

# Peer Review to Ensure Quality in Forensic Mental Health Publication

Alan R. Felthous, MD, and Robert M. Wettstein, MD

Peer reviewers have been called the gatekeepers of science. For journal publications in forensic psychiatry, as well as other disciplines, the purposes of peer review are to assist in the selection of manuscripts to publish, improve the quality of manuscripts before their publication, and promote the fairness of the process. In this article, we examine, in particular, characteristics of high-quality peer reviewers, selection of peer reviewers, recruitment and retention of peer reviewers, desired quality of peer-reviewer ratings, and the value of peer review. We conclude with specific, albeit largely untested, recommendations for improvements in peer review of forensic mental health publications.

**J Am Acad Psychiatry Law 42:305–14, 2014**

Peer review of scientific publications is critical to the development of science. Defined as “the expert assessment of materials submitted for publication in scientific and technical journals” (Ref. 1, p 76; see also Ref. 2, p 654), peer review largely funnels the communication of scientific information into scientific journals. Because of their critical role in determining which scientific reports and reviews will be recorded and disseminated through publication, peer reviewers have been designated the gatekeepers of science,<sup>1,3,4</sup> although the editors who select the reviewers and are ultimately responsible for publication decisions share in this interactive responsibility. The peer review process arguably promotes high-quality science by filtering out manuscripts of little scientific merit and research that is flawed in concept, design, or execution; is of trivial or no scientific significance; or is uninterpretable.<sup>2,5</sup> This goal is best achieved if the peer review process itself is not flawed and does not have the appearance of being flawed.

The literature stresses the critical gatekeeper function of peer reviewers and the peer review process,

but there is another important function that is much more than a by-product: improvement of the manuscript.<sup>5,6</sup> The substance of a manuscript may sufficiently advance knowledge or scholarly discourse, but without improvement, the manuscript could be of less than optimal quality or even unpublishable. Thus, the hoped-for high quality of published manuscripts is the product, not only of the skillful work of authors, but also of the specific improvements recommended by reviewers. As in selecting manuscripts for publication, however, this quality-improvement function can be performed well, or it too can be suboptimal or even counterproductive.

Peer-reviewed empirical studies in clinical medicine now dictate the practice of medicine. Clinical practice parameters and medical society guidelines are formulated on the basis of existing empirical studies that are rated according to the quality of the research. Of special relevance to forensic psychiatry, the American Academy of Psychiatry and the Law (AAPL) Guidelines<sup>7–9</sup> are based on peer-reviewed publications, although generally not on empirical evidence.

Federal case law regarding the quality of peer review and scientific evidence explicitly acknowledges the importance of peer review when judges determine whether scientific data and expert opinions can be admitted as evidence in litigation. One of the four factors proposed by the United States Supreme

---

Dr. Felthous is Professor and Director, Division of Forensic Psychiatry, Department of Neurology and Psychiatry, the Saint Louis University School of Medicine, Saint Louis, MO. Dr. Wettstein is Clinical Professor, Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA. Address correspondence to: Alan R. Felthous, MD, 1438 South Grand Blvd., St. Louis, MO 63104. E-mail: felthous@slu.edu.

Disclosures of financial or other potential conflicts of interest: None.

Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*<sup>10</sup> to determine admissibility of scientific evidence is whether the theory or technique has been subjected to peer review and publication. In progeny cases, the Supreme Court clarified that the four *Daubert* factors, including peer-reviewed publications, apply as well, even if more flexibly, to judicial discretion in deciding whether to admit evidence that is technical or specialized and not purely scientific.<sup>11,12</sup>

Beyond shaping the direction of future scientific research, peer review outcomes influence authors, journal readers, and society when published reports inflame citizens and impel or temper public policy.<sup>1</sup> For university faculty members, peer review publication can be essential for career development, including academic track assignment, promotion in rank, tenure attainment, salary level, receipt of external grants, and success in scientific pursuit. As expected, authors are often dissatisfied with the review process when their manuscripts are rejected. Authors whose manuscripts are rejected are least satisfied when their works are not peer reviewed.<sup>13</sup> Thus, the fairness, or lack thereof, of a peer review process affects not only the scientific process, but, on a personal level, the careers of individual scientists and academicians.

