# Common Reasons for Rejecting Manuscripts at Medical Journals: A Survey of Editors and Peer Reviewers

**Daniel W Byrne**


## Abstract

**Background.** Most manuscripts submitted to medical journals are rejected for publication, but scientific data on the reasons are sparse. Some suggest that flaws occur with random frequency; others claim that deficiencies are too diverse to measure.

**Methods.** In 1995, an eight-page questionnaire that contained 83 questions was mailed to a random sample of 50 peer reviewers for the *Journal of the American Medical Association*, 67 editors-in-chief of a sample of prominent medical journals, and 25 Nobel laureates in physiology or medicine of the preceding decade.

**Results.** The overall response rate was 22% (29 of 130); 12 undeliverable questionnaires were excluded. Respondents reported that study-design problems were the most common cause of rejection (71%). The survey results showed that the methods section contained the most flaws, was most often responsible for rejection, and was often too brief. Other common deficiencies were "conclusions unsupported by data" and "results unoriginal, predictable, or trivial". The most common writing problem was "wordiness".

**Conclusions.** Peer reviewers and editors reported that the manuscript deficiencies that most frequently led to rejection were related to study design and to the methods section. Scientific communication and likelihood of acceptance could be improved by more careful planning of study design and by including more reproducible detail in methods sections.


*Daniel Byrne, previously of Byrne Research, is Director of Biostatistics at the General Clinical Research Center at Vanderbilt University Medical Center. He is the author of* Publishing Your Medical Research Paper: What They Don't Teach in Medical School *(Baltimore: Lippincott Williams & Wilkins; 1998).*

## Introduction

Surprisingly little peer-reviewed research has been published on the aspects of the peer-review process that are of concern to potential authors.[1] In particular, the relative frequencies of deficiencies in manuscripts that lead to their rejection by medical journals have received little investigation. Qualitative editorials and subjective accounts of the most common flaws have appeared, but leave largely untouched the relative significance of the various cited flaws in rejected manuscripts. The few quantitative studies available are generally limited to the analysis of manuscripts submitted to a single journal.[2-5] The lack of scientific information on the reasons for manuscript rejection causes frustration and inefficiency for authors, editors, reviewers, and educators. In some cases those effects can prevent or delay the publication of important medical research findings.

The aim of this study was to obtain data on the opinions and recollections of leading medical-research experts regarding the relative frequencies of deficiencies in medical-research manuscripts that proved fatal to their acceptance for publication. The null hypothesis was that the frequencies of such fatal deficiencies would be randomly distributed among the respondents.

## Methods

The study was based on a questionnaire developed as part of the research for my recent book.[6] The project was privately funded and thus independent of any journal or organization. The primary purpose of this survey was to estimate the frequencies of the most common reasons for rejection of manuscripts by medical journals. The eight-page questionnaire contained 83 questions. To enhance comparison with previous research, some questions were worded to match reasons for rejections described in previous publications.[2,7]

The questionnaire was sent to three groups: a selection of editors-in-chief of English-language medical journals, a random sample of peer reviewers for the *Journal of the American Medical Association (JAMA)*, and 25 Nobel laureates in physiology or medicine in 1985-1995.

The sample of 67 medical journals was based on three criteria: the 25 journals with the largest circulation (according to *Ulrich's International Periodical Directory*, 34th ed), the 25 with the highest impact factors (according to the *1994 Science Citation Index Journal Citation Report*), and 17 selected from the stacks of various medical-school and hospital libraries in the New York area. I drew a random sample of 50 *JAMA* peer reviewers from a table of 3023 *JAMA* reviewers.[8] I selected column and row numbers in the table of names by using a random-number table generated with the software True Epistat (Richardson, TX). (The random number 1311, for example, was used to denote the 13th column from the left and the 11th row from the top.) I surveyed reviewers for *JAMA* rather than reviewers for other journals for two reasons: *JAMA* reviews are generally regarded by the medical community as high-quality and thorough, and *JAMA* reviewers' names were readily available. The addresses of the *JAMA* reviewers were found in the American Medical Association's *Directory of Physicians in the United States* (34th edition) or various Marquis *Who's Who* directories. I obtained the names of Nobel laureates from an almanac and their addresses from *Who's Who*. If the addresses were not available from those sources, a MEDLINE search was conducted to find recent publications and contact information.

