AHA Scientific Statement

Peer Review Practices for Evaluating Biomedical Research Grants

A Scientific Statement From the American Heart Association

Lucy Liaw, PhD, FAHA, Chair; Jane E. Freedman, MD, FAHA; Lance B. Becker, MD, FAHA; Nehal N. Mehta, MD, FAHA; Laura Liscum, PhD; on behalf of the Peer Review Subcommittee of the American Heart Association National Research Committee; Council on Cardiovascular and Stroke Nursing; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Surgery and Anesthesia; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; Council on Hypertension; Council on Quality of Care and Outcomes Research; and Stroke Council

Abstract—The biomedical research enterprise depends on the fair and objective peer review of research grants, leading to the distribution of resources through efficient and robust competitive methods. In the United States, federal funding agencies and foundations collectively distribute billions of dollars annually to support biomedical research. For the American Heart Association, a Peer Review Subcommittee is charged with establishing the highest standards for peer review. This scientific statement reviews the current literature on peer review practices, describes the current American Heart Association peer review process and those of other agencies, analyzes the strengths and weaknesses of American Heart Association peer review practices, and recommends best practices for the future. (Circ Res. 2017;120:e000–e000. DOI: 10.1161/RES.0000000000000158.)

Key Words: AHA Scientific Statements ■ American Heart Association ■ biomedical research ■ financing, organized ■ foundations ■ National Institutes of Health ■ peer review

The biomedical research enterprise depends on the fair and objective peer review of research grants, leading to the distribution of scarce resources through efficient and robust competitive methods. In the United States, federal funding agencies and foundations such as the American Heart Association (AHA) collectively distribute approximately $117 billion annually to support biomedical research.1 It is important that we periodically examine our peer review policies and procedures to ensure that the best methods are being used.

For the AHA, a Peer Review Subcommittee is charged with establishing the highest standards for peer review. This review includes initializing policies and procedures that ensure that each application is reviewed by a committee whose membership has the expertise to provide optimal critical evaluation and feedback and is free of conflict or bias. To determine whether the AHA is achieving this goal, the Peer Review Subcommittee established a Review of Peer Review Task Force charged with reviewing current literature on peer review practices, describing the current AHA peer review process and those of other agencies, analyzing the strengths and weaknesses of AHA peer review practices, and recommending best practices for the future.
Review of Existing Literature on Peer Review Practices

The grant peer review cycle begins when an applicant submits an application to a funding agency, which convenes a panel of reviewers knowledgeable in the specific scientific area to critique the proposal. Once the science has been assessed, a second panel of evaluators typically will consider the critiques alongside the mandate of the funding agency and determine which applications should receive funding. Unsuccessful applications are often revised to address the reviewers’ comments and resubmitted. To ensure a fair grant review process, a funding agency must prevent conflicts of interest, promote confidentiality, promote transparency, allow opportunity for rebuttal, and encourage oversight and accountability.

Do funding agencies achieve the goal of fair review and optimal funding? Unfortunately, there is little empirical evidence that all aspects of current peer review mechanisms lead to the highest-quality biomedical research. In 2007, Demicheli and Di Pietrantonj\(^1\) articulated the need for a comprehensive assessment of peer review practices to determine their usefulness, soundness of methods and ethics, and accuracy. Such an all-inclusive assessment is still lacking. The quest for alternative grant review mechanisms is fueled in part by the impression that, although review panels are able to distinguish between good and poor proposals, they are less able to discriminate among good proposals.\(^4\) Success of the grant process is sometimes measured by the success of funded scientists. Indeed, scientists who receive funding have been shown to perform on a higher level (as measured by quantity and quality of publications) than those whose applications are not funded.\(^5\) However, that is a self-fulfilling prophecy. An additional concern is whether overarching mandates issued by funding agencies such as the National Institutes of Health (NIH) Roadmap lead to the highest-quality biomedical research capable of improving human health. Kuller\(^6\) argued that this systems biology approach will improve our description of disease but will not reduce disease burden. He maintained that a successful research agenda requires high-quality multidisciplinary research that links etiologic, phenomenological, and social/physical environment approaches. Ioannidis\(^7\) stated, “It is a scandal that billions of dollars are spent on research that has been shown to perform on a higher level (as measured by quantity and quality of publications) than those whose applications are not funded.” However, that is a self-fulfilling prophecy.

Most review panels select a subset of panelists to serve as primary, secondary, and tertiary reviewers for each application. Even if they are not primary or secondary reviewers, it is felt that all members of a grant review panel should have the opportunity to read the grant application and participate in the discussion and scoring.\(^6\) This removes the impact of extreme reviews and produces more consistency. At the same time, analysis has shown that if all members of the panel are equally responsible for reviewing an application, reviews can be more superficial than thorough and critical.

