Center for Scientific Review
Peer Review Advisory Committee Meeting
National Institutes of Health
U.S. Department of Health and Human Services

September 26, 2005

The third meeting of the Peer Review Advisory Committee (PRAC) convened at 8:00 a.m. on September 26, 2005, at the Bethesda Marriott, Bethesda, Maryland. The entire meeting was held in open session. Dr. Antonio Scarpa and Dr. Jeremy Berg presided as Co-Chairs.

Members Present

Jeremy Berg, Ph.D., Co-Chair
Antonio Scarpa, M.D., Ph.D., Co-Chair
Faye Calhoun, Ph.D.
Joe Martinez, Ph.D.
Craig McClain, M.D.
Edward N. Pugh, Jr., Ph.D.
Louise Ramm, Ph.D.
Anne P. Sassaman, Ph.D.
Beverly Torok-Storb, Ph.D. (phone)

Ad hoc Member
Dean E. Brenner, M.D.

Dr. Karl Malik was the Executive Secretary for the meeting.

Welcome, Introductions, and Approval of the May 2005 PRAC Meeting Minutes

Dr. Scarpa welcomed participants to the third Peer Review Advisory Committee meeting and asked PRAC members to introduce themselves. He asked for approval of the minutes from the May 2005 PRAC meeting. Dr. Joe Martinez moved to approve the minutes, which was seconded and then voted on unanimously.

Dr. Scarpa extended a special thank you for his work for PRAC to Dr. Karl Malik, who is moving to another position within the National Institutes of Health (NIH).

Electronic Receipt of Grant Applications

Ms. Megan Columbus, NIH Program Manager for Electronic Receipt of Grant Applications, presented a timeline for NIH receipt of electronic submissions, which also involves a conversion from PHS 398 to Standard Form 424 for grant applications submitted to NIH. The transition to electronic application submission, combined with this new set of application forms, is a huge initiative with an aggressive timetable.

Timeline
By the end of May 2007, the goal is that all grants to NIH will be submitted electronically through the online portal Grants.gov. Ms. Columbus said that the transition to electronic receipt

will benefit the applicant community and NIH by eliminating the burden of paper-based data collection, possibly shortening the cycle from application receipt to award, creating a comprehensive repository of data, improving data quality, saving costs, and improving the graphic quality of the scanned applications. The switch to the SF 424 family of forms means that applicants will use standard forms to apply to many federal agencies, and it reduces the administrative burden on the federal grants community. SF 424(R&R) is the government-wide form for research grant applications, and will be used at NIH.

Several factors drive this timeline. First, Public Law 106-107 and the President's Management Agenda both call for consolidation of Federal grant-making. Further, the goal of the Office of Management and Budget (OMB) is that 75 percent of agencies' funding opportunities in the "Find" section of the Grants.gov Web site be available to apply for through the portal in FY 2006. In addition, OMB clearance for the existing PHS398 expires in September 2007. Ms. Columbus described the components of the SF424(R&R) application, as well as agency-specific components. These latter are called PHS398, but differ from the PHS398 in current use.

NIH's strategy is to transition mechanism-by-mechanism, beginning with requiring that SBIR/STTR applications, due December 1, 2005 be submitted electronically on an SF424. This will be followed by R13/U13s, and then R15s. Mechanisms not yet transitioned will continue to come in on the PHS 398, either on paper or electronically. The goal is to give the affected communities as much notice as possible as each mechanism transitions. Looking ahead, Ms. Columbus said the R03, R21, and then R01 grant announcements will be on Grants.gov in 2006, with all mechanisms submitted via Grants.gov by June 2007.

Ms. Columbus emphasized that applicants' institutions need to register with Grants.gov, and that 4 weeks should be allowed for this process. Applicants also should be registered in the NIH Commons. Applicant institutions can use Pure Edge, a free software tool, to submit applications to Grants.gov, or they can create or use a service provider's XML-capable electronic system. Pure Edge is not Macintosh-compatible, which necessitates a workaround for Mac users.

