


Fall 1993

Science and Ethics in Conducting, Analyzing, and Reporting Social Science Research: Implications for Social Scientists, Judges, and Lawyers

Robert Rosenthal
Harvard University

Peter David Blanck
University of Iowa College of Law

Follow this and additional works at: <https://www.repository.law.indiana.edu/ilj>

 Part of the [Judges Commons](#), [Legal Ethics and Professional Responsibility Commons](#), [Legal Profession Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Rosenthal, Robert and Blanck, Peter David (1993) "Science and Ethics in Conducting, Analyzing, and Reporting Social Science Research: Implications for Social Scientists, Judges, and Lawyers," *Indiana Law Journal*: Vol. 68 : Iss. 4 , Article 9.

Available at: <https://www.repository.law.indiana.edu/ilj/vol68/iss4/9>

This Article is brought to you for free and open access by the Law School Journals at Digital Repository @ Maurer Law. It has been accepted for inclusion in Indiana Law Journal by an authorized editor of Digital Repository @ Maurer Law. For more information, please contact rvaughan@indiana.edu.

Science and Ethics in Conducting, Analyzing, and Reporting Social Science Research: Implications for Social Scientists, Judges, and Lawyers

ROBERT ROSENTHAL*
PETER DAVID BLANCK**

INTRODUCTION

This Article describes a number of scientific, ethical, and legal issues relevant to conducting, analyzing, and reporting social science research. Legal literature has devoted little attention to examining the scientific and ethical issues in evaluating empirical social science research. This is particularly troubling in light of the fact that empirical research is presented with increasing regularity to judges and juries as trial evidence.¹

There is a strong relationship between the ethics, the scientific quality, and the evidentiary importance of the way social science research is conducted, analyzed, and reported. Everything else being equal, social science research of higher scientific quality is likely to be more ethically defensible and more relevant to the resolution of a legal dispute. The higher the quality of the research program, the better invested has been the time of the research participants, the funds of the granting agency, and the space of the journals that report the results, as well as the general investment that a legal fact finder (that is, a judge or jury) has made in evaluating the importance of the

* Professor of Social Psychology, Harvard University. Ph.D., 1956, UCLA. Robert Rosenthal's preparation of this Article was supported in part by the Spencer Foundation.

** Professor of Law, University of Iowa College of Law; Senior Fellow, The Annenberg Washington Program. Ph.D., 1982, Harvard University; J.D., 1986, Stanford University. The authors thank Michael Saks for his comments on earlier drafts of this Article. Peter David Blanck's preparation of this Article was supported in part by The Annenberg Washington Program.

1. See, e.g., Linda Greenhouse, *Court Hears Case on Science Rules*, N.Y. TIMES, Mar. 31, 1993, at A19 (reporting U.S. Supreme Court's difficulty in deciding when scientific theory is sound enough to be admitted into evidence when it heard oral arguments in *Daubert v. Merrell-Dow Pharmaceuticals, Inc.*, No. 92-102 (U.S. argued Mar. 30, 1993)); *Communicating with Juries, Panel One: Judge-Jury Communications: Improving Communications and Understanding Bias*, *The Annenberg Washington Program Conference, April 10, 1992*, 68 IND. L.J. 1037, 1042 (1993) (statement of Robert Rosenthal) (cautioning against accepting results of one study without looking at results of similar studies).

For examples of discussions of the legal relevance of research methodology and ethics, see JOHN MONAHAN & LAURENS WALKER, *SOCIAL SCIENCE IN LAW* (1989); Peter D. Blanck, *Empirical Study of the Employment Provisions of the Americans with Disabilities Act: Methods, Preliminary Findings, and Implications*, 22 N.M. L. REV. 119 (1992); Peter D. Blanck, *On Integrating Persons with Mental Retardation: The ADA and ADR*, 22 N.M. L. REV. 259 (1992); Peter D. Blanck, *The Emerging Work Force: Empirical Study of the Americans with Disabilities Act*, 16 J. CORP. L. 693 (1991).

research to the resolution of a dispute. The lower the research quality, the less justified the researcher is in wasting not only the research participant's time, but also the funding agency's money, the journal's space, and the court's time.

Part I of this Article explores several scientific, ethical, and legal issues related to conducting social science research. Part II then highlights the importance to social scientists and to the legal community of viewing the analysis of empirical social science data as tied to the ethics of conducting research. Finally, Part III discusses how the reporting of social science research, whether in professional journals or in the courtroom, has implications for social scientists and for the evaluation of research in the adversarial process.

I. SCIENTIFIC, ETHICAL, AND LEGAL ISSUES IN CONDUCTING SOCIAL SCIENCE RESEARCH

This Part explores several considerations of research design—that is, the methods and procedures employed in a study—and recruitment of human research participants. In evaluating the ethical employment of human participants, issues of safety can be distinguished from more subtle issues of research ethics.² Research that is unsafe for participants is ethically questionable. Perfectly “safe” research, however, which puts no participant at risk, also may be ethically questionable because of the shortcomings of the design.

A. Issues of Design

Imagine that a court or jury in a school desegregation case is to evaluate as evidence the results of a research project. The study hypothesizes that private schools improve children's intellectual functioning more than public schools. In the study, children from selected private and public schools are tested extensively. The research hypothesis is examined by comparing scores earned by students from private versus public schools.

Assume that the safety of the children tested is not at issue. It can be argued, however, that the school research raises important ethical issues because of the inadequacy of its design. The study's goal is to determine the causal impact on student performance of private versus public schooling.

2. Peter D. Blanck et al., *Scientific Rewards and Conflicts of Ethical Choices in Human Subjects Research*, 47 AM. PSYCHOLOGIST 959 (1992); Thomas Grisso et al., *Standards in Research: APA's Mechanism for Monitoring the Challenges*, 46 AM. PSYCHOLOGIST 758 (1991) (discussing ethical issues studied by the American Psychological Association's Committee on Standards in Research).

