Statistical Reviewing Policies of Medical Journals

Caveat Lector?

Steven N. Goodman, MD, PhD, Douglas G. Altman, DSc, Stephen L. George, PhD

OBJECTIVE: To describe the current policies regarding statistical review of clinical research in biomedical journals.

DESIGN: Cross-sectional survey.

PARTICIPANTS: Editors of biomedical journals that publish original clinical research.

MEASUREMENTS: General policies on statistical review, types of persons used for statistical reviewing, compensation of statistical reviewers, percentage of articles subject to such review, percentage of time statistical review makes an important difference, journal circulation, and selectivity.

MAIN RESULTS: Of 171 journals, 114 (67%) responded to the survey. About one third of journals had policies that guaranteed statistical review for all accepted manuscripts. In approximately half of the journals, articles were sent for statistical review at the discretion of the editor. There was some evidence that statistical review policies differed between journals of different circulation size. In journals in the top quartile of circulation ($> 25,000$) the probability of definitely having a statistical review before an acceptance decision was 52%, but it was only 27% in journals in the lower three quartiles ($p = .09$). The probability of a statistical consultant on staff ranged from 31% in the bottom quarter, to 58% in the middle two, to 82% in the highest quarter ($p < .001$). Editors judged that statistical review resulted in an important change in a manuscript about half of the time.

CONCLUSIONS: Except in the largest circulation medical journals, the probability of formal methodologic review of original clinical research is fairly low. As readers and researchers depend on the journals to assess the validity of the statistical methods and logic used in published reports, this is potentially a serious problem. This situation may exist because the cost of such statistical review can be considerable, and because finding appropriate reviewers can be difficult. It may also exist partly because editors or publishers may not regard such review as important. The professions of medical publishing, statistics, epidemiology, and other quantitative disciplines should work together to address this problem.

KEY WORDS: medical journals; statistical review.

J GEN INTERN MED 1998;13:753-756.

The relative lack of knowledge of most physicians about statistical methods has been repeatedly documented.¹⁻⁵ For many, the de facto standard for statistical methodology is set by the medical journals, which publish research articles making use of a wide range of statistical techniques.⁶ Most clinicians and researchers depend on the journals, via the editorial and peer review processes, to ensure that statistical methods in published articles are being used and interpreted appropriately.

Given this state of affairs, a natural question to ask is whether the journals themselves are using a process that

is adequate to ensure that the statistical methods and data interpretation are appropriate for the articles they publish. George first addressed this question in 1984 by surveying journal editors about their policies regarding specialized statistical review.⁷ This article reports the results of a similar survey of a larger number of medical journals conducted approximately 10 years later.

METHODS

Identification of Journals

Journals were selected for inclusion in the study by a combination of formal and informal criteria. First, all journals publishing medically related research with the 100 highest "impact factors" were selected, using the 1992 Institute for Scientific Information (ISI) journal ranking.⁸ The impact factor measures the number of times articles from a journal are cited in other scientific articles, but does not capture a range of subspecialty journals that have influence on the small community of specialists within a field, but little outside it.⁹ We therefore augmented the impact factor list by consulting several medical librarians in the United States and United Kingdom and examining the approximately 1,500 journals on display at the Welch Library of the Johns Hopkins School of Medicine. The goal of the effort was to create a list that was a representative cross section of the journals that communicate research results to practicing and research physicians. Journals that published no clinical research (i.e., research using human subjects) were not included in the survey.

Survey Instrument

The survey instrument was an expanded version of one used by George,⁷ a one-page form with questions relating to statistical reviewing policy, how reviewers are chosen, the presence of guidelines, and a variety of journal characteristics.

Received from Johns Hopkins University School of Medicine, Division of Biostatistics, Oncology Center, Baltimore, Md. (SNG); Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, England (DGA); and Division of Biometry, Duke University Medical Center, Durham, N.C. (SLG).

Address correspondence and reprint requests to Dr. Goodman: Johns Hopkins University, 550 N. Broadway, Suite 409, Baltimore, MD 21205.