Ms. Columbus explained the steps for finding grant opportunities and submitting electronic applications via Grants.gov. She noted that eRA software checks applications against NIH business rules and applicants are notified if their application is accepted or if changes are needed before acceptance. Several Web sites will provide applicants with more information about the new process.

Discussion

Impact on institutions: Dr. Edward Pugh asked Ms. Columbus whether the information about electronic receipt is going out to institutions with enough lead-time. He was particularly concerned about smaller institutions and wanted to ensure that none felt at a disadvantage. Ms. Columbus said that a communications plan is being developed. For example, Institute and Center (IC) officials will make presentations at scientific meetings and targeted e-mails will be sent. Dr. Pugh suggested finding a way for institutions to sign off to affirm that they were notified about the new procedures.

Dr. Anne Sassaman asked the extent to which institutions already use Grants.gov, perhaps with other agencies. Ms. Columbus was unsure of exact numbers, but offered to research the figure.

Dr. Louise Ramm expressed concern that the R15 is the second mechanism to use Grants.gov, since recipients are typically smaller institutions. Ms. Columbus replied that targeted communications are going out to them. In response to a follow-up question from Dr. Ramm, Ms. Columbus explained the ways that Macintosh users could submit a grant without Pure Edge.

Contingency plans: Dr. Beverly Torok-Storb asked about contingency plans if the system fails. Ms. Columbus asked Dr. Suzanne Fisher, Director of the CSR Division of Receipt and Referral and head of the contingency planning team, to describe her team's efforts. Dr. Fisher also observed that, in her experience, smaller institutions are usually very careful with their grant submissions. Dr. Torok-Storb asked for clarification on the "business rules" against which an application is checked. Dr. Fisher explained that the rules are somewhat mechanism-specific, such as whether a modular budget is required, but that applicants are alerted to go to the NIH Commons if a problem is detected that needs their attention.

Dr. Faye Calhoun suggested regional, face-to-face meetings, given the scope of the changes. Ms. Columbus said that plans were under way to attend professional meetings at the national and regional levels, but NIH-convened regional meetings might also be a good idea.

Dr. Dean Brenner expressed concern that the system may not be able to handle the large number of applications submitted at the deadline. Ms. Columbus said that contingency plans would handle that concern. Dr. Scarpa said that staggering submission might help relieve the pressure. Dr. Brenner stressed the importance of fairness so that no applicant receives extra time to prepare an application. Ms. Columbus said that NIH has been working with Grants.gov to alert them to the potential volume that could occur at each deadline. In addition, ramping up slowly, starting with the SBIR, will be helpful. She said that fairness is a core principle in all planning.

Users' reactions: Ms. Columbus concluded by stressing that the delivery system and application package is different, but the same information is required. Dr. Fisher called electronic submission an essential piece to have in place to shorten review cycles. In response to a question about training workshops from Dr. Craig McClain, Ms. Columbus described a test in which groups of investigators, ranging from post-docs to very senior researchers, filled out the SF424, and used the application guide. The message was that training is essential, so this is being looked at, perhaps at national meetings and through train-the-trainer approaches.

Dr. Sassaman suggested making sure that NIH Program staff feel comfortable with the new system so they can educate applicants with whom they interact. Dr. Ramm said that a review group at the National Center for Research Resources (NCRR) looks at applications that use PHS398 and SF424 formats. The reviewers have had no problem with the SF424 and could easily focus on the research plan. Dr. Pugh also suggested using all NIH websites to alert people to the changes.

Report from the New CSR Director: Vision and Future Directions

Dr. Scarpa began his presentation by expressing what an honor and privilege it is to work at NIH, and his belief in peer review to provide the best research and cures to the country. CSR will process about 78,000 applications this year and review about 60 percent of the total. In addition, said Dr. Scarpa, CSR has two unique characteristics: to serve as a major portal between investigators and NIH and to fulfill its mission statement of ensuring that NIH applications "receive fair, independent, expert and timely reviews—free from inappropriate influences-so NIH can fund the most promising research." This mission statement is understood and supported by all stakeholders.

