

Finding the Missing Science: The Fate of Studies Submitted for Review by a Human Subjects Committee

Harris Cooper
University of Missouri—Columbia

Kristina DeNeve
Baylor University

Kelly Charlton
University of Missouri—Columbia

Publication bias, including prejudice against the null hypothesis, and other biasing filters may operate on researchers as well as journal editors and reviewers. A survey asked 33 psychology researchers to describe the fate of 159 studies approved by their departmental human subjects review committee. About two thirds of completed studies did not result in published summaries. About half of the unpublished studies fell out of the process for reasons other than methodological quality. Among these, lack of interest and aims that did not include publication were cited more often than nonsignificant results as the reasons why publication was not pursued. However, significant findings were more likely than nonsignificant findings to be submitted for meeting presentation or publication. These results indicate attention needs to be paid to improving how psychological scientists communicate, especially to the creation of prospective research registers.

The term *publication bias* has been given various definitions in the literature on scientific research and communication. For example, Chalmers, Frank, and Reitman (1990) defined publication bias as “basing acceptance or rejection of a manuscript [submitted for publication] on whether it supports the treatment tested” (p. 1392). Begg (1994) defined the term more broadly to include “selective publication, in which the decision to publish is influenced by the results of the study” (p. 400). Thus, Begg’s definition permits publication bias both for and against a treatment. It also allows for the type of bias most often associated with the publication process, that is, bias in favor of statistically significant results, regardless of their direction.

Harris Cooper and Kelly Charlton, Department of Psychology, University of Missouri—Columbia; Kristina DeNeve, Department of Psychology, Baylor University.

Support for this research was provided by the Center for Research in Social Behavior, University of Missouri—Columbia.

Correspondence concerning this article should be addressed to Harris Cooper, Department of Psychology, McAlester Hall, University of Missouri, Columbia, Missouri 65211. Electronic mail may be sent via Internet to psychhc@showme.missouri.edu.

Publication bias is of concern to the scientific community because it acts as a filter of research findings that is not based on methodological quality. Therefore, the existence of publication bias means that the published literature may not give the best representation of the state of scientific evidence on the particular topic or hypothesis. Instead, the published evidence may (a) overrepresent studies that support the prevailing zeitgeist or (b) overestimate the magnitude of a relation or treatment effect. The latter may occur because prejudice against the null hypothesis leads to the systematic exclusion from publication of studies that revealed small magnitudes of effect due to chance or sampling error alone.

Greenwald (1975) demonstrated that prejudice against the null hypothesis is not restricted to journal reviewers and editors only. In a survey of authors of journal articles, he found that 41% said they would submit a study for publication if the initial full-scale test of the focal hypothesis permitted rejection of the null hypothesis. Only 5% said they would submit a similar test that did not permit rejection of the null. Thus, Greenwald’s survey suggests that some biasing filters may be operating on the originating researcher, not just on the participants in the journal review.

This article presents the results of a survey of psychology researchers who were asked to describe the

