

Peer Review and Publication

PRESIDENTIAL ADDRESS BEFORE THE 70TH ANNUAL MEETING
OF THE AMERICAN SOCIETY FOR CLINICAL INVESTIGATION,
SAN FRANCISCO, CALIFORNIA, 30 APRIL 1978

Jean D. Wilson

*From the Department of Internal Medicine, The University of Texas
Southwestern Medical School, Dallas, Texas 75235*

The evaluation of performance by colleagues permeates all academic life; by way of illustration one quickly thinks of such diverse activities as appointment of faculty, assessment for promotion, publication of manuscripts, awarding of funds by granting agencies, and, directly or indirectly, admission to many societies including our own. Indeed, Merton (1) and Storer (2) have developed the thesis that the concern of scientists for an appropriate recognition by their peers is central to the workings of science as a social institution. However, in the past several years serious questions have been raised both in regard to its fundamental value and to the institutional arrangements by which peer review functions in specific instances, and the whole question of peer review is currently the subject of intense scrutiny including that of a recent commission to evaluate National Institutes of Health study sections. I propose today to focus on one narrow aspect of the subject, namely the process by which most journals assess the suitability of papers for publication. My preoccupation with this problem stems from my tenure as editor of *The Journal of Clinical Investigation*. Two issues will be addressed. First, is the manuscript review system fair? Second, what does it accomplish?

From their beginnings in the mid-17th century scientific journals were subjected to criticism about the quality of what they put into print (3). As a result, from the outset they began to develop referee systems for the express purpose of controlling the quality of the papers accepted. The result was an institutionalized mechanism for the application of standards to scientific work, which has changed little in the ensuing centuries.

When *The Journal of Clinical Investigation* (JCI) was founded in 1924 a system of editing was adopted in which manuscripts were assigned to one member of an editorial committee who in turn decided the paper's suitability for publication and, in the case of acceptance, assumed responsibility for its editing. This system functioned substantially unchanged until 1942 when, faced with the increasing complexity of scientific papers, Editor James L. Gamble made the decision to send manuscripts to scientists outside the Editorial Board to obtain the advice of a group of critics more diversified in scientific expertise than afforded by the editors or the members (4). The system used by the JCI (and most other journals) ultimately came to have two functions: one, to provide reliable advice to assist the editors in making decisions; the other, to provide a critique for the author by which the good or potentially suitable paper can be made better. The second function has not received much recognition although it has been aptly described by Cornforth (5):

"Sometimes one has to tell an author where and why he is being illogical, inaccurate, sloppy or trivial: and there is no doubt that adverse criticisms rouse personal animus at the time. Having been at the receiving end of referees' comments, I've usually found on reflection that they were worth heeding; and I am grateful now to these anonymous colleagues for their candour. Anonymous advice is both franker and more effective: as the old saying goes, your best friends won't tell you."

The growth of the outside reviewer system—abetted or acquiesced to over the years by the Society, its editorial committees, and its various groups of editors—was based upon the assumption that the

system provides an assurance that what appears in *The Journal* can generally be relied upon. It has been pointed out that the functional significance of the referee system increases with the growing differentiation of science into extensive networks of specialties. The more specialized the paper, the fewer there are who can responsibly appraise its worth. But while only a few may be fully competent to assess, many more on the periphery of the subject and in other related fields may find the paper relevant to their work. It is for them that the role of referee as deputy takes on special importance. When a scientist is working on a problem in a published article, he can serve as his own referee. He may in fact be better qualified to assess its worth than the official referee who helped usher it into print. It is not so much the fellow-specialist as the others making use of published results in fields tangential to their own who particularly depend upon the referee system (3, 6). Under this system the impact and influence of *The Journal of Clinical Investigation* have been high. This is attested by the remarkably high rate of citation of its articles in diverse journals throughout the world (7). It is natural, therefore, that succeeding editorial groups have supported these editorial practices and have resisted making major changes.