Peer review also contributes to the favorable reputation of the journal. Editors, reviewers, and authors are increasingly aware of Thomson Reuters' impact factor, viewed as a quality measure of journals. It is defined as "a measure of the frequency with which 'the average article' in a journal has been cited in a particular year or period."<sup>14</sup> Thomson Reuters compiles a ranking of scientific publications for the Science Citation Index (SCI) and the Social Science Citations Index (SSCI), and this ranking is regarded as a measure of a journal's quality. The ranking is based on the journal's impact factor, as well as other elements, but also on the journal's use of peer review in its editorial process.<sup>15</sup>

Critical analysis of the peer review process has caused some authors to raise serious questions about its value. Much criticism is based on research suggesting poor inter-referee reliability, which is defined as, "the extent to which two or more independent reviews of the same [manuscript] agree" (Ref. 16, p 120). Mahoney<sup>17</sup> conducted a controlled experiment in which manuscripts that described identical experimental procedures but reported different, positive, negative, mixed, or no results were reviewed by

75 journal reviewers. The results of the study showed that peer-reviewed evaluations can be heavily influenced by factors such as the outcome of the reported research. Positive findings were valued more than negative ones, yet the latter may be just as important to the advancement of scientific knowledge. Second, his experiment showed poor agreement on assessment parameters. Although poor inter-referee agreement suggests poor reliability, it can also reflect divergent, confirmatory, or ideological bias. There may be value in the reviewers' not holding the same (potentially wrong) assumptions. Given the results of his study, Mahoney concluded that there may be "little to defend [peer review] other than tradition" (Ref. 17, p 174). Moreover, even where there is acceptance and decisional congruence among peer reviewers, such agreement does not necessarily translate into quality reviews and should not obviate careful assessment of manuscripts by the journal editor.<sup>18</sup> A study of the validity of peer review concluded that even with low inter-rater reliability, peer review contributes to winnowing high- from low-impact articles.<sup>19</sup> In contrast to the impression of low inter-rater reliability among peer reviewers, a recent article showed sufficient concordance among the reviewers to assist in editorial decisions.<sup>20</sup>

In maintaining the focus of this article, we will not discuss the important question of fairness and the double-blind review process, the value of which has been questioned. We should point out, however, that a recent online survey found that most academics favor a peer review process that is double blind.<sup>21</sup>

One might hope that the peer review process would efficiently identify those manuscripts that are the product of fraudulent work and plagiarism. The authors are familiar with instances wherein the editor or a reviewer recognized plagiarism. Unfortunately, however, current evidence suggests that the peer review process does not sufficiently identify fraudulent work.<sup>22</sup>

The present discussion examines potential and demonstrated deficiencies in peer review as well as purposes of review that ought to be pursued. Is peer review a worthy enough enterprise to be supported? If not, by what process should the gatekeeping and improvement objectives be carried out? If peer review is to be supported, how can its flaws be overcome? We attempt to draw special relevance to journals in forensic psychiatry and mental health law with the

realization that there has been no empirical research on peer review for these journals.

Journals in forensic psychiatry and psychology include reports of original scientific research on matters relevant to the field, but peer-reviewed articles are not limited to empirical research reports. Forensic mental health journals include, as well, literature reviews and scholarly analyses of forensically relevant clinical conditions, ethics-based principles, legal doctrines, case law, public policies, and jurisprudence. Thus, the peer review process shapes not only the nature and direction of forensically relevant science, but also of forensic practice, education, scholarly discourse, and public policy through legislation, *amicus* briefs, and trial and appellate court decisions, including court-generated law. *Amicus* briefs and landmark cases in mental health law not uncommonly cite peer-reviewed articles published in journals of forensic psychiatry or psychology.

In this review, we begin with the recognition that the peer review process for forensic mental health journals is unstudied. Using neurology journals as an example, Wong<sup>23</sup> stressed the inadequacy of applying peer review research to an unstudied specialty. We must therefore draw on studies of peer review of other professional and scientific journals with awareness that empirical findings may be imperfectly generalized to include forensic mental health journals. Attempts to identify characteristics that will predict which reviewers will produce superior peer reviews have not been fruitful. This shortcoming, of course, creates a challenge for editors in identifying, selecting, recruiting, and training superior peer reviewers. Implications of these findings are analyzed with the aim of developing suggestions for improving the peer review process without overburdening publishers, editors, or reviewers.

Other factors, beyond the assessed quality of the manuscript, that can influence the decision of whether to accept a manuscript for publication include manuscript load at the time of the dispositional decision, priority of the topic, originality of the topic, length of the manuscript, institution of affiliation, negative or positive results, sample size, generalizability of the research results, whether the report was solicited and the inventory level of the category at the time of the dispositional decision,<sup>24</sup> usefulness of the findings, interest of journal's readership, available space in the journal, justification of the manuscript's length,<sup>16</sup> and identity and status of the authors.