Snell\(^7\) analyzed postdoctoral fellowship applications to the Canadian Institutes for Health Research and found that luck plays an increased role in determining outcome when applications have few reviewers. He concluded that using 5 reviewers was optimal to avoid random effects, which are most problematic for applications in the middle range of competitiveness. Increasing the number of reviewers may add reliability, but it also increases direct and indirect costs to the process and adds burdens to the scientific community.

The benefit of better training for peer reviewers was demonstrated by Sattler et al.,\(^11\) who developed a training video that focused on specific, fundamental aspects of the grant review process. The 11-minute video covered the importance of the review process, how scores influence funding decisions, the meaning of each value on the rating scale, how to assign scores, and why it is important to understand the funding criteria of the specific agency. Novice and experienced reviewers were randomly assigned to view the video or to visit a website with information on the grant program and funding criteria. They then assigned ratings to specific examples of grant criteria and answered a questionnaire on the rating scale. The study showed that even this minimal training increased the reliability of grant rating for both novice and experienced reviewers. Abdoul et al.\(^14\) underscored the importance of reviewer training in grant assessment criteria.

The Review Panel

Most funding agencies have standing review panels that are chosen before applications are submitted, although some panels are assembled in response to the specific grants submitted. In either setting, these peer reviewers are selected with the expectation that the panel will be able to competently review any grant that is assigned. In addition, in the setting of multidisciplinary grants, the panel as a whole needs to have adequate scientific breadth.

It is generally accepted that applicants should not play a role in choosing the peer reviewers of their grant applications. However, applicants occasionally are able to suggest reviewers who have specific knowledge of their research area. An evaluation of Australian Research Council grant panels suggested that this practice is not advisable because reviewers who were nominated by the applicants typically submitted more favorable critiques than reviewers who were nominated by the panels. This finding was interpreted as evidence that reviews that are biased because of a personal connection between the reviewer and applicant.\(^8\) However, another interpretation is that the applicant-nominated reviewers were more knowledgeable about the subject matter or thought that the scientific area is more important than others.

Sometimes a standing review panel needs to solicit the opinions of ad hoc reviewers who have specific expertise. Often, these reviews have less impact on the final committee score than the reviews of the standing committee members. This conclusion was drawn from a retrospective analysis of grant-in-aid proposals to the Heart and Stroke Foundation of Ontario, which found that the mean score of standing committee members had a larger impact on final committee score than the mean score of ad hoc reviewers.\(^9\)

Panel Discussion of Applications

Does panel discussion of a grant application have an impact on its final priority score? We found different opinions on this issue. There appears to be consensus that the outcomes of highly ranked and poorly ranked applications are unchanged...
by discussion; however, discussion has an impact on the applications that are on the borderline of funding in most cases. A study by the Heart and Stroke Foundation of Ontario concluded that discussion of grants by the review panel is a critical part of the process. This study found that final committee scores were significantly different from initial scores, indicating that discussion is an important step.

A study of NIH applications by Martin et al found an effect of discussion on final grant scores. Examination of the peer review process for Research Project Grant (R01) applications at the NIH indicated that preliminary scores have a positive correlation with final outcomes. In most cases, the preliminary scores were not markedly different from the final scores of discussed grants. For 13% of the grants, however, the score adjustment, although small, had the potential to have an impact on funding.

Carpenter et al compared the effect of discussion in teleconference and face-to-face peer review panels for an NIH R01-like program offered by a US federal agency from 2009 to 2012. They found that the effect of discussion was small, on average, for both types of panels. However, like Martin et al, they found that discussion was important for ≈10% of the applications in both settings. In those cases, discussion more often resulted in a poorer score.

Carpenter et al also noted that the major differences between teleconference and face-to-face panels are the virtual versus in-person communication and the level of trust among reviewers. Trust forms through visual cues and the ability to socialize during breaks and meals. They cite previous studies showing that the communication setting can play a role in the panel’s commitment to the process.

Other analyses found that discussion had less impact on funding outcomes. Pina et al analyzed peer review of applications to a major European Union research program, Marie Curie Actions, and found very high consistency between scores before and after discussion. Given this high agreement, they suggested that the meeting itself could be eliminated without affecting the reliability of the grant review process. Fogelholm et al found that reviewer disagreement could be reduced by discussion, but they also found that using the mean of reviewers’ preliminary scores from their individual readings of the proposal, rather than a consensus score reached after discussion by the panel, does not compromise the reliability of the funding decisions and is both cost-effective and practical.

The Grant Itself

Several aspects of the grant application itself have been analyzed for their impact on funding and success of the research. These include the length of the application and consideration of the applicant’s track record.

Herbert et al assessed agreement on funding decisions on applications to the Australian National Health and Medical Research Council between (1) a 12-panelist review of 100-page applications; (2) a 7-panelist review of abbreviated 9-page versions, with 2-page summaries of applicant track record and limited discussion; and (3) a 2-panelist review of the 9-page versions, with references and synopses, similar to journal manuscript reviews. Review processes 2 and 3 each agreed with process 1 about as often as multiple reviews by full peer review systems have agreed in other studies. They concluded that simplifying the peer review process could save grant applicants and reviewers considerable time and resources, which could be devoted to the actual direct costs of research.