However, the research design does not permit reasonable causal inference because of the absence of randomization or a reasonable attempt to consider plausible rival hypotheses.³

How does the poor quality of the hypothetical school study design raise ethical and legal objections to the proposed research? The ethical objections stem in part from the fact that students', teachers', and administrators' time will have been taken from potentially more beneficial educational experiences. In the context of an adversarial process, the problem is that the poorly designed study is likely to lead to unwarranted and inaccurate conclusions by a fact finder.⁴ Had the hypothetical research question been appropriate to the research design, the ethical and evidentiary issues would have been less acute. For example, if the investigators had set out only to learn whether performance differences exist between students in private versus public schools, their design would have been perfectly appropriate to their question.

B. Issues of Recruitment

Much attention has been devoted to considering a variety of ethical issues in the selection and recruitment of human participants in social science research.⁵ The following comments address the implications of participant selection and recruitment choices for the use of social science research in the courtroom.

First, based on several reviews of the literature, it has been shown that a number of research methods or procedures may be designed to reduce participant volunteer bias. In other words, certain research methods may actually increase the generality or external validity⁶ of social science research results.⁷ To provide one example, the external validity of psychological

3. THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN & ANALYSIS ISSUES FOR FIELD SETTINGS* 341-83 (1979).

4. In addition, the time and money allocated to poor quality science serve to keep those finite resources of time and money from better quality science in an undeniably zero-sum world.

5. AMERICAN PSYCHOLOGICAL ASS'N, *ETHICAL PRINCIPLES IN THE CONDUCT OF RESEARCH WITH HUMAN PARTICIPANTS* (1982); see also HERBERT C. KELMAN, *A TIME TO SPEAK: ON HUMAN VALUES AND SOCIAL RESEARCH* 202-25 (1968); Blanck et al., *supra* note 2, at 962-63; Grisso et al., *supra* note 2, at 760-61.

6. "External validity is the problem of interpreting the difference, the problem of generalization. To what other populations, occasions, stimulus objects, and measures may the obtained results be applied?" EUGENE J. WEBB ET AL., *UNOBTRUSIVE MEASURES: NONREACTIVE RESEARCH IN THE SOCIAL SCIENCES* 10-11 (1966) (emphasis in original).

7. ROBERT ROSENTHAL & RALPH L. ROSNOW, *ESSENTIALS OF BEHAVIORAL RESEARCH: METHODS AND DATA ANALYSIS* 228-29 (2d ed. 1991) [hereinafter *ESSENTIALS OF BEHAVIORAL RESEARCH*]; ROBERT ROSENTHAL & RALPH L. ROSNOW, *THE VOLUNTEER SUBJECT* 119-20 (1975) [hereinafter *THE VOLUNTEER SUBJECT*]; RALPH L. ROSNOW & ROBERT ROSENTHAL, *BEGINNING BEHAVIORAL RESEARCH: A CONCEPTUAL PRIMER* (1993) [hereinafter *BEGINNING BEHAVIORAL RESEARCH*].

research on jurors' comprehension of pattern instructions in a death penalty case in a particular jurisdiction may be enhanced substantially by employing research methods that maximize the recruitment of potential jurors from that jury pool.⁸

Employment of procedures that enhance the generality of research results has also led some social scientists to view human research participants as yet another "granting agency," which, in reality, they are, since research participants must decide whether to grant social scientists their time, attention, and cooperation.⁹ Treating research participants as another granting agency also provides the researcher incentives to give participants more information about the long-term benefits of the research.

In giving prospective participants enhanced information about the study—and ultimately its findings and conclusions—social scientists, expert witnesses, and legal practitioners should avoid "hyperclaiming" or overstating the implications of the research. An example of hyperclaiming might be an attempt by an expert witness to convey to a fact finder that the results of a research project are likely to demonstrate broader conclusions than, in fact, they are likely to reflect—as might happen in the hypothetical school study¹⁰

Presumably, through the adversarial process—the presentation of experts on both sides—fact finders are able to evaluate research claims and hyperclaims fairly well.¹¹ Absent more systematic information about research design and method, however, fact finders may not be able to fully evaluate research claims. Therefore, one component of the researcher's ethical obligation may be to explain to the participants and to a fact finder what the research can actually accomplish. Again, in the hypothetical school study, this approach contrasts with the claim that the research may demonstrate the superiority of private over public schooling.

Closely related to the phenomenon of hyperclaiming is the phenomenon of "causism." Causism refers to the tendency to imply a causal relationship where none has been established, that is, where the research data do not support it. Understanding the nature of causism is particularly important to the presentation of social science research in the courtroom. This is because one party typically presents one study to evidence a cause-and-effect relationship,

8. See, e.g., *United States ex rel. Free v. Peters*, 806 F. Supp. 705 (N.D. Ill. 1992) (vacating death penalty on grounds that jurors did not comprehend pattern instructions).

9. See, e.g., *THE VOLUNTEER SUBJECT*, *supra* note 7.

10. See *supra* text accompanying notes 3-4.

11. See, e.g., Kenneth R. Kreiling, *Scientific Evidence: Toward Providing the Lay Trier with the Comprehensible and Reliable Evidence Necessary to Meet the Goals of the Rules of Evidence*, 32 ARIZ. L. REV. 915, 929-35 (1990).

while the opposing party tries to demonstrate the opposite with a different study.

The following characteristics of causism are relevant to the discussion here:

- (a) the absence of an appropriate evidentiary base supporting the causal inference;
- (b) the presence of language directly implying cause; for example, "the effect of," "the impact of," "the consequence of," or "as a result," where more appropriate language would have been "is related to," "is predictable from," or "may be inferred from", and
- (c) the self-serving benefits to the "causist" (*e.g.*, social scientist or legal proponent).