Dr. Scarpa said that his understanding of what needs to be done to improve peer review has changed as a result of many discussions he has had with IC Directors, leaders of scientific societies and others in this country and abroad. Beyond electronic and other changes in peer review itself, there have been changes in diseases of the American people to more chronic conditions, and changes in how research is done. More people now collaborate on studies, and they have less protected time, especially in clinical departments. CSR must keep up with these changes to fulfill its mission. Because peer review is CSR's sole mission, it should be a "laboratory" to find better ways of doing peer review. He divided potential changes into three matrices depending on complexity, impact, needs of different stakeholders, and time:
(1) Changes within CSR; (2) Changes within the current system of grants; and (3) Possible new systems.

Changes in CSR Operations
Possible changes center on strategies to increase communications, increase uniformity, increase efficiency, and facilitate the work of IC program staff:

Increasing communications: Dr. Scarpa described efforts to improve internal and external communications, including an open-door policy for staff and an expanded CSR Review Notes newsletter.

Increasing uniformity: More consistency across study sections is needed in, for example, how resumes are written and critiques are posted on the Internet. CSR is in the process of developing policies to make these processes more uniform.

Increasing efficiency: Knowledge management tools, such as Collexis software now used for disease coding, are being looked at in four pilots. For example, these tools may facilitate assigning applications to Integrated Review Groups (IRGs) and selecting reviewers.

Facilitating work of program staff: As an example, Dr. Scarpa explained how secure speakerphones will allow Program staff to "attend" study section meetings efficiently. They can phone into a meeting, rather than spend time traveling from meeting to meeting.

Changes in Systems
Larger systemic changes that concern Dr. Scarpa include: shortening the review cycle, addressing the concern that clinical research is not properly evaluated, improving the assessment of innovative research, and doing more to recruit and retain more high-quality reviewers.

A trans-NIH committee that includes CSR staff is looking at various options to shorten the review cycle, and its recommendations are expected soon. He shared his concern that the time between initial submission and award is too long, especially for new investigators.

Dr. Scarpa said that he has heard many individuals say that clinical research is not properly evaluated. He noted that the data evaluated to date show a small but significant difference between the review outcomes of clinical vs. nonclinical research. Dr. Scarpa emphasized that CSR should work to address this issue regardless of whether the problem is a review issue or not. CSR will continue to assess this issue with its Clinical Research Advisor, Dr. Theodore Kotchen, from the Medical College of Wisconsin.

Generally, Dr. Scarpa said, a culture has developed that values more conservative research. He noted that he used to advise his faculty to write conservatively and he did himself. There is no silver bullet to change the culture, but it is being openly discussed in the community.

Yet, even with all these changes, noted Dr. Scarpa, CSR would fail its mission without high-quality, experienced reviewers. Comparing 1998 and 2005 figures, he pointed out a decrease in reviewers who are full professors and an increase in assistant professors. While he welcomed younger reviewers, he said that study sections need the broad backgrounds of more experienced reviewers. He also warned against assistant professors taking on too much reviewing responsibility at a critical time in their careers. To relieve the burden on reviewers, pilots are under way to assess new electronic review platforms, such as video-enhanced discussion and/or asynchronous electronic discussion.

Possible New Systems
What would we design from scratch if the peer review system did not already exist? Dr. Scarpa said that NIH peer review has served science and the public well. The system, however, was developed in the 1950s to meet the needs of the time. It might be useful to step back and consider a whole new design. In any event, it will be important to stay true to CSR's mission to provide fair, independent, expert, and timely reviews. Dr. Scarpa explained that the current system is great to start with. He added that he has been impressed with the dedication and commitment of CSR staff and their contributions to peer review and biomedical research.

Discussion
Dr. Pugh joined in praising the CSR staff noting that the whole system depends on the peer review process. Dr. Ramm agreed that the review cycle should be shortened, but that ICs also need to look at ways to shorten the entire cycle—from application to award.

