Quantitative Methods and Ethics

Citation

Rosnow, Ralph L., and Robert Rosenthal. 2013. Quantitative Methods and Ethics. In The Oxford Handbook of Quantitative Methods in Psychology Vol. 1, ed. Todd D. Little. New York: OUP.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11718215>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Quantitative Methods and Ethics

Ralph L. Rosnow and Robert Rosenthal

Abstract

The purpose of this chapter is to provide a context for thinking about the role of ethics in quantitative methodology. We begin by reviewing the sweep of events that led to the creation and expansion of legal and professional rules for the protection of research subjects and society against unethical research. The risk–benefit approach has served as an instrument of prior control by institutional review boards. After discussing the nature of that approach, we sketch a model of the costs and utilities of the “doing” and “not doing” of research. We illustrate some implications of the expanded model for particular data analytic and reporting practices. We then outline a 5×5 matrix of general ethical standards crossed with general data analytic and reporting standards to encourage thinking about opportunities to address quantitative methodological problems in ways that may have mutual ethical and substantive rewards. Finally, we discuss such an opportunity in the context of problems associated with risk statistics that tend to exaggerate the absolute effects of therapeutic interventions in randomized trials.

Key Words: Accountability, American Psychological Association (APA) Ethics Code, Belmont Report, ethical principles, health statistics, institutional review board (IRB), moral dilemmas, Nuremberg (Nuremberg) Code, quantitative methodology, risk–benefit assessment, statistical illiteracy, transparency, volunteer bias

Introduction

In this chapter we sketch an historic and heuristic framework for assessing certain ethical implications of the term *quantitative methods*. We use this term in the broadest sense to include not only statistical procedures but also what is frequently described as quantitative research (in contrast to qualitative research) in psychology and some other disciplines. As defined in the *APA Dictionary of Psychology*, the traditional distinction between these two general types of research rests on whether “the approach to science” does (*quantitative research*) or does not (*qualitative research*) “employ the quantification (expression in numerical form) of the observations made” (VandenBos, 2007, pp. 762–763). Of course, quantitative and qualitative methods should

not be seen as mutually exclusive, as it can often be illuminating to use both types in the same research. For example, in the typical psychological experiment in which the observations take a numerical form, it may be edifying to ask some of the participants in postexperimental interviews to reflect on the context in which the experiment was conducted and to speculate on the ways in which it may have influenced their own and other participants’ behaviors (Orne, 1962, 1969). By the same token, it is usually possible to quantify nonquantitative observations by, for example, decomposing the qualitative subject matter element by element and then numerically and visually analyzing and summarizing the results. Blogs and online discussion groups are currently a popular source of qualitative subject

matter, which researchers have trolled for patterns or relationships that can be quantified by the use of simple summary statistics (e.g., Bordia & Rosnow, 1995) or coded and visually mapped out using social network analysis to highlight links and nodes in the observed relationships (e.g., Kossinets & Watts, 2006; *see also* Wasserman & Faust, 1994). Whether blogs and online discussion groups' data are treated quantitatively or qualitatively, their use may raise ethical questions regarding the invasion of privacy. The fact that bloggers and participants in online discussion groups are typically fully aware that their communications are quite public minimizes the risk of invasion of privacy.

The term *ethics* was derived from the Greek *ethos*, meaning "character" or "disposition." We use the term here to refer to the dos and don'ts of codified and/or culturally ingrained rules by which morally "right" and "wrong" conduct can be differentiated. Conformity to such rules is usually taken to mean *morality*, and our human ability to make ethical judgments is sometimes described as a *moral sense* (a tradition that apparently goes back to David Hume's *A Treatise of Human Nature* in the eighteenth century). Philosophers and theologians have frequently disagreed over the origin of the moral sense, but on intuitive grounds it would seem that morality is subject to societal sensitivities, group values, and social pressures. It is not surprising that researchers have documented systematic biases in ethical judgments. For example, in a study by Kimmel (1991), psychologists were asked to make ethical judgments about hypothetical research cases. Kimmel reported that those psychologists who were more (as compared to less) approving in their ethical judgments were more often men; had held an advanced degree for a longer period of time; had received the advanced degree in an area such as experimental, developmental, or social psychology rather than counseling, school, or community psychology; and were employed in a research-oriented context as opposed to a service-oriented context. Citing this work of Kimmel's (1991), an American Psychological Association (APA) committee raised the possibility that inconsistent implementation of ethical standards by review boards might result not only from the expanded role of review boards but also from the composition of particular boards (Rosnow, Rotheram-Borus, Ceci, Blanck, & Koocher, 1993). Assuming that morality is also predicated on people's abilities to figure out the meaning of other people's actions and underlying

intentions, it might be noted that there is also empirical evidence of (1) individual differences in this ability (described as *interpersonal acumen*) and (2) a hierarchy of intention-action combinations ranging from the least to most cognitively taxing (Rosnow, Skleder, Jaeger, & Rind, 1994).

Societal sensitivities, group values, and situational pressures are subject to change in the face of significant events. On the other hand, some moral values seem to be relatively enduring and universal, such as the golden rule, which is frequently expressed as "Do unto others as you would have them do unto you." In the framework of quantitative methods and ethics, a categorical imperative might be phrased as "Thou shalt not lie with statistics." Still, Huff, in his book, *How to Lie with Statistics*, first cautioned the public in 1954 that the reporting of statistical data was rife with "bungling and chicanery" (Huff, 1982, p. 6). The progress of science depends on the good faith that scientists have in the integrity of one another's work and the unbiased communication of findings and conclusions. Lying with statistics erodes the credibility of the scientific enterprise, and it can also present an imminent danger to the general public. "Lying with statistics" can refer to a number of more specific practices: for example, reporting only the data that agree with the researcher's bias, omitting any data not supporting the researcher's bias, and, most serious of all, fabricating the results of the research. For example, there was a case reported in 2009 in which an anesthesiologist fabricated the statistical data that he had published in 21 journal articles purporting to give the results of clinical trials of a pain medicine marketed by the company that funded much of the doctor's research (Harris, 2009). Another case, around the same time, involved a medical researcher whose accounts of a blood test for diagnosing prostate cancer had generated considerable excitement in the medical community, but who was now being sued for scientific fraud by his industry sponsor (Kaiser, 2009). As the detection of lying with statistics is often difficult in the normal course of events, there have been calls for the public sharing of raw data so that, as one scientist put it, "Anyone with the skills can conduct their own analyses, draw their own conclusions, and share those conclusions with others" (Allison, 2009, p. 522). That would probably help to reduce some of the problems of biased data analysis, but it would not help much if the shared data had been fabricated to begin with.

In the following section, we review the sweep of events that led to the development and growth of restraints for the protection of human subjects and society against unethical research.¹ A thread running throughout the discussion is the progression of the APA's code of conduct for psychological researchers who work with human subjects. We assume that many readers of this Handbook will have had a primary or consulting background in some area of psychology or a related research area. The development of the APA principles gives us a glimpse of the specific impact of legal regulations and societal sensitivities in an area in which human research has been constantly expanding into new contexts, including "field settings and biomedical contexts where research priorities are being integrated with the priorities and interests of nonresearch institutions, community leaders, and diverse populations" (Sales & Folkman, 2000, p. ix). We then depict an idealized risk–benefit approach that review boards have used as an instrument of prior control of research, and we also describe an expanded model focused on the costs and utilities of "doing" and "not doing" research. The model can also be understood in terms of the cost–utility of adopting versus not adopting particular data analytic and reporting practices. We then outline a matrix of general ethical standards crossed with general data analytic and reporting standards as (1) a reminder of the basic distinction between ethical and technical mandates and (2) a framework for thinking about promising opportunities for ethical and substantive rewards in quantitative methodology (cf. Blanck, Bellack, Rosnow, Rotheram-Borus, & Schooler, 1992; Rosenthal, 1994; Rosnow, 1997). We discuss such an opportunity in the context of the way in which a fixation on relative risk (RR) in large sample randomized trials of therapeutic interventions can lead to misconceptions about the practical meaning to patients and health-care providers of the particular intervention tested.

The Shaping of Principles to Satisfy Ethical and Legal Standards

If it can be said that a single historical event in modern times is perhaps most responsible for initially galvanizing changes in the moral landscape of science, then it would be World War II. On December 9, 1946 (the year after the surrender of Germany on May 8, 1945 and the surrender of Japan on August 14, 1945), criminal proceedings against Nazi physicians and administrators who had participated

in war crimes and crimes against humanity were presented before a military tribunal in Nuernberg, Germany. For allied atomic scientists, Hiroshima had been an epiphany that vaporized the old iconic image of a morally neutral science. For researchers who work with human participants, the backdrop to the formation of ethical and legal principles to protect the rights and welfare of all research participants were the shocking revelations of the war crimes documented in meticulous detail at the Nuernberg Military Tribunal. Beginning with the German invasion of Poland at the outbreak of World War II, Jews and other ethnic minority inmates of concentration camps had been subjected to sadistic tortures and other barbarities in "medical experiments" by Nazi physicians in the name of science. As methodically described in the multivolume report of the trials, "in every one of the experiments the subjects experienced extreme pain or torture, and in most of them they suffered permanent injury, mutilation, or death" (*Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, p. 181). Table 3.1 reprints the principles of the Nuernberg Code, which have resonated to varying degrees in all ensuing codes for biomedical research with human participants as well as having had a generative influence on the development of principles for the conduct of behavioral and social research.

