ETHICS & BEHAVIOR, *16*(4), 305–318

Misconduct in the Analysis and Reporting of Data: Bridging Methodological and Ethical Agendas for Change

Sonya K. Sterba
Department of Psychology
University of North Carolina at Chapel Hill

Fraudulent analysis and reporting of psychological data have the potential to contaminate the scientific knowledge base and eventuate in the unjustified expenditure of public money and scientific effort (Koocher & Keith-Spiegel, 1998). Traditionally, the field has relied on quantitative methodologists to educate researchers in proper analysis and reporting practices, and to examine these via peer review. The field has also relied on psychologists with training or board service in ethics to establish standards and implement strategies to discourage misconduct. However, this division of responsibility for examination, standard setting, and deterrence is shown to compromise the effectiveness of methodologists' and ethicists' respective gatekeeping efforts. Methodologically and ethically trained specialists instead need to coordinate efforts to safeguard analysis and reporting procedures. Researchers also need to increase self-monitoring. Potential obstacles to achieving these ends are considered, and three tactics are proposed to overcome them.

Keywords: methodology, research ethics, data analysis, quantitative psychology, misconduct

The analysis and reporting of data are widely acknowledged to be high-stakes endeavors. Fraudulent analysis and reporting of psychological data have the potential to contaminate the scientific knowledge base and eventuate in the unjustified ex-

Correspondence should be addressed to Sonya K. Sterba, L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, Davie Hall, Campus Box 3270, Chapel Hill, NC 27599–3270. E-mail: ssterba@email.unc.edu

penditure of public money and scientific effort (Koocher & Keith-Spiegel, 1998). Seen in this light, improper analysis and reporting practices constrain not only the collective advancement and integrity of the field but also the professional effectiveness of the individual researcher.

Although its importance is difficult to understate, practically speaking, assuring proper analysis and reporting practices is, indeed, a considerable undertaking. According to Gibson and Pope (1993, p. 335), it entails (a) the *examination* of the profession's own practices, (b) the establishment of *standards* to which psychologists hold themselves accountable, and (c) the implementation of *strategies* to foster ethical behavior. Traditionally, the field has largely relied on specialists in two subfields of psychology for said examination, standard setting, and strategy implementation. Quantitative methodologists have been relied on both to educate applied researchers in proper analysis and reporting practices and to examine finished analyses via peer review. Psychologists with specialized training or board service in ethics have been relied on to establish standards and to implement strategies to discourage misconduct.

I argue herein that this manner of separating responsibility actually serves to prevent both methodologists and ethicists from adequately achieving their intended gatekeeping functions. After first describing types of analysis and reporting misconduct and presenting examples of each, I survey the effectiveness of current examination of, standards for, and strategies to discourage analysis and reporting misconduct. This serves to highlight drawbacks of the current division of gatekeeping responsibilities. It also highlights the need for (a) methodologically and ethically trained specialists to do collaborative gatekeeping and (b) researchers to collectively perform more self-monitoring. I provide possible reasons for why the latter has not yet occurred. I conclude with three tactics for bridging ethical and methodological efforts and promoting collective involvement. These, I claim, will serve to better monitor, detect, and deter misconduct in the analysis and reporting of data.

TYPES OF MISCONDUCT IN THE ANALYSIS AND REPORTING OF DATA

As alluded to earlier, misconduct in data analysis and reporting can be classified in multiple ways. For example, misconduct can be overt or covert and intentional or unintentional.[1] *Overt* misconduct includes improper analytic practices a researcher re-

---

[1]The presence or absence of intent to deceive is not a focus of discussion here. Unintentional–overt and intentional–overt misconduct are both detectable by methodological reviewers. It is argued elsewhere in this article that both should be handled similarly, regardless of intent. Intentional–covert errors could be prevented by increased availability of ethics training pertinent to analysis and reporting, and unintentional–covert errors due to lack of awareness or sloppiness could be prevented by increased availability of methodological training.

ports in the results section of an article. In contrast, *covert* misconduct is not detectable just from reading the results section of an article. I illustrate each type of misconduct with multiple examples. These examples highlight junctures in the data analysis process where each type of misconduct can arise.