Dr. Calhoun concurred that increasing the participation of full and associate professors as reviewers is needed. Dr. Sassaman said review administrators are finding it hard to recruit reviewers because of the time commitment. Yet, ensuring that those with sufficient experience

review an application is important as a core value. Dr. Scarpa said that many study sections are very broad, so that maybe only a subset of people could be called upon to review a particular application. He reiterated that it is important to get young reviewers involved, but not to overburden them. Dr. Martinez said that workload has to be considered for older reviewers, too. Dr. Brenner suggested creating "educational slots" in study sections for younger reviewers. With limited loads on a one-time basis, younger reviewers could benefit from study section participation without being overburdened.

Dr. Brenner asked Dr. Scarpa about his priorities for the many potential changes he described. Dr. Scarpa said strategies to shorten the review cycle are being aggressively pursued. As soon as the pilots are analyzed, changes can be proposed. In terms of the clinical research issue, more data are needed, and Dr. Kotchen will be involved in that effort. The tendency toward more conservative grants needs to be discussed, but he said he did not know of a solution at this point.

Dr. McClain praised the idea of electronic discussions as a way to save time. Dr. Scarpa noted that a number of platforms would probably be needed, because of different needs. He explained that face-to-face review meetings would continue to be the ideal, and CSR would seek to use the new electronic platforms only when they address represent the best way to convene the best reviews or when a particular group of reviewers express a preference.

Core Values of NIH Peer Review: Feedback from Stakeholders and Implementation Plan

Dr. Alan Willard, Chair of the Review Policy Committee and Chair of the Scientific Review Branch of the National Institute of Neurological Disorders and Stroke, updated PRAC after his presentation on core values at the May 2005 meeting. He noted that two core values emerged, after much discussion: (1) scientific and technical competence, and (2) fairness and objectivity.

SRA Retreat
The core values are the theme of a trans-NIH SRA retreat, scheduled for September 29, 2005, with more than 400 SRAs signed up to participate. This is the first such retreat in about 10 years, he noted. Dr. Norka Ruiz Bravo and Dr. Scarpa will make keynote addresses. Small groups will discuss topics that relate to the core values, such as writing the resume of discussion. Dr. Willard also described the large group and plenary sessions that are planned.

Discussion
Dr. Sassaman praised the retreat agenda and asked how issues that emerge will be subsequently implemented. Dr. Willard explained the logistics to make the retreat not just a sharing, but also a problem-solving, session. He recognized the work of the team who put together the program, headed up by Dr. Francisco Calvo, Chief of the Review Branch of the National Institute of Diabetes and Digestive & Kidney Diseases.

Dr. Martinez said that the retreat is in keeping with Dr. Scarpa's vision of CSR at the frontier of peer review and hoped that retreats could occur more frequently. Dr. Torok-Storb asked about reviewer participation. Dr. Willard noted that Dr. Lloyd Fricker, an experienced reviewer, would talk on the academic perspective on peer review in a plenary session. He also said that a suggestion might emerge to hold reviewer retreats, which some individual sections have had.

Dr. Torok-Storb said that training junior reviewers to see the big picture would be worthwhile. Dr. Willard said that at a training session held for K awardees, there was much discussion about whether the trend toward funding conservative grants comes more from senior or junior members of a study section. A video of a mock study section has been a helpful training tool. Dr. Brenner noted that an important function of the study sections on which he has served is to educate clinical researchers in how to prepare a competitive grant. Dr. Willard said that some ICs have been aggressive in using K mechanisms to train physician-scientists in grantsmanship.

Follow-up: Quantitative Aspects of Peer Review

Dr. Berg introduced this section of the meeting by noting that four presentations would be made. He requested that discussion be held until the end of the four.

Introduction and Orientation
Dr. Bravo stressed that the NIH peer review system is one of the best, if not the best, in the world. It has allowed NIH to identify excellent science to fund. But, she said, no system is perfect, and needs change over time. Thus, NIH should continually seek to improve, such as by looking at the scoring system. She introduced the topics of the rest of the session: how NIH arrived at the current scoring system, advances in the field of decision-making that might affect the NIH system, and potential next steps.

History of Priority Scores at NIH
Dr. Michael R. Martin, Director of CSR Division of Physiology and Pathology, thanked Richard Mandel and Tony Dempsey as his sources for his presentation.