Bolli argued that the track record of an established applicant should be strongly considered when grant applications are judged. He noted that past productivity is the best gauge of what a grant applicant is capable of achieving. Several granting agencies concur with this and fund the applicant rather than a single, defined project (eg, Howard Hughes Medical Institute’s Investigator Award, NIH Outstanding Investigator Award). Early-career investigators, however, cannot rely on past productivity and need the traditional peer review process.

Several studies have found that evaluation of an applicant’s citation metrics (number of articles and citations, impact factor of journals, h-index) has, in most cases, no correlation with the peer review score of a grant. Lindner and Nakamura found that the bibliometric indexes of publications from funded projects are not an appropriate measure of the scientific impact of an application. Similarly, Kaltman et al and Doyle et al found no association between grant percentile ranking and an applicant’s bibliometric indexes. Derrick et al evaluated whether citation metrics correlated with peer assessment of a researcher’s influence on his/her field. They found that there was modest positive correlation in some fields but no correlation in others. They concluded that metrics could be combined with peer review only if the specific field has been analyzed. Clearly, better understanding of whether or how bibliometric indexes should be used requires further clarification. For instance, perhaps a specific range of years or narrowing types of publications (ie, excluding reviews) would lead to better value for bibliometric indexes.

Li and Agha found that better peer review scores correlate with better research outcomes. They analyzed publication and patenting outcomes of >130000 R01 grants funded by the NIH and concluded that the system works. They determined that percentile rank scores remain predictive of research success after adjustment for other factors such as experience of the principal investigator, previous funding awards, institutional quality, sex, and ethnicity. Gallo et al found a similar correlation between low (positive) peer review scores and high citation impact of the grant project. Li and Agha noted that some applications (≈1%) have poor priority scores but are nonetheless funded. These applications may be funded “out of order” because they are in subject areas that the NIH Advisory Council deems to be of high scientific interest. Seemingly contradictorily, Li and Agha found that these grants also have higher citation counts. They suggest that program officers who make these exceptions are able to identify applications capable of performing better than their initial score would suggest.

There are concerns about using citation counts or impact as a metric for research success. For example, Ioannidis noted that most published research findings are false. He postulated that research findings are less likely to be true when there is a financial stake. Competition for grant funds may lead investigators to publish research that cannot be replicated and will be refuted by subsequent studies. Furthermore, grant applications can be distorted by citation bias, for example, by
Choosing citations that support the proposed research, misrepresenting findings from cited research, and stating invested claims as fact.28

Consistent with the above concerns, Kaltman et al25 analyzed the outcomes of 1492 National Heart, Lung, and Blood Institute–funded R01s and found no association between percentile rank scores and subsequent citation impact of the project. They found, however, that the prior citation impact of the applicant was predictive of the citation impact of the grant project. The authors stated, “Analyses such as this one may identify factors, such as number of prior publications or prior citation impact, that more accurately predict the potential for future scientific impact.”

A more accurate prediction of potential research success might be achieved from analyzing “peer use” of an applicant’s published research.29 Peer use refers to research that has led to and thus been validated by successful research by scientific peers. Unfortunately, as noted by Charlton,27 testing science by use is retrospective and has the time scale of years. It is also incomplete because more findings are published than actually are evaluated by peer use.

Although an investigator’s track record, productivity, and environment were discussed above as potential predictors of achievement, potential bias caused by these factors may contribute to a lack of reliability of the peer review process.3 This has led to discussions of double-blind peer review, in which authors or applicants are anonymous to the reviewer and vice versa.30 The consequences of double-blind review of grants have not been rigorously studied. The Department of Defense uses double-blind review for preproposals to determine which should be invited to the full proposal state; however, review of full proposals is not double-blind. It is likely that current trials of different models of manuscript peer review, including double-blind review,31 will be informative for future discussions on improving the grant peer review process.

Comprehensive Proposals to Reform Peer Review

In his discussion of reforming peer review, Daniels32 offers steps that should be considered by the NIH. He recommends that the NIH encourage qualified and experienced senior scientists to serve on review panels, establish multidisciplinary review panels that may accept risky ideas, and remove outlier best and worst scores from consideration, which may allow riskier proposals that cannot achieve consensus to go forward.