Survey Implementation

The survey was initially sent by mail to the addresses indicated on the journal masthead, with a cover letter signed by the three authors explaining its purpose, and a promise of confidentiality (i.e., individual journal results would not be cited). Nonresponding journals were followed up by mail, fax, and telephone call. The survey process was initiated in July 1993, and last efforts at follow-up were finished in 1995. Respondents were typically the managing editors or editors-in-chief of the journal.

RESULTS

The original list to be surveyed included 171 journals, and 114 (67%) ultimately responded. Of the 57 that did not respond, almost all were small subspecialty journals. Table 1 shows the characteristics of the responding journals. The median journal circulation was a rather modest 10,500, although the range extended to 700,000. The median number of clinical research articles published each year was 120 (60% of 200), representing a median of 40% of the manuscripts submitted.

Figure 1 shows a strong inverse relation of two indices that relate to journal selectivity, prestige, and resources: the circulation size and the percentage of submitted research that is published. These factors will be seen to have a strong relation to statistical review policies.

Table 2 shows the responses with respect to statistical review policy. One third of journals had a policy that guaranteed a statistical review before an acceptance decision, but for about half, an editor made an individual decision about whether statistical review was needed. Fifty-four percent of journals either retained a statistical consultant or used an associate editor with statistical expertise.

Several aspects of statistical review policies related strongly to journal circulation (and therefore to selectivity and presumably journal resources). The circulation quartiles were 4,100, 10,500, and 25,000. The probability of a statistical consultant on staff ranged from 31% in the bot-

FIGURE 1. Percentage of submitted research that is published, by the circulation of the journal. The spacing on the circulation axis is logarithmic. Linear regression performed without the outlier in the upper right showed a 5.3% reduction (95% confidence interval 4.1%, 6.5%) in probability of acceptance with each doubling of circulation ($p < .001$).

tom quarter, to 58% in the middle two, to 82% in the highest quarter ($p < .001$, trend test for 4 proportions). Similarly, the fraction who compensated these reviewers increased from 12% in the lowest quarter, to 29% in the

Table 1. Characteristics of Journals

Characteristic	Median (n = 114)	10th percentile, 90th percentile
Circulation	10,500	2,500, 80,200
Percentage of submitted research published	40	20, 60
Annual number of original research articles	200	31, 590
Percentage of published research that is "clinical"	60	10, 95
Journal type	%	n
General	16	18
Specialty	66	75
Subspecialty	18	21

Table 2. Policies on Statistical Review

Policy	%
General review policy (n = 114)	
All papers statistically reviewed before acceptance decision	16
Statistical review after initial review, but before decision	18
Statistical review only after acceptance decision	3
Editor decides if and when statistical review is needed	52
Statistical review rarely needed	9
Other	4
Source of statistical reviewers (n = 104)	
Associate editor with statistical expertise	21
Retained statistical consultants	33
Ad hoc basis	41
Other	6
Journals that compensate statistical reviewers (n = 103)	33
Journals in which statistical reviewer sees manuscript after revision (n = 103)	50
Journals with written statistical guidelines (n = 103)	25
Original research statistically reviewed, median (n = 94)	20
Statistical review is judged to produce an important change, median (n = 83)	50

middle two, to 79% in the highest ($p < .001$). The percentage of articles subject to statistical review ranged from a median of 15% in the lowest three quarters to 40% in the highest ($p < .001$, linear regression). Finally, among the 103 journals that published articles that could be subject to review, the proportion having a general policy that guaranteed review before an acceptance decision was 27% in the lowest three quarters and 52% in the highest quarter ($p = .09$, trend test). Written guidelines were most common in the largest journals.