Nevertheless, from its outset there have been difficulties with the referee system, and scientists have complained about its inequities. Here is the young Thomas Henry Huxley writing in 1852 (8):

"You have no idea of the intrigues that go on in this blessed world of science. . . .

For instance, I know that the paper I have just sent in (to the Royal Society) is very original and of some importance, and I am equally sure that if it is referred to the judgement of my 'particular friend' . . . that it will not be published. He won't be able to say a word against it, but he will pooh-pooh it to a dead certainty.

You will ask with some wonderment, Why? Because for the last twenty years . . . has been regarded as the great authority on these matters, and has had no one to tread on his heels, until at last, I think, he has come to look upon the Natural World as his special preserve, and 'no poachers allowed.' So I must manoeuvre a little to get my poor memoir kept out of his hands."

Such criticisms of the system have continued over the years. However, as Gordon (9) has noted, the level at which this discussion usually takes place is surprising. Generally, challenges are made to the orthodox system by noting particular instances of aberration in its workings followed by speculative extrapolation as to the frequency of such abuses. Cumulatively, these questions are serious, but the vigorous debate that ensues after such challenges is inevitably limited because of the lack of empirical evidence about how the system actually functions. Despite the importance of the manuscript review system to scientists (if not to

science) there are few objective data scrutinizing the actual record of performance. Nevertheless, recently some information has been accumulated, both published and unpublished, about the review system, and I think it is now possible to formulate tentative answers about some of the frequently raised criticisms of the system and to design appropriate studies to answer some of the remainder.

It is generally agreed that the mean acceptance rate of manuscripts for journals in the biological sciences is approximately 50%, whereas in physics it is much higher (3). For the purposes of this discussion I assume that everyone will agree that some papers should not be published and should be rejected by someone (editor or reviewer) and that sound papers should be published. The substantive question has to do with the problematical papers where uncertainties exist. Should they be published routinely, as in the physical sciences, or rejected, as is thought to be the practice in the biological sciences? As I interpret the literature on the subject the arguments against the conservative system presently used in the biological sciences can be grouped into three main categories:

(a) It is not possible to evaluate the potential importance of a given paper; the fact that the frequency of agreement among referees in the biological sciences is no greater than would be expected by chance alone proves that the system is invalid.

(b) Even if usually correct, the referee system is subject to a variety of biases so that individuals, groups, or viewpoints are discriminated against systematically.

(c) Even when the system functions at its best reviewers and editors serve as the gatekeepers of science and consequently act as a screen to limit innovation and enforce conformity, thus distorting science.

Let us consider each of these issues in detail. The first is the argument that the entire system is scientifically invalid. Everyone is aware that glaring errors are made by editors. These are frequently well publicized and embarrassing, but they do not necessarily mean the system is worthless if the mistakes are exceptions rather than the rule. It has been argued that the fact that referees themselves do not usually agree in their recommendations proves that the practice has no valid basis. Everyone who has studied this problem (3, 9), including Ingelfinger (10) for *The New England Journal of Medicine*, has found that agreement among reviewers for biological journals is no greater than would be expected by chance alone. Therefore, it is argued that the biomedical reviewing system, as presently constituted and used, is not a dependable screen that sieves out the inferior and saves the superior.

In considering this problem two issues are germane. First, reviewers are frequently if not generally chosen