There have been calls for more transparency in the grant application process. Currently, most funding agencies publish a list of funded projects; many agencies publish the abstracts of funded projects. Gurwitz et al33 proposed 3 measures that all agencies could implement quickly that would enhance the robustness of the peer review system. These measures are a published list of the members of the review panels and any external reviewers, a published impact statement of each funded project, and a public final report for each funded project. The last 2 suggestions would strengthen the public’s knowledge of and trust in scientific research. This sentiment was echoed by Mietchen.34

More radical changes to our current system of grant peer review were proposed by Bollen et al,35 who enumerated problems with our current system. These include the costs in time and money to assess grants and situations in which a minority of grants are funded, a process that inhibits serendipitous discovery and instead selects safe proposals. Bollen et al proposed a novel, highly decentralized mechanism in which each scientist would receive an unconditional, equal amount of money per year, with the requirement that each scientist pass on a fixed percentage to others who would make the best use of the money. The authors believe that this “simple, highly distributed, self-organizing process can yield sophisticated behavior at a global level.” It funds people rather than projects. Scientists would dynamically adjust funding levels as they assess each other’s merits. A computer simulation of the proposed system showed that it approximated the present distribution of NIH and National Science Foundation funding at a fraction of the cost. However, Avin et al36 worried that such a system would not distribute funds on the basis of scientific merit; instead, it would be unreliable and less effective than peer review. The only way to know whether such a decentralized mechanism would work is to empirically test it.

Ioannidis7 assessed various options for revamping the funding system: funding everyone, funding some according to merit (past accomplishments or automated indexes of scientific impact), or funding a few who articulate broad goals. A consideration of scientific citizenship could reward researchers who share data, are open to collaboration, share negative findings, and publish reproducible data, protocols, and algorithms. At this time, there are calls for open sharing of data from clinical trials; however, organized mechanisms for doing so are just emerging. Ioannidis noted that many agencies use a variety of funding mechanisms. The current system has led to scientists spending an inordinate amount of time writing, reviewing, and administering grants rather than expanding our knowledge base. Unfortunately, when the funding rate falls below 10% to 15%, investigators must submit many proposals to maintain research funding.38

Everyone agrees that changes to peer review should take place methodically because rapid changes can have unforeseen repercussions. Significant problems were identified when the Canadian Institutes for Health Research reformed its research grant programs and peer review process.39–41 The funding agency moved from face-to-face meetings to a virtual process with anonymous online reviews. The changes were said to “have undeniably compromised the quality and accuracy of the recommendations obtained.”39 The new process was riddled with conflicts of interest and resulted in a larger-than-expected variance in reviewer scores. The institute is working to correct these problems.

Description of AHA Peer Review Practices

Peer review practices at the AHA strive to provide fair and equitable review of scientific proposals from a diverse group of scientists who and disciplines that deal with improving health and better understanding heart disease and stroke. What does it mean to be a grant applicant’s peer? A peer must have expertise in the scientific area of the grant and familiarity with the techniques and approaches that will be used. A peer without this experience may not fully understand the challenges and significance of the application. At the same time, a peer who is too close to the subject may be biased...
toward or against the applicant, entrenched in his/her own views, not willing to support an alternative hypothesis, or focused on the application’s experimental detail. The AHA aims not only to conduct fair and competent peer review but also to perform ongoing self-evaluation and revision of existing peer review practices.

Peer Review at the AHA

The AHA currently reviews applications through a national process that aims to provide scientific rigor and transparency to rank scientific proposals. There are 2 general routes of peer review and oversight. The first is the standard process for long-standing portfolio opportunities for grants and fellowships (blue boxes in the Figure). In this scenario, the AHA maintains standing peer review committees that cover specified scientific areas. These committees evaluate the following funding opportunities: Predoctoral Fellowship, Postdoctoral Fellowship, Mentored Clinical and Population Research Award, Beginning Grant-in-Aid, Scientist Development Grant, and Grant-in-Aid. Program-specific review committees (in green) are established for the Innovative Research Grant and the Established Investigator Award, as well as for the recently added Strategically Focused Research Networks, Collaborative Science, Mentor-AHA Mentee Award, and the Cardiovascular Genome-Phenome Study research programs.

The Research Committee of the AHA has oversight for the research portfolio and reviews and recommends policies and strategies related to funding and peer review. Its ongoing assessment of peer review conduct and policies allows relatively rapid modification and improvement of processes. For instance, this has recently included standardization of the biosketch to be consistent with the NIH and adapting NIH guidelines on inclusion on the basis of sex, race, ethnicity, and age in clinical research. The Research Committee carefully considers NIH policy changes as they occur to simplify the submission and review processes for applicants and reviewers. The Research Committee reports to the Science Advisory and Coordinating Committee and to the Board of Directors. Within the Research Committee, there are topical working groups, including the Peer Review Subcommittee. These groups are made up of professional researchers, clinicians, and, at times, qualified lay reviewers who serve as volunteers in furthering the mission of the AHA. The Science Advisory and Coordinating Committee and the Research Committee oversee the peer review committees, which are divided by scientific category and by research type. These science classifications are re-evaluated periodically. In 2016, there were 121 peer review committee meetings covering basic science (64 in the spring and 57 in the fall); 41 peer review committee meetings covering clinical and translational research, population, outcomes, and epidemiology (21 in the spring and 20 in the fall); and 20 peer review committee meetings (12 in the spring, 8 in the fall) that spanned basic, clinical, and population areas in specialized topics that received a lower volume of applications; for example, cardiac arrest and resuscitation.