I took several steps to improve response rates: I used first-class postage, included a self-addressed stamped return envelope

with each questionnaire (for addresses outside the United States, a prepaid coupon for the return postage), printed the questionnaires on light green paper (research has shown that this improves the response rate[9]), included the respondent's name in the cover-letter salutation, and promised a report of the results as an incentive for participating. No prenotification, repeat mailings, or follow-up reminders were used.

The sample size was based on time and financial limitations. No prestudy sample-size calculations were performed. I mailed a total of 142 questionnaires between 20 June 1995 and 8 October 1995. Twelve questionnaires were returned by the post office, so the denominator for the response rate was 130, not 142. Returns were postmarked between 26 June 1995 and 25 October 1995, with a median of 8 days after the initial mailing. (I included returned questionnaires in the study if they were received by 9 November 1995.) The overall response rate was 22% (29 of 130), but it was substantially higher among editors (26%, 17 of 66; one undeliverable) and reviewers (27%, 11 of 41; nine undeliverable) than among Nobel laureates (4%, one of 23; two undeliverable). Most questions were answered by all 29 respondents, but some were answered by as few as 24. The results are presented in percentages of those who answered the individual questions.

Categorical nominal variables were assessed with the Pearson chi-square without Yates's correction. A one-sample chi-square test was used to test actual nominal variable values against the expected values. Ordinal variables were assessed with the chi-square test for a trend or the Mann-Whitney *U* test. No continuous variables were recorded in this study. All tests were two-tailed, and *P* values smaller than 0.05 were considered statistically significant. The statistical software package SPSS version 9.0 (SPSS Inc, Chicago) was used for the statistical analysis.

**Table 1**
**General Deficiencies in Medical Manuscripts**

| | | | *P* |
|---|---|---|---|
| **Question 1[12]** | | | |
| What is the single most common type of flaw that results in outright rejection of a manuscript? | | | <0.001 |
| Design of study | 20 | (71%) | |
| Interpretation of the findings | 4 | (14%) | |
| Importance of the topic | 4 | (14%) | |
| Presentation of the results | 0 | (0%) | |
| **Question 2** | | | |
| Which section usually contains the most flaws? | | | <0.001 |
| Methods | 16 | (55%) | |
| Discussion | 7 | (24%) | |
| Results | 6 | (21%) | |
| Introduction | 0 | (0%) | |
| **Question 3** | | | |
| Which section is most often responsible for outright rejection? | | | <0.001 |
| Methods | 15 | (52%) | |
| Results | 8 | (28%) | |
| Discussion | 6 | (21%) | |
| Introduction | 0 | (0%) | |

## Results

"Design of study" was reported by 71% (20 of 28) of the respondents as the most common type of flaw resulting in rejection of four possible answers to the question, this result is nearly 3 times as frequent as would be expected by chance (71%/25%=2.8; $P<0.001$). Respondents also reported that the methods section was the most problematic section of a manuscript (Table 1, Question 2) and that this section is often too short (Figure). In response to the question "Which section is generally too short?" one reviewer answered, "NONE!"

The most frequent deficiency in interpretation was "Conclusions unsupported by data" (Table 2, Question 4). The most frequent deficiency with respect to the importance of the research was "Results unoriginal, predictable, or trivial" (Table 2, Question 5). The leading flaw in presentation was "Inadequate or inappropriate presentation of the data" (Table 2, Question 6).

Respondents were asked about flaws in design and interpretation. Research-variables, use of biased samples, and failure to take confounding factors into account were judged to be the most frequent flaws that led to rejection of a manuscript (Table 3, Questions 8 to 10). Respondents reported that the most common of nine suggested writing problems was "Verbiage, wordiness" (Table 4, Question 11).