As a general matter, causism is self-serving because it makes the causist's result appear more important or fundamental to a fact finder than it really is. Cases in which causism is present, yet its perpetrator (for example, the expert witness) is unaware of the causism, reflect poor scientific training or legal preparation. Voir dire of experts at trial may be one way to focus increasingly on the nature of self-serving causism.¹²

Where causism is found, and its perpetrator is aware of the causism, it reflects unethical misrepresentation and deception, either to the scientific community or to the fact finder. The expert witness caught perpetrating causism during testimony has likely dealt a fatal blow to his or her credibility before a fact finder.¹³

Whereas well-trained social scientists can readily differentiate causist language from inferentially more accurate language, research participants, policy makers, and fact finders ordinarily cannot. Additionally, when a proposed research study is described in causal language, it may often represent an unfair participant recruitment device. This recruitment approach is at best inaccurate, when employed out of ignorance, and at worst dishonest, when employed as hype to increase participation rates. One consequence of this knowledge for trial attorneys is to invite more careful probing of the testifying expert's evaluation of research participation, of participant drop-out rates in the evaluation of the generalizability of a study's results, and of its subsequent usefulness to the fact finder.

12. *Cf.* FED. R. EVID. 702 (providing for when an expert witness may testify in the form of opinion).

13. Of course, the broader ethical question raised by deliberate use of causism exists independent of the perpetrator being caught.

C. Bad Science Makes for Bad Ethics

Causism is only one example of how social scientists and attorneys can use bad science to misrepresent research results to a finder of fact. Additionally, poor quality of research design, poor quality of data analysis, and poor quality of research reporting lessen both the ethical justification and legal relevance of any research project. This applies not only where deception, discomfort, or embarrassment of the participants is involved but, in many cases, to the most benign research experience for participants.

If, because of the poor quality of the science, no benefit can come of the research study, how are social scientists to justify the use of participants' time and effort, and the money, space, and other resources that have been expended on the research project? Likewise, how are trial attorneys to justify the presentation of bad science at trial? When we add the inescapable zero-sum nature of time, attention, effort, money, and other resources to the "no good can come of it" argument, it becomes difficult to justify poor quality research on any ethical or professional basis.¹⁴ It also becomes increasingly difficult for attorneys to justify the presentation of bad science to a finder of fact. One reason, among others, is that such presentation conflicts with the attorney's code of professional responsibility.¹⁵

Poor research quality makes for poor quality of education as well. This is true especially where student participant pools are employed so that participation is quasi-coercive. In fact, the use of such pools is typically justified by the fact that participants will benefit educationally.

But if individuals are required to participate in poor quality research, they are likely to acquire only misconceptions about the nature of the science. To stay with the school study hypothetical, if the personality scale scores for student participants from public and private schools are correlated with their

14. Elsewhere, we have argued that Institutional Review Boards (IRBs) must consider the technical scientific competence of the investigators whose proposals they are asked to evaluate. That will increase the work required of IRB members and change their composition somewhat to include a degree of methodological expertise. Even then it will not always be easy to come to a decision about the scientific competence of the investigator and of the particular proposal; but then, it is not always easy to come to a decision about the more directly ethical aspects of the proposal either. See Robert Rosenthal, Science and Ethics in Conducting, Analyzing, and Reporting Psychological Research, Presentation at the Meeting of the American Psychological Association, Washington, D.C. (Aug. 15, 1992); see also Ralph L. Rosnow et al., *The Institutional Review Board as a Mirror of Scientific and Ethical Standards*, 48 AM. PSYCHOLOGIST (forthcoming 1993) (discussing costs of doing and of not doing research).

15. See, e.g., MODEL RULES OF PROFESSIONAL CONDUCT Rule 3.1 (1983) (requiring that lawyers' claims and contentions be meritorious); MODEL CODE OF PROFESSIONAL RESPONSIBILITY DR 7-102(A)(1) (1968) (prohibiting a lawyer from taking an unwarranted action merely to harass or maliciously injure another); FED. R. CIV. P. 11; see also *infra* note 36.

SAT scores or their course grades, and they are then told that "this research is designed to learn the impact of personality on cognitive functioning," they have been poorly served educationally as part of having been misled scientifically. The same would hold true for the fact finder presented with bad science.

D. Costs and Utilities of Doing or Not Doing Research

When social scientists are confronted with a questionable research proposal, they ordinarily employ a cost-utility analysis. In such an analysis, the costs of doing a study include the possible negative effects on participants, time, money, effort, and other resources. These are evaluated simultaneously against such utilities as benefits to participants, to other people at other times, to science, to the investigator, or to the resolution of a particular legal dispute.¹⁶ The potential benefits of higher quality studies and research addressing more important topics exceed the potential benefits of lower quality studies and studies addressing less important topics. Of course, what constitutes an "important" study is open to considerable debate.

Elsewhere, this type of cost-utility analysis has been diagrammed as a two-dimensional plane in which costs are one dimension and utilities the other.¹⁷ Any study with high utility and low cost should likely be carried out. Conversely, any study with low utility and high cost should not be carried out. Studies in which costs approximate utilities are more difficult to evaluate.

The cost-utility model, however, is often insufficient to answer questions of whether to conduct the study because this calculus fails to consider the costs and utilities of *not* conducting a particular study.¹⁸ The failure to conduct a study that could be conducted is as much an act to be evaluated on ethical grounds as is the conducting of a study. For instance, the research group that could find a cancer preventive, but feels the work would be dull and a distraction from their major interest is making a decision that is to be evaluated on ethical grounds. This decision is as surely an ethical decision as is the decision of a researcher to investigate tumors with a procedure that

16. See, e.g., Peter D. Blanck, *The "Process" of Field Research in the Courtroom: A Descriptive Analysis*, 11 LAW & HUM. BEHAV. 337, 345-46 (1987) (discussing how judges' behavior affects jurors).

