Annals of Internal Medicine | Academia and Clinic

Reproducible Research: Moving toward Research the Public Can Really Trust

Christine Laine, MD, MPH; Steven N. Goodman, MD, PhD, MHS; Michael E. Griswold, PhD; and Harold C. Sox, MD

A community of scientists arrives at the truth by independently verifying new observations. In this time-honored process, journals serve 2 principal functions: evaluative and editorial. In their evaluative function, they winnow out research that is unlikely to stand up to independent verification; this task is accomplished by peer review. In their editorial function, they try to ensure transparent (by which we mean clear, complete, and unambiguous) and objective descriptions of the research. Both the evaluative and editorial functions go largely unnoticed by the public—the former only draws public attention when a journal publishes fraudulent research. However, both play a critical role in the progress of science. This paper is about both functions. We describe the evaluative processes we use and announce a new policy to help the scientific community evaluate, and build upon, the research findings that we publish.

Ann Intern Med. 2007;146:450-453. www.annals.org
For author affiliations, see end of text.

The context for our paper is every journal's nightmare: publishing fraudulent research. Cases of scientific misconduct have surfaced with alarming frequency, tarnishing public trust in journals, researchers, and science itself. To much fanfare, the journal *Science* published promising developments in stem cell research only to learn months later that the findings were bogus (1). *Lancet* editors retracted a paper on oral cancer when the editors learned that the researchers had fabricated their data (2). The *New England Journal of Medicine* published a statement of concern on a drug toxicity study when evidence suggested that the authors had withheld important adverse outcomes (3). Incidents of authors who have not disclosed potential conflicts of interest have prompted *JAMA* to reinforce its disclosure policies (4–6). Here at *Annals*, we retracted an article after learning that the lead author had fabricated the data. (7–9). In the midst of these accumulating scandals, a *New York Times* article declared, "Reporters Find Science Journals Harder to Trust, but not Easy to Verify" (10). Reporters are not alone; journal editors also find biomedical research harder to trust and not easy to verify.

Fraud gets headlines and drives editors to despair but fortunately occurs rarely. We at *Annals* spend much of our time evaluating the validity and generalizability of every article that we consider for publication. Like our colleagues at other biomedical journals, we strive to assure ourselves and our readers that the information we publish is accurate. Yet, because we do not personally witness every step of the research process, we can never be entirely certain of validity. We are less concerned with outright fraud, which can elude even the most astute reviewers and editors, than with the subtler but much more prevalent biases that are present in one form or another in most clinical research. Prompted by both the recent wave of scientific misconduct and our own experience in trying to minimize the impact of bias in our pages, we review here how we protect against publishing research that is misleading or outright wrong.

Existing Strategies to Guard against Research Impropriety

Annals relies on a variety of processes to evaluate the validity of work that we consider for publication (**Table**). These processes span the life of the article from submission through postpublication.

To facilitate the detection of potential sources of bias or threats to validity, *Annals* requires authors to include several pieces of information with submitted manuscripts. They must disclose the contributions and potential conflicts of interest of all authors on the byline, including but not limited to any financial relationship that involves diseases, tests, or treatments discussed in the manuscript or competing products. This alerts editors, reviewers, and readers to situations in which competing interests might bias content or threaten validity (11). Authors know that this information will be public, which we hope increases their sense of public accountability for unbiased reporting.

We also require authors of clinical trials to register the trials (12) and submit copies of the trial protocol. These requirements give editors access to the detailed, prespecified research plan, which can be useful if the manuscript lacks key details or when we need to reconcile what the researchers planned with what they reported.

Annals' next layer of defense against publishing erroneous information is a multilayered review that includes scrutiny by both journal staff and external reviewers. We obtain external review on about 50% of submitted manuscripts, turning to the more than 12 000 individuals in our reviewer database to augment the range of skills and knowledge of our editors. Although the peer-review process is not perfect, these reviews provide important information

See also:

Web-Only
Conversion of table into slide

Table. **Existing and New Strategies That *Annals* Uses to Guard against the Publication of Biased or Invalid Research**