In response to the question "Do you feel that you have adequate skills to evaluate the statistical aspects of most medical manuscripts you are asked to review?" 29% of editors and 58% of reviewers answered no.

Editors and reviewers rated several deficiencies differently. More editors than reviewers said that manuscripts were frequently poorly written or had excessive jargon ($P=0.005$). Editors also were more likely than reviewers to report that "poor presentation" was a frequent problem ($P=0.028$). "Failure to adhere to journal format and policy" was cited by more edi-

*Common Reasons continued*

**Table 2**
**Specific Deficiencies**

| | | | *P* |
|---|---|---|---|
| **Question 4. Deficiencies in Interpretation[12]** | | | |
| Which of the following eight deficiencies is most often responsible for outright rejection? | | | <0.001 |
| Conclusions unsupported by data1 | 7 | (61%) | |
| Data inconclusive | 7 | (25%) | |
| Data too preliminary | 2 | (7%) | |
| Unconvincing evidence of cause and effect | 2 | (7%) | |
| Results not generalizable | 0 | (0%) | |
| Excessive bias in interpretation | 0 | (0%) | |
| Insufficient recognition of previous research | 0 | (0%) | |
| Economic consequences ignored or overinterpreted | 0 | (0%) | |
| **Question 5. Questions About Importance of Research[12]** | | | |
| Which of the following four deficiencies is most often responsible for outright rejection? | | | <0.001 |
| Results unoriginal, predictable, or trivial1 | 9 | (79%) | |
| Few or no clinical implications | 3 | (13%) | |
| Results of narrow interest, highly specialized | 2 | (8%) | |
| Issues outdated or no longer relevant | 0 | (0%) | |
| **Question 6. Deficiencies in Presentation[12]** | | | |
| Which of the following eight deficiencies is most often responsible for outright rejection? | | | 0.002 |
| Inadequate or inappropriate presentation of the data | 9 | (32%) | |
| Rationale confused, contradictory | 7 | (25%) | |
| Failure to give a detailed explanation of the experimental design | 7 | (25%) | |
| Essential data omitted or ignored | 2 | (7%) | |
| Poorly written; excessive jargon | 2 | (7%) | |
| Boring | 1 | (4%) | |
| Important work by others ignored | 0 | (0%) | |
| Excessive zeal and self-promotion | 0 | (0%) | |
| **Question 7[2]** | | | |
| Which of the following eight deficiencies is most often responsible for outright rejection? | | | 0.006 |
| Poor methods | 10 | (36%) | |
| Inadequate results | 7 | (25%) | |
| Poor presentation | 3 | (11%) | |
| Inappropriate statistical analysis | 3 | (11%) | |
| Weak discussion | 2 | (7%) | |
| Lack of originality | 2 | (7%) | |
| Weak conclusions | 1 | (4%) | |
| Failure to adhere to journal format and policy | 0 | (0%) | |

tors than reviewers as a frequent problem ($P$=0.012). But editors and reviewers provided similar answers for most questions. For instance, 50% of reviewers and 59% of editors reported that the methods section contained the most flaws. Design of the study was identified by 67% of reviewers and 75% of editors as the most common type of flaw.

## Discussion

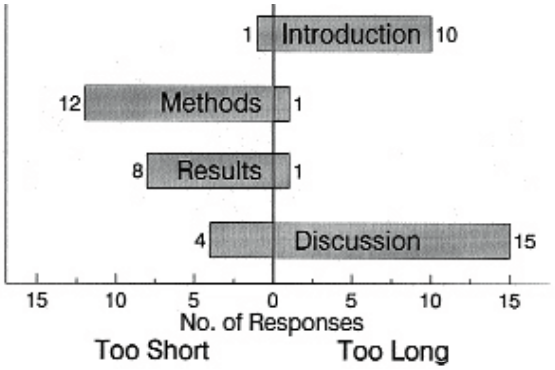
This study demonstrates that the flaws in rejected manuscripts do not occur with random frequency. Some deficiencies occur significantly more frequently than would be expected by chance alone. Although the sample is small, Tables 1 to 4 provide sufficient evidence to reject the null hypothesis. This study meets the criteria set forth by Bailar and Patterson,[1] who called for more scientific research of peer review. Their criteria were: pertinence to biomedical science, a design that would test a hypothesis or examine a specific issue, and use of a clearly defined sample.