We pick up the story again in the 1960s in the United States, a period punctuated by the shocking assassinations of President John F. Kennedy in 1963 and then of Dr. Martin Luther King, Jr., and Senator Robert F. Kennedy in 1968. The 1960s were also the beginning of the end of what Pattullo (1982) called "the hitherto sacrosanct status" of the human sciences, which moved "into an era of uncommonly active concern for the rights and welfare of segments of the population that had traditionally been neglected or exploited" (p. 375). One highly publicized case in 1963 involved a noted cancer researcher who had injected live cancer cells into elderly, noncancerous patients, "many of whom were not competent to give free, informed consent" (Pattullo, p. 375). In 1966, the U.S. Surgeon General issued a set of regulations governing the use of subjects by researchers whose work was funded by the National Institutes of Health (NIH). Most NIH grants funded biomedical research, but there was also NIH support for research in the behavioral and social sciences. In 1969, following the exposure of further instances in which the welfare of subjects had been ignored or endangered in biomedical research (cf. Beecher, 1966, 1970; Katz,

Table 3.1. The Nuernberg Principles of 1946–1949 for Permissible Medical Experiments*

1. The voluntary consent of the human subject is absolutely essential.
This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment.
The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.
2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.
3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.
4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
5. No experiment should be conducted where there is an *a priori* reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.
8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.
9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.
10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

* Reprinted from pp. 181–182 in *Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, October 1946–April 1949, Vol. II. Washington, DC: U.S. Government Printing Office.

1972), the Surgeon General extended the earlier safeguards to all human research. In a notorious case (not made public until 1972), a study conducted by the U.S. Public Health Service (USPHS) simply followed the course of syphilis in more than 400 low-income African-American men residing in Tuskegee, Alabama, from 1932 to 1972 (Jones, 1993). Recruited from churches and clinics with the promise of free medical examinations and free health care, the men who were subjects in this study were

never informed they had syphilis but only told they had “bad blood.” They also were not offered penicillin when it was discovered in 1943 and became widely available in the 1950s, and they were warned not to seek treatment elsewhere or they would be dropped from the study. The investigators went so far as to have local doctors promise not to treat the men in the study with antibiotics (Stryker, 1997). As the disease progressed in its predictable course without any treatment, the men experienced damage to

their skeletal, cardiovascular, and central nervous systems and, in some cases, death. In 1972, the appalling details were finally made public by a lawyer who had been an epidemiologist for the USPHS, and the study was halted (Fairchild & Bayer, 1999). The following year, the Senate Health Subcommittee (chaired by Senator Edward Kennedy) aired the issue of scientific misconduct in public hearings.

The early 1960s was also a period when emotions about invasions of privacy were running high in the United States after a rash of reports of domestic wiretapping and other clandestine activities by federal agencies. In the field of psychology, the morality of the use of deception was being debated. As early as the 1950s, there had been concerned statements issued about the use of deception in social psychological experiments (Vinacke, 1954). The spark that lit a fuse in the 1960s in the field of psychology was the publication of Stanley Milgram's studies on obedience to authority, in which he had used an elaborate deception and found that a majority of ordinary research subjects were willing to administer an allegedly dangerous level of shock to another person when "ordered" to do so by a person in authority, although no shock was actually administered (cf. Blass, 2004; Milgram, 1963, 1975). Toward the end of the 1960s, there were impassioned pleas by leading psychologists for the ethical codification of practices commonly used in psychological research (Kelman, 1968; Smith, 1969). As there were new methodological considerations and federal regulations since the APA had formulated a professional code of ethics in 1953, a task force was appointed to draft a set of ethical principles for research with human subjects. Table 3.2 shows the final 10 principles adopted by the APA's Council of Representatives in 1972, which were elucidated in a booklet that was issued the following year, *Ethical Principles in the Conduct of Research with Human Participants* (APA, 1973). An international survey conducted 1 year later found there were by then two dozen codes of ethics that had been either adopted or were under review by professional organizations of social scientists (Reynold, 1975). Although violations of such professional codes were supported by penalties such as loss of membership in the organization, the problem was that many researchers engaged in productive, rewarding careers did not belong to these professional organizations.

By the end of the 1970s, the pendulum had swung again, as *accountability* had become the watchword of the decade (National Commission on Research, 1980). In 1974, the guidelines provided

by the Department of Health, Education, and Welfare (DHEW) 3 years earlier were codified as government regulations by the National Research Act of July 12, 1974 (Pub. L. 93-348). Among the requirements instituted by the government regulations was that institutions that received federal funding establish an institutional review board (IRB) for the purpose of making prior assessments of the possible risks and benefits of proposed research.² This federal act also created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Following hearings that were held over a 3-year period, the document called "The Belmont Report" was issued in April, 1979 (available online and also reprinted in Sales & Folkman, 2000). Unlike other reports of the Commission, the Belmont Report did not provide a list of specific recommendations for administrative action by the DHEW, but the Belmont Report recommended that the report be adopted in its entirety as a statement of DHEW policy. In the preamble, the report mentioned the standards set by the Nuernberg ("Nuremberg") Code as the prototype of many later codes consisting of rules, some general and others specific, to guide researchers and assure that research involving human participants would be carried out in an ethical manner. Noting that the rules were often inadequate to cover complex situations, that they were often difficult to apply or interpret, and that they often came into conflict with one another, the National Commission had decided to issue broad ethical principles to provide a basis on which specific rules could then be formulated, criticized, and interpreted. As we track the development of the APA principles in this discussion, we will see that there has been a similar progression, and later we will emphasize some broad ethical principles when we discuss the interface of ethical and technical standards in quantitative methodology. For now, however, it can be noted that the Belmont Report proposed that (1) respect for persons, (2) beneficence, and (3) justice provide the foundation for research ethics. The report also proposed norms for scientific conduct in six major areas: (1) the use of valid research designs, (2) the competence of researchers, (3) the identification of risk-benefit consequences, (4) the selection of research participants, (5) the importance of obtaining informed voluntary consent, and (6) compensation for injury.³

In 1982, the earlier APA code was updated, and a new version of *Ethical Principles in the Conduct of Research with Human Participants* was published

Table 3.2. The Ethical Principles Adopted in December, 1972, by the Council of Representatives of the American Psychological Association*

The decision to undertake research rests upon a considered judgment by the individual psychologist about how best to contribute to psychological science and to human welfare. The responsible psychologist weighs alternative directions in which personal energies and resources might be invested. Having made the decision to conduct research, psychologists must carry out their investigation with respect for the people who participate and with concern for their dignity and welfare. The Principles that follow make explicit the investigator's ethical responsibilities toward participants over the course of research, from the initial decision to pursue a study to the steps necessary to protect the confidentiality of research data. These Principles should be interpreted in terms of the context provided in the complete document offered as a supplement to these Principles.

1. In planning a study the investigator has the personal responsibility to make a careful evaluation of its ethical acceptability, taking into account these Principles for research with human beings. To the extent that this appraisal, weighing of scientific and humane values, suggests a deviation from any Principle, the investigator incurs an increasingly serious obligation to seek ethical advice and to observe more stringent safeguards to protect the rights of the human research participants.
2. Responsibility for the establishment and maintenance of acceptable ethical practice in research always remains with the individual investigator. The investigator is also responsible for the ethical treatment of research participants by collaborators, assistants, students, and employees, all of whom, however, incur parallel obligations.
3. Ethical practice requires the investigator to inform the participant of all features of the research that reasonably might be expected to influence willingness to participate and to explain all other aspects of the research about which the participant inquires. Failure to make full disclosure gives added emphasis to the investigator's responsibility to protect the welfare and dignity of the research participant.
4. Openness and honesty are essential characteristics of the relationship between investigator and research participant. When the methodological requirements of a study necessitate concealment or deception, the investigator is required to ensure the participant's understanding of the reasons for this action and to restore the quality of the relationship with the investigator.
5. Ethical research practice requires the investigator to respect the individual's freedom to decline to participate in research or to discontinue participation at any time. The obligation to protect this freedom requires special vigilance when the investigator is in a position of power over the participant. The decision to limit this freedom increases the investigator's responsibility to protect the participant's dignity and welfare.
6. Ethically acceptable research begins with the establishment of a clear and fair agreement between the investigator and the research participant that clarifies the responsibilities of each. The investigator has the obligation to honor all promises and commitments included in that agreement.
7. The ethical investigator protects participants from physical and mental discomfort, harm, and danger. If the risk of such consequences exists, the investigator is required to inform the participant of that fact, secure consent before proceeding, and take all possible measures to minimize distress. A research procedure may not be used if it is likely to cause serious and lasting harm to participants.
8. After the data are collected, ethical practice requires the investigator to provide the participant with a full clarification of the nature of the study and to remove any misconceptions that may have arisen. Where scientific or human values justify delaying or withholding information, the investigator acquires a special responsibility to assure that there are no damaging consequences for the participant.

Table 3.2. (Continued)

9. Where research procedures may result in undesirable consequences for the participant, the investigator has the responsibility to detect and remove or correct these consequences, including, where relevant, long-term aftereffects.
10. Information obtained about the research participants during the course of an investigation is confidential. When the possibility exists that others may obtain access to such information, ethical research practice requires that this possibility, together with the plans for protecting confidentiality, be explained to the participants as a part of the procedure for obtaining informed consent.