Overt Misconduct

One overt problem is dichotomizing continuous data (i.e., median splitting), which drastically reduces variability and can create significant results. Dichotomization has been denounced in both the quantitative literature and applied literature for decades, yet 12% of a recent sample of applied studies employed this procedure (see MacCallum, Zhang, Preacher, & Rucker, 2002). Other overt problems routinely detected by methodologists include crossvalidating exploratory data procedures with confirmatory data procedures on the same data set (found in 83% of structural equation modeling [SEM] articles reviewed) or not testing alternative models that equivalently fit the data but imply different theoretical conclusions (found in 98% of articles reviewed; Breckler, 1990). The former practice capitalizes on chance variation in the data set and can lead to erroneous conclusions, and the latter prematurely thwarts the consideration of competing theories. These two types of SEM misconduct have been denounced in the methodology literature for decades, yet are still quite prevalent in substantive applications.

Covert Misconduct

There are at least three common types of covert misconduct: trimming, capitalization on chance, and selective reporting. The first type of covert misconduct is trimming data in a systematic way. For example, several authors have reanalyzed data sets used in published studies and inferred that the original authors were statistically significantly more likely to omit outlier data points if doing so swayed the significance level in the direction of their stated hypotheses (see Dunnette, 1965; Kimmel, 1996; Rosenthal, 1994).

The second type of covert misconduct involves capitalizing on chance variations in data sets. For example, consider the practice of entering predictors stepwise into a regression and retaining those that explain the most variability in the outcome. "This is legitimate," commented Robert MacCallum (personal communication, February 24, 2005), past member of the editorial board of *Psychological Methods,* "if you proclaim 'my goal was to account for as much variance as possible with as few variables as possible.'" If, however, the final set of predictors was reported as if it was theoretically conceived a priori, with no mention of other predictors tried and eliminated, this would contribute to the *file drawer problem* (see Rosenberg, 2005). Most generally, when the data-driven nature of exploratory analyses is not acknowledged and the Type I error rate is not adjusted, researchers risk overinterpreting chance variation in data sets (Holland & Copenhaver, 1988).

For example, consider a researcher who builds and tests a structural equation model, but the output indicates that the model is a poor fit for the data. The researcher might change parameters, add correlated errors, and so forth—even though these were not hypothesized a priori—in order to obtain acceptable model fit. The researcher might then claim that the final model was what was hypothesized in the first place. Commenting explicitly on the ethical nature of this infraction, MacCallum noted, "I don't know the prevalence because authors don't actually say it; authors just report it as if it is the model they started with. *It is an ethical issue* and there is no question that it occurs" (personal communication, February 15, 2005 [italics added].

The third type of covert misconduct involves selective reporting, which takes several forms. Selective reporting of background literature in the introductory section of a grant application presents an unfounded inflation of the importance of a researcher's proposed project (Hoey, 2003) and could sway funding allocations away from projects that have more substantial empirical backing. Selective reporting of model fit indexes, in addition, is a problem in applications of factor analysis and SEM. Many computer programs that perform these procedures output pages of model fit criteria, and Bollen (1989) has denounced the covert practice of skimming the list of model fit criteria for acceptable values and selectively reporting those. A similarly covert practice is selectively reporting the parameters of the model that was fit to the data. This practice, claimed Breckler (1990), is common when researchers have made a number of data-driven model modifications and do not want readers to realize this. To prevent readers from finding out, the researchers might neglect to describe their model specification in detail, knowing that this effectively prevents readers from determining the precise model that was fit.

This overview of analysis and reporting misconduct conveys how easily one or more incidents of misconduct could compromise the theoretical import of research findings. I next investigate the extent to which methodologists' examinations and ethicists' standards succeed in detecting and deterring misconduct.

EXAMINATION OF DATA ANALYSIS AND REPORTING

Applied researchers in psychology have historically relied upon quantitative methodologists to examine analysis and reporting practices, identify problems, and propose and explicate correctives for these. However, methodologists are able to recognize only those problems that are *overtly* observable in the results sections of applied articles.