The main findings of this study were that study-design flaws and inadequate detail in the methods section are often responsible for rejection of a manuscript. One possible explanation might be that, when reading a medical journal, many busy physicians skip or skim the methods section and therefore give this section too little attention when they write their own papers.

The idea that deficiencies in medical manuscripts are so numerous and diverse that it is impossible to identify the most common ones is false. This new information can help aspiring authors improve their manuscripts, and this improvement should save much time for the many involved in the medical-journal publication process.

My findings have several practical implications. Authors who wish to have manuscripts accepted for publication would be wise to pay more attention to their study designs and provide more detail in their methods sections. The findings suggest that most manuscripts could be improved by lengthening the middle two sections (methods and results) and shortening the

**Figure. Sections of a manuscript that are too long and too short. Which section is usually too short? Which section is usually too long?**




**Byrne DW. Publishing your medical research paper: what they don't teach in medical school. Baltimore: Lippincott Williams & Wilkins; 1998. p 58.**

two on the ends (introduction and discussion), as illustrated in the figure.

My study has several limitations that should be discussed. The information is more than 4 years old, and perhaps the relative frequencies of deficiencies have shifted. However, the changes in the peer-review process during these years would probably not affect the conclusions. The CONSORT statement,[10] which has been adopted by more than 90 journals, might have improved reporting and reviewing of randomized clinical trials.

Another limitation was that there were only 29 respondents and the response rate was only 22%. The 101 nonresponders could be systematically different from the 29 responders, and this could result in a response bias. Late responders are generally more similar to nonresponders than to early responders,[11] and this can be used to assess the response bias. Statistical analysis of these groups, however, showed few differences in answers between early and late responders. Perhaps the responders were more organized and methodical than nonresponders and weighted these qualities—methods and study design—as more important in the questionnaire than nonresponders would have. Several nonresponders who did not complete the questionnaire wrote or called to explain that they no longer review manuscripts and were therefore not qualified to respond. Thus, the low response rate is unlikely to be a source of significant bias.

Other surveys in this field have had higher response rates, but many have been limited to reviewers for a single journal and have asked a smaller set of questions. My eight-page survey was more comprehensive than most others on the topic.

The validity and reliability of my questionnaire were not tested. The internal validity would be difficult to measure. For instance, a reviewer might report that study design presents the most common type of flaw that results in rejection, but importance of the topic might be the most common type of flaw among the manuscripts that this reviewer rejects. Nevertheless, respondents' opinions about

**Table 3**
**Deficiencies in Design and Interpretation**

| | | | *P* |
|---|---|---|---|
| **Question 8** | | | |
| Which of the following six deficiencies is most often responsible for outright rejection? | | | 0.070 |
| Research design problems | 8 | (30%) | |
| Deficiency in methodology | 7 | (26%) | |
| Poor conceptualization of problem or approach | 7 | (26%) | |
| Inadequate control of variables | 2 | (7%) | |
| Duplication of previous work | 2 | (7%) | |
| Inadequate protection of human subjects | 1 | (4%) | |
| **Question 9** | | | |
| Which of the following four deficiencies is most often responsible for outright rejection? | | | 0.003 |
| Failure to collect data on variables that could influence the interpretation of results | 15 | (52%) | |
| Poor response rates in surveys | 8 | (28%) | |
| Subjects lost to follow-up and inadequate duration of follow-up in long-term studies | 3 | (10%) | |
| Extensive missing data and quality-control problems | 3 | (10%) | |
| **Question 10[7]** | | | |
| Which of the following seven deficiencies is most often responsible for outright rejection? | | | <0.001 |
| Biased sample which reduced the representativeness of the population studied | 10 | (34%) | |
| Confounding factors that were not taken into account | 10 | (34%) | |
| Inadequate sample size | 6 | (21%) | |
| Insufficient information about the patient population | 1 | (3%) | |
| Vague endpoints, such as "much improved", without explanation | 1 | (3%) | |
| Straying from the hypothesis or changing the objective | 1 | (3%) | |
| Poor control of numbers (errors or inconsistencies) | 0 | (0%) | |

what they perceive as the most common flaws are useful.