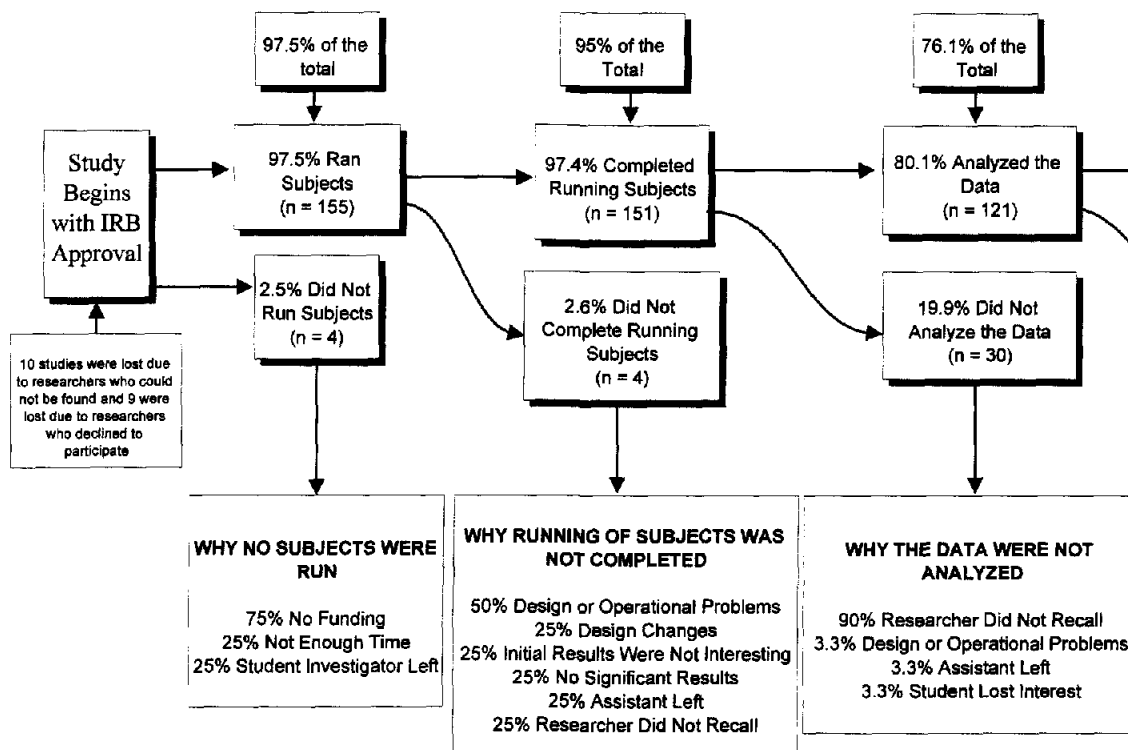


Figure 1. Flow diagram of the fate of instructional review board (IRB)-approved research projects.

fate of research projects submitted for approval to their departmental human subjects review committee or institutional review board (IRB). Through a series of questions, we hoped to discover many of the reasons why planned research is not completed and, especially, why completed research is not submitted for publication.

Method

Sample

A total of 178 studies approved by the Department of Psychology Human Subjects Committee at a large, midwestern, research-oriented, state university during the academic years 1986–1987 and 1987–1988 were eligible for inclusion in the study.

Investigator Contact and Interview

Each form requesting access to human subjects specified a faculty sponsor and an investigator, who could have been the faculty sponsor or a graduate or undergraduate student.

Telephone contact was first made with the faculty sponsor of each proposed investigation. Faculty who had departed the university were contacted at their

new institutions. During the initial phone contact, faculty sponsors were informed of the purpose of the study and were asked if they would be willing to be interviewed about the history of the research project (or projects) of interest. If the faculty member agreed, they were first asked to provide any written summaries of the research and were provided by us with copies of the human subjects review forms they originally submitted for approval. A time was then set up for the formal interview.¹

In several instances, the faculty sponsor suggested that they were no longer cognizant of the progress of the study and suggested we get in touch with the student who served as the investigator. Either the faculty sponsor or the department office provided addresses and phone numbers of students. The student investigators were then contacted.

Using this procedure, we were able to gain the cooperation of 33 faculty sponsors or investigators for

¹ Self-reports of reasons for the noncontinuation of studies can be influenced by both self-enhancing attributions for failure and concerns about self-presentation. The reader should keep these processes in mind as the results of our survey are interpreted.