Although instances of overt misconduct are potentially detectable by methodologists, I could find no published surveys of overt misconduct that were both comprehensive and randomized. This may be because the need for comprehensive examination outstrips what the relatively few methodologists can supply. For example, there are fewer than 10 quantitative graduate programs in the United States and a ratio of hundreds of applied psychological researchers to each methodologist (Clay, 2005). These relatively few methodologists are responsible not only for examining overt analysis and reporting practices in ongoing, submitted, and published research, but also for training applied researchers. Hence, prevalence estimates of overt misconduct tend to be largely based on periodic surveys of specific techniques used in certain journals. Several of these were mentioned earlier. They showed that after a methodologist publishes a standard, overt errors in violation of that standard decrease very slowly over time in applied research (e.g., MacCallum et al., 2002), or stay the same (e.g., Cohen, 2003). It may be that methodologists' directives for improving analysis and reporting accuracy are slow to affect the status quo of research practice outside of quantitative psychology because these directives lack an ethical backbone that would serve as a motivating force for change. This slow adoption of methodological standards and guidance is becoming even more problematic. This is because rapid software developments have propelled the use of sophisticated analytic procedures in journal and grant submissions by "researchers and clinicians [who] often lack the skills they need to interpret ever more sophisticated science" (Clay, 2005, p. 26).

Compounding this problem is the realm of *covert* analysis and reporting misconduct that is not detectable by methodologists alone. Covert misconduct has traditionally been considered to fall within an ethical, rather than a methodological, purview. Yet, academic texts and courses on research ethics have devoted little attention to this issue (e.g., Sales & Folkman, 2000). Transgressions in data analysis and reporting are commonly only briefly mentioned in ethics texts or not discussed at all. Deception, confidentiality, and dispersion of risks versus benefits with human research participants take center stage in such volumes. This mirrors the rigor with which institutional ethics boards scrutinize recruiting, collecting, and storing data from human participants—and with which applied researchers attend to the same. In contrast, we know little about the prevalence of covert misconduct in data analysis or reporting. A sense of rates of covert misconduct can be obtained from the most extensive survey to date of perceived transgressions in data analysis and reporting. Kimmel (1996) surveyed 2,600 faculty and graduate students in the social and medical sciences and found that fully 43% of the students and 50% of faculty were aware of misconduct in their own laboratories—that included forging data and withholding data (p. 298).

In sum, thorough examination of data analysis and reporting practices is lacking. Methodologists are limited to examining overt practices, and they are too few in number to adequately address even these overt practices. Psychologists with specialized training in ethics also are limited in number and are expected to multitask service on ethics boards with training researchers. Ethicists may neglect examination of analysis and reporting misconduct due to this overload of responsibilities or due to the misperception that analysis and reporting misconduct falls solely within the jurisdiction of methodologists. Either way, this review shows that the field cannot feasibly rely on small subgroups of specialists with ethical or methodological training to safeguard data analysis and reporting practices. Increased outreach efforts to disseminate appropriate methodological and ethical knowledge to other leaders in the field (e.g., journal editors, principal investigators on research studies, and professors) are needed. Given the rapid pace of methodological developments, involving the collective body of researchers in enhanced self-monitoring is necessary to sustain confidence in published analyses.

STANDARDS FOR DATA ANALYSIS AND REPORTING

Ethics boards at the institutional, state, and national association level are the traditional standard-setting bodies for research practices in psychology. These boards issue ethical codes containing standards that apply to data analysis and reporting (e.g., the American Psychological Association's [APA's] *Ethical Principles of Psychologists and Code of Conduct,* 2002). The standards featured in ethics codes, however, differ markedly from those published by methodologists on the same topic area. Methodologists' standards are much more specific.

Motives for compliance also differ for methodological and ethical standards. Compliance with standards issued by ethics boards is aided by the threat of sanctions and by appeals to integrity and ethical imperatives (i.e., "Their … intent is to … inspire psychologists toward the very highest ethical ideals of the profession;" APA, 2002, p. 2). In contrast, compliance with methodologists' standards is typically motivated neither by explicit threat nor by an appeal to a higher ethical imperative. As discussed earlier, compliance with the methodologists' standards is low. The key issue here is whether these ethical standards for data analysis and reporting are too vague to be of practical use to guide researchers. This issue is examined through in-depth study of the data analysis and reporting portions of the *Code* (APA, 2002), as follows.

Readers familiar with the *Code* (APA, 2002) will recall that the document includes both general principles—nonobligatory, aspirational ideals to consider in arriving at an ethical course of action—as well as ethical standards—enforceable rules for conduct (p. 2). Both the general principles and most of the ehical standards are written with intentional generality, to ensure applicability to a broad swath of contexts and infractions (APA, 2002, p. 2). Yet this inclusive format is better suited to some areas of the discipline than others. In the analysis and reporting area, I contend that the vague nature of these principles does not prevent them from "becoming quickly outdated" (APA, 2002, p. 2), as intended. Instead, they are so disconnected from the detailed directives given in the methodological literature that their practical utility is limited, even in the present. In other words, the vague nature of the *Code*'s principles and standards for data analysis and reporting also makes it difficult to identify infractions. This, in turn, limits the effectiveness of the *Code*'s formal and informal strategies for reporting and resolving identified infractions (described in Standards 1.04 and 1.05).