Each Peer Review Committee is administered at the national level, led by a chair and co-chair, supported by an AHA
Several important changes have been made to the peer review structure over the past decade. Unified peer review administered at the national level rather than by AHA affiliates led to a critical mass of applications for each science classification and has been shown statistically to lead to fairer funding when rankings between groups are compared. To maximize the funding dollars, beginning in 2010, in-person study sections at the AHA began to transition to teleconference meetings. Holding teleconference peer review meetings allowed the AHA to increase the number of panels, to add scientific specificity, and to optimize dollars to grantees. In 2009, $1,544,760 was spent on in-person meetings, and that money is now put toward grant programs. Next, after a successful pilot program, an initiative in spring 2014 led to the inclusion of lay reviewers as members of select peer review committees. A lay reviewer is defined as an individual without formal science training or research background who has a strong interest in the prevention or management of heart disease and stroke. Lay reviewers are individuals who endorse the mission of the AHA and volunteer to review applications or applicants for the purpose of funding science. They receive appropriate training and are charged, as part of the peer review process, to focus on the lay summary within each application and to evaluate how the proposal will affect the mission of the AHA. It is anticipated that lay stakeholders will continue to serve on select study sections and strategic committees.

The majority of members on each peer review committee are scientific experts who are standing members on the committee and are chosen on the basis of several criteria. A scientific peer reviewer must be at a minimum career level of assistant professor or equivalent, have an independent research program and recognized competence in his/her field of research, have a history of research funding, and be actively publishing scientific articles. The credentials of potential peer reviewers are vetted by AHA staff and reviewed by the Peer Review Subcommittee to determine suitability for any particular committee. A peer reviewer must be willing to commit to attending meetings and reviewing up to 10 applications per meeting. Appointment as a regular member involves 3 to 4 years of service. After completion of 4 years of service, a peer reviewer must complete a 1-year hiatus before returning to a peer review committee. These requirements are online as public information. The composition of the science experts in a peer review committee is decided by consideration of the subject area covered by the review panel. In addition, the population of the committee should be diverse, including representation from all geographic regions of the country, career levels, sexes, races/ethnicities, and professional affiliations.

On occasion, ad hoc or consultant reviewers are invited to the peer review committee for a particular review cycle to provide additional expertise. An ad hoc reviewer has the same basic criteria for selection as a standing committee member but is appointed on a temporary basis for ≥1 meetings. An ad hoc reviewer has the same reviewer load as a regular standing member. By policy, ad hoc reviewers cannot constitute more than one third of the committee roster at any meeting. A consultant reviewer must also fulfill the same criteria for selection as a standing committee member and may be appointed on a temporary basis to provide needed expertise for just a few applications (1–3 applications). Consultant reviewers may be brought in for particular applications for which expertise does not exist within the committee.

AHA reviewers (both scientific and lay) undergo extensive online training that orients participants to a wide range of topics, including how to review a grant and understanding conflict of interest. This online training is particularly important for early, first-time, and lay reviewers, but it provides a framework for all reviewers to standardize the review process. In addition, new reviewers are mentored by the vice chairs of the study sections. Each application is assigned 3 reviewers who provide a score and written critique. Impact on the AHA mission is considered a specific review criterion. Final scoring of each application occurs at the peer review meeting after discussion of the applications. The funding decisions that result from peer review are determined by the AHA Funding Subcommittee, which makes decisions about the distribution of research funds. Applications to standard study sections that fail to receive funding can be revised and resubmitted 2 additional times with a response to reviewers. In those cases, applicant comments are taken into consideration in the review process. Both the original application and critiques can be viewed by the reviewers of the resubmission application.

The AHA has made every effort to make its peer review process transparent to all stakeholders. The peer review committee members, success rates for each award, and award recipients are available on the AHA website.

Other Organizations

We performed a comprehensive review of similar foundations and institutes supporting biomedical research programs in 2012 with the goal of identifying and sharing best practices. These organizations were selected on the basis of their administration of significant research dollars (from $5 million to >$500 million annually), types of funding opportunities similar to those of the AHA (training awards, early-career awards, investigator-initiated research grants, targeted research areas), and their willingness to be interviewed for information on peer review practices. Major points of comparison are summarized in the Table. Many organizations share common practices of peer review, and some had some unique aspects. This section briefly covers common shared practices and highlights unique practices.