## Characteristics of High-Quality Peer Reviewers

If journal editors could select reviewers based on characteristics known to predict high-quality reviews, they could improve the reviews' consistency. Unfortunately, research on the characteristics of superior reviewers is limited and inconclusive. A study of reviewers for the *Annals of Emergency Medicine*<sup>25</sup> showed that most variables tested, such as status as principal investigator on a grant, formal training in critical analysis of statistics, and high academic rank, did not predict high-quality reviews. Relative youth and inexperience and employment at a university-operated hospital favored quality reviews. The predictive power of all variables, however, was weak. Another study also found that younger reviewers produce better reviews.<sup>23</sup> The more time given to reviewing a manuscript, the better the product, up to three hours, after which additional time is not contributory.<sup>26</sup> Callaham and Tercier concluded, "... almost none of the experiences and training that might logically be thought to make for a high-quality reviewer ... actually predict subsequent performance of high-quality reviews" (Ref. 25, p 0036). Neither can the qualities of a superior reviewer be taught, as indicated by the available research.<sup>25,27-29</sup> For example, attendance at a training workshop designed to improve the quality of their reviews had no significant effect on the reviewers.<sup>27</sup> The crucial elements that make a reviewer superior remain elusive.<sup>27</sup>

As an exception to these pessimistic findings regarding the value of training, van Rooyen and colleagues<sup>30</sup> achieved favorable results with their Review Quality Instrument (RQI), only after providing guidelines and training in its use. It was not used for peer review, but for rating the quality of the reviewer. More to the point, they found that training in epidemiology and statistics was associated with the quality of reviews ( $B = 0.201$ ;  $SEB = 0.05$ ;  $p < .001$ ). Similarly, Black and colleagues<sup>26</sup> found that the only characteristic significantly associated with high-quality reviews was prior training in epidemiology or statistics. In this regard, a special challenge for editors of forensic psychiatry and psychology journals is selecting and recruiting a sufficient number of regular reviewers who have the necessary special expertise in both research methodology and the specific subspecialty subject matter. It may be the case that peer

reviewer training is more valuable in some fields or topic areas than in others.

Another exception to the above-cited reports was a study of reviewers of meta-analyses that found remarkable consistency among reviewers in assessing the scientific quality of the submitted papers.<sup>31</sup> In this study, reviewers were selected on the basis of their experience and interest in research methodology. Oxman and colleagues<sup>31</sup> posited that the excellent inter-rater agreement was the result of the reviewers' training in research methodology and the explicit rating criteria that were based on process, not outcome. The downside of stressing research methodology to the exclusion of other measures of quality is that worthy dimensions such as the importance of the question to be addressed and accuracy and sophistication of the treatment of the subject matter will be overlooked. Such other dimensions can be most important considerations for forensic articles and analyses of legal topics that are not research reports or meta-analyses.

Reviewers with expertise in scientific methodology should help ensure a satisfactory level of reliability and validity of scientific submissions. Also useful for forensic journals in particular are subject matter experts who are familiar with the field of knowledge and its extant literature. Reviewers need not be in perfect agreement to be useful, as they can make salient observations that complement one another and enhance the value of the composite review.

Research on the characteristics of quality reviewers does not necessarily establish the characteristics of quality reviews. In testing for inter-rater agreement among reviewers, Oxman and colleagues<sup>31</sup> focused on four criteria for assessment of the quality of the scientific process of meta-analyses, but they acknowledged that other scientific and nonscientific characteristics contribute to the overall quality of scientific articles. Van Rooyen and colleagues<sup>30</sup> developed and validated an instrument designed for assessing the quality of manuscript peer reviewers. Those who investigate the characteristics of quality reviewers<sup>29</sup> have used this or similar instruments. We will comment later on the characteristics of quality reviews, but for assessment of the peer reviewers or the reviews themselves, such instruments are not entirely satisfactory, having lent themselves to highly subjective valuations. Consider the following item from Callahan's scoring system<sup>29</sup> for example: "The reviewer provided the author with useful suggestions for im-

provement of the manuscript" (Ref. 29, p 2781). No doubt, this is an important feature of many peer reviews, but one without an objective, operationalized standard. Moreover, suggestions for improvement may be disingenuous and unedifying if the submission is actually beyond either redemption or improvement.

Although not tested empirically, the best screening method for superior reviewers would reasonably be through a system of selection and retention that includes an ever-changing pool of *ad hoc* reviewers in addition to the members of the journal's review board. Regular reviewers are selected on the basis of the consistently high quality of their *ad hoc* reviews. Editors must monitor the quality of reviews from individual reviewers. Instruments used in research on the quality of reviews may be useful in this regard (e.g., the Review Quality Instrument<sup>30</sup>), although the routine use of such instruments can be editorially time consuming.<sup>25</sup>

### Selection of Peer Reviewers

In our experience, we suspect that medical journals do not select their reviewers by standardized methods, such as evaluating their expertise in critical review of research or research methodology. Indeed, selection of peer reviewers should be based not only on the quality of their reviews, but on other characteristics, as well, including their established expertise in the desired subject area.

