

APs OBSERVER

Published by the American Psychological Society

Volume 14, Number 8

October 2001

- [Front Page](#)
- [Presidential Column](#)
- [Teaching Tips](#)
- [Student Notebook](#)
- [Department Profile](#)
- [People](#)
- [Members in the News](#)
- [Announcements](#)
- [Letters](#)
- [Meeting Calendar](#)
- [Browse all Classifieds](#)
- [Search the Classifieds](#)
- [Rates and Deadlines](#)
- [Submit a Classified](#)
- [Contact the Observer](#)
- [Subscriptions](#)
- [Previous Issue](#)

The IRB Review System:
How Do We Know It Works?

By John H. Mueller
and John J. Furedy

This is the second of a two-part series in which the authors consider the effectiveness of the research proposal ethics review process as it has evolved in psychological research in North America. They raise a fundamental question: Is there any evidence that these reviews are effective at reducing risk to the public? In Part I, they defined the situation and reviewed some irrelevant measures. In Part II, they discuss approaches and benefits to answering this question.

EVIDENCE OF REVIEW EFFECTIVENESS

In Part I ([Observer, September, 2001](#)), we argued that "problems found" do not constitute an acceptable measure of review effectiveness. The per-decade incident-rate analysis mentioned in Part I is one approach that would measure effectiveness. Here are at least two other ways one might assess the success of the proposal review process:

1. Conduct an experiment where for a year, a random half of the applications to an IRB are approved without review, whereas the other half undergo conventional review. At the end of the year we look at the number of problems arising during the actual experiments. Would the number of problems arising in the unreviewed group be any different than in the reviewed group? It seems doubtful, yet that problem-actually-arising rate difference is the only adequate evidence for the success of risk avoidance by proposal review activity.
2. Another experiment would be to take proposals approved at one research site and submit them to an IRB elsewhere. Would the prospect of approval at the second (third, etc.) IRB be different from 50:50? Alternatively, proposals rejected at one research site could be reviewed elsewhere. Perhaps the strongest test of this type would be to take the method section from published articles and submit them for review to various ethics review boards.

Analogous research has been done in the area of peer review (e. g., Peters & Ceci, 1982). The results were not popular; conventional wisdom about peer review proved to be less than robust. This re-submission procedure begs to be applied to the ethics review process.

Do we have repeat-reliability for the ethics review decisions? What we know is not encouraging. Eaton (1983) reported reliability to be 8 percent. Ceci, Peters, and Plotkin (1985) found reliability to vary as a function of the "sensitivity" of the research proposed. The values obtained by Ceci et al. were obtained in a context where the IRB reviewers knew they were involved in an experiment, perhaps a best case scenario because the reviewers can be expected to be especially thorough when there is an audience. This is quite different from the circumstances under which review boards operate on a daily basis, where there is no audience. Furthermore, these values derive from an era when reviews were still based on everyday risk, and when consideration of social issues was explicitly discouraged, and it seems unlikely that 20 years of obfuscation of review criteria have improved reliability.

(It may occur to some that such tests would have to be submitted to the IRBs and that the committees would not be interested in finding out the answers. But to the contrary, IRB review would not be an issue. Here's why: The alleged purpose of the review of proposals is to "protect the public," and this project would never involve the public, just the review boards. If the review process is about protecting the public, one of these re-review projects could be done by anyone at anytime).

MISSED OPPORTUNITIES

Finally, we should note that research opportunities have been squandered if, as we surmise, no incident data have been collected.

Profile of the Offender - Social scientists would normally inquire about the "profile of the offender," that is, what are the common characteristics of those proposals that result in public risk. Actually, even without systematic data, the offender profile seems fairly clear: When the research involves a vested interest (e.g., drug profits) the probability of misconduct is increased. In other words, it is likely that the typical offender is not Bob and Sally Neuprof in social sciences, individuals struggling to start a career, yet it is Bob and Sally who nonetheless are forced through all the "preventive" review. Bob and Sally may forget to lock a file cabinet, just as someone may fax medical records to the wrong number, but IRBs can not prevent such human lapses. We can do a much better job of using watchdog resources than by pretending we are all equal-opportunity offenders to-be.