The foundations and institutes we surveyed support a variety of research funding mechanisms. Most offer research grants to early-career independent investigators and fellowships to scientists in training. Some have unique grant opportunities. These include the Innovative Research Grant and the Mentored Clinical and Population Research Award offered by the AHA, American Cancer Society Professorships...
<table>
<thead>
<tr>
<th>Entity</th>
<th>Research Budget per Year</th>
<th>Types of Awards</th>
<th>Applications/ Review Cycles per Year</th>
<th>Peer Reviewer Roles</th>
<th>Review Overview</th>
<th>Scoring and Review Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Heart Association</td>
<td>$149 million</td>
<td>Research Fellowships</td>
<td>6000/2</td>
<td>Review up to 10 proposals Teleconference 2 times a year 3- to 4-y term</td>
<td>3 Reviewers per proposal Streamlining system for lower half</td>
<td>1–5 Scoring scale Review criteria similar to that of the NIH</td>
</tr>
<tr>
<td>American Diabetes Association</td>
<td>$34 million</td>
<td>Research Training</td>
<td>1000/2</td>
<td>Review up to 10 proposals In-person meeting 2 times a year 3-y term</td>
<td>2-Phase review 20%–25% advance 5-y retrospective analysis every 3 y</td>
<td>1–5 Scoring scale Same review criteria as NIH</td>
</tr>
<tr>
<td>American Cancer Society</td>
<td>$90 million</td>
<td>Research Mentored awards</td>
<td>2000/2</td>
<td>Review 6–8 proposals In-person meeting All proposals discussed</td>
<td>3 Reviewers per proposal All proposals discussed</td>
<td>1–5 Scoring scale</td>
</tr>
<tr>
<td>American Lung Association</td>
<td>$5 million</td>
<td>Early career Senior investigator</td>
<td>200</td>
<td>In-person meeting 2 Reviewers per proposal Streamlining system 25%–30% funded</td>
<td>2 Reviewers per proposal Streamlining system 25%–30% funded</td>
<td>1–5 Scoring scale Same review criteria as NIH</td>
</tr>
<tr>
<td>Howard Hughes Medical Institute</td>
<td>$80 million (grants)</td>
<td>Science education fellowship Early career Investigators</td>
<td>3200</td>
<td>Review panel of senior researchers Investigators Reevaluated every 5 y</td>
<td>4 Reviewers per proposal Competitions held every 2–3 y ≤2.5% funded</td>
<td>1–5 Scoring scale for new applications A, B, C scale and category ballot for renewals Applicants present to a panel</td>
</tr>
<tr>
<td>Leukemia &amp; Lymphoma Society</td>
<td>$76 million</td>
<td>Career development SCOR</td>
<td>1000</td>
<td>Review 8–10 proposals Letters of intent reviewed by phone</td>
<td>Letters of intent reviewed on phone Invited full applications reviewed in person</td>
<td>1–5 Scoring scale Additional mission-related score</td>
</tr>
<tr>
<td>March of Dimes</td>
<td>$25 million</td>
<td>Scholarships Research</td>
<td>1000</td>
<td>Reviewers selected by VP of Research and Global Programs and chair Review 10–20 proposals In-person review of grants 6-y term</td>
<td>Letters of intent reviewed 80%–90% invited for full proposal Streamlining process</td>
<td>1–5 Scoring scale Same review criteria as NIH</td>
</tr>
<tr>
<td>Robert Wood Johnson Foundation</td>
<td>$300 million</td>
<td>7 Program areas</td>
<td>2900</td>
<td>Review 10–15 proposals In-person and teleconference reviews 4–5 d/y 2-y term</td>
<td>National program office selects proposals for invitation to full proposal</td>
<td>Scoring scale varies depending on program Innovation is a major criterion</td>
</tr>
</tbody>
</table>
and International Fellowships, Howard Hughes Medical Institute Science Education Grants, special initiatives from the Leukemia & Lymphoma Society, and Transdisciplinary Research Centers funded by the March of Dimes.

The peer review process used by these foundations is quite similar. Experts in the field serve on review panels that meet once or twice a year for 1 to 2 days per meeting. Most reviewers critique up to 10 proposals per meeting, although some foundation reviewers have a heavier load. Most proposals are evaluated by 3 reviewers, who assign numerical scores to each grant according to specific criteria. Scores and written critiques are submitted before the meeting. Most foundations use a 1-to-5 scoring scale, with review criteria similar to that of the NIH. Most also use a streamlining, or triage, system in which only the most meritorious proposals are discussed. The American Cancer Society is an exception; its reviewers discuss all proposals. The Howard Hughes Medical Institute is unique in that its review process focuses on the applicant rather than details of the project.

Most of the foundations we surveyed conduct their reviews at a face-to-face meeting. The AHA switched from all face-to-face meetings to primarily teleconference meetings for the fall 2010 cycle. This has yielded a considerable cost savings with no apparent detrimental effect on the ability of the review panel to identify the strongest applications. The AHA conducted a survey of 980 peer reviewers in spring 2013, and the comments indicated that teleconferences did not affect the quality of the process. Some reviewers miss the collegial interactions they experienced at face-to-face meetings, whereas others appreciate the lack of travel.