*Quoted from pp. 1–2 in *Ethical Principles in the Conduct of Research with Human Participants*. Washington, DC: American Psychological Association. Copyright © 1973 by the American Psychological Association.

by the APA. In the earlier version and in the 1982 version, the principles were based on actual ethical problems that researchers had experienced, and extensive discussion throughout the profession was incorporated in each edition of *Ethical Principles*. The principles in the 1982 code are reprinted in Table 3.3. Notice that there were several new terms (*subject at risk* and *subject at minimal risk*) and also an addendum sentence to informed consent (referring to “research with children or with participants who have impairments that would limit understanding and/or communication”). The concept of *minimal risk* (which came out of the Belmont Report) means that the likelihood and extent of harm to the participants are presumed to be no greater than what may be typically experienced in everyday life or in routine physical or psychological examinations (Scott-Jones & Rosnow, 1998, p. 149). In actuality, the extent of harm may not be completely anticipated, and estimating the likelihood of harm is frequently difficult or impossible. Regarding the expanded statement on deception, the use of deception in research had been frowned upon for some years although there had long been instances in which active and passive deceptions had been used routinely. An example was the withholding of information (passive deception). Randomized clinical trials would be considered of dubious value in medical research had the experimenters and the participants not been deprived of information regarding which condition was assigned to each participant. On the other hand, in some areas of behavioral experimentation, the use of deception has been criticized as having “reached a ‘taken-for-granted’ status” (Smith, Kimmel, & Klein, 2009, p. 486).⁴

Given the precedence of federal (and state) regulations since the guidelines developed by the DHEW were codified by the National Research Act in 1974 (and revised as of November 6, 1975), researchers

were perhaps likely to take their ethical cues from the legislated morality and its oversight by IRBs as opposed to the aspirational principles embodied in professional codes, such as the APA code. Another complication in this case is that there was a fractious splintering of the APA in the late-1980s, which resulted in many members resigning from the APA and the creation of the rival American Psychological Society, subsequently renamed the Association for Psychological Science (APS). For a time in the 1990s, a joint task force of the APA and the APS attempted to draft a revised ethics code, but the APS then withdrew its participation following an apparently irresolvable disagreement. In 2002, after a 5-year revision process, APA adopted a reworked ethics code that emphasized the five general principles defined (by APA) in Table 3.4 and also “specific standards” that fleshed out these principles.⁵ The tenor of the final document was apparently intended to reflect the remaining majority constituency of the APA (practitioners) but also the residual constituency of psychological scientists who perform either quantitative or qualitative research in fundamental and applied contexts. Of the specific principles with some relevance to data analysis or quantitative methods, there were broadly stated recommendations such as sharing the research data for verification by others (Section 8.14), not making deceptive or false statements (Section 8.10), using valid and reliable instruments (Section 9.02), drawing on current knowledge for design, standardization, validation, and the reduction or elimination of bias when constructing any psychometric instruments (Section 9.05). We turn next to the risk–benefit process, but we should also note that ethical values with relevance to statistical practices are embodied in the codes developed by statistical organizations (e.g., American Statistical Association, 1999; see also Panter & Sterba, 2011).

Table 3.3. The Revised Ethical Principles Adopted in August, 1982, by the Council of Representatives of the American Psychological Association for Research with Human Participants*

The decision to undertake research rests upon a considered judgment by the individual psychologist about how best to contribute to psychological science and human welfare. Having made the decision to conduct research, the psychologist considers alternative directions in which research energies and resources might be invested. On the basis of this consideration, the psychologist carries out the investigation with respect and concern for the dignity and welfare of the people who participate and with cognizance of federal and state regulations and professional standards governing the conduct of research with human participants.

- A. In planning a study, the investigator has the responsibility to make a careful evaluation of its ethical acceptability. To the extent that the weighing of scientific and human values suggests a compromise of any principle, the investigator incurs a correspondingly serious obligation to seek ethical advice and to observe stringent safeguards to protect the rights of human participants.
- B. Considering whether a participant in a planned study will be a “subject at risk” or a “subject at minimal risk,” according to recognized standards, is of primary ethical concern to the investigator.
- C. The investigator always retains the responsibility for ensuring ethical practice in research. The researcher is also responsible for the ethical treatment of research participants by collaborators, assistants, students, and employees, all of whom, however, incur similar obligations.
- D. Except in minimal-risk research, the investigator establishes a clear and fair agreement with research participants, prior to their participation, that clarifies the obligations and responsibilities of each. The investigator has the obligation to honor all promises and commitments included in that agreement. The investigator informs the participants of all aspects of the research that might reasonably be expected to influence willingness to participate and explains all other aspects of the research about which the participants inquire. Failure to make full disclosure prior to obtaining informed consent requires additional safeguards to protect the welfare and dignity of the research participants. Research with children or with participants who have impairments that would limit understanding and/or communication requires special safeguarding procedures.
- E. Methodological requirements of a study may make the use of concealment or deception necessary. Before conducting such a study, the investigator has a special responsibility to (1) determine whether the use of such techniques is justified by the study’s prospective scientific, educational, or applied value; (2) determine whether alternative procedures are available that do not use concealment or deception; and (3) ensure that the participants are provided with sufficient explanation as soon as possible.
- F. The investigator respects the individual’s freedom to decline to participate in or to withdraw from the research at any time. The obligation to protect this freedom requires careful thought and consideration when the investigator is in a position of authority or influence over the participant. Such positions of authority include, but are not limited to, situations in which research participation is required as part of employment or in which the participant is a student, client, or employee of the investigator.
- G. The investigator protects the participant from physical and mental discomfort, harm, and danger that may arise from research procedures. If risks of such consequences exist, the investigator informs the participant of that fact. Research procedures likely to cause serious or lasting harm to a participant are not used unless the failure to use these procedures might expose the participant to risk of greater harm or unless the research has great potential benefit and fully informed and voluntary consent is obtained from each participant. The participant should be informed of procedures for contacting the investigator within a reasonable time period following participation should stress, potential harm, or related questions or concerns arise.
- H. After the data are collected, the investigator provides the participant with information about the nature of the study and attempts to remove any misconceptions that may have arisen. Where scientific or human values justify delaying or withholding this information, the investigator incurs a special responsibility to monitor the research and to ensure that there are no damaging consequences for the participant.

Table 3.3. (Continued)

- I. Where research procedures result in undesirable consequences for the individual participant, the investigator has the responsibility to detect and remove or correct these consequences, including long-term effects.
- J. Information obtained about a research participant during the course of an investigation is confidential unless otherwise agreed upon in advance. When the possibility exists that others may obtain access to such information, this possibility, together with the plans for protecting confidentiality, is explained to the participant as part of the procedure for obtaining informed consent.

*Quoted from pp. 5–7 in *Ethical Principles in the Conduct of Research with Human Participants*. Washington, DC: American Psychological Association. Copyright ©1982 by the American Psychological Association.

Expanding the Calculation of Risks and Benefits

After the Belmont Report, it seemed that everything changed permanently for scientists engaged in human subject research, and it made little difference whether they were engaged in biomedical, behavioral, or social research. As the philosopher John E. Atwell (1981) put it, the moral dilemma was to defend the justification of using human subjects as the means to an end that was beneficial in some profoundly significant way (e.g., the progression of science, public health, or public policy) while protecting the moral “ideals of human dignity, respect for persons, freedom and self-determination, and a sense of personal worth” (p. 89). Review boards were now delegated the responsibility of making prior assessments of the future consequences of proposed research on the basis of the probability that a certain magnitude of psychological, physical, legal, social, or economic harm might result, weighed against the likelihood that “something of positive value to health or welfare” might result. Quoting the Belmont Report, “risk is properly contrasted to probability of benefits, and benefits are properly contrasted with harms rather than risks of harms,” where the “risks and benefits of research may affect the individual subjects, the families of the individual subjects, and society at large (or special groups of subjects in society).” The moral calculus of benefits to risks was said to be “in a favorable ratio” when the anticipated risks were outweighed by the anticipated benefits to the subjects (assuming this was applicable) and the anticipated benefit to society in the form of the advancement of knowledge. Put into practice, however, researchers and members of review boards found it difficult to “exorcize the devil from the details” when challenged by ethical guidelines that frequently conflicted with traditional technical criteria (Mark, Eyssell, & Campbell, 1999, p. 48). As human

beings are not omniscient, there was also the problem that “neither the risks nor the benefits . . . can be perfectly known in advance” (Mark et al., 1999, p. 49).

These complications notwithstanding, another catch-22 of the risk–benefit assessment is that it focuses only on the *doing of research*. Some years ago, we proposed a way of visualizing this predicament—first, in terms of an idealized representation of the risk–benefit assessment and, second, in terms of an alternative model focused on the costs and benefits of both the *doing* and *not doing* of research (Rosenthal & Rosnow, 1984). The latter model also has implications for the risk–benefit (we prefer the term *cost–utility*) of using or not using particular quantitative methods (we return to this idea in a moment). First, however, Figure 3.1 shows an idealized representation of the traditional risk–benefit assessment. Risk (importance or probability of harm) is plotted from low (C) to high (A) on the vertical axis, and the benefit is plotted from low (C) to high (D) on the horizontal axis. In other words, studies in which the risk–benefit assessment is close to A would presumably be *less likely* to be approved; studies close to D would be *more likely* to be approved; and studies falling along the B–C “diagonal of indecision” exist in a limbo of uncertainty until relevant information nudges the assessment to either side of the diagonal. The idea of “zero risk” is a methodological conceit, however, because all human subject research can be understood as carrying some degree of risk. The potential risk in the most benign behavioral and social research, for example, is the “danger of violating someone’s basic rights, if only the right of privacy” (Atwell, 1981, p. 89). However, the fundamental problem of the traditional model represented in Figure 3.1 is that it runs the risk of ignoring the “not doing of research.” Put another way, there are also moral costs when potentially useful research is forestalled, or if the design or implementation is

Table 3.4. General Principles Adopted in 2003 by the American Psychological Association*

General Principles

General Principles, as opposed to Ethical Standards, are aspirational in nature. Their intent is to guide and inspire psychologists toward the very highest ethical ideals of the profession. General Principles, in contrast to Ethical Standards, do not represent obligations and should not form the basis for imposing sanctions. Relying upon General Principles for either of these reasons distorts both their meaning and purpose.

Principle A: Beneficence and Nonmaleficence

Psychologists strive to benefit those with whom they work and take care to do no harm. In their professional actions, psychologists seek to safeguard the welfare and rights of those with whom they interact professionally and other affected persons, and the welfare of animal subjects of research. When conflicts occur among psychologists' obligations or concerns, they attempt to resolve these conflicts in a responsible fashion that avoids or minimizes harm. Because psychologists' scientific and professional judgments and actions may affect the lives of others, they are alert to and guard against personal, financial, social, organizational, or political factors that might lead to misuse of their influence. Psychologists strive to be aware of the possible effect of their own physical and mental health on their ability to help those with whom they work.