17. Robert Rosenthal & Ralph L. Rosnow, *Applying Hamlet's Question to the Ethical Conduct of Research: A Conceptual Addendum*, 39 AM. PSYCHOLOGIST 561, 562 (1984); BEGINNING BEHAVIORAL RESEARCH, *supra* note 7; ESSENTIALS OF BEHAVIORAL RESEARCH, *supra* note 7, at 232; Ralph L. Rosnow, *Teaching Research Ethics Through Role-Play and Discussion*, 17 TEACHING OF PSYCHOL. 179, 181 (1990).

18. Rosenthal & Rosnow, *supra* note 17; BEGINNING BEHAVIORAL RESEARCH, *supra* note 7; ESSENTIALS OF BEHAVIORAL RESEARCH, *supra* note 7; Rosnow et al., *supra* note 14.

carries a high risk to human participants. Similarly, the hypothetical school researcher whose study might help reduce violence, racism, or sexism, but who refuses to conduct the study because it involves deception, has not solved an ethical problem. This researcher has only traded one ethical dilemma for another.

These ideas relating to ethical tradeoffs have been applied with great eloquence by the late Professor John Kaplan of the Stanford University Law School.¹⁹ The context of his remarks was that of the use of animals in research and of the efforts of "animal rights" activists to chip away "at our ability to afford animal research. [I]t is impossible to know the costs of experiments not done or research not undertaken. Who speaks for the sick, for those in pain, and for the future?"²⁰ To extend Kaplan's point to the adversarial context, the cost of relying on bad or inadequate research (or on hypothetical findings) raises important ethical questions concerning the usefulness of social science research that is subsequently presented in the courts.

In the preceding examples, the costs of failing to conduct research have accrued to future generations (or to future legal disputes) or to present generations not including the research participants themselves. But sometimes incidental benefits to research participants are so important that they must be considered in the calculus of "good science."

An example helps to illustrate the point. The first author was asked once to testify before an Institutional Review Board about the implications of his studies on interpersonal expectations for the ethics of a proposed project on karyotyping (that is, the characteristics of cell structure). The proposed study was designed to test young children for the presence of the XYY chromosome, which had been hypothesized to be associated with criminal behavior. The participating children would be studied until adulthood, so that the correlation between chromosome type and criminal behavior could be determined. The first author was asked to testify about his research on interpersonal expectancy effects because it was believed that the proposed project should be conducted "double-blind," that is, it was suggested that the participants and the researchers should not be told the results of the chromosome analysis. This method was believed to be important because of the possibility that the parents' or researchers' expectations for increased

19. John Kaplan, *The Use of Animals in Research*, 242 SCIENCE 839, 839-40 (1988).

20. *Id.* at 839.

criminal behavior by the XYY males might in fact become a self-fulfilling prophecy²¹

A double-blind design should have solved that problem, but the Institutional Review Board decided not to permit the research. The Board did not consider the costs to the participants themselves or to their families of *not* doing the study. What were those costs? The costs were the loss of twenty years of free, high-quality pediatric care to children whose parents could never have afforded any high-quality pediatric care. Was it an ethically defensible decision to deprive hundreds of children of medical care that they would otherwise not have received to avoid a double-blind design that had little potential for actually harming the participants? At the very least, these costs of failing to do the research should have received a fuller discussion.

II. DATA ANALYSIS AS AN ETHICAL ARENA AND IMPLICATIONS FOR SOCIAL SCIENCE RESEARCH IN THE COURTROOM

This Part explores several of the ethical and legal consequences of the analysis of social science data. Issues discussed here relate to data-dropping choices, subject and variable selection procedures, and the importance of meta-analytic techniques to the presentation of research results in the courtroom.

A. Data Dropping

Ethical issues in the analysis of data range from the obvious to the subtle. Probably the most obvious and serious transgression in data analysis is the analysis of fabricated data. Perhaps more frequent is the dropping of data that contradict the data analyst's theory, prediction, or commitment. This strategy of data rejection warrants discussion, as it has important implications for the potential usefulness of a social science study in the adversarial context.

There is a venerable tradition in data analysis of dealing with outliers, or extreme scores, a tradition going back over 200 years.²² Both technical and ethical issues are involved. The technical issues relate to the best ways of dealing with outliers without reference to the implications for the tenability of the data analyst's theory or hypothesis. The ethical issues have to do with

21. For more on expectancy and the use of double-blind tests, see ROBERT ROSENTHAL, *EXPERIMENTER EFFECTS IN BEHAVIORAL RESEARCH* 367-68 (expanded ed. 1976); ROBERT ROSENTHAL & LENORE JACOBSON, *PYGMALION IN THE CLASSROOM* 18-19 (expanded ed. 1992).

22. See VIC BARNETT & TOBY LEWIS, *OUTLIERS IN STATISTICAL DATA* (1978).

the relationship between the data analyst's theory and the choice of method for dealing with outliers.

Some social science findings suggest that data outliers might be rejected more often if they contradict the data analyst's theory, but treated less harshly if they support the data analyst's theory.²³ At the very least, when outliers are rejected, the investigator or the expert testifying about the results of a study ethically must inform the evaluators of that fact. In addition, it would be useful for investigators or experts to report the results that would have been obtained had the outliers not been rejected. In the adversarial setting, if the proponent expert of the study does not report data-dropping procedures, certainly, cross-examination should raise this issue.

B. Subject Selection

A different type of data dropping is subject selection, where a subset of the data is not included in the analysis. Here, too, there are technical and ethical issues that have implications for the accurate presentation of social science research in the courtroom.

Often, there may be good technical reasons for setting aside a subset of the data. For example, a researcher may drop a subset because its size is especially small or because dropping the subset would make the data more comparable to some other research. There are, of course, ethical and evidentiary issues raised when just those subsets are dropped that do not support the data analyst's or expert's theory or hypotheses. At a minimum, when subsets of subjects are dropped, social scientists or the fact finder should be informed of that fact and, potentially, of the results for that subset.²⁴

C. Variable Selection

Another form of data dropping is variable selection. This is found where the results for one or more variables are not reported in the journal publication or to the fact finder. Once again, there are technical and ethical issues that have implications for the evidentiary value of the research results presented to a fact finder.