Strategy	Purpose
Prepublication	
Conflict disclosures	Alerts editors and reviewers to potential sources of bias
Peer review	Provides an opportunity for content experts to examine the work for potential threats to validity
Protocol submission	Enables editors to reconcile what researchers planned with what they report
Statistical review by an in-house team (may include requests for data and alternative analyses)	Enables independent statisticians to critique analysis and look for threats to validity
Trial registration	Permits public access to research plans; discourages suppression of unfavorable results
Publication	
Publication of conflicts of interest	Alerts readers to potential sources of bias
Publication of author contributions	Establishes accountability for specific components of the research process
Publication of detailed methods	Enables readers to critique methods and look for threats to validity
Postpublication	
Letters to the editor/rapid responses	Provides a venue for readers to air their concerns
New safeguards	
Publication of information about availability of protocol, statistical code, and data	Increases potential for reproducibility, allows greater scrutiny for potential threats to validity Permits confirmation of results by independent individuals

that we weigh during the decision-making process; reviewers influence our decisions on a daily basis by identifying problems that we missed. Many authors believe (incorrectly) that reviewers' opinions always dictate the editorial decision. Sometimes, the external reviewer's opinion is decisive; however, usually the reviewer's comments are incorporated into the insights and judgments of our own editorial team, which includes scientists and clinicians with a range of methodological and clinical skills. We then make a preliminary decision at an editorial conference, during which we discuss all papers under consideration by the associate and deputy editors. After the preliminary decision, the journal's processes are a mix of evaluative and editorial functions, with the editorial function taking on greater importance as we near final acceptance of the manuscript.

We never rely solely on the statistical expertise of the authors. Any papers selected for possible publication must pass through yet another filter, a "methods" conference in which *Annals*' 6 statistical editors, 4 deputy editors, and editor-in-chief scrutinize a study's methodology. The statistician assigned to the manuscript presents the main statistical issues for discussion at this conference. Often, we ask for statistical code or data to clarify or validate methods. We may ask authors to provide supplementary analyses to assess the sensitivity of results to different assumptions or methods, work with authors to improve analyses, or on occasion, perform our own analyses of the data with the authors' cooperation. Usually, the revised paper reflects these clarifications. These discussions, or subsequent inquiries, occasionally diminish our confidence in the work to the point that we choose not to publish it. More often, they improve the clarity and completeness of the manuscript, making the strengths and weaknesses of the study more evident to the reader and ensuring that the tone of the conclusions matches the strength of the evidence. These outcomes of editing and peer review have been documented in empirical studies at *Annals* (13) and, more recently, in a preliminary report of a study that included several journals (14).

Arguably, the most important function of prepublication peer review and editing is to make effective postpublication peer review possible. The most stringent internal processes cannot guarantee truth in reporting or reliably detect fraud; however, they can provide the scientific community with the tools to assess the strength and validity of published evidence. We strive to publish methods sections that provide sufficient detail so that readers can assess all the potential threats to validity. Reporting standards like CONSORT for randomized, controlled trials (15, 16) also facilitate thorough reporting of study methodology. When the methods are transparent, fraud, bias, and weak research methods are easier to spot. In this age of instant electronic dissemination and reader feedback, features like rapid response letters to the editors provide a medium for postpublication peer review and discussion of the research.

Reproducible Research: A New Strategy to Increase Confidence in Published Research

Even with these safeguards, experience teaches us that we must seek more innovative strategies to promote truth in publication. Independent replication by independent scientists in independent settings provides the best assurance that a scientific finding is valid; however, the resources and time required for high-quality clinical studies makes literal replication of published studies a slow corrective to any errors in the original publication. However, scientists and journal editors can promote "reproducible research," described recently by Peng and colleagues (17). The term

has its roots in highly technical forms of statistical programming and data presentation, but its requirement for sharing data and analytic code is applicable to clinical research. Reproducibility involves methods to ensure that independent scientists can reproduce published results by using the same procedures and data as the original investigators. It also requires that the primary investigators share their data and methodological details. These include, at a minimum, the original protocol, the dataset used for the analysis, and the computer code used to produce the results.

Reproducibility requires a level of openness that is rare in the current competitive environment of biomedical research. Investigators might be understandably concerned that sharing data would make them vulnerable to unwanted scrutiny or enable other researchers to make discoveries using the data that took them years to assemble. Recognizing this, Peng and colleagues describe forms of data sharing that are between unlimited access and no access (17). Their proposals, adapted in part from commercial models of intellectual property protection, involve limitations on the permitted use of the data set, as set forth in a license or written agreement between an investigator and a requestor. At one extreme, researchers could only have access to portions of the analytic data set and solely for the purpose of reproducing published results; at the other extreme, researchers could have unrestricted access to the entire file of raw data. Programs incorporated into commercial software that create technical barriers to unauthorized use or widespread dissemination could also protect authors' interests.