Principle B: Fidelity and Responsibility

Psychologists establish relationships of trust with those with whom they work. They are aware of their professional and scientific responsibilities to society and to the specific communities in which they work. Psychologists uphold professional standards of conduct, clarify their professional roles and obligations, accept appropriate responsibility for their behavior, and seek to manage conflicts of interest that could lead to exploitation or harm. Psychologists consult with, refer to, or cooperate with other professionals and institutions to the extent needed to serve the best interests of those with whom they work. They are concerned about the ethical compliance of their colleagues' scientific and professional conduct. Psychologists strive to contribute a portion of their professional time for little or no compensation or personal advantage.

Principle C: Integrity

Psychologists seek to promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology. In these activities psychologists do not steal, cheat, or engage in fraud, subterfuge, or intentional misrepresentation of fact. Psychologists strive to keep their promises and to avoid unwise or unclear commitments. In situations in which deception may be ethically justifiable to maximize benefits and minimize harm, psychologists have a serious obligation to consider the need for, the possible consequences of, and their responsibility to correct any resulting mistrust or other harmful effects that arise from the use of such techniques.

Principle D: Justice

Psychologists recognize that fairness and justice entitle all persons to access to and benefit from the contributions of psychology and to equal quality in the processes, procedures, and services being conducted by psychologists. Psychologists exercise reasonable judgment and take precautions to ensure that their potential biases, the boundaries of their competence, and the limitations of their expertise do not lead to or condone unjust practices.

Principle E: Respect for People's Rights and Dignity

Psychologists respect the dignity and worth of all people, and the rights of individuals to privacy, confidentiality, and self-determination. Psychologists are aware that special safeguards may be necessary to protect the rights and welfare of persons or communities whose vulnerabilities impair autonomous decision making. Psychologists are aware of and respect cultural, individual, and role differences, including those based on age, gender, gender identity, race, ethnicity, culture, national origin, religion, sexual orientation, disability, language, and socioeconomic status and consider these factors when working with members of such groups. Psychologists try to eliminate the effect on their work of biases based on those factors, and they do not knowingly participate in or condone activities or others based upon such prejudices.

*Quoted from the American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* (<http://www.apa.org/ethics/code2002.html>). Effective date June 1, 2003, copyrighted in 2002 by the American Psychological Association.

compromised in a way that jeopardizes the integrity of the research (cf. Haywood, 1976).

Figure 3.2 shows an alternative representing a cost–utility assessment of both the doing and not doing of research. In Part A, the decision plane model on the left corresponds to a cost–utility

appraisal of the “doing of research,” and the model on the right corresponds to an appraisal of the “not doing of research.” We use the terms *cost* and *utility* each in a collective sense. That is, the cost of doing and the cost of not doing a particular research study include more than only the risk of psychological or

Figure 3.1 Idealized decision-plane model representing the relative risks and benefits of research submitted to a review board for prior approval (after Rosenthal & Rosnow, 1984; Rosnow & Rosenthal, 1997).

physical harm; they also include the cost to society, funding agencies, and to scientific knowledge when imagination and new scientifically based solutions are stifled. As one scientist observed, “Scientists know that questions are not settled; rather, they are given provisional answers for which it is contingent upon the imagination of followers to find more illuminating solutions” (Baltimore, 1997, p. 8). We also use *utility* in a collective sense, not just in the way that a “tool” can immediately be instrumentally useful, but in a way that may have no immediate application and instead “speaks to our sense of wonder and paves the way for future advances” (Committee on Science, Engineering, and Public Policy, 2009, p. 3). These figurative definitions of cost and utility aside, Part B of Figure 3.2 suggests a way of transforming the three dimensions of Part A to a two-dimensional model. Suppose an A–D “decision diagonal” for each of the decision planes

Figure 3.2 Decision-planes representing the ethical assessment of the costs and utilities of doing and not doing research (after Rosenthal & Rosnow, 1984, 2008). (A) Costs and utilities of doing (left plane) and not doing (right plane) research. (B) Composite plane representing both cases in Part A (above).

in Part A (in contrast to B–C and B'–C', the diagonals of indecision). For any point in the plane of *doing*, there would be a location on the cost axis and on the utility axis, where any point could be translated to an equivalent position on the decision diagonal. Thus, if a point were twice as far from A as from D, the transformed point would then be located two-thirds of the way on the decision diagonal A–D (closer to D than to A). Similar reasoning is applicable to *not doing*, with the exception that closeness to A would mean “do” rather than “not do.” Points near D tell us the research *should* be done, and points near D' tell us the research should *not* be done.⁶

Figure 3.2 can also be a way of thinking about cost–utility dilemmas regarding quantitative methods and statistical reporting practices. In the 2009 edition of the U.S. National Academy of Sciences (NAS) guide to responsible conduct in scientific research, there are several hypothetical scenarios, including one in which a pair of researchers (a post-doctoral and a graduate student) discuss how they should deal with two anomalous data points in a graph they are preparing to present in a talk (Committee on Science, Engineering, and Public Policy, 2009). They want to put the best face on their research, but they fear that discussing the two outliers will draw people's attention away from the bulk of the data. One option would be to drop the outliers, but, as one researcher cautions, this could be viewed as “manipulating” the data, which is unethical. The other person comments that if they include the anomalous points, and if a senior person then advises them to include the anomalous data in a paper they are drafting for publication, this could make it harder to have the paper accepted by a top journal. That is, the reported results will not be unequivocal (a potential reason for rejection), and the paper will also then be too wordy (another reason to reject it?). In terms of Figure 3.2, not including the two anomalous data points is analogous to the “not doing of research.” There are, of course, additional statistical options, which can also be framed in cost–utility terms, such as using a suitable transformation to pull in the outlying stragglers and make them part of the group (cf. Rosenthal & Rosnow, 2008, pp. 310–311). On the other hand, outliers that are not merely recording errors or instrument errors can sometimes provide a clue as to a plausible moderator variable. Suppressing this information could potentially impede scientific progress (cf. Committee on Science, Engineering, and Public Policy, 2009, p. 8).

Unfortunately, there are also cases involving the suppression of data where the cost is not only that it impedes progress in the field, but it also undermines the authority and trustworthiness of scientific research and, in some instances, can cause harm to the broader society, such as when public policy is based on only partial information or when there is selective outcome reporting of the efficacy of clinical interventions in published reports of randomized trials (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008; Vedula, Bero, Scherer, & Dickersin, 2009). In an editorial in *Science*, Cicerone (2010), then president of the NAS, stated that his impression—based on information from scattered public opinion polls and various assessments of leaders in science, business, and government—was that “public opinion has moved toward the view that scientists often try to suppress alternative hypotheses and ideas and that scientists will withhold data and try to manipulate some aspects of peer review to prevent dissent” (p. 624). Spielmans and Parry (2010) described a number of instances of “marketing-based medicine” by pharmaceutical firms. Cases included the “cherry-picking” of data for publication, the suppression or understatement of negative results, and the publication (and distribution to doctors) of journal articles that were not written by the academic authors who lent their names, titles, and purported independence to the papers but instead had been written by ghost writers hired by pharmaceutical and medical-device firms to promote company products. Spielmans and Parry displayed a number of screen shots of company e-mails, which we do not usually get to see because they go on behind the curtain. In an editorial in *PLoS Medicine* (2009) lamenting the problem of ghost writers and morally dubious practices in the medical marketing of pharmaceuticals, the editors wrote:

How did we get to the point that falsifying the medical literature is acceptable? How did an industry whose products have contributed to astounding advances in global health over the past several decades come to accept such practices as the norm? Whatever the reasons, as the pipeline for new drugs dries up and companies increasingly scramble for an ever-diminishing proportion of the market in “me-too” drugs, the medical publishing and pharmaceutical industries and the medical community have become locked into a cycle of mutual dependency, in which truth and a lack of bias have come to be seen as optional extras. Medical journal editors need to decide whether they want to

roll over and just join the marketing departments of pharmaceutical companies. Authors who put their names to such papers need to consider whether doing so is more important than having a medical literature that can be believed in. Politicians need to consider the harm done by an environment that incites companies into insane races for profit rather than for medical need. And companies need to consider whether the arms race they have started will in the end benefit anyone. After all, even drug company employees get sick; do they trust ghost authors?

Ethical Standards and Quantitative Methodological Standards

We turn now to Table 3.5, which shows a matrix of general ethical standards crossed with quantitative methodological standards (after Rosnow & Rosenthal, 2011). We do not claim that the row and column standards are either exhaustive or mutually exclusive but only that they are broadly representative of (1) aspirational ideals in the society as a whole and (2) methodological, data analytic, and reporting standards in science and technology. The matrix is a convenient way of reminding ourselves of the distinction between (1) and (2), and it is also a way of visualizing a potential clash between (1) and (2) and, frequently, the opportunity to exploit this situation in a way that could have rewarding ethical and scientific implications. Before we turn specifically to the definitions of the row and column headings in Table 3.5, we will give a quick example of what we mean by “rewarding ethical and scientific implications” in the context of the recruitment of volunteers. For this example, we draw on some of our earlier work on specific threats to validity (collectively described as *artifacts*) deriving from the volunteer status of the participants for research participation. Among our concerns when we began to study the volunteer was that ethical sensitivities seemed to be propelling psychological science into a science of informed volunteers (e.g., Rosenthal & Rosnow, 1969; Rosnow & Rosenthal, 1970). It was long suspected that people who volunteered for behavioral and social research might not be fully adequate models for the study of behavior in general. To the extent that volunteers differ from nonvolunteers on dimensions of importance, the use of volunteers could have serious effects on such estimated parameters as means, medians, proportions, variances, skewness, and kurtosis. The estimation of parameters such as these is the principal goal in survey research, whereas in experimental

research the focus is usually on the magnitude of the difference between the experimental and control group means. Such differences, we and other investigators observed, were sometimes affected by the use of volunteers (Rosenthal & Rosnow, 1975, 2009).