23. Robert Rosenthal & Donald B. Rubin, *Pygmalion Reaffirmed*, in JANET D. ELASHOFF & RICHARD E. SNOW, *PYGMALION RECONSIDERED* 139 app. C (1971); Robert Rosenthal, *How Often Are Our Numbers Wrong?*, 33 AM. PSYCHOLOGIST 1005, 1007 (1978).

24. As part of the trial examination of the expert, it may be important to perform analyses of the dropped subset data and to present the results to the fact finder.

A simple technical issue would be having more data than the journal editor or the court will allow to be reported,²⁵ or even than the reviewing body would be capable of digesting properly.²⁶ In these instances, the social scientist or expert would be forced to present some of the data. Results would then often be extrapolated inappropriately or overgeneralized to cases in which no actual data are considered. For instance, data extrapolation has occurred in cases involving mass torts. In such cases, it is sometimes impossible to present all the data relating to a class of individuals who are raising legal claims before a court.²⁷ In prior work, the use of such sampling procedures in mass trials has been evaluated under several theories of justice.²⁸ Importantly, sampling and aggregation of data and legal claims, when done well, may actually produce more precise and reliable trial outcomes than the justice afforded by more traditional case-by-case determinations.²⁹

As a general matter, it is safe to say that ethical and legal issues are raised in selecting which data to present, just as when choosing those variables not to report. This is particularly true in cases where the results are less favorable to the proponent's theory. At a minimum, when variables are dropped, social scientists and legal fact finders should be informed of that decision, as well as of the implications of those decisions.

D. Data Exploitation Can Be Beautiful

That data-dropping and variable selection have ethical and legal implications is fairly obvious. We turn now to an issue that has ethical and legal implications that are more subtle. The issue is data exploitation.

Exploiting research participants, students, staff, and colleagues to achieve certain results is, of course, reprehensible. There is, however, a kind of exploitation to be cherished by social scientists and fact finders alike: data exploitation. Social scientists are taught that it is technically improper and probably unethical to analyze and re-analyze their data in many ways. It is not proper to snoop around in the data, performing a "fishing expedition."

Likewise, social scientists are taught to test the experimental prediction with one particular pre-planned statistical test and to take a result significant at the

25. For instance, at some point, repeated presentation of case studies to the trier of fact may be deemed repetitive.

26. See generally Michael J. Saks & Peter D. Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815 (1992).

27. *Id.* at 816-21 (discussing the problem of courts having too many cases to try).

28. *Id.* at 826-41.

29. *Id.* at 851.

magic “ p less than .05 level”³⁰ as the true finding. In evaluating the importance of social science research, many lawyers also believe that should a statistical result not be significant at the .05 level, it is somehow not useful or relevant to resolving the dispute at hand.

Directly put, the lack of careful and regular attention by social scientists and lawyers to research findings that do not yield results significant at the .05 level makes for bad science, bad ethics, and uninformed uses of social science in the courtroom. It makes for bad science because, while snooping does affect statistical significance or “ p ” values, it is likely to turn up something new, interesting, and important.³¹ It makes for bad ethics because, as mentioned above, data collection is expensive in terms of time, effort, money, and other resources, and lawyers’ clients are increasingly willing to attest to this fact.³² Finally, it makes for uninformed uses of social science research in the courtroom because the full value of the study and its potential implications are not presented to the fact finder.

If social science research is worth doing, it is worth analyzing thoroughly. It is worthy of holding up to the light in many different ways so that research participants, funding agencies, and fact finders get their full value. The fundamental utility of data exploitation is therefore knowledge gained. About the only cost of exploitation is the loss of the “sense of sin” over looking at data for which some null hypothesis³³ may not have been rejected at the .05 level.³⁴ As Tukey pointed out so brilliantly over twenty years ago, social science would be further ahead if data analysts thought of themselves less as clergy blessing data with p less than .05 and more as detectives solving mysteries.³⁵ The same is often true for lawyers presenting and evaluating social science research findings in court.³⁶

30. See generally Peter D. Blanck, *Calibrating the Scales of Justice: Studying Judge's Behavior in Jury and Bench Trials*, 68 IND. L. J. 1119 (1993).

31. JOHN W. TUKEY, *EXPLORATORY DATA ANALYSIS* (1977).

32. This is particularly true when the client's bill for professional services arrives.

33. The null hypothesis is the hypothesis that there is no relationship between two or more variables.

34. It is also the loss of our Judeo-Christian-Buddhist guilt and suffering that was designed to be our lot when p was greater than .05; a suffering that was to be borne as an adult—mature, resigned, and devout.

35. John W. Tukey, *Analyzing Data: Sanctification or Detective Work*, 24 AM. PSYCHOLOGIST 83 (1969). Before leaving this topic, it should be repeated that snooping in the data can indeed affect the p value obtained, depending on how the snooping is done. But statistical adjustments can be helpful here (for example, Bonferroni adjustments). See ESSENTIALS OF BEHAVIORAL RESEARCH *supra* note 7; Robert Rosenthal & Donald B. Rubin, *Multiple Contrasts and Ordered Bonferroni Procedures*, 76 J. EDUC. PSYCHOL. 1028 (1984). Most importantly, replications of the research will be needed—whether the data were snooped or not.