The peer reviewer selection process must first be informed by the composite needs of the journal; if the journal's main focus is forensic psychiatry, a substantial complement of forensic psychiatrists would be appropriate but also with one or more mental health attorneys, forensic psychologists, epidemiologists, or statisticians with advanced expertise in research design. Reviewers might include both generalists within the subspecialty and subject matter experts in the particular areas likely to be emphasized in the journal. The diversity of types of expertise to be represented by the reviewers is determined by the editor's vision for the journal and the interests and needs of the readership to be served.

Beyond regular reviewers, a journal editor may select guest situational or consultant reviewers, who may be depended on to handle a surge of submissions when regular reviewers are already overextended. Ideally, such reviewers are called on for their known specialized expertise in a particular topic or method-

ology. The use of nonregular reviewers also provides an opportunity for the editor to assess the quality of reviews of potential regular reviewers. Selection of guest reviewers should be based on similar considerations as for regular reviewers except where such reviewers are solicited for expertise not available among the pool of regulars.

A potential reviewer's discipline and areas of interests should be apparent from the *curriculum vitae* and prior publications, but as emphasized earlier, studies have shown that training, academic rank accomplishments, and prolific publishing record are poor predictors of quality of peer reviews.<sup>25</sup> On the other hand, postgraduate training in epidemiology or statistics has been shown to be associated with reviews of good quality.<sup>26,30</sup> Eminence in one's field also helps identify the individual's field. The Thorndike halo<sup>32</sup> effect or Matthew effect,<sup>33,34</sup> of inflated impressions of an author's work as judged by a reviewer, may apply as well to an editor's presumption about a well-known author or investigator's ability to provide quality reviews. Editors may select notable experts in a particular field to enhance the prestige of the journal and its marketing success. The finding that academic status and professional reputation do not predict superior peer reviewers may not be reason enough to avoid such appointments. No factor is a strong predictor, and some individuals who are renowned in their fields are also excellent reviewers. Also, it may be more helpful for some submissions to have one review by someone with a high level of expertise in the subject matter and another by a reviewer who is otherwise of superior quality but without unique knowledge of the subject matter. Our knowledge of what determines a quality peer review is limited.

Some journal editors offer the authors the opportunity to nominate or suggest potential reviewers for their manuscripts. An author's recommendations can help to identify experts in a subject for which the editor's known choices are limited. They can also assist the editor in expanding the pool of reviewers. On the negative side, they can give the author an opportunity to select a reviewer who will be lenient or one with an undisclosed conflict of interest. Reducing the level of blinding and objectivity in selecting reviewers may diminish actual or perceived fairness. To our knowledge, all of these questions, as well as the overall success and utility of this approach, have not been evaluated empirically.

Beyond having expertise in the field and proficiency in conducting reviews of high quality, the desirable reviewer should be available and responsive. He should be willing and able to complete reviews in a timely manner. The superior reviewer also strives for fairness and impartiality and submits reviews that are constructive and free of degrading, abrasive, sarcastic, competitive, and gratuitously critical remarks.<sup>1</sup> The recommendation as to whether the manuscript should be published should not be shared with the author. Unless selected for a specific purpose such as assessing the appropriateness of the statistical method used in the research, it is essential that reviewers be familiar with the professional area that is the subject of the manuscript as well as the relevant scientific literature on that subject and the weaknesses of the extant literature.

### Recruitment and Retention of Peer Reviewers

Regardless of how excellent in quality an individual's reviews can be, if the reviewer does not perform the review, it is of no use. The timeliness of submission is also important. Although in some cases a late review is better than none, delay can result in a shortened time frame for the editor and for the authors if revision is needed and can cause the submission to fall off the publication track. Thus, useful reviewers complete their work in a timely manner.

Many potential reviewers either do not agree to review a given manuscript or simply do not respond to the soliciting editor. Research by Pitkin and Burmeister<sup>35,36</sup> showed that requesting a potential reviewer in advance leads to an increased risk that the request will be rejected. Sending the manuscript to the reviewer without first seeking permission results in a higher rate of review. Those reviewers who agreed in advance, however, complete their reviews more promptly.<sup>35</sup> When reviewers were late, reminders were effective in prompting completion within seven days in about two-thirds of late reviews. This improved responsiveness with a reminder occurred regardless of whether the reminder was delivered by telephone, fax, or e-mail.<sup>36</sup>

Peer reviewers report many reasons for agreeing to review a scientific manuscript. According to Tite and Schroter, the most common reasons were that the paper seemed to make a contribution to the subject area, the topic pertained to the reviewer's own interest, and the review provided an opportunity for the

reviewer to learn something new from the manuscript (Ref. 37, p 10).

Although the possibility has not been examined empirically to our knowledge, some individuals may be willing reviewers if named to the journal's editorial board, but remain unwilling without such acknowledgment. Other expressed reasons for both regular and guest reviewers to decline are time constraints and insufficient expertise on the subject of the manuscript.