because they provide different backgrounds of expertise for the manuscript in question. This feature of the reviewing process is implicit in Gamble's original rationale for the use of reviewers outside narrow editorial groups (4). In the biological sciences and in particular in a journal which aims to serve as a bridge between basic and clinical science, many if not most manuscripts transcend more than one, frequently more than two, traditional disciplines. It is therefore necessary to choose reviewers who represent different viewpoints for the proper evaluation of individual papers. Even if this were not the case, reviewers inevitably bring to their reading different values and different types of expertise. In these instances concurrence of opinion may be irrelevant. What is more important in editorial decisions is the weight of the scientific arguments mustered by the individual reviewers in question and the judgment with which the editorial decision is reached. Second, evidence that the system is valid on an average has been obtained by comparing citation rates for papers accepted by the JCI as compared with papers during the same interval that were rejected by the Journal but published promptly elsewhere (Fig. 1). As can be seen the citation rate for papers accepted by the JCI was approximately twice as great as that for those that were rejected and published elsewhere. The difference between the two would be greater of course if we included those papers rejected by the JCI and ultimately not published. This is not to say that mistakes of two types were not made: some papers that were accepted should probably have been rejected; others that were rejected have proved ultimately to be important. But, whatever the individual errors that were made by this particular reviewer-editor setup, the system was effective on average in separating high- from low-impact papers.

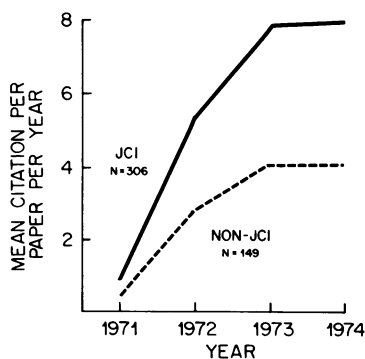


FIGURE 1 Comparison of the citation rates for papers accepted and rejected by *The Journal of Clinical Investigation* but published elsewhere. The mean citation rates for manuscripts rejected by The Journal in 1970 and published elsewhere in 1971 are compared with those for the papers published by The Journal during the same year. (Data were compiled by Dr. Morton Malin of the Institute for Scientific Information.)

A second criticism is that regardless of the median effectiveness of the system bias enters into reviews in such a way as to discriminate against individual scientists or groups of scientists; as a result, the frequency of editorial mistakes is so large as to negate the positive value of the reviewer system. Bias can be of several types. One obvious form occurs when the conclusions of the manuscript do not agree with the preconceived views of the reviewer. For example, Mahoney sent manuscripts to 75 reviewers for a single psychological journal (11). All manuscripts contained identical introductions, method sections, and bibliographies. However, some referees received manuscripts with a results section that agreed with their presumed theoretical viewpoint, and another group received manuscripts that contradicted their viewpoint. Referee evaluations were dramatically affected by the direction of the data in that manuscripts were recommended for publication much more frequently when the results were in agreement with those of the reviewer (11). Although possibly not of the same magnitude this type of bias also influences reviews in the biological sciences.

Another potential bias is discrimination against unknowns in favor of established scientists. Stated in a different way, the eminence of the author may influence the communication system of science (as well as the visibility of papers when published). Of course this is not just a problem for reviewers and editors but permeates every aspect of the reward and recognition system of science, the so-called "Matthew Effect" which results in the enhancement of the position of already eminent scientists at every stage from credit for multiauthored contributions to the awarding of grants (12). The question that concerns us, however, is the extent to which this type of bias influences the reviewing system. It has been assumed to be widespread (7, 10, 11). Designing studies to answer this question is difficult because single authored papers in the biological sciences are uncommon, particularly among established scientists. However, Zuckerman and Merton (3) have documented that there is no such bias in physics where the younger the author, the higher the acceptance rate, regardless of the eminence of the author in the hierarchy of science. This implies that access to publications in distinguished journals in physics derives largely from continuing scientific performance and not necessarily from the assessment of quality in the past. Whether the biological sciences are equally free of this type of bias is not known.