With problems such as these serving as beginning points for empirical and meta-analytic investigations, we explored the characteristics that differentiated volunteers and nonvolunteers, the situational determinants of volunteering, some possible interactions of volunteer status with particular treatment effects, the implications for predicting the direction and, sometimes, the magnitude of the biasing effects in research situations, and we also thought about the broader ethical implications of these findings (Rosenthal & Rosnow, 1975; Rosnow & Rosenthal, 1997). For example, in one aspect of our meta-analytic inquiry, we put the following question to the research literature: *What are the variables that tend to increase or decrease the rates of volunteering obtained?* Our preliminary answers to this question may have implications for both the theory and practice of behavioral science. That is, if we continue to learn more about the situational determinants of volunteering, we can learn more about the social psychology of social influence processes. Methodologically, once we learn more about the situational determinants of volunteering, we should be in a better position to reduce the bias in our samples that derives from the volunteer subjects being systematically different from nonvolunteers in a variety of characteristics. For example, one situational correlate was that the more important the research was perceived, the more likely people were to volunteer for it. Thus, mentioning the importance of the research during the recruitment phase might coax more of the “nonvolunteers” into the sampling pool. It would be unethical to exaggerate or misrepresent the importance of the research. By being honest, transparent, and informative, we are treating people with respect and also giving them a well-founded justification for asking them to volunteer their valuable time, attention, and cooperation. In sum, the five column headings of Table 3.5 frequently come precorrelated in the real world of research, often with implications for the principles in the row headings of the table.

Turning more specifically to the row headings in Table 3.5, rows A, B, C, and E reiterate the three “basic ethical principles” in the Belmont Report, which were described there as respect for persons, beneficence, and justice. *Beneficence* (the ethical

Table 3.5. General Ethical Standards Crossed with Quantitative Methodology Standards (after Rosnow & Rosenthal, 2011)

Ethical standards	Quantitative methodological standards				
	1. Transparency	2. Informativeness	3. Precision	4. Accuracy	5. Groundedness
A. Beneficence					
B. Nonmaleficence					
C. Justice					
D. Integrity					
E. Respect					

ideal of “doing good”) was conflated with the principle (b) of *nonmaleficence* (“not doing harm”), and the two were also portrayed as *obligations* assimilating two complementary responsibilities: (1) do not harm and (2) maximize possible benefits and minimize possible harms. Next in Table 3.5 is *justice*, by which we mean a sense of “fairness in distribution” or “what is observed” (quoting from the Belmont Report). As the Belmont Report went on to explain: “Injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly.” Conceding that “what is equal?” and “what is unequal?” are often complex, highly nuanced questions in a specific research situation (just as they are when questions of justice are associated with social practices, such as punishment, taxation, and political representation), justice was nonetheless considered a basic moral precept relevant to the ethics of research involving human subjects. Next in Table 3.5 is *integrity*, an ethical standard that was not distinctly differentiated in the Belmont Report but that was discussed in detail in the NAS guide (Committee on Science, Engineering, and Public Policy, 2009). Integrity implies honesty and truthfulness; it also implies a prudent use of research funding and other resources and, of course, the disclosure of any conflicts of interest, financial or otherwise, so as not to betray public trust. Finally, *respect* was described in the Belmont Report as assimilating two obligations: “first, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection.” In the current APA code, respect is equated with civil liberties—that is, privacy, confidentiality, and self-determination.

Inspecting the column headings in Table 3.5, first by *transparency*, we mean here that the quantitative

results are presented in an open, frank, and candid way, that any technical language used is clear and appropriate, and that visual displays do not obfuscate the data but instead are as crystal clear as possible. Elements of graphic design are explained and illustrated in a number of very useful books and articles, particularly the work of Tufte (1983, 1990, 2006) and Wainer (1984, 1996, 2000, 2009; Wainer & Thissen, 1981), and there is a burgeoning literature in every area of science on the visual display of quantitative data. Second, by *informativeness*, we mean that there is enough information reported to enable readers to make up their own minds on the basis of the primary results and enough to enable others to re-analyze the summary results for themselves. The development of meta-analysis, with emphasis on effect sizes and moderator variables, has stimulated ways of recreating summary data sets and vital effect size information, often from minimal raw ingredients. Third, the term *precision* is used not in a statistical sense (the likely spread of estimates of a parameter) but rather in a more general sense to mean that quantitative results should be reported to the degree of exactitude required by the given situation. For example, reporting the average scores on an attitude questionnaire to a high degree of decimal places is psychologically meaningless (*false precision*), and reporting the weight of mouse subjects to six decimal places is pointless (*needless precision*). Fourth, *accuracy* means that a conscientious effort is made to identify and correct mistakes in measurements, calculations, and the reporting of numbers. Accuracy also means not exaggerating results by, for example, making false claims that applications of the results are unlikely to achieve. Fifth, *groundedness* implies that the method of choice is appropriate to the question of interest, as opposed to using whatever is fashionable or having

a computer program repackaged the data in a one-size-fits-all conceptual framework. The methods we choose must be justifiable on more than just the grounds that they are what we were taught in graduate school, or that “this is what everyone else does” (cf. Cohen, 1990, 1994; Rosnow & Rosenthal, 1995, 1996; Zuckerman, Hodgins, Zuckerman, & Rosenthal, 1993).

Clinical Significance and the Consequences of Statistical Illiteracy

To bring this discussion of quantitative methods and ethics full circle, we turn finally to a problem that has been variously described as innumeracy (Paulos, 1990) and statistical illiteracy. The terms are used to connote a lack of knowledge or understanding of the meaning of numbers, statistical concepts, or the numeric expression of summary statistics. As the authors of a popular book, *The Numbers Game*, put it: “Numbers now saturate the news, politics, life. . . . For good or for evil, they are today’s preeminent public language—and those who speak it rule” (Blastland & Dilnot, 2009, p. x). To be sure, even people who are most literate in the language of numbers are prone to wishful thinking and fearful thinking and, therefore, sometimes susceptible to those who use numbers and gimmicks to sway, influence, or even trick people. The mathematician who coined the term innumeracy told of how his vulnerability to whim “entrained a series of ill-fated investment decisions,” which he still found “excruciating to recall” (Paulos, 2003, p. 1). The launching point for the remainder of our discussion was an editorial in a medical journal several years ago, in which the writers of the editorial lamented “the premature dissemination of research and the exaggeration of medical research findings” (Schwartz & Woloshin, 2003, p. 153). A large part of the problem is an emphasis on RR statistics that hook general readers into making unwarranted assumptions, a problem that may often begin with researchers, funders, and journals that “court media attention through press releases” (Woloshin, Schwartz, Casella, Kennedy, & Larson, 2009, p. 613). Confusion about risk and risk statistics is not limited to the general public (cf. Prasad, Jaeschke, Wyer, Keitz, & Guyatt, 2008), but it is the susceptible public (Carling, Kristoffersen, Herrin, Treweek, Oxman, Schünemann, Akl, & Montori, 2008) that must ultimately pay the price of the accelerating costs of that confusion. Stirring the concept of statistical significance into this mix can frequently produce a

truly astonishing amount of confusion. For example, writing in the *Journal of the National Cancer Institute*, Miller (2007) mentioned that many doctors equate the level of statistical significance of cancer data with the “degree of improvement a new treatment must make for it to be clinically meaningful” (p. 1832).⁷

In the space remaining, we concentrate on misconceptions and illusions regarding the concepts of RR and statistical significance when the clinical significance of interventions is appraised through the lens of these concepts in randomized clinical trials (RCTs). As a case in point, a highly cited report on the management of depression, a report that was issued by the National Institute for Health and Clinical Excellence (NICE), used RR of 0.80 or less as a threshold indicator of clinical significance in RCTs with dichotomous outcomes and statistically significant results.⁸ We use the term *clinical significance* here in the way that it was defined in an authoritative medical glossary, although we recognize that it is a hypothetical construct laden with surplus meaning as well (cf. Jacobson & Truax, 1991). In the glossary, clinical significance was taken to mean that “an intervention has an effect that is of practical meaning to patients and health care providers” (NICHSR, 2010; cf. Jeans, 1992; Kazdin, 1977, 2008). By *intervention*, we mean a treatment or involvement such as a vaccine used in a public health immunization program to try to eradicate a preventable disease (e.g., the Salk poliomyelitis vaccine), or a drug that can be prescribed for a patient in the doctor’s office, or an over-the-counter medicine (e.g., aspirin) used to reduce pain or lessen the risk of an adverse event (e.g., heart attack), or a medication and/or psychotherapy to treat depression. By tradition, RCTs are the gold standard in evidence-based medicine when the goal is to appraise the clinical significance of interventions in a carefully controlled scientific manner. Claims contradicted by RCTs are not always immediately rejected in evidence-based medicine, as it has been noted that some “claims from highly cited observational studies persist and continue to be supported in the medical literature despite strong contradictory evidence from randomized trials” (Tatsioni, Bonitsis, & Ioannidis, 2007). Of course, just as gold can fluctuate in value, so can conclusions based on the belief that statistical significance is a proxy for clinical significance, or when it is believed that given statistical significance, clinical significance is achieved only if the reduction in RR reaches some arbitrary fixed magnitude (*recall*, for example, NICE, 2004). The challenge

Table 3.6. Final results for myocardial infarction (MI) and hemorrhagic stroke (HS) for the aspirin (325 mg every other day) component of the Physicians' Health Study (Steering Committee of the Physicians' Health Study Research Group, 1989). The increase in relative risk (RRI) for HS was more than twice the reduction in relative risk (RRR) for MI. Having one more case of HS in the aspirin group would have yielded a chi-square significant at $p < 0.05$, $RR = 2.0$, and $RRI = 100\%$. In the combined samples, the event rate of MI ($378/22,071 = 0.0171$, or 1.71%) exceeded the event rate of HS ($35/22,071 = 0.0016$, or 0.16%) by a ratio of about 10:1, and a difference of $1.71\% - 0.16\% = 1.55\%$. In the subtable on the right, RRI is the relative risk increase, computed as RRR (see Table 3.7), but indicated as RRI when the treatment increases the risk of the adverse outcome.