For example, of all cases opened by the APA Ethics Committee in 2002, 2003, and 2004 for alleged ethical violations, only one case included "improper research techniques" and none included "biasing data" (APA, 2003, 2004, 2005). This rate of allegations for ethical violations is far lower than the perceived (Kimmel, 1996) or observed (MacCallum et al., 2002) prevalence of misconduct. It may be that the vague definitions of ethical standards impair identification of misconduct in some cases (see also Holaday & Yost, 1995). Example sections of the *Code* pertaining to analysis and reporting of results are presented here for illustrative purposes. The potential guidance afforded by more vaguely stated norms versus more concretely stated norms is contrasted.

General Principles

In General Principle B of the *Code*, psychologists are held to be "concerned about the ethical compliance of their colleagues' scientific and professional conduct" (APA, 2002, p. 3). As previously mentioned, approximately half of researchers are aware of examples of noncompliance incidences among colleagues (Kimmel, 1996), yet the *Code* provides unclear guidance for how researchers should demonstrate this concern or how they ought to act upon it. Is showing concern merely an emotive state, or does it entail taking action? In General Principle C, psychologists are told to "seek to promote accuracy, honesty and truthfulness in the science … of psychology" and not to "engage in … intentional misrepresentation of fact" (APA, 2002, p. 3). Again, because scientists are not given explicit, concrete directives but instead are told to "seek" to "promote" truthfulness, it is unclear the ends to which they need go to make sure that research is conducted and portrayed veraciously. The latter directive not to engage in the intentional misrepresentation of fact is relatively clear; however, it is difficult for colleagues to conclusively determine intent. (Intentional data errors could be relabeled by authors as unintentional mistakes after the fact. Moreover, even if a researcher denied intent, the ethical issue of negligence could apply if his or her analyses contain considerably more errors than expected by chance alone.)

Ethical Standards

In the Ethical Standards portion of the *Code*, psychologists are mandated in Standard 5.01 not to "knowingly make public statements that are false, deceptive, or fraudulent concerning their research" (APA, 2002, p. 8). The same critique applies here as was leveled at General Principle C with regard to the use of "knowingly" and associated burden of proof. Similarly, Standard 8.10 proclaims that "psychologists do not fabricate data" (APA, 2002, p. 12). Because there are a myriad of ways researchers can cook, trim, or fabricate data (see Rosenthal, 1994), the rules need more specificity here to evaluate the severity and reprehensibility of acts ranging from selectively eliminating outliers to blatant data forgery.

Standards 6.01 and 8.14a, however, are more concretely specified than those previously cited and seem to permit clearer identification of misconduct. Standard 6.01 calls for psychologists to "create, and to the extent the records are under their control, maintain, disseminate, store, retain, and dispose of records and data relating to their professional and scientific work in order to … allow for replication of research design and analyses" (APA, 2002, p. 9). Similarly, Standard 8.14a calls for psychologists not to "withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose" (APA, 2002, p. 13).

This survey of types of analysis and reporting standards and their limitations suggests that ethical standards, such as those found in the *Code*, sometimes may be too vague to be of practical use for guiding researchers. Ethical principles and standards found in the *Code* are intentionally kept general, which is helpful for domains such as exploitative relationships and confidentiality, where specific guidelines could not be feasibly concocted for the diversity of potential situations. In the data analysis and reporting domain, however, there are specific, clearly preferable choices for many decision points. The *Code*'s standards for data analysis and reporting do not help researchers to recognize specific instances of misconduct. For these reasons, I contend that methodologists ought to increase participation in ethics committees. Ethics committee members likely have not had the same level of quantitative training as have methodologists; collaboration would ensure that standards are meaningfully specific and current.