Perhaps the most serious type of potential bias and the one that is mentioned most commonly in the literature stems from the competitiveness that is an inevitable part of scientific research (9, 13). Manifestations of such competitiveness can take many forms and are often obvious to editors. Nevertheless, because

either jealousy or antipathy may influence the judgments of reviewers in subtle ways, it is possible, and indeed likely, that specific viewpoints and specific careers may be promoted or retarded as the result of the composition of specific editorial and review groups. The most frequently proposed reform of the current manuscript review system is to abolish the anonymity of reviewers, the proponents arguing that this would prevent the type of bias that is based upon competitiveness (7, 8, 13). However, most journals have been reluctant to abandon the anonymous manuscript review for the reasons expressed by the editors of *Nature (London)* (14). (a) Referees should not be subject to face-to-face or telephone encounters with irate authors (as happens on occasion to editors). The presentation of science is and must remain ultimately a written activity. (b) Not all the unpleasant things that are said from behind the cloak of anonymity are said from a desire to do mischief. Some need to be said. (c) The anonymous opinion to the conscientious author cannot easily be discounted. The value of the criticism can be diminished by knowing the identity of the referee and therefore dismissing the critique on personal grounds. As a consequence, it has been believed generally that while identification might promote more honest review it might also remove a lot that is good about the present system. Objective evidence for or against the anonymous review is not available, and it is time for some journal to design a study to assess its value.

Thus, whatever its median effectiveness it must be conceded that the review system is subject to error and to bias of several types. However, even if the most serious objections to the system are correct, they are of major importance only if the third criticism is also true. Do editors and reviewers in fact act as the gatekeepers of science; do they control what gets published? To attempt to determine if this is the case we performed a study of the editorial and review process for *The Journal of Clinical Investigation* for the year 1970. Contributors were polled as to the fate of rejected manuscripts. Of the papers rejected by The Journal in that year 85% were subsequently published elsewhere: 3% in 1970, 16% in 1971, 37% in 1972, and the remainder in 1973–76. The journals in which these papers were published constitute a distinguished list of publications, with 14 journals accounting for one-half of the papers. I have no doubt that the JCI in the same period also accepted its share of manuscripts rejected by other journals and thus corrected potential mistakes made by other editorial groups. Most of the authors of the 15% of the rejected papers that were not published elsewhere report that they were convinced by the review process that the paper was either unoriginal or wrong. Of the 85% of the papers that were published, the majority were either not changed or changed only

in minor ways before publication elsewhere. In only one-sixth of the papers was additional work done before publication. We assume that these figures would be similar for other journals. Thus, the peer review process of the JCI had a significant impact on only one-third of rejected papers.

The implications of these findings are of interest. First, it is clear that more than 90% of papers submitted to the JCI are ultimately published, roughly the same percentage of papers that are published in the physical sciences. Second, reviewers and editors for The Journal influence only where papers are published and when they are published, not what is published. They act as traffic officers but not as the gatekeepers of science. The unappreciated strength of the system lies in the fact that there are many journals available to serve as alternatives for the publication of a given paper. Whatever scientific errors are made by individual editorial groups and whatever biases influence specific journals, the diversity of the system means that these wrongs can be corrected and usually corrected quickly. It is inconceivable that all reviewers and editors make the same mistakes or are subject to the same biases. Consequently they cannot function (even if they wished to) as an effective screen of what gets published, conservative or otherwise. If papers are rejected by several journals, it is probable that this means something.

If one accepts the conclusion that the errors and biases in the manuscript review system, of whatever magnitude, do not ultimately do much harm, then it is appropriate to ask, "What is the point of a system in which the major end-result is a shuffling of manuscripts between journals that have different purposes, domains, and audiences? Does *The Journal of Clinical Investigation* have a unique purpose? Is it worth the enormous effort on the part of authors, reviewers, and editors to select papers for this single journal when the majority of papers submitted will be published anyway?" In his introductory essay to the first issue of The Journal, Alfred E. Cohn pointed out that there are two fundamental types of advance in medicine (15). One type stems directly from discoveries in the basic sciences. For example, the discovery of penicillin and the conquest of poliomyelitis are examples of advances that ultimately transformed the practice of medicine but in which clinicians played no direct role. The other type depends upon knowledge of disease in patients and on actual direct contact with diseased persons and is made by practitioners of medicine themselves. This explains the significance of Sydenham, Jenner, and Laennec for past generations and of Albright, Garrod, and Waldenstrom for the present. There is a special type of advance in medicine that in all probability can only be made by properly trained and equipped physicians. As Cohn phrased it, medicine has tasks of great