Myocardial infarction (Heart attack)			Hemorrhagic stroke				
	MI	No MI	Total		HS	No HS	Total
Aspirin	139	10,898	11,037	Aspirin	23	11,014	11,037
Placebo	239	10,795	11,034	Placebo	12	11,022	11,034
Total	378	21,693	22,071	Total	35	22,036	22,071
Chi-square 26.9, $p = 2.1 \times 10^{-7}$			Chi-square 3.46, $p = 0.063$				
RR	0.58			RR	1.92		
RRR	42%			RRI	92%		
r(phi)	0.035			r(phi)	-0.013		

is to reverse the accelerating cost curve of statistical illiteracy in an area that affects us all (see, for example, Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2008).

Table 3.6 helps us illustrate the folly of a delicate balancing act that is sometimes required between statistical significance and RR. The table shows a portion of the results from the aspirin component of a highly cited double-blind, placebo-controlled, randomized trial to test whether 325 milligrams of aspirin every other day reduces the mortality from cardiovascular disease and whether beta-carotene decreases the incidence of cancer (Steering Committee of the Physicians' Health Study Research Group, 1989). The aspirin component of the study was terminated earlier than planned on finding "a statistically significant, 44 [sic] percent reduction in the risk of myocardial infarction for both fatal and nonfatal events . . . [although] there continued to be an apparent but not significantly increased risk of stroke" (p. 132). RR (for relative risk) refers to the ratio of the incidence rate of the adverse event (the illness) in the treated sample to the control sample; RRR is the relative risk reduction; and RRI, is the relative risk increase (the computation of these indices is described in Table 3.7). When tables of independent counts are set up as shown in Tables 3.6 and 3.7, an RR less than 1.0 indicates that the treated sample fared better than the control sample

(thereby implying RRR), and an RR greater than 1.0 indicates the treated sample did more poorly than the control (thereby implying RRI). Observe that the "slightly increased risk of stroke" ($RRI = 92\%$) was actually more than twice the reduction in risk of heart attack ($RRR = 42\%$)! Suppose the study had continued, and one more case of stroke had turned up in the aspirin group. The p -value would have reached the 0.05 level, and the researchers might have arrived at a different conclusion, possibly that the benefit with respect to heart attack was more than offset by the increased risk in stroke. Apparently, a p -value only a hair's-breadth greater than 0.05 can trump a RR increase of 92%. On the other hand, the event rate of stroke in the study as a whole was only 0.16%, less than one-tenth the magnitude of the event rate of 1.7% of heart attack in the study as a whole.⁹ However, we would never know this from the RR alone.

The fact is that RR statements are oblivious to event rates in the total N . To give a quick example, suppose in a study with 100 people each in the treated and control samples that 1 treated person and 5 untreated people (controls) became ill. RR and RRR would be 0.20 and 80%, respectively. Stating there was an 80% reduction in risk of the adverse event conveys hope. However, suppose we increase each sample size to 1,000 but still assume 1 case of illness in the treated sample and 5 cases

Table 3.7. Six studies each with total sample size (*N*) of 2,000 and 1% event rates in Studies 1 and 4 (20 cases out of 2,000), 25% event rates in Studies 2 and 5 (500 cases out of 2,000), and 50% event rates in Studies 3 and 6 (1,000 cases out of 2,000). RR, the relative risk or risk ratio, indicates the ratio of the incidence rate of level of risk in the treated group to the level of risk in the control group. With cells labeled A, B, C, D from upper left (A) to upper right (B), to lower left (C), to lower right (D), $RR = [A/(A+B)]/[C/(C+D)]$, where $RR < 1.0$ favors the treatment effect (risk reduction) and $RR > 1.0$ favors the control effect (risk increase). OR, the odds ratio, also called relative odds or the cross-product ratio, is the ratio of A/B to C/D, or the cross-product AD/BC. RRR, the relative risk reduction, is the reduction in risk of the adverse outcome (e.g., illness) in the treated sample relative to the control, which is indicated as a percentage by dividing RD (defined next) by $[C/(C+D)]$ and then multiplying by 100. RD, the risk difference, also called the absolute risk reduction (ARR), is the reduction in risk of the particular adverse outcome (e.g., cancer, heart attack, stroke) in the treated group compared with the level of baseline risk in the control—that is, $[A/(A+B)] - [C/(C+D)]$. Multiplying RD (or ARR) times 10,000 estimates the number of people in a group of 10,000 that are predicted to benefit from the treatment. $NNT = 1/RD = 1/ARR$, is the number needed to treat to prevent a single case of the particular adverse outcome.

	Study 1 (<i>N</i> = 2,000)		Study 2 (<i>N</i> = 2,000)		Study 3 (<i>N</i> = 2,000)	
	Adverse outcome (1%)		Adverse outcome (25%)		Adverse outcome (50%)	
Condition	Yes	No	Yes	No	Yes	No
Treatment	1	999	25	975	50	950
Control	19	981	475	525	950	50
Chi-square	16.4 ($p < 0.0001$)		540.0 ($p < 0.0001$)		1,620.0 ($p < 0.0001$)	
RR	0.05		0.05		0.05	
OR	0.05		0.03		0.003	
RRR	94.7%		94.7%		94.7%	
<i>r</i> (phi)	0.090		0.52		0.90	
RD=ARR	0.018		0.45		0.90	
$NNT=1/ARR=1/RD$	55.6		2.2		1.1	
ARR(10,000)	180		4,500		9,000	
	Study 4 (<i>N</i> = 2,000)		Study 5 (<i>N</i> = 2,000)		Study 6 (<i>N</i> = 2,000)	
	Adverse outcome (1%)		Adverse outcome (25%)		Adverse outcome (50%)	
Condition	Yes	No	Yes	No	Yes	No
Treatment	9	991	225	775	450	550
Control	11	989	275	725	550	450
Chi-square	0.2 ($p = 0.89$)		6.7 ($p = 0.01$)		20.0 ($p < 0.0001$)	
RR	0.82		0.82		0.82	
OR	0.82		0.77		0.67	
RRR	18.2%		18.2%		18.2%	
<i>r</i> (phi)	0.01		0.06		0.10	
RD=ARR	0.002		0.05		0.10	
$NNT=1/ARR=1/RD$	500		20		10	
ARR(10,000)	20		500		1,000

Figure 3.3 Histograms based on the six studies in Table 3.7, in which the total sample size (N) was 2,000 in each study. Darkened areas of the bars indicate the number of adverse outcomes (event rates), which increased from 1% (20 cases out of 2,000) in Studies 1 and 4, to 25% (500 cases out of 2,000) in Studies 2 and 5, to 50% (1,000 cases out of 2,000) in Studies 3 and 6. However, the relative risk (RR) and relative risk reduction (RRR) were insensitive to these vastly different event rates. In Studies 1, 2, and 3, the RR and RRR remained constant at 0.05 and 94.7%, respectively, whereas in Studies 4, 5, and 6, the RR and RRR remained constant at 0.82 and 18.2%, respectively.

of illness in the control sample. We would still find $RR = 0.20$ and $RRR = 80\%$. It makes no difference how large we make the sample sizes, as RR and RRR will not budge from 0.20 and 80% so long as we assume 1 case of illness in the treated sample and 5 cases of illness in the control sample. Suppose we now hold the N constant and see what happens to the RR and RRR when the event rate in the overall N changes from one study to another. In Figure 3.3, we see the results of six hypothetical studies in which the event rates increased from 1% in Studies 1 and 4, to 25% in Studies 2 and 5, to 50% in Studies 3 and 6. Nonetheless, in Studies 1, 2, and 3, RR remained constant at 0.05 and RRR remained constant at an attention-getting 95%. In Studies 4, 5, and 6, RR and RRR stayed constant at 0.82 and 18%, respectively.

Further details of the studies in Figure 3.3 are given in Table 3.7. The odds ratio (OR), for the ratio of two odds, was for a time widely promoted as a measure of association in 2×2 tables of counts (Edwards, 1963; Mosteller, 1968) and is still frequently reported in epidemiological studies

(Morris & Gardner, 2000). As Table 3.7 shows, OR and RR are usually highly correlated. The absolute risk reduction (ARR), also called the risk difference (RD), refers to the absolute reduction in risk of the adverse event (illness) in the treated patients compared with the level of baseline risk in the control group. Gigerenzer et al. (2008) recommended using the absolute risk reduction (RD) rather than the RR. As Table 3.7 shows, RD (or ARR) is sensitive to the differences in the event rates. There are other advantages as well to RD, which are discussed elsewhere (Rosenthal & Rosnow, 2008, pp. 631–632). Phi is the product-moment correlation (r) when the two correlated variables are dichotomous, and Table 3.7 shows it is sensitive to the event rates and natural frequencies. Another useful index is NNT, for the number of patients that need to be treated to prevent a single case of the adverse event. Relative risk may be an easy-to-handle description, but it is only an alerting indicator that tells us that something happened and we need to explore the data further. As Tukey (1977), the consummate exploratory data analyst, stated: “Anything

that makes a simpler description possible makes the description more easily handleable; anything that looks below the previously described surface makes the description more effective" (p. v). And, we can add, that any index of the magnitude of effect that is clear enough, transparent enough, and accurate enough to inform the nonspecialist of exactly what we have learned from the quantitative data increases the ethical value of those data (Rosnow & Rosenthal, 2011).

Conclusion

In a cultural sphere in which so many things compete for our attention, it is not surprising that people seem to gravitate to quick, parsimonious forms of communication and, in the case of health statistics, to numbers that appear to speak directly to us. For doctors with little spare time to do more than browse abstracts of clinical trials or the summaries of summaries, the emphasis on parsimonious summary statistics such as RR communications in large sample RCTs may seem heavily freighted with clinical meaning. For the general public, reading about a 94.7% reduction in the risk of some illness, either in a pharmaceutical advertisement or in a news story about a "miracle drug that does wonders," is attention-riveting. It is the kind of information that is especially likely to arouse an inner urgency in patients but also in anyone who is anxious and uncertain about their health. Insofar as such information exaggerates the absolute effects, it is not only the patient or the public that will suffer the consequences; the practice of medicine and the progress of science will as well. As Gigerenzer et al. (2008) wrote, "Statistical literacy is a necessary precondition for an educated citizenship in a technological democracy" (p. 53). There are promising opportunities for moral (and societal) rewards for quantitative methodologists who can help us educate our way out of statistical illiteracy. And that education will be beneficial, not only to the public but to many behavioral, social, and medical researchers as well. As that education takes place, there will be increased clarity, transparency, and accuracy of the quantitative methods employed, thereby increasing their ethical value.