We believe that the reproducible research model is important. By strengthening the scientific process, it can also protect the public against any long-term adverse effects of research gone wrong. For this reason, *Annals* is launching a "reproducible research" initiative. Every original research article will include a statement that indicates whether the study protocol, data, or statistical code is available to readers and under what terms authors will share this information. Sharing will not be mandatory, but we will require authors to state whether they are willing to share the protocol, data, or statistical code. This new policy will apply to all articles containing original research (articles and brief communications) submitted on or after 1 April 2007. We hope that willingness to share these elements will promote trust in the authors' findings, similar to what has happened with our publishing of conflict disclosures, funding sources, and author contributions.

Whether authors make these materials available to the public will not influence *Annals'* editorial decisions. To assure authors of our good faith, we will not inquire about an author's willingness to make this information available until after we have accepted the manuscript for publication. Per our current practice, we will continue to ask authors to make these items available to the editors during the review process whenever necessary. We emphasize that the main purpose of this new initiative is to promote the scientific process, not to deter fraud. A recently published example at *Annals* documents how the sharing of data resulted in valuable insights about the findings of a highly visible study published several years ago (18).

The inclusion of statistical code in this initiative may also serve to remind investigators to keep an exact record of the statistical procedures performed and the portions of the data set that the authors analyzed with statistical software. Most statistical software packages have some method for storing the analysis code, which facilitates examination by statistical editors or other researchers and modification of the analysis for subsequent projects. Universal data archiving is an essential basic component of research accountability that researchers often neglect. We urge all authors, even those who are reluctant to share their data under any terms, to archive their data. Even those who never intend to share their data may need to address new findings or concerns that surface long after publication. Too often, researchers say that they are unable to respond to postpublication inquiries because the data are no longer available. This response weakens confidence in the original work, serves the needs of medical science poorly, and betrays the trust of the patients who agreed to participate in the research.

Annals is not taking this initiative in a vacuum. Other developments show that scientists recognize the value of transparency and reproducibility. Our colleagues at *JAMA* have established a policy requiring an independent academic statistician to corroborate analyses of industry-funded studies (19). In their recommendations to the editors of *Science*, a committee that the journal assembled to examine the process involved in the publication of the fraudulent stem cell papers by Hwang and colleagues states, "Primary data are essential and should be available to reviewers and readers. The General Information for authors should be modified to make it clear that, for example, requests for materials, methods, or data necessary to verify the conclusions may be required prior to acceptance (20)." In the basic sciences, the journal *Nature* requires protocol and data sharing (21), and most journals publishing bioinformatics information (for example, genetic sequences) require authors to deposit their data in public databanks when possible (22). In addition, the National Institutes of Health (NIH) requires data sharing plans for all large projects (23), and the Shelby amendment requires sharing health research data used as the basis for law or policy. Some NIH institutes have established their own data sharing policies (24). The recognition that reproducibility of research can be a defense against bias and scientific misconduct and a check on the robustness of conclusions is clearly growing among the institutions on which the public relies to conduct and disseminate medical research.

Major cultural shifts in research must occur before a world of completely reproducible research can exist. These shifts include increasing the technical capacity of many research teams, further developing acceptable data sharing

mechanisms, and supporting—both professionally and financially—the publishing of reproducible research. We believe that small demonstrations of the power of open sharing of methodological details and data may motivate such shifts in the clinical research culture. These changes are most likely to occur if other major biomedical journals join us in this effort. We hope that shining a spotlight on the availability of the study protocol, data, and statistical code for every *Annals* research report will be seen as a small but important step toward biomedical research that the public can really trust. At the same time, it will enhance what is perhaps the main function of a journal: to provide a transparent medium for a conversation about science.

Acknowledgments: The authors thank Drs. Michael Berkwits, Paul Epstein, and Cynthia D. Mulrow for their review of earlier drafts of the manuscript.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Customer Service, American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.

Current author addresses are available at www.annals.org.