STRATEGIES FOR ENFORCING DATA ANALYSIS AND REPORTING STANDARDS

Extant strategies for enforcing these *Code* mandates (referenced in Standards 1.4 and 1.5) entail initial attempts to resolve the issue with the individual in question and, if necessary, subsequent "referral to state or national committees on professional ethics, to state licensing boards, or to the appropriate institutional authorities" (APA, 2002, p. 4). I here discuss these enforcement agencies and strategies, and their limitations, proceeding from the national association level (mentioned in the *Code*) to the institutional level (mentioned in the *Code*) to the context of peer review (not mentioned in the *Code*) to the context of proximal individual monitoring (mentioned in the *Code*).

First, at the national association level, the strategy of referring analysis and reporting misconduct to the APA Ethics Committee was previously shown to be little utilized in practice (at least between 2002 and 2004). Second, at the institutional level, institutional review boards are typically less involved in the monitoring of analysis and reporting of data; they are predominantly consumed with the supervision of collection and storage of data and informed consent from human subjects. The National Institutes of Health and National Science Foundation do have offices that oversee the detection and investigation of scientific misconduct by institutions. These National Institutes of Health and National Science Foundation offices, however, typically remain in a position of distal oversight. For example, their involvement includes setting timelines for the investigation of allegations, with the potential for government intervention if these are not upheld (see Grisso et al., 1991). Hence, responsibility for proximal supervision and detection of analysis and reporting wrongdoing is placed on journal peer reviewers and on the scientific community itself.

Yet, at the peer review level, there are comparatively fewer qualified methodologists available to review manuscripts than there are manuscripts employing sophisticated analytic techniques (Clay, 2005). Thus, methodologically sophisticated analyses may be reviewed by professionals lacking training in the technique that was used. Compounding this issue, there are currently no systematic assessments of (a) the adequacy or methodological soundness of reviewers' comments or (b) the proportion of overt transgressions that are detected during peer review (Kimmel, 1996). In fact, some journal editors report incidents of reviewers insisting that authors conduct methodologically inopportune procedures (such as a median split; R. MacCallum, personal communication, February 15, 2005). Moreover, reviewers exert a bias for publishing studies with statistically significant findings (Koocher & Keith-Spiegel, 1998, p. 306). Finally, reviewers can detect only overt transgressions. Unmentioned data trimming or selectively reported fit indexes would remain undiscovered. Simply put, although peer reviewers are supposed to ensure the quality, accuracy, and honesty of reported findings, they do not have enough information to detect the broad variety of potential overt and covert transgressions.

Last, at the individual level, it is rare for fellow researchers to identify and confront each other about potential analysis and reporting misconduct. Fellow researchers are remiss in expressing suspicions about their colleagues' ethical conduct for fear of reprisals (Kimmel, 1996). They report finding it difficult to conclusively prove whether a given finding is the result of intentional fraud, the result of an honest error, or due to scientific disagreement (Grisso et al., 1991). In addition, colleague accusations generate negative media attention to reporting misconduct. They are perceived to threaten the public investment and trust in the research enterprise needed for quality research to progress (Koocher & Keith-Spiegel, 1998). In other words, researchers feel that "skepticism must be tempered by trust in the honesty of fellow researchers in order for scientific progress to proceed in a relatively smooth fashion" (Koocher & Keith-Spiegel, 1998, p. 305).

In lieu of such skepticism, researchers hope that the scientific practice of replication of results will automatically weed aberrant, forged findings from authentic results. Yet, far from serving a self-cleansing function, most journals are biased against publishing mere replication studies (Kazdin, 2002). Even when replication is performed, scientists usually cannot rule out the possibility that unique contextual factors or confounds—rather than some form of intentional or negligent misconduct—led to discrepant results. Thus, the potential for replication is not as potent a deterrent for researchers as is typically imagined.

This review of strategies for enforcing data analysis and reporting standards suggests that available strategies at the national association, institutional, peer review, and individual levels are sometimes underutilized or unreliable. Before considering whether these strategies need to be revised—and what resources this would entail—it is worthwhile to consider what is likely to happen to rates of misconduct if no changes are made. Although the true prevalence of overt and covert misconduct is poorly understood, one can estimate whether rates of misconduct are likely to change by attending to whether the *motives* for such misconduct are likely to change.

WHY MISCONDUCT IS LIKELY TO CONTINUE IN THE ABSENCE OF CHANGES

The publish-or-perish research climate, as well as the field's attachment to, and reliance on, point estimates and null hypothesis significance testing, may make analysis and reporting misconduct more likely to occur. I describe potential contributing factors and why they are likely to remain influential in the foreseeable future.