DISCUSSION

For a large proportion of published medical research, the main guarantor of the appropriateness of the quantitative methods is the medical journal itself. The basic logic of research and quantitative methods is typically taught briefly in a medical school curriculum,¹⁰⁻¹² and the reinforcement or accountability for this information during the years of clinical training is highly variable. A variety of studies have documented the lack of understanding by many physicians of basic statistical concepts.¹⁻⁵ In this context it may not be surprising that statistical errors in medical research articles are rife.¹³⁻²³

It is a common misconception that the main value of statistical reviewers is to ensure that analyses are appropriate and that various statistical conventions have not been flouted. However, a much more critical contribution is to assess the overall strength of the evidence—an amalgam of the strength of the design, the execution, the analysis, and the biological reasoning that ties them all together.^{14,24} A statistical reviewer, together with the editor, can play an important role in adjusting the language of the discussion and conclusions so that the limitations of the study are apparent, and that the tone of the conclusions reflects the weight one should put on the results, aspects that can greatly affect how the lay media report studies. One study showed these aspects changed the most between submission and publication.²⁵

It is not always a statistician that must play the role of statistical reviewer. Depending on the complexity and type of methods used in a particular report, a physician with graduate training in quantitative methods or persons trained in companion disciplines like economics could serve in this capacity. Because the focus of such a reviewer is considerably broader than the quantitative component, and because persons who are not professional statisticians can play this role, we prefer the term “methodologic” be used in place of “statistical” as a descriptor, and will henceforth use that designation.

This survey shows that it is only in the largest medical journals (circulation over 25,000) that expert methodologic review is likely. This is not in itself a guarantee of quantitative rigor, but several studies have suggested that such review reduces both the number and seriousness of errors in published articles. (Palmer CR, Johnson AL. Sta-

tistical review: lessons from the *Lancet*. Stat Med. Manuscript submitted.)^{18,26-28} Our study did not directly address that question, but editors perceived that statistical review made an “important change” about 50% of the time.

Comparison with George’s 1985 study provides some evidence that methodologic reviewing policies have been changing. Inferences must be made with considerable caution because he used a smaller and somewhat different sample (75 journals), although there was substantial overlap with journals we used, particularly the larger ones. Unfortunately, the unavailability of raw data from the earlier study prevents direct comparison of those journals that are common to the two. In George’s study 39% of journals had a statistical consultant on retainer or as an associate editor, compared with 54% in this study. In 1985, 14% of the journals George surveyed reported paying their consultants, versus 33% in this study. Finally, the number of journals with written guidelines was only 7% in 1985, versus 25% today. Because our survey included a higher percentage of small circulation journals (and many smaller journals did not respond), it is likely that these changes would be greater if response had been higher, and the figures could be adjusted for journal size.

An optimal methodologic review policy is difficult to define. Should all research be methodologically reviewed? Should this be before or after an acceptance decision? Does a case-by-case decision by the editor target resources most effectively? In the experience of these authors, more inclusive policies appear to be superior. Manuscripts with a heavy reliance on statistical procedures may be the ones flagged for review, but significant problems occur even when there is little use of formal statistics.

The main barrier to more widespread use of methodologic review is undoubtedly resources, in terms of both the money it takes to obtain such review and the number of qualified reviewers in relation to the need. Methodologic reviewers are typically called upon far more frequently than typical peer reviewers (often by several journals), and such a commitment to the review process often requires compensation, which smaller journals can be hard-pressed to afford.

Many major journals are viewing the importance of expert methodologic review more seriously. Journals that have incorporated routine statistical review into their evaluation process since 1990 include the *Journal of the National Cancer Institute*, *Obstetrics & Gynecology*,²⁸ and *Lancet*.²⁹ However, the need may be also increasing as investigators become more sophisticated themselves, and as computers and software enable investigators with minimal statistical training to utilize advanced statistical procedures.

This survey shows that readers need to be aware that many journals, particularly the ones with smaller circulations, may not have the resources to ensure the methodologic adequacy of published original research articles. This puts a burden on readers and consumers to have a level of critical reading skills that may not be realistic. One way to circumvent the economic barrier might be for journals with smaller circulations to pool resources and

draw from a common pool of methodologic reviewers. It is incumbent on the medical publishing enterprise to incorporate more methodologic expertise into the review of research articles, and for the professions of statistics, epidemiology, and other quantitative disciplines to be partners in this endeavor by providing to medical editors pools of competent and willing reviewers.