36. We recognize that lawyers always select evidence to present; that is advocacy. Certainly, it is not unethical to put forth the client's best case. Yet, for a scientist, the opposite is often true. Analysis of the parallels and divergences between legal and ethical presentation of evidence raises important

*E. Meta-Analysis as an Ethical Imperative in Presenting
Social Science Research in the Courtroom*

Meta-analysis is a set of concepts and procedures employed to summarize quantitatively any domain of research.³⁷ We know from statistical and empirical research that, compared to traditional reviews of the literature, meta-analytic procedures are more accurate, comprehensive, systematic, and statistically more powerful.³⁸

Meta-analytic procedures use more of the information in the data, thereby yielding:

- (a) more accurate estimates of the overall magnitude of the effect or relationship being investigated;
- (b) more accurate estimates of the overall level of significance of the entire research domain; and
- (c) more useful information about the variables moderating the magnitude of the effect or relationship being investigated.

Meta-analysis allows social scientists and fact finders to learn more from data. Meta-analysis has a unique ability to increase *retroactively* the benefits and lessen the costs of the studies being summarized. The costs of time, attention, and effort of human participants employed in the individual studies of the meta-analysis are more justified when their data enter into a meta-analysis. This is because meta-analysis increases the utility of all the individual studies that are being summarized. Other costs of individual studies—including funding, investigator time and effort, and other resources—are similarly more justified because the utility of individual studies is increased by the strength obtained when information from more studies is combined in a sophisticated way

issues in itself. For a recent discussion, see *Daubert v. Merrell-Dow Pharmaceuticals, Inc.*, 951 F.2d 1128 (9th Cir. 1991), *cert. granted*, 113 S. Ct. 320 (1992) (No. 92-102) (reviewing ways to evaluate the admissibility of scientific information); Greenhouse, *supra* note 1, at A19 (reporting on the oral arguments to the Supreme Court in *Daubert*).

37. See generally ROBERT ROSENTHAL, *META-ANALYTIC PROCEDURES FOR SOCIAL RESEARCH* (rev. ed. 1991); Robert Rosenthal, *Cumulating Psychology: An Appreciation of Donald T. Campbell*, 2 PSYCHOL. SCIENCE 213 (1991); see also Peter D. Blanck, *What Empirical Research Tells Us: Studying Judges' and Juries' Behavior*, 40 AM. U. L. REV. 775 (1991); Peter D. Blanck et al., *The Measure of the Judge: An Empirically-Based Framework for Exploring Trial Judges' Behavior*, 75 IOWA L. REV. 653 (1990).

38. Harris M. Cooper & Robert Rosenthal, *Statistical Versus Traditional Procedures for Summarizing Research Findings*, 87 PSYCHOL. BULL. 442 (1980); LARRY V. HEDGES & INGRAM OLKIN, *STATISTICAL METHODS FOR META-ANALYSIS* (1985); Frederick M. Mosteller & R.R. Bush, *Selected Quantitative Techniques*, in 1 HANDBOOK OF SOCIAL PSYCHOLOGY: THEORY AND METHOD 289-334 (Gardner Lindzey ed., 1954).

The failure to employ meta-analytic procedures when they could be used has important ethical and legal implications. In part, this is because the opportunity to increase the benefits of past individual studies has been foregone. In addition, when resources are employed by scientists or court-appointed experts to prepare literature reviews, it is fair to ask whether those resources are being used efficiently or ethically.³⁹ Now that it is possible to summarize literature meta-analytically, it hardly seems justifiable to review literature in a pre-meta-analytic, pre-quantitative manner.

In many types of legal cases, it also no longer seems acceptable, defensible, or professionally responsible to present to a fact finder the results of individual research studies that claim to contribute to the resolution of controversy.⁴⁰ This is most apparent in instances where the investigator or expert witness has not already conducted a meta-analysis to determine whether, in fact, there really is a controversy at all. To provide one example: Currently, a new study of the effects of psychotherapy may not be worth doing given prior meta-analytic results.⁴¹ Until meta-analytic work resolved the issue, the question of whether psychotherapy worked in general was controversial. It is controversial no longer.

Meta-analysis further helps to resolve courtroom controversies because it eliminates two common errors, or misleading strategies, by social scientists and lawyers in the evaluation and presentation of research replications. The first error is the belief that when one study obtains a significant effect, and a replication does not, there is a failure to replicate the particular effect. This belief is clearly untrue. A failure to replicate is properly measured by the magnitude of difference between the effect sizes (that is, the magnitude of the experimental effect) of the two studies.

The second error is the belief that if there is a real effect in nature, that each study of that effect will show a significant effect. Even if the effect is quite substantial in reality, there is a limited chance that all investigators will get results significant at the p less than .05 level.⁴²

39. Cf. FED. R. EVID. 403 (providing that evidence may be excluded "by considerations of undue delay, waste of time, or needless presentation of cumulative evidence").

40. For example, does psychotherapy treatment work? What is the reliability of eyewitness testimony? What are the characteristics of jurors' decision-making processes in death penalty cases? For other examples, see Frank L. Schmidt, *What Do Data Really Mean? Research Findings, Meta-Analysis, and Cumulative Knowledge in Psychology*, 47 AM. PSYCHOLOGIST 1173 (1992).

41. Gene V. Glass, *Primary, Secondary, and Meta-Analysis of Research*, 5 EDUC. RESEARCHER, Nov. 1976, at 3; MARY LEE SMITH ET AL., *THE BENEFITS OF PSYCHOTHERAPY* (1980).

42. For example, if the effect is substantial with a correlation coefficient of $r = .24$, and each study employs a sample size of 64, we have the typical situation in psychology of operating at a power level of .50. Given this typical situation, there is only one chance in four that two investigations will both get results significant at the .05 level. If three studies were carried out, there would be only one chance in

In these ways, meta-analytic procedures increase the utility of individual studies by highlighting implications for how and whether significance testing is performed and reported. Good meta-analytic practice shows little interest in whether the results of an individual study are significant. Rather than recording whether a study reached a critical level, say $p = .05$, two-tailed,⁴³ meta-analysts record the actual level of significance obtained. This is usually done not by recording the p value, but by recording the standard normal deviate that corresponds to the p value. Thus, a result significant at the .05 level, one-tailed test,⁴⁴ in the predicted direction is recorded as $Z = +1.645$ (that is, showing the standard normal deviate corresponding to the p value).