Letters of intent are currently required by the AHA for the Collaborative Science Award and the Merit Award. The letter of intent ensures that applications are responsive to the specific program and that reviewers can be assembled to assess the specific topics submitted. The Leukemia & Lymphoma Society and March of Dimes request letters of intent for all of their grants; the Susan G. Komen Foundation requires the submission of “preapplications,” which are scored such that only the top 20% to 25% of the applicants submit the full application. This screening process allows granting agencies to select the most responsive applications and to assemble appropriate review panels in advance and significantly reduces the number of grants that are reviewed.

The review process adopted by the American Cancer Society features a unique “parking system.” It distributes funds to the most highly ranked applications until it reaches the end of the list of meritorious applications or until funds run out. Any grant that was on the should-be-funded list but failed to receive funding is put on a pay-if list. The American Cancer Society pays that grant if funds become available from another source.

Several foundations have begun to conduct retrospective analyses of their past awardees to track their individual rates of success. Although tracking and evaluation are difficult and metrics are not standardized, common postaward analyses include progress in career development (promotions, awards, current position), subsequent grant awards, and publication record. These are intended to inform foundations about the impact of their training programs and to help them revise their peer review practices.

**Recommendations for Best Practices**

Our review of peer review found several themes. Peer review procedures and practices have a deep history and robust procedures. A remarkable amount of thought and energy go into peer review. Robust peer review is considered the state-of-the-art standard when it comes to the evaluation of scientific proposals for funding. However, it is not a perfect process, and legitimate concerns about specific areas of peer review exist.

The published evidence base for the effectiveness of peer review is quite weak. Whether current peer review processes actually select the most meritorious applications is an ongoing, unanswered question. There are relatively few studies on the actual effectiveness of peer review in the world’s literature. We were not able to identify any randomized controlled trials involving peer review. Specifically, we were unable to identify even a single study that compared peer-reviewed with non–peer-reviewed grant applications. No articles could be identified that compared commonly variable practices in peer review such as how many experts are required for peer review and whether there is an optimal weighting that can be used within peer review scoring to select for a particular type of discovery outcome.

The general practice of peer review at the AHA is quite similar in concept and practice to peer review by the NIH. A few organizations conduct peer review with some unique

---

**Table. Continued**

<table>
<thead>
<tr>
<th>Entity</th>
<th>Research Budget per Year</th>
<th>Types of Awards</th>
<th>Applications/Review Cycles per Year</th>
<th>Peer Reviewer Roles</th>
<th>Review Overview</th>
<th>Scoring and Review Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susan G. Komen Foundation</td>
<td>$58 million</td>
<td>Training, Career catalyst, Investigator initiated, Clinical trials</td>
<td>1500</td>
<td>Review by Komen scholars, In-person review of grants</td>
<td>Preapplications are reviewed, 20%-25% invited for full applications, 3 Reviewers and 1 advocate, 27% Funding rate</td>
<td>1-5 Scoring scale, Same review criteria as NIH, Breast cancer advocate involved in review</td>
</tr>
</tbody>
</table>

NIH indicates National Institutes of Health; SCOR, Specialized Center of Research Program; and VP, vice president.
practices whereby a higher value is placed on certain characteristics (eg, innovation); however, direct comparisons of these alternative styles of peer review and more traditional styles of peer review have not been published.

An important notion emerged from our review of peer review practices. It is difficult to compare grant applications that are very different in purpose in a single review committee. For example, it is difficult to compare the value (or score) of training grants that focus on the potential of a candidate with the value of original investigator-initiated research. This is particularly true if these different applications are rated by a single group on a similar scoring scale. Another important concept is that excellence and innovation are fundamentally quite different and may be difficult to evaluate simultaneously. A provocative presentation to the NIH as part of the 2006 Peer Review Advisory Committee observed that new ideas almost by definition will not gain consensus and suggested that we should evaluate true innovation by a metric related to the lack of agreement rather than a traditional score. The author proposed that we actively measure controversy as a surrogate for innovation with a new methodology, suggesting that we explore metrics such as variance or negative kurtosis (the degree that observations occur in the tails of the grading distribution) and place value on this when we seek proposals of high innovation.

After the process of interviews was completed for this study, the NIH, in partnership with Nature Publishing Group and Science, convened a workshop to address rigor and reproducibility in research. This assembly and the resulting recommendations were spurred by publications recognizing that much of the reported research results in the literature were not reproducible. The Principles and Guidelines in Reporting Preclinical Research have been endorsed by many editors and publishers of scientific literature. In addition, new requirements by the NIH for a discussion of rigor and reproducibility and authentication of key biological resources have brought awareness of these issues to the grant application and peer review process. We recommend that, moving forward, organizations that fund research should carefully consider how rigor and reproducibility can be incorporated as part of the peer review process.

