a Rule 50(a) motion made, but not granted, that “[t]he movant may renew its request for judgment as a matter of law by filing a motion no later than 10 days after entry to judgment”¹³⁸ Thus, either during trial or after trial, Rule 50 permits the judge to decisively enter a judgment under the appropriate conditions. The United States Supreme Court held in *Anderson v. Liberty Lobby, Inc.*,¹³⁹ that the practice under Rule 50 (judgment as a matter of law) was essentially the same as under Rule 56 (summary judgment),¹⁴⁰ the only difference being a matter of timing. The Court stated that “the inquiry under each is the same: whether the evidence presents a sufficient disagreement to require submission to a jury or whether it is so one-sided that one party must prevail as a matter of law.”¹⁴¹ Thus, the party benefited by a “high” mean on some set of sampled data may have a Rule 50 motion for judgment as a matter of law granted at the lower bound of the 95% confidence interval, either during or after trial.¹⁴²

VI. CONCLUSION

With the split between the Fifth and Ninth Circuits, and with *McLaughlin* now on interlocutory appeal before the Second Circuit, the use of sampling in mass tort litigation is at a crossroads. We have here strenuously defended Judge Weinstein’s plan to entirely eliminate proof of individual class member claims in the face of the enormous cost in time and money of gathering individualized evidence, and instead rely solely on

137. FED. R. CIV. P. 50(a).

138. FED. R. CIV. P. 50(b).

139. *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242 (1986).

140. *Id.* at 250–51 (“The inquiry performed is the threshold inquiry of determining whether there is the need for a trial—whether, in other words, there are any genuine factual issues that properly can be resolved only by a finder of fact because they may reasonably be resolved in favor of either party. . . . Petitioners suggest, and we agree, that this [Rule 56] standard mirrors the standard for a directed verdict under Federal Rule of Civil Procedure 50(a), which is that the trial judge must direct a verdict if, under the governing law, there can be but one reasonable conclusion as to the verdict. If reasonable minds could differ as to the import of the evidence, however, a verdict should not be directed.”) (citations omitted).

141. *Id.* at 251–52.

142. The non-moving party may have the same at the high bound of the 95% confidence interval.

evidence produced through proven methods of sampling. Such an aggregate strategy is not only much more cost-effective than case-by-case evidence gathering, but it is also much more likely to arrive at an accurate determination of damages. Courts should not only be permitted to rely on sampling evidence in mass tort cases, but they should also be required to do so whenever sampling is the most cost-effective way to resolve claims. By laying out a comprehensive proposal for the use of sampled evidence in litigation, we hope to hasten the day when legal decision makers finally accept a practice that has been a hallmark of scientific decision making for the past sixty years.