If the result above had been significant at the .05 level, one-tailed, but in the wrong or unpredicted direction, it would be recorded as $Z = -1.645$ (that is, with a minus sign to indicate that the result is opposite to the predicted direction). Signed normal deviates are informative to social scientists and fact finders as a characteristic of the result of a study that is presented in continuous rather than in dichotomous form. Typical null hypothesis decision procedures are then seen more appropriately as a commentary on the state of mind of the data analyst or expert witness. For instance, "I, plaintiff's expert data analyst, chose to reject the null hypothesis." Defendant's expert data analyst might not have chosen to reject the null hypothesis. In short, signed normal deviates, as indices of significance levels, change the emphasis from what the expert data analyst thinks or advocates, to what the study actually showed.

Getting rid of the dichotomous null-hypothesis decision procedure, and the related state of mind, in the use of social science in the courtroom thus increases: (a) the information value and legal relevance of a study, which, (b) increases the utility or real-world usefulness of the study and, therefore, (c) changes the cost-utility ratio and, hence, social scientists' and legal practitioners' evaluations of the ethical and evidentiary value of the study⁴⁵

eight that all three studies would yield significant effects, even though we know the effect in nature is both real and important in magnitude. See JACOB COHEN, *STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES* (2d ed. 1988); Jacob Cohen, *The Statistical Power of Abnormal-Social Psychological Research: A Review*, 65 J. ABNORMAL & SOC. PSYCHOL. 145 (1962); Robert Rosenthal, *On Being One's Own Case Study: Experimenter Effects in Behavioral Research—Thirty Years Later*, in *THE SOCIAL PSYCHOLOGY OF SCIENCE* (W.R. Shadish, Jr. & S. Fuller eds., forthcoming 1993); Peter Sedlmeier & Gerd Gigerenzer, *Do Studies of Statistical Power Have an Effect on the Power of Studies?*, 105 PSYCHOL. BULL. 309 (1989).

43. A two-tailed test is a test of significance in which the null hypothesis is rejected if the results are significant (e.g., at the $p < .05$ level) in either of two possible directions.

44. A one-tailed test is a test of significance in which the null hypothesis is rejected only if the results are significant (e.g., at the $p < .05$ level) in one of the two possible directions.

45. Getting rid of the dichotomous null-hypothesis decision procedure also rids us of the worthless issue of one- versus two-tailed significance testing. As illustrated herein, the sign of the standard normal

Another way in which meta-analysis increases research utility, therefore furthering the ethical justification and evidentiary base of research studies, is by providing accurate estimates of effect sizes. In fact, effect sizes can be of major importance even when they are so small as to have $r^2 = .00$. When well-estimated effect sizes are provided to a fact finder, it is valuable also to assess their practical importance. The traditional r^2 method of effect size estimation does a poor job of this since an r^2 of .00 can, by way of one example, be associated with an experimental treatment method that reduces death rates by as much as 7 per 100 lives lost.⁴⁶ Once social scientists, lawyers, and fact finders are aware that effect size, where r values of .05, .10, and .20 (with r^2 s of .00, .01, and .04, respectively), for example, are associated with benefits equivalent to saving 5, 10, or 20 lives per 100 people, they may more accurately weigh the costs and utilities of undertaking any particular study, as well as the ultimate value of the study

III. REPORTING SOCIAL SCIENCE RESEARCH

This final Part discusses several issues related to the reporting of social science research. The issues raised relate to the misrepresentation of findings and research credit and to the decision not to report or publish research findings. This Part discusses the relevance of each issue to the use of social science in the courtroom.

A. Misrepresentation of Findings

Mother nature makes it hard enough to learn her secrets, without the additional difficulty of being misled by social scientists or expert witnesses as to results that just are not found or by inferences that are prejudicial. Although all misrepresentations of findings are damaging to the progress of science and to the ability to provide justice in the courtroom, some misrepresentations are more obviously unethical than others.

The most blatant intentional misrepresentation is the reporting of fabricated data. That behavior, if detected, often ends the career of the perpetrator. Certainly, however, there are less egregious forms of misrepresentation.

deviate tells the tail, so to speak, of the direction of the result. How fully grown methodologists can argue over the use of one- versus two-tailed tests has long escaped the first author. Most psychologists are quite capable of doubling a p value or dividing it by two should they want to do so.

46. See ROSENTHAL, *supra* note 37; Rosenthal, *supra* note 37; Robert Rosenthal & Donald B. Rubin, *A Simple, General Purpose Display of Magnitude of Experimental Effect*, 74 J. EDUC. PSYCHOL. 166 (1982).

Counsel at trial and fact finders should be wary of other more subtle forms of intentional misrepresentation.

For instance, a more subtle form of intentional misrepresentation occurs when investigators knowingly allocate to experimental or control conditions those participants whose response is more likely to support the investigators' hypothesis. Another potential form of intentional misrepresentation occurs when investigators record the participants' responses without being blind to the participants' treatment condition, or when research assistants record the participants' responses knowing the research hypothesis and the participants' treatment condition. Of course, if the research report specifically notes the failure to run blind, it is no longer misrepresentation, but it is scientifically unwise if it could have been avoided.