Impact of Lay Input - Much has been made of the added value of having members of the public at large on the ethics review boards. Had we been collecting incident data over these decades, we might be able to see that lay input has indeed further reduced the number of untoward incidents, without relying on intuition that such things are effective. Any number of other ethics innovations could be validated as truly adding value in similar fashion, if only we had been collecting the incident data.

As it stands, reviews have been escalating with no regard for whether or not the added effort is worthwhile. This is akin to driving blindfolded or firing a gun in the dark, not normally considered ethical behaviors.

It is shoddy scholarship, and irresponsible bureaucracy, not to be collecting incident data. We need to know what aspects of the review process add value, in terms of actually reducing the number of problems arising, and just as important which practices merely waste time and money.

CONCLUSIONS

We have not been able to find any hard evidence that screening has had any effect in reducing problems arising in research involving human subjects. Thus we conclude that we need to take an honest look at the possibility that all this time and effort is not accomplishing anything in the way of protecting the public.

If there is no evidence supporting the effectiveness of the review process, we really should ask how much ineffective regulation we are willing to impose on researchers. If we are not even interested in measuring incident rate as it actually occurs in the experiments, in terms that would satisfy accountants, tax auditors, insurance actuaries, and the VP-Finance, why not?

Ceci et al. (1985) established 20 years ago that "socially sensitive" issues were troublesome for review boards, even though the guidelines of that day explicitly forbade using social criteria. Things likely have not improved in this regard as the review criteria have become more nebulous. The questions remain: What are we really trying to achieve here? How can we document success and failure?

Identifying alleged problems does not indicate that the ethics review process is successful at avoiding incidents (real or imagined) in the experimental setting. If hazard avoidance is the goal, a declining problem-arising rate in actual experiments is the only valid measure of success. We are aware that the measures that we have noted here have short-comings, but the purpose in raising them was to underscore the need to acknowledge and pursue the question of accountability for the proposal review process.

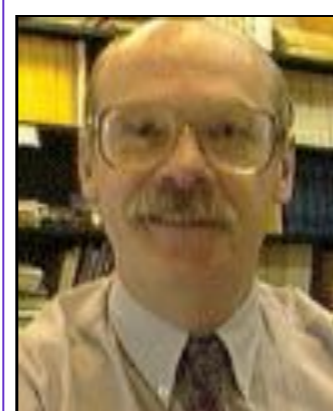
REFERENCES

Ceci, S. J., Peters, D., & Plotkin, J. (1985). Human subjects review, personal values, and the regulation of social science research. *American Psychologist*, 40, 994-1002.

Eaton, W. O. (1983). The reliability of ethics reviews: Some initial findings. *Canadian Psychologist*, 24, 269-270.

Peters, D., Ceci, S. (1982). Peer-review practices of psychological journals: The fate of submitted articles, submitted again. *Behavioral and Brain Sciences*. 5, 187-255.

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (1998, January). URL: <http://www.nserc.ca/programs/ethics/english/policy.htm>



JOHN H. MUELLER, a Fellow of APS, received his PhD in experimental psychology from St. Louis University in 1968, served on the faculty at the University of Missouri, and has been at the University of Calgary since 1990, where he is a professor in the division of applied psychology. His work has been in the areas of human learning, social cognition, and educational technology, with additional interests in the history of psychology and issues involving academic freedom. You can reach him via e-mail: mueller@ucalgary.ca.



JOHN J. FUREDY, a Fellow of APS, received his PhD in Psychology from the University of Sydney in 1965, and has been at the University of Toronto since 1967. His work has been in psychophysiology, philosophy of science, and sex differences in human and animal cognition. He also is engaged in advocacy, debate and scholarship concerning issues such as individual merit versus group-determined identity in the evaluation of faculty and students, and university autonomy, and on conceptual and empirical research regarding education vs. indoctrination in higher education. You can reach him at furedy@psych.utoronto.ca.