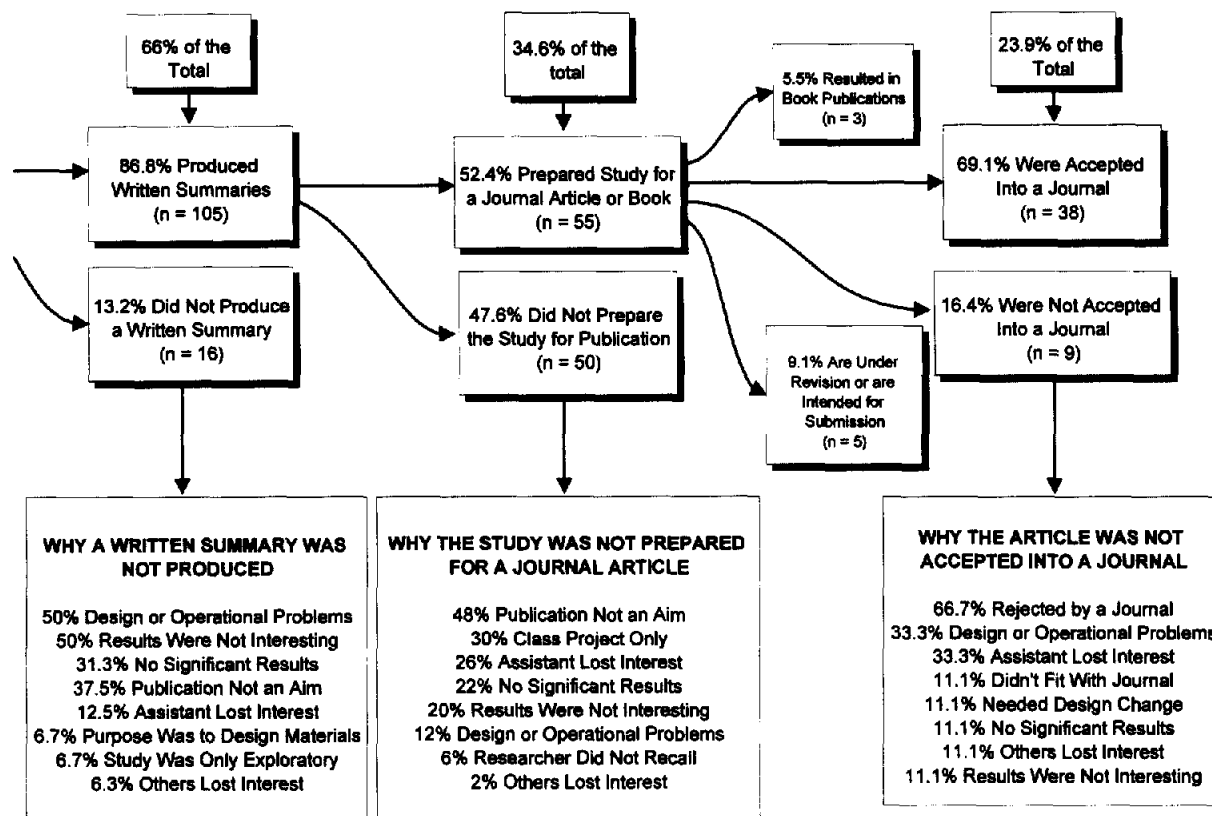


Figure 1 (continued). Flow diagram of the fate of institutional review board (IRB)-approved research projects.

159, or 89%, of the 178 studies. Twelve investigators for the remaining 19 studies could not be found or refused to help. In the statistics that follow, we will treat the 159 studies as the entire sample.

Information Collected About Each Study Approved by the IRB

Information about each study was obtained from three possible sources: (a) the original human subjects approval form, (b) the interview with the faculty sponsor or student investigator, and (c) the written study summaries. Considerably more information was obtained on each study than was eventually used in the analyses. For purposes of brevity, the study characteristics of interest will be described below as the results associated with them are presented.

Results

Description of the IRB Database

The investigators for 26 studies (17%) were undergraduates; investigators for 31 studies (20%) were graduate students; and for 102 studies (64%), the in-

vestigators were faculty members. About one third (33%, n = 51) of the studies were undertaken as class projects, another third (34%, n = 52) as theses or dissertations, and one quarter as independent research projects (25%, n = 39). The rest, about 8% (n = 12), were conducted as pilot tests for larger studies. Over 90% of studies were conducted without outside funding.

The Fate of Approved Studies

Figure 1 presents a flow diagram charting the fate of each study that gained approval from the IRB. Of the 159 studies that obtained approval, only 3% (n = 4) never commenced running subjects. In open-ended responses to a query about why no subjects were run, investigators cited a lack of funds in three instances, a lack of time in one instance, and the departure of the student investigator in one instance.

Once subjects began to be run, there was a very high probability that the project continued until the investigator felt that the data collection was complete. In only 3% (n = 4) of the cases did the running of subjects cease prior to the study's completion. Inves-

tigators were allowed to give more than one reason for why they did not complete the data collection. Design or operational problems were cited in two cases for halting the study at this juncture. Investigators once mentioned design changes, the departure of assistants, or both. Finally, investigators said in two instances that, based on analyses conducted on a portion of their data, the initial results were either uninteresting or nonsignificant.

A considerably larger portion of investigators completed the data collection process but never undertook formal statistical analysis. Of the 151 studies that completed running subjects, 20% ($n = 30$) of the data sets were never analyzed. Curiously, most researchers ($n = 27$) could not recall why this was so. We suspected that the majority of these cases might have involved research undertaken by undergraduates as part of a course requirement, rather than as part of the cognizant faculty members ongoing research program. A crossbreak analysis bore out this suspicion; 27 of the 30 unanalyzed data sets had been generated as undergraduate class projects. Thus, from this point on, the flow diagram data are relatively free of research undertaken as a course requirement.

Thus far, Figure 1 indicates that 76% ($n = 121$) of all approved studies were carried out at least until a statistical test was conducted on the main hypothesis. We could argue that this set of 121 studies constitutes the most unbiased representation of the support, or lack thereof, for the main hypotheses tested in psychological studies at this institution. That is, once investigators analyze their data we can reasonably assume they have determined the test is credible and they are ready to "throw the dice." Thus, the completion of data analysis can be viewed as the point at which investigator self-filters and publication bias begin to influence which hypothesis tests are and are not available through traditional retrieval mechanisms.