Chronology: Dr. Martin began with August 1946 when NIH began the peer review process. Early panels voted either to approve or disapprove grant applications. By 1949, as the number of applications increased compared to available funds, reviewers were asked to prioritize on a 1 to 5 scale in half point intervals [1.0, 1.5, 2.0, etc], with scores averaged and then multiplied by 100. Reviewers were asked to give the "top" priority a value of 1 and the lowest a value of 5.

A change in the locus of peer review occurred in 1962 when Institute review panels were established. Various bureaus, institutes, and divisions started developing their own review systems and practices, which raised concerns about comparability of priority scores across review groups and the development of funding plans. Thus, in 1971, the Executive Committee on Extramural Activities Priority Score Review Committee, the first of many study groups to look at the NIH scoring system, recommended retaining priority scores but standardizing the system. They suggested percentiling or normalization, with preference for the latter. After a pilot and evaluation, three ICs objected to normalization so it was not implemented across all of NIH. Finally in 1979, another NIH committee recommended displaying the normalized score using a 41-point scale [1.0, 1.1, 1.2, etc.]. In 1980, NIH Director Frederickson mandated the use of raw priority scores but let ICs develop alternatives. The 41 point scale was adopted.

In the early 1980s, Congress established 212 as the priority score for funding grants, thus adjusting downward the budget of ICs that used only normalized scores. Having an established pay line meant that over time the mean priority scores dropped.

During the next few years different study sections evolved very different scoring patterns even though some of these study sections were pairs reviewing identical sets of science. These factors and the global downward trend in priority scores raised concerns about the value of using priority scores for developing pay lines across study sections. In 1987, the EPMC Working Group on the Movement of Priority Scores adopted as its main recommendation system-wide percentiling. They also recommended a reversion to the nine-step 1 to 5 scale and showing scores to two digits, rather than three. Only the first recommendation was adopted.

The next main concern, expressed in a Government Accounting Office report in 1994, focused on whether reviewers were using different criteria in their scoring decisions. The 1996 Committee on Rating of Grant Applications recommended three criteria, with scoring on each criterion using an eight-step scale. The Peer Review Oversight Group and DRG Advisory Committee (PRAC predecessors) proposed four criteria to be addressed within a written critique. They rejected the proposals to score individual criteria and to change to an eight-step scale. Ultimately, NIH Director Harold Varmus approved five criteria as the basis for an evaluation, rating the global score in tenths, with criterion-based discussion and critiques but not individual scores.

Observations: Dr. Martin noted that a number of other studies have also taken place over the years that have generally validated the NIH system. He concluded with four overall observations:

- The utility and value of the priority score method is widely recognized
- Reviewers report being very comfortable with the 41 interval scale
- Priority scores are reliable indicators of relative scientific merit and rank order within a study section over time, but not between study sections
- Percentiling is a valuable tool that has allowed cross study section comparisons.

Dr. Martin concluded that study sections sometimes appear "messy" to outside observers, but the system is effective in producing a quality product.

Priority Scores and Percentiles

Dr. James Onken, Chief of the Office of Program Analysis and Evaluation, National Institute of General Medical Sciences, defined and provided examples of priority and percentile scores. The issues surrounding scoring revolve around three questions that are related to variation: How reliable are the scores? Are they being expressed at the appropriate level of precision? Do the average priority score and percentile provide enough information to applicants and to NIH staff?

In classic psychometric theory, variations in scores are based on true differences in applications and differences of opinion among reviewers, as well as on some error. Percentiling variations can result from variations in the priority scores as well as the base against which a given application is compared. For example, a priority score of 153 can be at the 20.7 percentile one round and the 22 percentile another round, depending on the applications against which it is being compared.

Priority scores: Dr. Onken explained that reliability is the proportion of the variance caused by true variation in the applications compared to the overall variation: thus, if the error variance is very large, there is a low reliability in the ratings. The reliability of a group average is a function not only of the reliability of the individual reviewers' scores, but also of the number of reviewers' scores that make up the average. With a typical study section size of 20 to 25 members, a very reliable average priority score can be obtained.