It is impractical to appoint all who would review on an *ad hoc* basis to the editorial board. Similarly, editors may wish to call on a reviewer who has rotated off the board, but once off the board or once in professional retirement, some would-be reviewers no longer desire to review. Publication and journal editors commonly reward guest reviewers by listing their names with appreciation in the final issue of the volume, but the incentive value of this measure is unknown.

### Desired Qualities of Peer-Reviewed Ratings

If peer review is to serve well its two-fold *raison d'être*, to select submissions worthy of publication and to improve the quality of papers before publication, attention must be directed to those qualities of peer review assessments that best support these purposes. Naturally, scientific reports should be examined for their scientific quality, to ensure reliability and validity and to diminish the risk of bias. Ensuring quality means attending to methodology and components of the process of designing and executing the study, including the concept of the research, subject recruitment, consistency testing, validity testing, appropriateness of methods, representativeness of the condition that is studied,<sup>31</sup> and conflicts of interest. A full exposition on the analysis and the quality of a scientific study is beyond the scope of this review, except to note that certain types of studies require additional methodological and ethics-related considerations. For example, review of a randomized drug trial report might best be accomplished using the 30-item scale based on the Consolidated Standards of Reporting Trials (CONSORT)<sup>38</sup> or a reasoned modification of this scale.<sup>39</sup>

The scientific process or methodology of the research or meta-analysis is not the only consideration in assessing the scientific quality of a manuscript. Features other than scientific methodology and data-

analysis strategies<sup>16</sup> include the manuscript's potential impact on future scientific developments, the degree of innovation in the approach, the author's imagination and originality,<sup>16</sup> the importance of the question that is addressed,<sup>28</sup> and the relevance and completeness of the literature review.<sup>16</sup> Additional considerations encompass the manuscript's clinical relevance and literary quality<sup>31</sup> (including clarity of written expression,<sup>40</sup> organization of information,<sup>16</sup> and style that is appropriately measured, focused, and based on evidence presented or cited); acknowledgment of limitations; and well-reasoned presentation.

Of the several instruments developed to assess the quality of reviews, only that designed by van Rooyen and colleagues,<sup>30</sup> who themselves recognize that there is no gold standard for criteria validation and comparison, has been empirically tested for internal consistency and reliability. The Review Quality Instrument (RQI) showed high internal consistency with a Cronbach's  $\alpha$  of .84 and good test-retest ( $\kappa_w = 1.00$ ) and inter-rater ( $\kappa_w = .83$ ) reliability (weighted  $\kappa$  statistic). Parameters incorporated into this instrument and therefore deemed useful in assessing the quality of a manuscript were importance of the research question, originality, methodological strength with minimal weaknesses, and presentation of the results. Characteristics of a quality review itself and not the manuscript are constructiveness and substantiation of written comments intended for the author(s). Although it is not always mentioned in studies of peer review, we believe that substantiation of the comments in the review is just as important as any criticism.

In the study by van Rooyen and colleagues,<sup>30</sup> small improvements in the level of internal consistency resulted from removing the items on originality and importance. However, these items were retained in the instrument because of their importance to editors. Indeed, we would add that an item may be no less relevant to quality simply because its internal consistency or reliability has not been empirically demonstrated. In some instances, the item can be reworded to achieve greater consistency and reliability. In any case, the valuation of specific parameters should be most useful to the editor and to the authors as suggestions for improvement, if examples are cited to substantiate the specific rating or comment.

The identification of the desired qualities of peer review assessments presupposes identification of

quality parameters of the manuscripts themselves. Dougherty and colleagues<sup>41</sup> conducted a 69-item online survey of 1,675 nursing journal peer reviewers, to identify which indicators of quality were most favored by reviewers. The most critical factors in assessing the value of a manuscript were whether the topic of the manuscript is of current interest, whether it represents a newly emerging area, and whether the manuscript contributes to knowledge or research evidence. Clinical relevance and research rigor also contributed to a manuscript's publication value. Thus, beyond its scientific vigor, the valued manuscript is also of interest to the journal's readership and is of material value to the knowledge base of the field.

In emphasizing the importance of inter-reviewer reliability, we want to add a word about creative disagreement. One reviewer may detect a flaw or suggest an improvement that does not occur to other reviewers,<sup>16</sup> resulting in inter-reviewer unreliability that is not undesirable. Though again untested, editors may constructively seek some divergence in types of expertise for a given manuscript without blindly pursuing maximum reliability. Such diversity of critical input can be of special, if unmeasured, importance for cross-disciplinary manuscripts, as many are in the interdisciplinary field of psychiatry and the law. Reviewers can and do have different preferences as to manuscript content and approach, and these differences can influence their ratings of manuscript quality and acceptability.

