

Are journals doing enough to prevent fraudulent publication?

Recent warnings by editors of 3 major journals that data contained in published papers were or may have been incomplete,¹ falsified² or fabricated³ has dismayed scientists and scientific editors around the world and added to the public's growing scepticism about the authority of science. How is it that flawed or fraudulent research can slip through the net of peer review and editorial scrutiny?

Reputable scientific journals use a systematic approach to reviewing and editing research papers. At *CMAJ*, submissions that are not intercepted after an initial screening for suitability and relevance are sent for peer review. Reviewers are chosen on the basis of their interest and expertise, publication record, and quality of previous reviews. Peer reviewers devote perhaps a few hours to reading the paper, consulting the existing literature and writing their review. About 20% of the completed reviews we receive are rated as excellent; we generally succeed in obtaining 2 "good" or "excellent" reviews for each manuscript.

After peer review, submissions are carefully reassessed by the scientific editors, and about 6% are selected for publication. Almost all require substantive editing, guided by a scientific editor working closely with the authors. Once this process of revision is complete, the manuscript is copyedited for clarity, precision, coherence and house style. Problems with the presentation and interpretation of data can come to light at any point in this process, even at the late stage of copyediting.

For the most part, this intensive series of editorial checkpoints works well. But it is not perfect. In 2005, PubMed received 67 notices of article retractions (Sheldon Kotzin, National Library of Medicine; personal communication, 2006.) This is undoubtedly an underestimate of the total number of flawed, grossly misleading or frankly fraudulent papers.

Editors (and peer reviewers) work from the submitted manuscript along with any other material supplied by the authors (e.g., survey instruments or additional tables, graphs and figures). In assessing randomized clinical trials, most editors examine the study protocol to try to ensure that the study report reflects the planned design and analysis. However, it is almost impossible to detect by these processes whether data have been fabricated, or if key elements are missing. Editors, particularly of general journals, rarely have the expertise in the particular topic of the research to enable them even to suspect fabrication when it occurs. Reviewers may have the expertise but not necessarily the time to examine findings in exhaustive detail; moreover, they can assess only those data that the authors actually disclose.

Alarmed by their own experiences with particular manuscripts, some journals are taking further steps to ensure that

authors are faithful to their data. For example, the *Journal of the American Medical Association (JAMA)* now requires independent statistical re-analysis of the entire raw data set of any industry-sponsored study in which the data analysis has been conducted by a statistician employed by the sponsoring company.⁴ The *Journal of Cell Biology* (www.jcb.org) has specific policies prohibiting the enhancement of images and scrutinizes submitted images for evidence of manipulation. It will be important to evaluate the effectiveness of these measures as time goes on, since their costs in time and resources are not trivial.

At *CMAJ* we are contemplating the steps that would be required to allow us to make available, as an online-only appendix, the entire data set on which a research paper is based. Doing so would enable more intensive post-publication peer review. Interested persons with the necessary expertise could confirm the published analyses, conduct further analyses and increase the efficiency of research by making it more widely used. Fraud might also be detected sooner, and perhaps the knowledge that their data set will be open to public scrutiny will deter some authors from fabricating or falsifying data (if it does not make others more clever in their deceits). Current online publishing systems enable authors to readily supplement their articles with data sets in any file format (spreadsheets, databases, jpegs, etc.) and to index these files for proper attribution and with helpful information (e.g., the open source Open Journal Systems at <http://pkp.sfu.ca/ojs>; Dr. John Willinsky, University of British Columbia; personal communication, 2006). The costs of posting additional data as appendices to manuscripts are trivial, and the ethical and legal obstacles (rendering the data anonymous when they involve patients, and protecting the intellectual property rights of investigators and sponsors) can be overcome.⁵

No editorial review system will ever be entirely impermeable to human error or deceit. But journals could do more to ensure the integrity of published scientific results; one place to start might be to publish all of the data on which research findings are based. — *CMAJ*

REFERENCES

1. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med* 2000;343:1520-8.
2. Hwang WS, Roh SI, Lee BC, et al. Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science* 2005;308:1777-83.
3. Sudbø J, Lee JJ, Lippman SM, et al. Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *Lancet* 2005;366:1359-66.
4. Fontanarosa PB, Flanagan A, DeAngelis CD. Reporting conflicts of interest, financial aspects of research and role of sponsors in funded studies. *JAMA* 2005;294:110-1.
5. Walter C, Richards EP. Public access to scientific research data. Available: <http://biotech.law.lsu.edu/IEEE/ieee36.htm> (accessed 2006 Jan 13).