Future Directions

An important theoretical and practical question remains to be addressed: To what extent is there agreement among quantitative methodologists in their evaluation of quantitative procedures as to the

degree to which each procedure in a particular study meets the methodological standards of transparency, informativeness, precision, accuracy, and groundedness? The research program called for to address these psychometric questions of reliability will surely find that specific research contexts, specific disciplinary affiliations, and other specific individual differences (e.g., years of experience) will be moderators of the magnitudes of agreement (i.e., reliabilities) achieved. We believe that the results of such research will demonstrate that there will be some disagreement (that is, some unreliability) in quantitative methodologists' evaluations of various standards of practice. And, as we noted above, that is likely to be associated with some disagreement (that is, some unreliability) in their evaluations of the ethical value of various quantitative procedures.

Another important question would be addressed by research asking the degree to which the specific goals and specific sponsors of the research may serve as causal factors in researchers' choices of quantitative procedures. Teams of researchers (e.g., graduate students in academic departments routinely employing quantitative procedures in their research) could be assigned at random to analyze the data of different types of sponsors with different types of goals. It would be instructive to learn that choice of quantitative procedure was predictable from knowing who was paying for the research and what results the sponsors were hoping for. Recognition of the possibility that the choice of quantitative procedures used might be affected by the financial interests of the investigator is reflected in the increased frequency with which scientific journals (e.g., medical journals) require a statement from all co-authors of their financial interest in the company sponsoring the research (e.g., pharmaceutical companies).

Finally, it would be valuable to quantify the costs and utilities of doing and not doing a wide variety of specific studies, including classic and not-so-classic studies already conducted, and a variety of studies not yet conducted. Over time, there may develop a disciplinary consensus over the costs and utilities of a wide array of experimental procedures. And, although such a consensus is building over time, it will be of considerable interest to psychologists and sociologists of science to study disciplinary differences in such consensus-building. Part of such a program of self-study of disciplines doing quantitative research would focus on the quantitative procedures used, but the primary goal would be to apply survey research methods to establish

the degree of consensus on research ethics of the behavioral, social, educational, and biomedical sciences. The final product of such a program of research would include the costs and utilities of doing, and of *not* doing, a wide variety of research studies.

Notes

1. Where we quote from a document but do not give the page numbers of the quoted material, it is because either there was no pagination or there was no consistent pagination in the online and hard copy versions that we consulted. Tables 3.1–3.4 reprint only the original material, as there were slight discrepancies between original material and online versions.

2. Patullo (1982) described the logical basis on which “rule-makers” (like DHEW) had proceeded in terms of a syllogism emphasizing not the potential benefits of research but only the avoidance of risks of harm: “(a) Research can harm subjects; (2) Only impartial outsiders can judge the risk of harm; (3) Therefore, all research must be approved by an impartial outside group” (p. 376).

3. Hearings on the recommendations in the Belmont Report were conducted by the President’s Commission for the Study of Ethical Problems in Medicine, Biomedical, and Behavioral Research. Proceeding on the basis of the information provided at these hearings and on other sources of advice, the Department of Health and Human Services (DHHS) then issued a set of regulations in the January 26, 1981, issue of the *Federal Register*. A compendium of regulations and guidelines that now govern the implementation of the National Research Act and subsequent amendments can be found in the DHHS manual known as the “Gray Booklet,” specifically titled *Guidelines for the Conduct of Research Involving Human Subjects at the National Institutes of Health* (available online at <http://ohsr.od.nih.gov/guidelines/index.html>).

4. Smith, Kimmel, and Klein (2009) reported that 43.4% of the articles on consumer research in leading journals in the field in 1975 through 1976 described some form of deception in the research. By 1989 through 1990, the number of such articles increased to 57.7%, where it remained steady at 56% in 1996 through 1997, increased to 65.7% in 2001 through 2002, and jumped to 80.4% in 2006 through 2007. The issue of deception is further complicated by the fact that active and passive deceptions are far from rare in our society. Trial lawyers manipulate the truth in court on behalf of their clients; prosecutors surreptitiously record private conversations; journalists get away with using hidden cameras and undercover practices to obtain stories; and the police use sting operations and entrapment procedures to gather incriminating evidence (cf. Bok, 1978, 1984; Saxe, 1991; Starobin, 1997).

5. The document, titled “Ethical Principles of Psychologists and Code of Conduct,” is available online at <http://www.apa.org/ETHICS/code2002.html>.

6. Adaptations of the models in Figures 3.1 and 3.2 have been used to cue students about possible ethical dilemmas in research and data analysis (cf. Bragger & Freeman, 1999; Rosnow, 1990; Strohmets & Skleder, 1992).

7. The confusion of statistical significance with practical importance may be a more far-reaching problem in science. In a letter in *Science*, the writers noted that “almost all reviews

and much of the original research [about organic foods] report only the statistical significance of the differences in nutrient levels—not whether they are nutritionally important” (Clancy, Hamm, Levine, & Wilkins, 2009, p. 676).

8. NICE (2004) also recommended that researchers use a standardized mean difference (SMD) of half a standard deviation or more (i.e., d or $g \geq 0.5$) with *continuous outcomes* as the threshold of clinical significance for initial assessments of statistically significant summary statistics (NICE, 2004). However, effects far below the 0.5 threshold for SMDs have been associated with important interventions. For example, in the classic Salk vaccine trial (Brownlee, 1955; Francis, Korns, Voight, Boisen, Hemphill, Napier, & Tolchinsky, 1955), $\phi = 0.011$, which has a d -equivalent of 0.022 (Rosnow & Rosenthal, 2008). It is probably the case across the many domains in which clinical significance is studied that larger values of d or g are in fact generally associated with greater intervention benefit, efficacy, or clinical importance. But it is also possible for large SMDs to have little or no clinical significance. Suppose a medication was tested on 100 pairs of identical twins with fever, and in each and every pair, the treated twin loses exactly one-tenth of 1 degree more than the control twin. The SMD will be infinite, inasmuch as the variability (the denominator of d or g) will be 0, but few doctors would consider this ES clinically significant. As Cohen (1988) wisely cautioned, “the *meaning* of any given ES is, in the final analysis, a function of the context in which it is embedded” (p. 535).

9. The high RR of HS in this study, in which participants (male physicians) took 325 milligrams every other day, might explain in part why the current dose for MI prophylaxis is tempered at only 81 milligrams per day.

Author Note

Ralph L. Rosnow is Thaddeus Bolton Professor Emeritus at Temple University, Philadelphia, PA (rosnow@temple.edu). Robert Rosenthal is a Distinguished Professor at the University of California at Riverside (robert.rosenthal@ucr.edu) and Edgar Pierce Professor of Psychology, Emeritus, Harvard University.

References

- Allison, D. B. (2009). The antidote to bias in research. *Science*, 326, 522.
- American Psychological Association. (1973). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- American Psychological Association. (1982). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- American Statistical Association (1999). *Ethical guidelines for statistical practice*. <http://www.amstat.org/profession/index.cfm?fusaction=ethical>
- Atwell, J. E. (1981). Human rights in human subjects research. In A. J. Kimmel (Ed.), *New directions for methodology of social and behavioral science: Ethics of human subject research* (No. 10, pp. 81–90). San Francisco, CA: Jossey-Bass.
- Baltimore, D. (January 27, 1997). Philosophical differences. *The New Yorker*, p. 8.