Another approach to assessing the value of peer review in improving the quality of manuscripts is to assess how they are changed from initial submission to final publication. When such a study was conducted concerning manuscripts for the *Annals of Internal Medicine*, the changes responded to the following types of initially identified problems: too much information (too detailed, redundant, extraneous), too little or missing information (not enough detail, emphasis needed), inaccurate information (incorrect, inappropriately synthesized, inadequately synthesized), misplaced information (alternative location preferred, inappropriate location), and structural problems.<sup>42</sup> Of these, changes were most frequently made because information was missing or extraneous.

As with other studies of peer review, these findings may or may not be extrapolated to forensic mental health journals. The extent to which these changes were driven by journal policy or judgment of the

editor, peer reviewers, or both is unclear. Important to consider, however, is that the peer review process can and perhaps often does result in substantial changes, assumed to be improvements, in the quality and quantity of information. Such changes need not conform in all respects to the scientific quality of the manuscript *per se*, and the inter-referee reliability for the need for such changes in information is unknown.

### Is Peer Review a Worthwhile Enterprise?

Is peer review a worthy enough endeavor to continue? Certainly, the two-fold aims of peer review are worthy ones: winnowing the wheat from the chaff in terms of publishable science and other scholarly work and improving manuscripts before publication to ensure satisfactory quality. There seems to be a consensus in the literature that peer review is the desired method for selection and quality improvement and that the scientific community currently has no viable alternative to its use. The literature that assesses peer review on the other hand also suggests that it does not uniformly meet these aims as well as it should, is costly in professional time and resources, and can in some ways be counterproductive; but, in fact, those journals that are not peer reviewed are often regarded less favorably.

In addition to the two-fold explicit purposes of peer review, it ought to support the perception of and *bona fide* presence of fairness by striving for it. Just as courts strive for fairness and expert witnesses strive for objectivity, peer review should strive for balanced assessments. With a peer review process in place, journal readers should have confidence that editorial selections are properly informed and based on merit. Authors and investigators who submit manuscripts for publication should also be reassured by the existence of a peer review process that selection is determined more by the merit of the submission than the uninformed whims and biases of a single editor. Similar to sound selection and quality improvement, however, fairness is enhanced by a process that works well and can be undermined by flawed procedures. The appearance and the actual presence of fairness are both requisite elements.

Thus, the aims of peer review—quality selection, quality improvement, and enhanced fairness—are desirable. There seems to be no better approach that can replace peer review, regardless of its imperfections. What must be addressed then is the matter of

improving the peer review process, so as to ensure that it usefully achieves its aims.

### Improving the Quality of Peer Review

Without attention to improvement, any human process risks unintended regression with worsening results. Improvement measures must be practical, however, and achievable without counterproductive expenditure of time and energy. Some authors have listed methods, often speculative, for improving peer review. We will limit our attention to potential improvement measures that address real or potential flaws identified in this review with special consideration for the unique needs and circumstances of journals in forensic psychiatry and psychology.

Studied attention should be given to the positive and negative value of reviewers' comments for the author, for whom they are arguably more influential, at least in the process of manuscript improvement, than the parameter ratings. Yet the individual comments are less studied and less subject to quantitative comparisons. It may be that the variety of comments without inter-rater reliability are more useful than carbon copy reviews.

Before venturing to suggest measures for improving peer review of forensic psychiatry and psychology journals, some caveats and limitations must be emphasized. If weaknesses, inconsistencies, and shortcomings in achievement of desired results cause some to doubt the value of peer review, published recommendations for improvement are even less empirically studied and substantiated. Idealistic measures for improvement can be impractical; the more onerous the measures for editors and reviewers, who typically are fully occupied with responsibilities of employment (clinical, forensic, administration, and research), the less likely the measure or the review itself will be accomplished. There may be financial or other practical limitations in available resources that preclude institution of peer reviewer quality improvement measures. Commercial journal publishers may not desire to institute changes in just one of their published periodicals. Improvement measures should be especially relevant to forensic psychiatry and psychology journals that have not been included in the literature on peer review:

The journal should select reviewers based on their expertise and experience in writing and editing in forensic mental health. Examining the

reviewer's *curriculum vitae* and examples of published or edited works can be useful in this regard. Familiarity with the special journal is essential, so that the reviewer can thereby gauge the journal's interests and standards. Of importance, assessing a reviewer's performance as a guest reviewer is a reasonable attempt at identifying his potential work product as a reviewer. Some variety in the skills and knowledge of the entire group of regular reviewers is desirable (e.g., health law attorneys and law professors, forensic psychologists, forensic psychiatrists, forensic social workers, experts in research methodology, and statistical analysis) is more useful than having everyone come from one of these fields. New reviewers should describe their areas of expertise and the limits thereof to the editor.