Written summaries were produced for 87% ($n = 105$) of the studies that had complete data analysis. For 8 of the 16 remaining studies (50%), the investigator claimed that design or operational problems were responsible for not describing the results in some text form. In fact, if this rationale is true then, contrary to our above reasoning, these 8 studies ought not be included in any estimate of the bias that might exist in published summaries because these data could not be considered credible tests of hypotheses.

However, there is reason to suggest that the decision not to write-up the eight data sets that were analyzed but were reported to have design or operational

flaws might also have been influenced by nonmethodological filters of study results. Conceptually, it is possible that design or operational flaws present in a study would not be evident to the investigator until after the conduct of the data analysis. It is also possible that the concern about flaws in these studies was at least in part generated by concern about the results of the statistical test. To test this possibility, we examined whether investigators who cited design flaws at this juncture also cited other reasons for not preparing a written summary. We found that in all eight instances where investigators cited flaws they also cited at least one of the nonmethodological reasons for not preparing a written summary. In two instances no significant results were mentioned, in three instances uninteresting results were mentioned, in one instance publication not being an aim was mentioned, in one instance nonsignificant and uninteresting results were mentioned, and in one instance all three nonmethodological reasons were mentioned.

For the eight studies not mentioning design flaws, the nonmethodological reasons given primarily related to a loss of interest, nonsignificant findings, and publication not being an aim. Most researchers would consider these to be troublesome filters for whether a study's results were obtainable when an assessment of the evidence was undertaken.

Of the 105 studies for which a written summary existed, just over half (52%, $n = 55$) were prepared for submission as a journal article or book chapter. For the 50 studies not prepared for publication, six investigators cited design or operational problems as the cause. In four of these instances, nonsignificance, loss of interest, or publication not being an aim were also listed as causes.

Finally, there were nine articles submitted for publication that had been withdrawn entirely from the review process. In four of these instances the investigator cited as the cause design or operations problems.

In sum then, 76% of the 159 studies that obtained human subjects approval produced a test of the main hypothesis. Of the 121 studies that resulted in a statistical test, 34% ($n = 41$) were ultimately published in a journal or book and 4% ($n = 5$) were still under revision or otherwise in process. About 15% ($n = 18$) were withdrawn because of methodological flaws, although in most of these instances nonmethodological reasons were cited as well. The remaining 47% ($n = 57$) were withdrawn from the publication process for strictly nonmethodological reasons.

For exploratory purposes, we examined whether several characteristics of studies were related to (a) whether the study ever resulted in a written document and (b) whether the written document was ever published.

Studies that produced written documents were significantly less likely to be initiated as class projects, $r(158) = -.28, p < .0003$, more likely to be initiated by a graduate student to obtain a master's or doctoral degree, $r(158) = .47, p < .0001$, more likely to have run a greater number of subjects, $r(158) = .18, p < .03$, and more likely to have been based on a questionnaire rather than an experimental or quasi-experimental design, $r(158) = .20, p < .02$. Studies that were eventually published were significantly less likely to be initiated as class projects, $r(158) = -.27, p < .0007$, and more likely to be initiated as independent research projects, $r(158) = .28, p < .0003$.

Because we were concerned that many of these results were influenced by the presence in our sample of studies undertaken as class projects or pilots, we reran the analyses without including studies undertaken for these purposes. None of the above results were replicated using this subsample. Thus, beyond the purpose of the study (i.e., class projects vs. other purposes), there appeared to be no information available on human subjects approval forms that predicted whether a proposed study would eventually result in a written summary or publication.

Evidence for Bias Against the Null Hypothesis

Finally, we examined whether studies that produced significant results were more likely to be submitted for presentation at a professional meeting, for publication in a journal or book, or both.

Table 1 presents the results of these analyses. Of the 117 studies that were carried out through data analysis and used significance testing, 72 found sig-

nificant results on either the main hypothesis, other hypotheses, or both.² The analyses indicated that half of studies with significant results were submitted for presentation at a professional meeting, while only 7% of nonsignificant results were submitted, $\chi^2(1, N = 117) = 23.40, p < .0001, r = .45$. Even more dramatically, 74% of studies with significant results were submitted for publication in a journal or book chapter while only 4% of nonsignificant results were submitted, $\chi^2(1, N = 117) = 53.18, p < .0001, r = .67$.