Summary and Recommendations

Peer review determines the allocation of scientific resources and thus has great influence on the direction and rate of scientific discovery. However, for both technical and political reasons, the effectiveness of peer review processes and whether peer review actually directs resources to maximal public benefit are difficult to evaluate scientifically. The difficulties involved are genuine, but with modest efforts, more progress could be made in this vital area. Particularly when research resources are scarce and constrained, there may be a high price for our failure to more aggressively perform research into these challenges.

1. Robust peer review requires a great deal of effort and time on the part of staff, reviewers, and applicants. Organizations that fund research should be prepared to assess and attempt to improve the effectiveness and value of peer review processes. It seems illogical to invest millions of dollars in research without investing funds to assess the strengths and weaknesses of the distribution methods.

2. Literature to support an evidence-based evaluation of peer review is lacking. Organizations should consider funding scientific study of peer review. Performing randomized controlled trials on innovative aspects of peer review should be considered. For example, comparing grant applications of differing lengths, using differing numbers of peer reviewers, developing specific metrics that better reflect innovation, or using a 2-stage peer review could be studied. Randomized controlled trials are feasible, the gold standard across the science community, and have yet to be performed on peer review methods.

3. A more formal method for sharing peer review practices between organizations should be encouraged. Most organizations have a common goal of funding the most meritorious grant applications, but only rarely do these organizations communicate on the methods by which they distribute funding.

4. Specific peer review procedures for special purposes (eg, to identify extremely innovative grants) should be considered experimental and undergo some academic rigor that would include evaluation and testing. As studies are performed on peer review, it may be possible in the future to more accurately match the goals of a particular grant purpose with the relative scoring and weighting of the grant review process.

5. The mathematical and technical aspects of scoring grants should also undergo some evaluation and scrutiny. Organizations may be able to improve their peer review scoring methods. Weighting, normalization, statistical analysis, and attention to large deviations in scoring may be methods that can be harnessed to improve peer review.
Disclosures

Writing Group Disclosures

<table>
<thead>
<tr>
<th>Writing Group Member</th>
<th>Employment</th>
<th>Research Grant</th>
<th>Other Research Support</th>
<th>Speakers’ Bureau/ Honoraria</th>
<th>Expert Witness</th>
<th>Ownership Interest</th>
<th>Consultant/ Advisory Board</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucy Liaw</td>
<td>Maine Medical Center Research Institute Center for Molecular Medicine</td>
<td>NIH*</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Lance B. Becker</td>
<td>Hofstra Northwell School of Medicine, North Shore University Hospital, Long Island Jewish Medical Center</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Jane E. Freedman</td>
<td>University of Massachusetts Medical/University of Massachusetts Medical School</td>
<td>NIH*</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Laura Liscum</td>
<td>Tufts University School of Medicine</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Nehal N. Mehta</td>
<td>National Heart, Lung, and Blood Institute</td>
<td>NIH*</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

This table represents the relationships of writing group members that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all members of the writing group are required to complete and submit. A relationship is considered to be “significant” if (a) the person receives $10,000 or more during any 12-month period, or 5% or more of the person’s gross income; or (b) the person owns 5% or more of the voting stock or share of the entity, or owns $10,000 or more of the fair market value of the entity. A relationship is considered to be “modest” if it is less than “significant” under the preceding definition.

Reviewer Disclosures

<table>
<thead>
<tr>
<th>Reviewer</th>
<th>Employment</th>
<th>Research Grant</th>
<th>Other Research Support</th>
<th>Speakers’ Bureau/ Honoraria</th>
<th>Expert Witness</th>
<th>Ownership Interest</th>
<th>Consultant/ Advisory Board</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steven D. Barger</td>
<td>Northern Arizona University</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>James H. Cauraugh</td>
<td>University of Florida</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Peter B. Imrey</td>
<td>Cleveland Clinic</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

This table represents the relationships of reviewers that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all reviewers are required to complete and submit. A relationship is considered to be “significant” if (a) the person receives $10,000 or more during any 12-month period, or 5% or more of the person’s gross income; or (b) the person owns 5% or more of the voting stock or share of the entity, or owns $10,000 or more of the fair market value of the entity. A relationship is considered to be “modest” if it is less than “significant” under the preceding definition.

References

15. Martin MP, Kopstein A, Janice JM. An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed by the
Peer Review Practices for Evaluating Biomedical Research Grants: A Scientific Statement From the American Heart Association

Lucy Liaw, Jane E. Freedman, Lance B. Becker, Nehal N. Mehta, Laura Liscum and on behalf of the Peer Review Subcommittee of the American Heart Association National Research Committee; Council on Cardiovascular and Stroke Nursing; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Surgery and Anesthesia; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; Council on Hypertension; Council on Quality of Care and Outcomes Research; and Stroke Council