

NEW SCORING PROCEDURES FOR EVALUATION OF RESEARCH APPLICATIONS

NOT-OD-09-024 12/2/2008

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html>

Background

The mission of the NIH is to support science in pursuit of knowledge about the biology and behavior of living systems and to apply that knowledge to extend healthy life and reduce the burdens of illness and disability. As part of this mission, applications submitted to the NIH for grants or cooperative agreements to support biomedical and behavioral research are evaluated for scientific and technical merit through the NIH peer review system. In June 2007, the NIH initiated a formal, agency-wide effort to review the NIH peer review system (<http://enhancing-peer-review.nih.gov/>). After careful deliberation and consideration of the recommendations resulting from this year-long effort, a number of key actions will be implemented in the NIH peer review system.

In current practice, each scored application is assigned a single, overall priority score that reflects the consideration of all review criteria. Individual reviewers assign scores on a 1 to 5 scale in 0.1 increments (e.g., 2.2), resulting in 41 possible rating discriminations for reviewers to make. The reviewers' individual scores then are averaged and multiplied by 100 to yield a single overall priority score for each scored application (e.g., 253).

Although this rating system has served the NIH and the research community well, several concerns led the NIH to consider a revised rating system for grant applications. Making 41 discriminations is difficult for reviewers to do reliably, and scores increasingly have become compressed toward the positive end of the scale. In addition, by averaging reviewer scores and multiplying by 100, the resulting priority score appears to have more precision than it actually has. To address these concerns, the NIH considered scoring systems with fewer rating options to increase potential reliability and with sufficient range and appropriate anchors to encourage reviewers to use the full scale. To increase transparency, the NIH also considered methods to communicate ratings from assigned reviewers even when the application is streamlined and not discussed, or discussed and scored by the full committee.

Implementation

New Scoring System. The new scoring system will be effective for all applications for research grants and cooperative agreements that are submitted for funding consideration for fiscal year 2010 (FY2010) and thereafter. The first standing due date for FY2010 is January 25, 2009; the new scoring system will be used for applications submitted in response to Parent Announcements and Program Announcements, including PARs and PASs published before or after this Guide Notice. An important aspect of the implementation of the new scoring system is to use it in a consistent manner for applications considered in a given fiscal year. Therefore, some RFAs and PARs for funding consideration in FY2010 have due dates before January 25, 2009, and responses to those will be evaluated using the new scoring system. Likewise some RFAs and PARs for FY2009 have due dates after January 25, 2009, and responses to those will be evaluated using the present scoring system.

The new scoring system will utilize a 9-point rating scale (1 = exceptional; 9 = poor).

Not Recommended for Further Consideration. An application may be designated Not Recommended for Further Consideration (NRFC) by the Scientific Review Group if it lacks significant and substantial merit; presents serious ethical problems in the protection of human subjects from research risks; or presents serious ethical problems in the use of vertebrate animals, biohazards, and/or select agents. Applications designated as NRFC do not proceed to the second level of peer review (National Advisory Council/Board) because they cannot be funded.

Percentile Rankings. Percentile rankings will be calculated anew, starting with scores from the May 2009 cycle of review, and reported to the nearest whole number.

Scores for Individual Criteria. Before the review meeting, each reviewer and discussant assigned to an application will give a separate score for each of five core review criteria (Significance, Investigator(s), Innovation, Approach, and Environment). For all applications, even those not discussed by the full committee, the scores of the assigned reviewers and discussant(s) for these criteria will be reported individually on the summary statement.

Priority Scores – Discussed Applications. Before the review meeting, each reviewer and discussant assigned to an application will give a preliminary impact score for that application. The preliminary impact scores will be used to determine which applications will be discussed. For each application that is discussed, a final impact score will be given by each eligible committee member (without conflicts of interest). Each member's impact score will reflect his/her evaluation of the overall impact that the project is likely to have on the research field(s) involved, rather than a weighted average applied to the reviewer's scores given to each criterion (see above).

The overall impact score for each discussed application will be determined by calculating the mean score from all the eligible members' impact scores, and multiplying the average by 10; the overall impact score will be reported on the summary statement. Thus, the 81 possible overall impact scores will range from 10 - 90. (Overall impact scores will not be reported for applications that are not discussed.)

Funding Decisions. The new scoring system may produce more applications with identical scores ("tie" scores). Thus, other important factors, such as mission relevance and portfolio balance, will be considered in making funding decisions when grant applications are considered essentially equivalent on overall impact, based on reviewer ratings.