complexity, and to solve them it is necessary to struggle constantly to master and adapt the methodologies of basic science—physics, physiology, and chemistry—to clinical investigation (15). Thus, the JCI was envisioned by its founders as an interface between the basic biomedical sciences and clinical investigation, and this I contend is a valid and continuing challenge. It was not meant to be a specialty journal but to be broadly based, encompassing all of biomedical science and clinical investigation (with the exception of case reports). Indeed, no other journal integrates these two critical areas so powerfully (16). It does not do everything equally well, but whatever its faults it symbolizes a high level of biomedical science and sets a standard for performance. The standard oscillates from time to time and from subject to subject. After breakthroughs it tends to publish “grey” articles to tell us where we are going and to be certain that the highest levels of science are used to apply new concepts to pathophysiology. In stable periods it veers in the direction of innovation. Depending upon the state of the art at the time, this oscillation between conservatism and innovation and between clinical and basic science serves to reflect development on the one hand and foster its advance on the other at the interface between clinical and basic science. Other journals have different purposes, different audiences, and different scientific domains. As a consequence, the selection of manuscripts for individual journals in the biomedical sciences serves to set standards for specific types of scientific endeavor.

To summarize, a referee system has been an integral portion of scientific publication almost since its inception, and it is difficult to envision publication in science that does not include some type of quality control. The system presently used in the biological sciences is not as bad as maintained by its most severe critics. It contains a self-correction mechanism that repairs most errors through diversity, and it probably serves a valid role in maintaining standards of scientific excellence. The system is costly, time

consuming, and subject to error and bias, and on occasion may result in delayed publication. We should all be concerned with means to make it more rapid, more efficient, and more equitable. On balance, however, it is probably pretty good and better than any alternative that has been proposed.

REFERENCES

1. Merton, R. K. 1957. Priorities in scientific discovery. *Am. Soc. Rev.* 22: 635–659.
2. Storer, N. 1966. *The Social System of Science*. Holt, Rinehart & Winston, New York. 19–27 and 66–73.
3. Zuckerman, H., and R. K. Merton. 1971. Patterns of evaluation in science: institutionalization, structure, and functions of the referee system. *Minerva*. 9: 66–100.
4. Brainard, E. R. 1959. History of the Journal of Clinical Investigation, 1924–1959. I. Personnel and policies. *J. Clin. Invest.* 38: 1865–1872.
5. Cornforth, J. W. 1974. Referees. *New Sci.* 62: 39.
6. Polanyi, M. 1958. *Personal Knowledge*. University of Chicago Press, Chicago, Ill. 163 and 216–222.
7. Garfield, E. 1972. Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies. *Science (Wash. D. C.)*. 178: 471–479.
8. Huxley, L. 1901. *Life and Letters of Thomas Henry Huxley*. D. Appleton and Co., New York, 1: 105.
9. Gordon, M. 1977. Evaluating the evaluators. *New Sci.* 73: 342–343.
10. Ingelfinger, F. J. 1974. Peer review in biomedical publication. *Am. J. Med.* 56: 686–692.
11. Mahoney, M. J. 1976. *Scientist as Subject: The Psychological Imperative*. Ballinger Publishing Co., Cambridge, Mass. 79–106.
12. Merton, R. K. 1968. The Matthew Effect in science. *Science (Wash. D. C.)*. 159: 56–63.
13. Jones, R. 1974. Rights, wrongs, and referees. *New Sci.* 61: 758–759.
14. In defense of the anonymous referee. *Nature (Lond.)*. 249: 601, 1974.
15. Cohn, A. E. 1924. Purposes in medical research. An introduction to the Journal of Clinical Investigation. *J. Clin. Invest.* 1: 1–11.
16. Garfield, E. 1974. Journal citation studies. VI. *Journal of Clinical Investigation*. How much clinical and how much investigation. *Curr. Contents*. 17: 5–8.