The journal should provide regular reviewers with written guidelines that explain how the manuscript rating form is to be used,<sup>27</sup> pitfalls to be avoided, and important qualities of good reviews. Editors can provide examples of improper reviewer conduct to potential manuscript reviewers, such as publicly discussing or citing a reviewed manuscript before its acceptance and publication or using the manuscript for the reviewer's own work. These written guidelines can advise reviewers in advance that they will receive periodic feedback on the quality of their reviews and suggestions for improvement. The journal should have those selected to be reviewers confirm in writing their understanding of their obligations.

The journal should design the manuscript review form to include clearly written, practical measures of manuscript quality, using those already recognized in the peer review literature as well as items specific to the subspecialty of forensic mental health publications. The journal should require commentary that substantiates the ratings, particularly the critical ratings, with corresponding suggestions for improvement of the manuscript.

Guest reviewers should be rewarded in any non-financial way that is feasible. At a minimum, they should be sent a letter of appreciation, recognized in an issue of the journal, and provided with a copy of the journal that contains the article that they reviewed.



The journal should adopt a mechanism for monitoring the performance of regular reviewers and providing them with periodic feedback. These comments should include specific recommendations for improvement and commendation on superior aspects of their reviews and expression of appreciation for their conscientious, diligent work.

Editors can provide peer reviewers with copies of reviews of the same manuscript by other reviewers. Reading the work of others may enhance the reviewers' motivation to perform comprehensive reviews and may serve a training function, especially for reviewers who tend to be outliers in their reviews.

The journal should periodically review appointments of regular reviewers, perhaps every four years, but stagger the turnover. Reappointment should be based on satisfactory peer review performance determined through a methodical and perhaps empirical performance review procedure. Options include replacing all regular reviewers within a specified period of tenure, retaining those who reliably provide timely reviews of high quality and replacing those whose work is less exemplary or vital, or designating as consultants some of the previously regular reviewers.

Because of the difficulty in extrapolating findings that may be unique to other disciplines to forensic psychiatry and forensic psychology, research on the peer review process in forensic journals would place any recommendations for change or improvement, such as these, on an empirically firmer and more useful footing.

## Conclusions

Krassier and Campion quipped in the title of their article that peer review is "crude and understudied, but indispensable."<sup>5</sup> Peer review is even less studied, but just as indispensable for journals in forensic psychiatry and psychology, we contend. If peer review in forensic psychiatry and psychology publication is to serve effectively the purposes of selecting good from bad science and scholarship to be published, by a transparently fair process, its problems must be unpacked and addressed. In initiating this enterprise for publications in forensic psychiatry, we turned to the general literature on peer review in science and health

care. Aspects of crudeness that can be addressed and applied to forensic journals for quality improvement include deficiencies in quality of peer reviews and peer reviewers and challenges in selecting and retaining high-performing reviewers. Although specific measures for improvement in peer review procedures are even less empirically established than the identification of strengths and flaws in current peer review practices, we offer several suggestions that could be useful.

In a closed loop of quality improvement, our suggested procedures should help identify and measure the strengths and flaws of peer review. Strengths should be more easily supported and sustained, and flaws should be more readily corrected or at least ameliorated. Moreover, with such procedures in place, the method of assessing the quality of peer review should itself become more accessible to study. Then, research on the peer review process for publications in forensic psychiatry and psychology could be particularly instructive.

## References

1. Hojat M, Gonnella J, Caellegh AS: Impartial judgment by the "gatekeepers" of science: fallibility and accountability in the peer review process. *Adv Health Sci Educ* 8:75–96, 2003
2. Bailar JC III, Patterson K: Journal peer review: the need for a research agenda. *N Engl J Med* 312:654–7, 1985
3. Crane D: The gatekeepers of science: some factors affecting the selection of articles for scientific journals. *Am Sociol* 32:195–201, 1967
4. DeGrazia A: The scientific reception system and Dr. Velikovsky. *Am Behav Sci* 7:38–56, 1963
5. Krassier JP, Campion EW: Peer review: crude and understudied, but indispensable. *JAMA* 272:96–97, 1994
6. van Rooyen S, Godlee F, Evans S, *et al*: Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA* 280:234–7, 1998
7. Gold LH, Anfang SA, Drukteinis AM, *et al*: AAPL Practice Guideline for the Forensic Evaluation of Psychiatric Disability. *J Am Acad Psychiatry Law* 36(Suppl 4):S3–50, 2008
8. Mossman D, Noffsinger SG, Ash P, *et al*: AAPL Practice Guideline for the Forensic Psychiatric Evaluation of Competence to Stand Trial. *J Am Acad Psychiatry Law* 35(Suppl 4):S3–72, 2007
9. Giorgi-Guarnieri D, Janofsky J, Keram E, *et al*: AAPL Practice Guideline for Forensic Psychiatric Evaluation of Defendants Raising the Insanity Defense. *J Am Acad Psychiatry Law* 30(Suppl 2):S3–40, 2002
10. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
11. *General Electric Co. v. Joiner*, 522 U.S. 136 (1997)
12. *Kumho Tire Co. Ltd. v. Carmichael*, 526 U.S. 137 (1999)
13. Weber EJ, Katz PP, Waeckerle JF, *et al*: Author perception of peer review: impact of review quality and acceptance on satisfaction. *JAMA* 287:2790–3, 2002
14. Garfield E: The Thomson Reuters Impact Factor. Available at [wokinfo.com/essays/impact-factor/](http://wokinfo.com/essays/impact-factor/). Accessed July 16, 2014
15. Testa J: The Thomson Reuters journal selection process. Updated 5-2012 Available at [http://thomsonreuters.com/products\\_](http://thomsonreuters.com/products_)