REFERENCES

- Altman DG, Bland JM. Improving doctors' understanding of statistics. *J R Stat Soc A*. 1991;154:223-67.
- Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *J Med Educ*. 1980;55:692-7.
- Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? *Stat Med*. 1987;6:3-10.
- Borak J, Veillieux S. Errors of intuitive logic among physicians. *Soc Sci Med*. 1982;16:1939-47.
- Friedman SB, Phillips S. What's the difference? Pediatric residents and their inaccurate concepts regarding statistics. *Pediatrics*. 1981;68:644-6.
- Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med*. 1987;317:426-32.
- George SL. Statistics in medical journals: a survey of current policies and proposals for editors. *Med Pediatr Oncol*. 1985;13:109-12.
- Garfield E. *SCI Journal Citation Reports: A Bibliometric Analysis of Science Journals in the ISI Database*. Philadelphia, Pa: Institute for Scientific Information; 1993.
- Hansson S. Impact factor as a misleading tool in evaluation of medical journals. *Lancet*. 1995;346:906. Letter.
- Dawson-Saunders B, Azen S, Greenberg RS, Reed AH. The instruction of biostatistics in medical schools. *Am Statistician*. 1987;41:263-6.
- Hacker CS. Medical statistics at the Medical School of the University of Texas Health Center at Houston. *Am Statistician*. 1987;41:266-7.
- Appleton D. What statistics should we teach medical undergraduates and graduates? *Stat Med*. 1990;9:1013-21.
- Gore S, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *BMJ*. 1977;1:85-7.
- Murray GD. The task of a statistical referee. *Br J Surg*. 1988;75:664-7.
- McGuigan SM. The use of statistics in the *British Journal of Psychiatry*. *Br J Psychiatry*. 1995;167:685-8.
- Gore SM, Jones G, Thompson SG. The *Lancet's* statistical review process: areas for improvement by authors. *Lancet*. 1992;340:100-2.
- Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med*. 1993;118:201-10.
- Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *JAMA*. 1990;263:1355-8.
- Welch GE II, Gabbe SG. Review of statistics usage in the *American Journal of Obstetrics and Gynecology*. *Am J Obstet Gynecol*. 1996;175:1138-41.
- Mainland D. Statistical ritual in clinical journals: is there a cure?—II. *BMJ (Clin Res Educ)*. 1984;288(6421):920-2.
- Mainland D. Statistical ritual in clinical journals: is there a cure?—I. *BMJ (Clin Res Educ)*. 1984;288(6420):841-3.
- Avram MJ, Shanks CA, Dykes MHM, Ronai AK, Stiers WM. Statistical methods in anesthesia articles: an evaluation of two American journals during two six-month periods. *Anesth Analg*. 1985;64:607-11.
- White SJ. Statistical errors in papers in the *British Journal of Psychiatry*. *Br J Psychiatry*. 1979;135:336-42.
- Vaisrub N. Manuscript review from a statistician's perspective. *JAMA*. 1985;253:3145-7.
- Goodman SN, Berlin JA, 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.
- Schor S, Karten I. Statistical evaluation of medical manuscripts. *JAMA*. 1966;195:1123-8.
- Gardner MJ, Altman DG, Jones DR. Is the statistical assessment of papers submitted to the *British Medical Journal* effective? *BMJ*. 1983;286:1485-8.
- Pitkin RM. Statistical evaluation of manuscripts: it's all in the numbers. *Obstet Gynecol*. 1994;83(6):1043-4. Editorial.
- Statistical review for journals. *Lancet*. 1991;337(8733):84. Editorial. See comments.

ANNOUNCEMENT

JGIM Website — Visit us online today!

Please visit the JGIM World-Wide Website:

<http://www.blacksci.co.uk/products/journals/xjgim.htm>