Some of the changes proposed in the 1990s were aimed at increasing individual reviewer reliability. However, according to Dr. Onken, research shows that such efforts may have diminishing returns when dealing with group averages. He acknowledged that this seems counterintuitive; indeed, research shows that people consistently underestimate the power of averaging. However, presenting an average, while reliable, may result in a loss of information. Options are to include the standard deviation or the range of scores. He presented the scores of several sample applications to illustrate the information that these methods would provide.

Percentiles: Sampling errors can arise in percentiling an application against a different base every round. He showed data from a study that compared priority scores from the same study sections in May 1995 and May 1996. Some variation was noted. A coefficient of variation can help quantify the extent.

Decision-making Research and NIH Peer Review

Dr. Jane Steinberg, Director of Extramural Activities, National Institute of Mental Health, presented information on decision-making research and its usefulness in looking at NIH peer review. The research shows that just providing information or instructions is not sufficient for people to make complex or high-risk decisions. Instead, people tend to turn to certain heuristics or biases to make complex decisions, even if that leads to non-optimal decisions. However, the research shows that these biases can be minimized or used to strengthen decision-making and behavior change. These findings have been used to help people make financial, medical, and security-related decisions.

Another finding is that people are much happier using their biases to make decisions than a structured decision-making tool. As an example, she cited interviewing a job candidate in person, even though relying on paper credentials might result in a better, more objective decision.

Application to NIH: Peer review is a series of judgments, involving first individual decisions, then a group decision, then back to individual decisions as each reviewer assigns a score. Consultants to the Rating of Grant Applications (RGA) subcommittee made recommendations that apply decision research to peer review, such as testing to see the effects of scoring by criterion. In addition, although weighting criteria has many detractors, Dr. Steinberg said, it is in fact what individual reviewers do in their heads to come up with an overall score.

The research also shows that decisions are very context-specific and sensitive to framing. For example, there might be a different result if reviewers were asked to focus on identifying "important or innovative research" compared with identifying "what we just can't afford to lose." In addition, the research shows that groups prefer to discuss information that all members know, rather than situations where one person raises unshared information. In a peer review

situation, if there were a bias to discuss only shared information, this could lead to less optimal decisions since critical unshared information would not inform the discussion. Many protections are in place to prevent this in peer review, but emphasizing the need for the chair to elicit or ask for any unshared information, instructing reviewers to consider all information, and allowing adequate time for discussion would minimize this potential bias.

Dr. Steinberg concluded that—

- People and groups use heuristics (biases) to make easy work of difficult decisions
- Most are successful, but some systematically provide less than optimal decisions
- Thirty years' worth of decision research has identified how to minimize these biases and strengthen decision-making
- The review process involves very complex decisions that may benefit from applying this knowledge base.

Discussion

To wrap up, Dr. Bravo said the intention of the preceding session was to take a critical look at what has been occurring and see if additional information could help inform scoring decisions.

The Big Picture: Dr. Pugh praised the presentations and warned against focusing on the details of scoring at this point. Rather, he stressed larger issues, such as how to deal with the five different criteria and push for innovation and new ideas. The focus should be on embedding the goals of the system into the decision-making process.

Scoring Criteria: Dr. Berg said that he felt that the audience for this issue has changed from applicants to program staff. If the goal is to extract information from an application and compare it to other applications, the question becomes how different Institutes use scores. For example, if innovation received a separate score and Congress or NIH itself said that not enough innovative science was funded, there would be a clear way to fund more innovative research. Dr. Bravo said this issue highlights whether additional information can be provided to staff and applicants. Dr. Martinez commented on the value of the background article [Yaniv A. The Benefit of Additional Opinions. J American Psychological Society 13(2):75-78, 2004] distributed to PRAC and also spoke in favor of some sort of scoring disaggregation. Dr. Ramm noted that in some big applications, review criteria are scored independently and contained in the summary statements, which gives program staff and applicants helpful information.

Dr. Scarpa raised questions about the statistical analysis when only two people read through an application. He also questioned how to capture situations in which two readers have very different opinions about the same application. Often, innovative research reveals the greatest discrepancy in opinion.