First, the aforementioned lack of investigations, safeguards, and potent deterrents can interact with the pervasive publish-or-perish academic climate to increase the temptation of analysis and reporting dishonesty. This temptation pertains not only to young researchers seeking promotion and tenure. Publication records, of course, facilitate maintenance of grants for established scientists as well. Left unchecked, there is little reason to believe that senior researchers would spontaneously abandon ethically questionable methodological practices that secured recognition and advancement for them early in their careers. In addition, researchers sometimes build their careers on the defense of a certain theory (e.g., the two-factor self-discrepancy theory; Higgins, 1987), which makes it all the more difficult to accept discrepant findings.

Second, the field's pervasive emphasis on, and utilization of, statistical significance testing may also increase the likelihood of data analysis and reporting misconduct. The ready supply of point estimates for making inferential decisions (e.g., $p < .05$, root mean square error of approximation < .05, Tucker-Lewis Fit Index, Comparative Fit Index > .95) can eclipse the more appropriate focus on strength and precision of relationships (see Cumming & Finch, 2005) and can tempt researchers to manipulate data to achieve gold-standard significance cutoffs. According to Dunnette (1965), this emphasis leads "most of us to remain content to build our theoretical castles on the quicksand of merely rejecting the null hypothesis" (p. 345).

Power analyses are a case in point. It is now commonplace for granting agencies to require applicants to demonstrate power of .80 to detect hypothesized effects (as it naturally does not make sense for agencies to fund studies whose sample size limitations preclude the detection of an effect, if one in fact exists; Hallahan & Rosenthal, 1996). Yet, submitting a grant proposal is, thus, in some ways predicated on achieving this .80 cutoff. Anecdotal evidence of researchers with archival samples working backward in power calculations—and reappraising and inflating hypothesized effect size estimates until the output power calculation inches up to .80—is not lacking (A. Panter, personal communication, January 28, 2004; R. MacCallum, personal communication, February 15, 2005). This practice undermines the whole point of a power analysis. This practice is further fueled by the lack of consensus on empirical effect size estimates for many substantive areas (Rosenthal, 1994, p. 131).

Schmidt (1996) has argued that unethical data manipulations aimed at clipping and preening a sample to move a finding of $p = .06$ or power = .70 into the golden area of $p < .05$ or power ≥ .80 would, likely, gradually become extinct if researchers switched emphasis to confidence intervals and power curves. Confidence intervals and power curves include all of the information captured in point estimates, yet also give estimates of precision, which are less amenable to tweaking (Cumming & Finch, 2005; Wilkinson & the Task Force on Statistical Inference, 1999). Such shifts in emphasis are slow to occur, however. In the meantime, psychologists ought to consider how to better monitor, detect, and deter analysis and reporting misconduct.

TACTICS FOR BETTER DETECTING AND DETERRING ANALYSIS AND REPORTING MISCONDUCT

In conclusion, ethical and methodological specialists' gatekeeping efforts in the area of data analysis and reporting have remained strikingly disparate and insular to date. They neither coordinate with each other nor involve the research community in outreach efforts aimed at engendering self-monitoring. Their independent efforts have led to insufficient examination of the prevalence of overt and covert misconduct, and to inconsistent standards that are unreliably enforced. Yet the quality control of data analyses and reporting practices is of prime importance. Thus, I propose three tactics to improve the prevention, detection, and deterrence of analysis and reporting misconduct that each involve melding of the methodological and ethical arenas.

First, psychologists need to better coordinate ethical and methodological standards pertaining to data analysis and reporting. Published methodological standards can lack the ethical imperative to motivate change, and published ethical standards can lack the specificity to direct that change. One first step toward coordinating standard setting across ethical and methodological specialties is offered here. Methodologists could be included on the committees of psychologists who create and revise research ethics codes and who respond to allegations of research ethics misconduct. In turn, committees disseminating methodological guidance, such as the APA Task Force on Statistical Inference, could include psychologists with research ethics expertise to aid in integrating an ethical perspective.