Another issue to consider is the external validity of the findings, that is, whether they are generalizable to a larger population of medical-journal editors and reviewers. *JAMA* reviewers and reviewers for a specialist journal might report different flaws. The mixture of editors, reviewers, and Nobel laureates should provide a balance of answers suitable for most medical journals.

Although I did not use a random sample of biomedical journals, I did use a sample of the more prestigious and popular journals. The existing published lists and directories of medical journals are unreliable and difficult to use for a mailing to obtain a random sample. A random sample of editors of all biomedical journals might not be as valuable as a nonrandom sample of editors of the leading medical journals.

In 1994 Kassirer and Campion[12] devised four major categories of deficiencies (Table 1, Question 1) and, within these, several specific reasons for rejection (Table 2, Questions 4 to 6). The results of my survey indicate that study-design deficiencies are by far the most common of the four (Table 1, Question 1). My study did not attempt to quantify the specific reasons for rejection on the basis of design deficiencies listed by Kassirer and Campion, because this would have required extensive explanation of the various forms of bias. Sackett[13] provides definitions of the various forms of bias for those who wish to study the frequency of rejection based on bias.

Also in 1994 Abby and others[2] reported that weak conclusions and discussion were the most common reason for rejected manuscripts submitted to the *American Journal of Surgery*. In contrast, the findings of my cross-journal survey found that poor methods are a more common problem. The differences in methods could explain the different findings. Abby and others identified flaws by analyzing the letters sent to authors. However, many manuscripts were rejected for several reasons. When questions do not require a mutually exclusive answer, the results are difficult to interpret. Asking for the most common reason for rejection from a list of deficiencies provides more useful information.

**Table 4**
**Writing Deficiencies**

| **Question 11** | | | ***P*** |
|---|---|---|---|
| Of the following 9 writing problems listed below which is most common? | | | <0.001 |
| Verbiage, wordiness | 12 | (43%) | |
| Poor flow of ideas | 6 | (21%) | |
| Poor syntax, poor grammar | 5 | (18%) | |
| Redundancy | 3 | (11%) | |
| Excessive abstraction | 1 | (4%) | |
| Unnecessary complexity | 1 | (4%) | |
| Wrong words | 0 | (0%) | |
| Excessive compression | 0 | (0%) | |
| Unnecessary qualification | 0 | (0%) | |

Morgan[7] listed and discussed the seven flaws in question 10 (Table 3). About 90% of the respondents to the present survey selected three of the seven. This information can help potential authors to focus more attention on the critical problems.

***Future surveys in this field could be improved by making them much***

Future research in this field could expand on the most common flaws reported here. For instance, if wordiness is indeed the leading writing problem, it should be studied in depth. "Study design", "forms of bias", and "conclusions unsupported by data" should also be studied in more detail and probably should be emphasized more in medical-school curricula. Fiske and Fogg[14] have an extensive list of manuscript flaws from the psychology literature that could be studied in the medical literature.

Future surveys in this field could be improved by making them much shorter. At eight pages, my questionnaire was too long—many respondents and nonrespondents were kind enough to point this out. The 57 questions related to the frequency of deficiencies yielded little information and made the questionnaire unnecessarily long. For example, for "How frequently do you encounter the following deficiencies? –Poor methods", the possible answers were 0% = never, 1%-25% = seldom, 26%-75% = occasionally, 76%-99% = frequently, and 100% = always. The "occasionally" choice was too broad (26% to 75%). Questions that forced the respondent to select one deficiency from a list provided more practical information.