Various errors in the process of data collection can lead to unintentional misrepresentation. Recording errors, computational errors, and data analytic errors can all lead to inaccurate results that are misrepresentations, albeit only inadvertent ones. Social scientists might not normally think of unintentional errors as constituting ethical issues, except for the fact that errors in the data decrease the utility and real-world validity of the research. This, in turn, moves the cost-utility ratio to justify the research on ethical grounds in the unfavorable direction. Some cases of misrepresentation (usually unintentional) are even more subtle. The use of causist language, discussed earlier, is one example.⁴⁷

B. Misrepresentation of Credit

So far, we have been discussing misrepresentation of findings or the issue of "what was really found?" Some discussion is needed also of the issue "who really found it?" Where so many papers-in social science, and the sciences generally, are multi-authored, it seems inevitable that there will be difficult problems in allocating authorship credit. Who becomes a co-author and who becomes a footnote? Among the co-authors, who is assigned first, last, or any other serial position in the listing? In social science journals, such questions have been discussed in depth and general guidelines have been offered.⁴⁸ In the courtroom, the issue is, "Who is the real expert?"

47. Even more subtle is the case of questionable generalizability. For a discussion of this, see Rosenthal, *supra* note 37.

48. American Psychological Ass'n, *Ethical Principles of Psychologists*, 36 AM. PSYCHOLOGIST 633 (1981); AMERICAN PSYCHOLOGICAL ASS'N, CASEBOOK ON ETHICAL PRINCIPLES OF PSYCHOLOGISTS (1987).

It seems that lawyers who regularly employ and confront courtroom experts could profit from further empirical studies in which authors, editors, referees, practitioners, and professors are asked to allocate authorship credit to people performing various functions in a scholarly enterprise. How much credit should be allocated to the person(s) who came up with the idea, collected the data, analyzed the data, prepared the first draft, organized the revisions of the draft, and so on. It would be useful to fact finders, courts, and lawyers to learn whether different groups of social scientists have different views of authorship allocation, and the degree of reliability of such ratings within groups. As a practical matter, problems of authorship are usually problems existing within research groups. Problems of priority are usually problems existing between research groups.⁴⁹ Answers to these questions also may prove useful to lawyers seeking to retain expert witnesses.

C. Failing to Report or Publish

Sometimes the ethical and evidentiary question is not about the accuracy of what was reported, or how credit should be allocated for what was reported, but rather about what was not reported and why it was not reported. The two major forms of failure to report, or censoring, may be identified as self-censoring and external censoring.

Some self-censoring is admirable and often advisable. When the study has been done badly, it may be a service to the science to simply start over. Some self-censoring is done for these admirable motives, but it often seems a waste of information. For example, some researchers believe that they should not cite their own or others' unpublished data because the data have not gone through peer review. Consistent with the discussion above, we believe that such data should be cited and employed in meta-analytic computations, as long as the data are well-collected.

There are also less admirable reasons for self-censoring. The expert failing to report data that contradicts his or her earlier research, theory, or even values related to the trial outcome, is practicing poor science and poor ethics. It is always possible to find or invent reasons why a study with unfavorable results should not be reported: the subjects were just starting the undergraduate course; the participants were subjected to prejudicial information; and so on. A sound general policy, good for science and for its integrity in the courtroom process, is to report all results. If it is important enough to do, it is important enough to report.

49. Further discussion of this issue appears in Rosenthal, *supra* note 37.

There is no denying that some results are more ground-breaking than others. If a new treatment prevents or cures mental illness, that fact may be worth more space in prestigious journals than the result that a new treatment does no good whatsoever. But that less thrilling finding should also be reported and made retrievable by other researchers who may need to know that finding.

Both the progress and the slowing of progress in science depend on external censoring. It seems likely that the social sciences would be more chaotic than they already are if not for the censorship exercised by peer review: by editors, by reviewers, and by program committees. All these gatekeepers help to keep bad science from clogging the pipelines of mainstream journals.

There are two major bases for external censorship. The first is based on an evaluation of the methodology employed in a research study. For example, if our hypothetical school study is flawed from a methodological point of view, it probably should not be reported or given much weight in the courtroom. The second major reason for external censorship is based not on an evaluation of the methodology, but on an evaluation of the results. It is often said of a study, "those results aren't possible" or "those results make no sense." Often, upon examination of such studies, the results are indeed implausible. This is a poor basis, however, on which to censor the results. Censoring or suppressing results that are not liked or that are believed not to have high prior probability is bad science and ethics.⁵⁰

CONCLUSION

This Article has highlighted several scientific, ethical, and legal issues in conducting, analyzing, and reporting social science research. An organizing theme has been that the ethical quality of research is often not independent of the scientific quality of research.⁵¹ We hope that detailing some of the

50. Robert Rosenthal, *On Balanced Presentation of Controversy*, 30 AM. PSYCHOLOGIST 937 (1975); Rosenthal, *supra* note 42.

51. The New Jersey Supreme Court has recently recognized this view, concluding that admissibility of scientific testimony depends on:

the expert's ability to explain pertinent scientific principles and to apply those principles to the formulation of his or her opinion. Thus, the key to admission of the opinion is the validity of the expert's reasoning and methodology. The court's function is to distinguish scientifically sound reasoning from that of the self-validating expert, who uses scientific terminology to present unsubstantiated personal beliefs.

Landrigan v. Celotex Corp., 605 A.2d 1079, 1084 (N.J. 1992) (citation omitted); *see also Caterinicchio v. Pittsburg Corning Corp.*, 605 A.2d 1092, 1094 (N.J. 1992). For a general review, see STEPHEN A. SALTZBURG, AM. BAR ASS'N & BROOKINGS INST., IMPROVING THE QUALITY OF JURY DECISION MAKING: EXECUTIVE SUMMARY, SYMPOSIUM ON THE FUTURE OF THE CIVIL JURY SYSTEM IN THE UNITED STATES (1992).

specifics of this theme has served two functions for social scientists, lawyers, fact finders, and policy makers. First, we hope it has shown how it is possible simultaneously to improve the quality of science and the quality of ethics in the presentation of social science research in the courtroom. Second, we hope that, in raising more issues than it has answered, this Article illustrates to social scientists, lawyers, fact finders, and policy makers that, in the matter of improving science and ethics, many issues are yet to be resolved.