Dr. Pugh wondered about the effect when few people read an application and the effects if one individual is forceful in his or her views, particularly in relation to innovation. Dr. McClain warned about changing the rating system at the same time that electronic submission is being introduced. He described his own method of reviewing applications as identifying what he thinks is the best grant application and becoming a major advocate for it within a study section.

Dr. Brenner suggested that a short session on decision-making for study sections would be helpful. He also concurred with Dr. Pugh that the real challenge is in figuring out how to weight the criteria, especially innovation. Dr. Scarpa said that training would help, but a framework is needed. Study sections tend to merge scores, which provides less information to program staff.

Dr. Torok-Storb said reviewers and applicants would welcome the information provided by scoring each of the five criteria. Dr. Martin clarified a point from his presentation: the earlier groups dealt with both scoring and weighting criteria. Either one, or both, could be done.

Dr. Sassaman said that more information from reviewers on application scores would be helpful in making funding decisions. Dr. Martinez said that more information would also help explain large variances in the scores. Dr. Berg suggested it would also be helpful to know what would the best possible score that an application could receive, given the problem selected for research. In some cases, a score can reflect a technically sound application. The applicant wants guidance on how to improve it, but the reality is that the score will probably not greatly improve given the research problem. Related to that, he suggested an interesting pilot would be along the lines of assigning to grant applications something like a "degree of difficulty" in diving. Dr. Torok-Storb questioned whether that could involve the significance criterion. Dr. Berg said that the meaning of "significance" was unevenly applied in his experience on/with study sections.

Dr. Bravo suggested several next steps. A working group within PRAC could consider what to do about scoring, or it could be sent to the Extramural Activities Working Group (EAWG) for staff to work through possible actions. At this point, she said, PRAC seems to have enthusiasm for providing more information on the individual criteria. A possibility is scoring individual criteria, but that would be a longer-term decision. The immediate issue is what additional information to include that would be helpful to applicants and staff. She underlined the need to think through the effects on all the various players. Dr. Pugh suggested that an internal staff working group develop specific recommendations. It was agreed that PRAC would ask the EAWG to study (1) additional information around priority scores, and (2) rating of individual criteria. Dr. Scarpa said that their work could be circulated before the next PRAC meeting.

Review of Member Conflict Applications

Dr. Sally Rockey, Deputy Director of the Office of Extramural Research, discussed how reviews of member applications are handled when a conflict of interest arises. She said she based her presentation on these assumptions:

- Reviewers are the engine that drives the peer review system;
- Reviewers must sacrifice extraordinary personal time to serve NIH;
- The system should be such that all applications receive an unbiased, fair, and high-quality review;
- Reviewers must know that service on a review committee is not going to jeopardize their own chances for NIH support;
- Member applications should be handled in a consistent manner across all study sections.

Current policy prohibits a review group from reviewing an application of one of its members. In such cases, the IRG chief arranges to have the application reviewed by another group with appropriate expertise, either another scientific review group or a Special Emphasis Panel (SEP). A SEP can contain a maximum of 50 percent of its members from the conflicting group. The SRA of the conflicting group cannot be in charge of the SEP, nor can a member of the conflicting group chair the SEP.

A "member conflict" is one in which the PI is a member of the assigned study section, or the applicant may have worked with or been otherwise connected with a member of the study section. In 2004, 4,067 of 27,605 applications were submitted by current reviewers. Of these, 702 were referred to a SEP. If a SEP is composed of 40 percent or fewer chartered members, percentile rankings are made against the "CSR All" base; otherwise an application's percentile is ranked against the parent committee. According to Dr. Rockey, some members question whether their applications are getting a fair and equitable review. Another issue that has arisen is whether CSR policies are followed in terms of such issues as SRA conflicts. Finally, so-called boutique SEPs, which review seven applications or fewer, have proliferated.

She asked PRAC members to consider three questions: Are member applications treated fairly? Are current practices appropriate and consistent? What improvements can be made?

Discussion

Dr. Martinez questioned the extent of the problem. He noted that the overall success rate for members' competing R01 renewals is about 35 percent, which compares well with the total. Dr. Rockey agreed, but noted that perception makes this a topic worth examining. Dr. Brenner pointed out that those who are not funded tend to be the ones who speak out.