Although some of the repetition was intentional to duplicate published tables, future studies could be improved by the elimination of redundant questions and choices. Future studies could also be less abstract if they included a series of examples and asked "What would you criticize?"

The purpose of my survey was not to compare the responses of editors with those of reviewers. Such a study would require a much larger sample. Because editors have much more say than reviewers in the decision to accept a manuscript, these differences might be useful to know. However, the conclusions of this paper are supported by results from both editors and reviewers.

Only in about one third of medical journals does a statistician review is every accepted manuscript. Many editors and reviewers reported that they do not have the skills needed to evaluate the statistics in papers, and this is another part of the review process that needs further investi-

gation. When editors and authors rated the quality of reviews, the only significant factor was "reviewers trained in epidemiology or statistics".[16] Furthermore, reviewers can miss two thirds of the major flaws in a manuscript.[17] Solving that problem might require long-term training; short-term training has not proved effective.[18] Although considerable research has been conducted on the misuse of statistics in published medical papers and on the ineffectiveness of the peer-review process to correct such misuse, there is little evidence that this research has led to improvements.

In conclusion, my findings suggest that authors could improve their odds of publishing their papers in medical journals by anticipating the most common flaws, particularly by investing more time in planning their study designs and providing more detail in their methods sections. An awareness of the most common fatal flaws is probably necessary if authors are to improve the quality of their manuscripts. Applying this knowledge should improve communication about medical science and save everyone involved a little time and frustration.

## References


1. Bailar JC III, Patterson K. Journal peer review: the need for a research agenda. In: Bailar JC, Mosteller F, editors. Medical uses of statistics. Waltham (MA): NEJM Books; 1986. p 349-69.
2. Abby M, Massey MD, Galandiuk S, Polk HC Jr. Peer review is an effective screening process to evaluate medical manuscripts. JAMA 1994;272: 105-7.
3. Frank ED. Publishing a manuscript: analysis of reviewed manuscripts with discussion. Part I. Radiol Technol 1988 Nov-Dec;60(2):125-9.
4. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. JAMA 1990;263:1355-7.
5. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: areas for improvement by authors. Lancet 1992;340: 100-2.
6. Byrne DW. Publishing your medical research paper: what they don't teach in medical school. Baltimore: Lippincott Williams & Wilkins; 1998. p 225-37.
7. Morgan P. An insider's guide for medical authors and editors. Philadelphia: ISI Press; 1986. p 30.
8. The 1994 JAMA peer reviewers. JAMA 1995;273:1309-16.
9.Fox RJ, Crask MR, Kim J. Mail survey response rate: a meta-analysis of selected techniques for inducing response. Public Opin Q 1988;52:467-91.
10. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. JAMA 1994;272:1926-31.
11. Babbie ER. Survey research methods. 2nd ed. Belmont: International Thomson; 1990.
12. Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. JAMA 1994;272:96-7.
13. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51-63.
14. Fiske DA, Fogg L. But the reviewers are making different criticisms of my paper. Am Psychol 1990;45:591-8.
15. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? J Gen Intern Med 1998;13: 753-6.
16. Black N, Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? JAMA 1998;280:231-3.
17. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med 1998;32:310-7.
18. Callaham ML, Wears RL, Waeckerle JF. Effect of attendance at a training session on peer reviewer quality and performance. Ann Emerg Med 1998;32:318-22.



## What's New

a new job?
a prize?
an award?
publication of a book?
a degree or certificate?
recent retirement?

If you have something new in your professional life, we invite you to share the news with readers of *Science Editor*. Please supply a brief announcement to Barbara Gastel, Editor,
*Science Editor*, Department of Journalism, 230 Reed McDonald Building, Texas A&M University, College Station TX 77843-4111, fax 979-845-6887, e-mail *b-gastel@tamu.edu*.



## What It Means To Be an Editor

**7th General Assembly and Con gress**
**European Association of Science Editors**
**21-24 May 2000**
**Tours, France**

For information, see *www.ease.org.uk* or contact *secretary@ease.org.uk*.