Discussion

This survey indicates that, at least at one research-oriented university, about two thirds of psychological studies that received IRB approval and that performed formal statistical tests of the main hypotheses did not result in published summaries. About half of the unpublished studies fell out of the process for reasons independent of methodological quality. Among these, a lack of interest and aims other than publication were cited more often than nonsignificant results as the reasons why a written document or publication was not pursued. Concurrently however, significant findings were dramatically more likely than nonsignificant ones to be submitted for presentation at professional meetings or for publication.

Thus, this study confirms the existence of filters in the research process other than bias against the null hypothesis that act to select studies out of the publication process. Although bias against the null hypothesis was clearly found to exist, half of significant results are not submitted for presentation and about a quarter are not submitted for publication. Therefore, statistical corrections for publication bias based solely on assumptions about unpublished null findings cannot be relied on to completely restore estimates of relationship strengths to the value that would be ob-

Table 1
Frequency of Studies With Significant and Nonsignificant Results That Were Submitted for Presentation at Professional Meetings or for Publication

Submission	Results	
	Significant	Not significant
Publication and presentation	26	1
Publication only	27	1
Presentation only	10	2
None	9	41

² These analyses assume that the decision about whether to submit each study is an independent event. This assumption may not strictly be true in that there were fewer investigators surveyed than there were studies. Hypothetical analyses that reduced the number of cases to the number of surveyed researchers ($n = 33$) but retained the cell proportions revealed significant results similar to those reported in the text. We hoped to look at whether significance on the main hypothesis of a study led to different submission decisions from significance on other hypotheses. However, too few researchers ($n = 11$) reported they obtained significance on only other hypotheses.

tained if access to every credible test of a given hypothesis were possible.

Our research, and numerous projects before it, point out many serious flaws that exist in the systems that psychological researchers use to communicate with one another. One partial solution to the problem involves the creation of research registers. Research registers are databases that catalogue studies from the time they are initiated rather than from when they are published.

In medicine, research registers have been proposed as an adjunct to reference databases of published studies, such as PsycINFO or Medline, as the starting point for a literature search when a review of the state of evidence on a particular hypothesis is undertaken. Because research registers do not screen the studies entered in them on the basis of results, publication bias cannot affect their contents. For example, the International Cancer Research Data Bank (ICRDB) contains the majority of cancer clinical trials funded by the National Institute of Health as well as other trials from outside the United States. Simes (1986) showed that the effects of cancer treatments based on retrieval of studies through the ICRDB led to smaller estimates of effect than retrieval of published studies using Medline.

Few research registers exist in psychology. However, although they are rarely thought of in this context, the human subjects approval forms that provided the data for this study can be viewed as forming a research register. IRBs and their departmental counterparts are mandated by law to review and approve most psychology research before it is initiated. Psychology-related IRBs routinely collect information that, if made accessible to the broader scientific community, could help surveyors of the research obtain a more representative assessment of an empirical literature. IRB registers could be used to survey ongoing research projects and the current state of evidence on a particular psychological relation or treatment.

The results of this study point out the potential benefits that might accrue if IRBs were used as registers for psychological research. However, although IRB databases may address the problem of publication bias and other research filters, they are not with-

out their own difficulties. Primarily, IRBs encompass research in only isolated parts of a single institution. Therefore, to be truly effective a database of multiple IRBs, perhaps national or international in scope, would have to be constructed before the database could produce searches of research activities more trustworthy than the techniques presently available.

An attempt to integrate multiple IRBs to form a more encompassing research register of psychological studies would be expensive and time consuming. Such an effort would need to focus on rules for the standardization of information, keywording, and protection of research priority and mechanisms for updating and accessing files. However, recent improvements in computer and networking technology have made registers of federally funded research projects accessible to individuals searching for information at desktop personal computers.

It is critical to the credibility of our science that we focus serious professional attention on improving our research-retrieval systems. An international, standardized, computer-accessible register of IRB-approved research projects, with all its problems and pitfalls, could be an important tool for helping surveyors of research find the missing science.

References

- Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Foundation.
- Chalmers, T. C., Frank, C. S., & Reitman, D. (1990). Minimizing the three stages of publication bias. *Journal of the American Medical Association*, *263*, 1392-1395.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1-20.
- Simes, R. J. (1986). Publication bias: The case for an international registry of clinical trials. *Journal of Clinical Oncology*, *4*, 1529-1541.

Received July 25, 1996

Revision received February 10, 1997

Accepted February 19, 1997 ■