## Peer Review to Ensure Quality

- services/science/free/essays/journal\_selection\_process/. Accessed January 9, 2013
16. Cicchetti DV: The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behav Brain Sci* 14:119–35, 1991
  17. Mahoney M: Publication prejudices: an experimental study of confirmatory bias in peer review system. *Cogn Ther Res* 1:161–75, 1977
  18. Callaham ML, Baxt WG, Waeckerle JF, *et al*: Reliability of editors' subjective quality ratings of peer reviews of manuscripts. *JAMA* 280:229–31, 1998
  19. Jackson J, Srinivasan M, Rea J, *et al*: The validity of peer review in a general medicine journal. *PLoS One* 6:e22475, 2011
  20. Baethge C, Frankling J, Mertens S: Substantial agreement of referee recommendations at a general medical journal: a peer review evaluation at *Deutsches Ärzteblatt International*. *PLoS One* 8:e61401, 2013
  21. Ho RC, Mak KK, Tao R, *et al*: Views on the peer review system of biomedical journals: an online survey of academics from high-ranking universities. *BMC Med Res Methodol* 13:74, 2013
  22. Stroebe W, Postmes T, Spears R: Scientific misconduct and the myth of self-correction in science. *Perspect Psychol Sci* 7:670–88, 2012
  23. Wong VSS: The examination of peer review and publication in neurology. *J Child Neurol* 25:1298–301, 2010
  24. Gilbert JR, Williams ES, Lundberg GD: Is there gender bias in JAMA's peer review process? *JAMA* 272:139–42, 1994
  25. Callaham ML, Tercier J: The relationship of previous training and experience of journal peer reviews to subsequent review quality. *PLoS Med* 4:e40, 2007
  26. Black N, van Rooyen S, Godlee F, *et al*: What makes a good reviewer and a good review for a general medical journal. *JAMA* 280:231–3, 1998
  27. Callaham ML, Wears RL, Waeckerle JF: Effect of attendance at a training session on peer reviewer quality and performance. *Ann Emerg Med* 32:318–22, 1998
  28. Callaham ML, Schriger DL: Effect of structured workshop training on subsequent performance of journal peer reviewers. *Ann Emerg Med* 40:323–8, 2002
  29. Callaham ML, Knopp RK, Gallagher EJ: Effect of written feedback by editors on quality of reviews: two randomized trials. *JAMA* 287:2781–3, 2002
  30. van Rooyen S, Black N, Godlee F: Development of the Review Quality Instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol* 52:625–9, 1999
  31. Oxman AD, Guyatt GH, Singer J, *et al*: Agreement among reviewers of review articles. *J Clin Epidemiol* 44:91–8, 1991
  32. Thorndike EL: A constant error in psychology rating. *J Appl Psychol* 4:25–9, 1920
  33. Merton RK: The Matthew effect in science. *Science* 159:56–63, 1968
  34. Peters DP, Ceci SJ: Peer review practices of psychological journals: the fate of published articles, submitted again. *Behav Brain Sci* 5:187–95, 1982
  35. Pitkin PM, Burmeister LF: Identifying manuscript reviewers: randomized comparison of asking first or just sending. *JAMA* 287:2795–6, 2002
  36. Pitkin RM, Burmeister LF: Prodding tardy reviewers: a randomized comparison of telephone, fax and email. *JAMA* 287:2794–5, 2002
  37. Tite L, Schroter S: Why do peer reviewers decline to review? A survey. *J Epidemiol Community Health* 61:9–12, 2006
  38. Beggs C, Cho M, Eastwood S, *et al*: Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 276:637–39, 1996
  39. Huwiler-Müntener K, Jünker C, Egger M: Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA* 287:2801–4, 2002
  40. Hames I: *Peer Review and Manuscript Management in Scientific Journals*. Malden, MA: Blackwell, 2007
  41. Dougherty MC, Freda MC, Kearney MH, *et al*: Online survey of nursing journal peer reviewers: indicators of quality in manuscripts. *West J Nurs Res* 33:506–21, 2011
  42. Purcell GP, Donovan SL, Davidoff F: Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA* 280:227–47, 1998