- Beecher, H. K. (July 2, 1966). Documenting the abuses. *Saturday Review*, 45–46.
- Beecher, H. K. (1970). *Research and the individual*. Boston: Little Brown.
- Blanck, P. D., Bellack, A. S., Rosnow, R. L., Rotheram-Borus, M. J., & Schooler, N. R. (1992). Scientific rewards and conflicts of ethical choices in human subjects research. *American Psychologist*, 47, 959–965.
- Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram*. New York: Basic Books.
- Blastland, M., & Dilnot, A. (2009). *The numbers game*. New York: Gotham Books.
- Bok, S. (1978). *Lying: Moral choice in public and private life*. New York: Pantheon.
- Bok, S. (1984). *Secrets: On the ethics of concealment and revelation*. New York: Vintage Books.
- Bordia, P., & Rosnow, R. L. (1995). Rumor rest stops on the information highway: A naturalistic study of transmission patterns in a computer-mediated rumor chain. *Human Communication Research*, 25, 163–179.
- Bragger, J. D., & Freeman, M. A. (1999). Using a cost–benefit analysis to teach ethics and statistics. *Teaching of Psychology*, 26, 34–36.
- Brownlee, K. A. (1955). Statistics of the 1954 polio vaccine trials. *Journal of the American Statistical Association*, 50, 1005–1013.
- Carling, C., Kristoffersen, D. T., Herrin, J., Treweek, S., Oxman, A. D., Schünemann, H., et al. (2008). How should the impact of different presentations of treatment effects on patient choice be evaluated? *PLoS One*, 3(11), e3693.
- Cicerone, R. J. (2010). Ensuring integrity in science. *Science*, 327, 624.
- Clancy, K., Hamm, M., Levine, A. S., & Wilkins, J. (2009). Organics: Evidence of health benefits lacking. *Science*, 325, 676.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304–1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997–1003.
- Committee on Science, Engineering, and Public Policy (2009). *On being a scientist: A guide to responsible conduct in research* (3rd ed.). Washington, DC: National Academies Press.
- Edwards, A. W. F. (1963). The measure of association in a 2×2 table. *Journal of the Royal Statistical Society*, 126, 109–114.
- Fairchild, A. L., & Bayer, R. (1999). Uses and abuses of Tuskegee. *Science*, 284, 918–921.
- Francis, T., Jr., Korn, R. F., Voight, R. B., Boisen, M., Hemphill, E., Napier, J., & Tolchinsky, E. (1955). An evaluation of the 1954 poliomyelitis vaccine trials: A summary report. *American Journal of Public Health*, 45(5), 1–63.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.
- Harris, G. (March 10, 2009). Doctor admits pain studies were frauds, hospital says. *The New York Times*. Retrieved March 20, 2009, www.nytimes.com/2009/03/11/health/research/11pain.html?ref=us.
- Haywood, H. C. (1976). The ethics of doing research . . . and of not doing it. *American Journal of Mental Deficiency*, 81, 311–317.
- Huff, D. (1982). *How to lie with statistics*. New York: Norton.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Jeans, M. E. (1992). Clinical significance of research: A growing concern. *Canadian Journal of Nursing*, 24, 1–2.
- Jones, H. H. (1993). *Bad blood: The Tuskegee syphilis experiment* (Revised edition). New York: Free Press.
- Kaiser, J. (September 18, 2009). Researcher, two universities sued over validity of prostate cancer test. *Science*, 235, 1484.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427–451.
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146–159.
- Katz, J. (1972). *Experimentation with human beings*. New York: Russell Sage.
- Kelman, H. C. (1968). *A time to speak: On human values and social research*. San Francisco, CA: Jossey-Bass.
- Kimmel, A. J. (1991). Predictable biases in the ethical decision making of psychologists. *American Psychologist*, 46, 786–788.
- Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, 311, 88–90.
- Mark, M. M., Eysell, K. M., & Campbell, B. (1999). The ethics of data collection and analysis. In J. L. Fitzpatrick & M. Morris (Eds.), *Ethical issues in program evaluation* (pp. 47–56). San Francisco, CA: Jossey-Bass.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67, 371–378.
- Milgram, S. (1975). *Obedience to authority: An experimental view*. New York: Harper Colophon Books.
- Miller, J. D. (2007). Finding clinical meaning in cancer data. *Journal of the National Cancer Institute*, 99(24), 1832–1835.
- Morris, J. A., & Gardner, M. J. (2000). Epidemiological studies. In D. A. Altman, D. Machin, T. N. Bryant, & M. J. Gardner (Eds.), *Statistics with confidence* (2nd ed., pp. 57–72). London: British Medical Journal Books.
- Mosteller, F. (1968). Association and estimation in contingency tables. *Journal of the American Statistical Association*, 63, 1–28.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (April 18, 1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved April 8, 2009, <http://ohsr.od.nih.gov/guidelines/belmont.html>.
- National Commission on Research. (1980). Accountability: Restoring the quality of the partnership. *Science*, 207, 1177–1182.
- NICE. (2004). *Depression: Management of depression in primary and secondary care* (Clinical practice guideline No. 23). London: National Institute for Health and Clinical Excellence.
- NICHSR. (2010). *HTA 101: Glossary*. Retrieved March 3, 2010 from <http://www.nlm.nih.gov/nichsr/hta101/ta101014.html>
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783.
- Orne, M. T. (1969). Demand characteristics and the concept of quasi-control. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). New York:

- Academic Press. (Reissued in Rosenthal & Rosnow, 2009, pp. 110–137).
- Panther, A. T., & Sterba, S. K. (2011). *Handbook of ethics in quantitative methodology*. Taylor & Francis.
- Pattullo, E. L. (1982). Modesty is the best policy: The federal role in social research. In T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., & L. Walters (Eds.), *Ethical issues in social research* (pp. 373–390). Baltimore, MD: Johns Hopkins University Press.
- Paulos, J. A. (1990). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Vintage Books.
- Paulos, J. A. (2003). *A mathematician plays the stock market*. New York: Basic Books.
- PLoS Medicine Editors (September, 2009). Ghostwriting: The dirty little secret of medical publishing that just got bigger. *PLoS Medicine*, 6, Issue 9, e1000156. Accessed September 19, 2009 from <http://www.plosmedicine.org/static/ghostwriting.action>.
- Prasad, K., Jaeschke, R., Wyer, P., Keitz, S., & Guyatt, G. (2008). Tips for teachers of evidence-based medicine: Understanding odds ratios and their relationship to risk ratios. *Journal of General Internal Medicine*, 23(5), 635–640.
- Reynold, P. D. (1975). Value dilemmas in the professional conduct of social science. *International Social Science Journal*, 27, 563–611.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127–134.
- Rosenthal, R., & Rosnow, R. L. (1969). The volunteer subject. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 59–118). New York: Academic Press. (Reissued in Rosenthal & Rosnow, 2009, pp. 48–92).
- Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject*. New York: Wiley-Interscience. (Reissued in Rosenthal & Rosnow, 2009, pp. 667–862).
- Rosenthal, R., & Rosnow, R. L. (1984). Applying Hamlet's question to the ethical conduct of research. *American Psychologist*, 45, 775–777.
- Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). New York: McGraw-Hill.
- Rosenthal, R., & Rosnow, R. L. (2009). *Artifacts in behavioral research*. New York: Oxford University Press.
- Rosnow, R. L. (1990). Teaching research ethics through role-play and discussion. *Teaching of Psychology*, 17, 179–181.
- Rosnow, R. L. (1997). Hedgehogs, foxes, and the evolving social contract in psychological science: Ethical challenges and methodological opportunities. *Psychological Methods*, 2, 345–356.
- Rosnow, R. L., & Rosenthal, R. (1970). Volunteer effects in behavioral research. In K. H. Craik, B. Kleinmuntz, R. L. Rosnow, R. Rosenthal, J. A. Cheyne, & R. H. Walters, *New directions in psychology* (pp. 211–277). New York: Holt, Rinehart & Winston.
- Rosnow, R. L., & Rosenthal, R. (1995). "Some things you learn aren't so": Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science*, 6, 3–9.
- Rosnow, R. L., & Rosenthal, R. (1996). Contrasts and interactions redux: Five easy pieces. *Psychological Science*, 7, 253–257.
- Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: W. H. Freeman.
- Rosnow, R. L., & Rosenthal, R. (2008). Assessing the effect size of outcome research. In A. M. Nezu & C. M. Nezu (Eds.), *Evidence-based outcome research* (pp. 379–401). New York: Oxford University Press.
- Rosnow, R. L., & Rosenthal, R. (2011). Ethical principles in data analysis: An overview. In A. T. Panther & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 37–58). New York: Routledge.
- Rosnow, R. L., Rotheram-Borus, M. J., Ceci, S. J., Blanck, P. D., & Koocher, G. P. (1993). The institutional review board as a mirror of scientific and ethical standards. *American Psychologist*, 48, 821–826.
- Rosnow, R. L., Skleder, A. A., Jaeger, M. E., & Rind, B. (1994). Intelligence and the epistemics of interpersonal acumen: Testing some implications of Gardner's theory. *Intelligence*, 19, 93–116.
- Sales, B. D., & Folkman, S. (Eds.). (2000). *Ethics in research with human participants*. Washington, DC: American Psychological Association.
- Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist*, 46, 409–415.
- Schwartz, L. M., & Woloshin, S. (2003). On the prevention and treatment of exaggeration. *Journal of General Internal Medicine*, 18(2), 153–154.
- Scott-Jones, D., & Rosnow, R. L. (1998). Ethics and mental health research. In H. Friedman (Ed.), *Encyclopedia of mental health* (Vol. 2, pp. 149–160). San Diego, CA: Academic Press.
- Smith, M. B. (1969). *Social psychology and human values*. Chicago, IL: Aldine.
- Smith, N. C., Kimmel, A. J., & Klein, J. G. (2009). Social contract theory and the ethics of deception in consumer research. *Journal of Consumer Psychology*, 19, 486–496.
- Spielmans, G. I., & Parry, P. I. (2010). From evidence-based medicine to marketing-based medicine: Evidence from internal industry documents. *Journal of Bioethical Inquiry*: doi:10.1007/s11673-010-9208-8.
- Starobin, P. (January 28, 1997). Why those hidden cameras hurt journalism. *The New York Times*, p. A21.
- Steering Committee of the Physicians' Health Study Research Group. (1989). Final report on the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine*, 321, 129–135.
- Strohmetz, D. B., & Skleder, A. A. (1992). The use of role-play in teaching research ethics: A validation study. *Teaching of Psychology*, 19, 106–108.
- Stryker, J. (April 13, 1997). Tuskegee's long arm still touches a nerve. *The New York Times*, p. 4.
- Tatsioni, A., Bonitsis, N. G., & Ioannidis, J. P. A. (2007). Persistence of contradicted claims in the literature. *Journal of the American Medical Association*, 298, 2517–2526.
- Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, October 1946–April 1949, Vol. II. Washington, DC: U.S. Government Printing Office.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tufte, E. T. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- Tufte, E. T. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3), 32–40.
- VandenBos, G. R. (Ed.). (2007). *APA dictionary of psychology*. Washington, DC: American Psychological Association.
- Vedula, S. S., Bero, L., Scherer, R. W., & Dickersin, K. (2009). Outcome reporting in industry-sponsored trials of Gabapentin for off-label use. *New England Journal of Medicine*, 361(20), 1963–1971.
- Vinacke, W. E. (1954). Deceiving experimental subjects. *American Psychologist*, 9, 155.
- Wainer, H. (1984). How to display data badly. *American Statistician*, 38, 137–147.
- Wainer, H. (1996). Depicting error. *American Statistician*, 50(2), 101–111.
- Wainer, H. (2000). *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. Mahwah, NJ: Erlbaum.
- Wainer, H. (2009). *Picturing the uncertain world*. Princeton, NJ: Princeton University Press.
- Wainer, H., & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology*, 32, 191–241.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, England: Cambridge University Press.
- Woloshin, S., Schwartz, L. M., Casella, S. L., Kennedy, A. T., & Larson, R. J. (2009). Press releases by academic medical centers: Not so academic? *Annals of Internal Medicine*, 150(9), 613–618.
- Zuckerman, M., Hodgins, H. S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science*, 4, 49–53.