References

1. **Kennedy D.** Editorial expression of concern [Letter]. Science. 2006;311:36. [PMID: 16373531]
2. **Horton R.** Retraction—Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. Lancet. 2006;367:382. [PMID: 16458751]
3. **Curfman GD, Morrissey S, Drazen JM.** Expression of concern reaffirmed [Editorial]. N Engl J Med. 2006;354:1193. [PMID: 16495386]
4. **Kurth T, Gaziano JM, Cook NR, Logroscino G, Diener HC, Buring JE.** Unreported financial disclosures in a study of migraine and cardiovascular disease [Letter]. JAMA. 2006;296:653-4. [PMID: 16896106]
5. **DeAngelis CD.** Unreported financial disclosures in a study of migraine and cardiovascular disease [Letter; author reply]. JAMA. 2006;296:654. [PMID: 16896106]
6. **Flanagin A, Fontanarosa PB, DeAngelis CD.** Update on JAMA's conflict of interest policy [Editorial]. JAMA. 2006;296:220-1. [PMID: 16835430]
7. **Garrel DR, Rabasa-Lhoret R, Boyle P.** Preventing scientific fraud [Letter]. Ann Intern Med. 2006;145:473; author reply 473-4. [PMID: 16983141]
8. **Poehlman ET.** Notice of retraction: final resolution [Letter]. Ann Intern Med. 2005;142:798. [PMID: 15845951]
9. **Sox HC.** Notice of retraction: final resolution [Letter]. Ann Intern Med. 2005;142:798. [PMID: 15845952]
10. **Bosman J.** Reporters find science journals harder to trust, but not easy to verify. New York Times. 13 February 2006.
11. **Davidoff F, DeAngelis CD, Drazen JM, Hoey J, Højgaard L, Horton R, et al.** Sponsorship, authorship, and accountability [Editorial]. Ann Intern Med. 2001;135:463-6. [PMID: 11560460]
12. **De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al.** Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors [Editorial]. Ann Intern Med. 2005;143:146-8. [PMID: 16027458]
13. **Goodman SN, Berlin J, Fletcher SW, Fletcher RH.** Manuscript quality before and after peer review and editing at Annals of Internal Medicine. Ann Intern Med. 1994;121:11-21. [PMID: 8198342]
14. **Lee KP, Boyd EA, Bero LA.** Editorial changes to manuscripts published in major biomedical journals. Accessed at www.ama-assn.org/public/peer/abstracts.html#peer on 25 January 2007.
15. **CONSORT GROUP (Consolidated Standards of Reporting Trials).** The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med. 2001;134:657-62. [PMID: 11304106]
16. **CONSORT GROUP (Consolidated Standards of Reporting Trials).** The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663-94. [PMID: 11304107]
17. **Peng RD, Dominici F, Zeger SL.** Reproducible epidemiologic research. Am J Epidemiol. 2006;163:783-9. [PMID: 16510544]
18. **Sylvestre MP, Huszti E, Hanley JA.** Do OSCAR winners live longer than less successful peers? A reanalysis of the evidence. Ann Intern Med. 2006;145:361-3; discussion 392. [PMID: 16954361]
19. **Fontanarosa PB, Flanagin A, DeAngelis CD.** Reporting conflicts of interest, financial aspects of research, and role of sponsors in funded studies [Editorial]. JAMA. 2005;294:110-1. [PMID: 15998899]
20. **Kennedy D.** Responding to fraud [Editorial]. Science. 2006;314:1353. [PMID: 17138870]
21. **Nature editorial policies.** London and New York: Nature. Accessed at www.nature.com/nature/authors/editorial_policies/index.html on 25 January 2007.
22. **Journal of the National Cancer Institute instructions to authors.** Bethesda, MD: Journal of the National Cancer Institute. Accessed at www.oxfordjournals.org/our_journals/jnci/for_authors/ on 25 January 2007.
23. **NIH data sharing policy.** Bethesda, MD: National Institutes of Health Office of Extramural Research. Accessed at http://grants.nih.gov/grants/policy/data_sharing/ on 25 January 2007.
24. **National Cancer Institute data sharing policy.** Bethesda, MD: National Cancer Institute Accessed at ctep.cancer.gov/guidelines/data_sharing_policy.pdf on 25 January 2007.

Current Author Addresses: Drs. Laine and Sox: Customer Service, American College of Physicians, 190 N. Independence Mall West, Philadelphia, PA 19106.

Dr. Goodman: Department of Oncology, Division of Biostatistics, Johns Hopkins University, 550 North Broadway, Suite 1103, Baltimore, MD 21209.

Dr. Griswold: Welch Center, Johns Hopkins University, 2024 E. Monument Street, Baltimore, MD 21205.