Second, we need to increase applied researchers' access to coordinated training in quantitative methods and research ethics. This will afford them the detailed methodological knowledge and the ethical imperative to better self-monitor their own analysis and reporting. Specifically, a cross-fertilization of ethics and methods instruction needs to take place throughout undergraduate and graduate training, and also at the faculty level. Currently, statistical and methodological courses are typically devoid of research ethics discussions, and vice versa. In fact, these ethics courses and methods courses are typically offered in different departments, by faculty members who rarely interact. Faculty guest lectures from the companion discipline can begin to bridge these fields. In addition, short quantitative workshops (such as those offered by the Interuniversity Consortium for Political and Social Research) and ethics workshops (such as those sponsored by the APA Ethics Committee) are outlets for reaching researchers who may not have access to methodological or ethical specialists at their home institutions. (Neither the Interuniversity Consortium for Political and Social Research nor the Ethics Committee currently lists ethics in data analysis and reporting as a topic area covered in their educational outreach efforts.)

It is essential that undergraduate and graduate psychology students be made mindful of the intersection of their methodological practices with ethical imperatives as they begin to conduct their own investigations—before poor habits become ingrained. We cannot expect students to completely autonomously make the connections between ethical and methodological imperatives; we need to scaffold them in this endeavor. This type of blended educational effort would increase the pool of journal and grant reviewers qualified to detect and enforce standards for analysis and reporting conduct. This, in turn, would render the field's examination of data analysis and reporting practices more pervasive and more reliable.

Third, psychologists need to more consistently implement strategies for preventing and deterring data analysis and reporting misconduct. Random auditing of analyses in articles submitted for peer review, and perhaps also systematic surveying of peer reviews themselves, are potential preventative deterrents (Kimmel, 1996). These deterrents would essentially be an expansion of the *Code*'s mandate to keep data available for potential reanalysis. If an audit of a given analysis reveals errors or discrepancies, the response would not be to try to determine whether this error was intentional or accidental. Instead, journal editors and reviewers would take it as their responsibility to inform authors of the ethical or methodological standards that were violated and issue a penalty—such as a request for reanalysis or replication—regardless of intent. This removes some of the professional hesitancy, fear of reprisals, and time involved in trying to prove intentional misconduct. This suggestion is in line with Snow's (1959) argument that "if we do not penalize false statements made in error, we open up the way, don't you see, for false statements by intention" (quoted in Kimmel, 1996, p. 273).

ACKNOWLEDGMENTS

The author thanks Bertram H. Raven, Elizabeth V. Swenson, and Erica H. Wise for helpful comments on an earlier version of this article.

REFERENCES

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060–1073.

American Psychological Association. (2003). Report of the Ethics Committee, 2002. *American Psychologist, 58,* 650–657.

American Psychological Association. (2004). Report of the Ethics Committee, 2003. *American Psychologist, 59,* 434–441.

American Psychological Association. (2005). Report of the Ethics Committee, 2004. *American Psychologist, 60,* 523–528.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107,* 260–273.

Clay, R. (2005). Too few in quantitative psychology. *Monitor on Psychology, 36,* 26–29.

Cohen, J. (2003). A power primer. In A. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (3rd ed., pp. 427–436). Washington, DC: American Psychological Association.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Dunnette, M. D. (1965). Fads, fashions, and foderol in psychology. *American Psychologist,* 343–352.

Gibson, W., & Pope, K. S. (1993). The ethics of counseling: A national survey of certified counselors. *Journal of Counseling and Development, 71,* 330–336.

Grisso, T., Baldwin, E., Blanck, P., Rotheram-Borus, M., Schooler, N., & Thompson, T. (1991). Standards in research: APA's mechanism for monitoring the challenges. *American Psychologist, 46,* 758–766.

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behavioral Research and Therapy, 34,* 489–499.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94,* 319–340.

Hoey, J. (2003). Unwanted results: The ethics of controversial research [Editorial]. *Canadian Medical Association Journal, 169,* 93.

Holaday, M., & Yost, T. (1995). A preliminary investigation of ethical problems in publication and research. *Journal of Social Behavior and Personality, 10,* 281–291.

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin, 104,* 145–149.

Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey.* Cambridge, MA; Blackwell.

Koocher, G. P., & Keith-Spiegel, P. (1998). *Ethics in psychology: Professional standards and cases* (2nd ed.). Oxford, England: Oxford University Press.

MacCallum, R., Zhang, S., Preacher, K., & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

Rosenberg, M. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59,* 464–468.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological results. *Psychological Science, 5,* 127–134.

Sales, B., & Folkman, S. (2000). *Ethics in research with human participants.* Washington, DC: American Psychological Association.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115–129.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.