Dr. Calhoun suggested that the system might err on the side of overprotection. NIH is trying to recruit senior researchers as reviewers, she said, many of whom may be submitting applications. Agreeing to serve on a study section may place their own applications at a disadvantage.

Dr. Torok-Storb asked whether experiments have been done on the effect of applications staying in the conflicted study section. Dr. Malik responded that this experiment could not take place, since it would conflict with NIH rules and regulations. One way to experiment would be to take a conflicted application to a SEP and to another standing study section, and compare the results. Dr. Ramm said that an "accidental" experiment occurred when an application by a member was considered within his review group. Even though he left the room when his application was discussed and his application was triaged out, the results of the review were voided.

Dr. Vicki Levin, an SRA in the Risk, Prevention, and Health Behavior IRG, said that she has heard the complaint that SEPs take place after the bulk of the review occurs, which means less time for the applicant if his or her application must be resubmitted. Dr. Dharam Dhindsa, Deputy Chief of the Surgical Sciences, Biomedical Imaging, and Bioengineering IRG, said that it does not seem to make a difference whether a member conflict application is percentiled against the study section or all CSR. Very few members have resigned because of this issue. As for the timing issue, the members' applications may be reviewed later because of delays by the SRA getting the applications from members.

Dr. Rockey concluded that the sense from PRAC was that handling member conflict applications was not a serious concern. Instead, there are a few things to look at, such as the timing of the SEPs. To remedy the delay, Dr. Calhoun asked about the feasibility of holding a SEP right after a study section meeting, with a portion of the membership augmented by ad hocs and a different SRA. This arrangement could be looked at in terms of the regulations and group dynamics.

Innovation Review Criterion

Dr. David Armstrong, Chief of the NIMH Review Branch, said he would address three areas: definition of innovation; NIH efforts to promote receipt and review of innovative applications; and possible initiatives. Citing a recent article in the Journal of the American Medical Association as an example, he said that NIH is perceived as risk-averse and conservative in its funding. Many innovative applications are not even submitted, which can deplete NIH of a vital set of resources critical to future success. However, such efforts as the Pioneer Awards, NIH Roadmap, and NIH Neuroscience Blueprint are changing the perception.

Definition
In addition to the dictionary definition, Dr. Armstrong shared some key concepts of "innovation" developed with the help of Dr. Morton Fleming, Director of the Lemelson-MIT program. Creativity involves the ability to solve problems, generate possibilities, or create products, within a specific domain, in an initially novel but eventually broadly accepted way. Inventiveness is a form of creativity that can lead to innovation. He defined invention as the process of devising something that is useful and not previous known or existing, developed through independent investigation, experimentation, and mental activity. Dr. Armstrong said that a key word to define innovation is "process." Innovation, which may or may not include an invention, is the complex process of introducing novel ideas into use or practice, and includes entrepreneurship as an integral part. As an example, he cited the process from defining a problem (slow transportation), to the invention of the car, to the assembly line that made cars a standard of transportation.

NIH Efforts
Dr. Armstrong extracted from guides to reviewers how NIH defines innovation. He noted that the definition is mechanism-independent. To learn how NIH currently promotes innovative science, Dr. Chana Rabiner [a participant in the emerging leaders program (ELP)] and he conducted interviews with 13 ICs on their current and past efforts to promote receipt and review of innovative applications; future initiatives under consideration; major impediments; and recommendations. They found that ICs use different RFAs and program announcements to try to solicit applications in areas that they feel are underrepresented or underserved. Often, a program officer recognizes a need, either in conjunction with an IC-sponsored workshop or independently. Some expressed dissatisfaction with the R21 in meeting its intended goal. ICs' authority to fund poorly scored applications is rarely used.

IC recommendations included more explicit language about innovation in program announcements and in reviewers' guides. Many of those interviewed said CSR's procedure is too conservative. SRAs should educate reviewers in the different mechanisms, especially those that have an innovation emphasis.

Dr. Armstrong suggested that a working group look at developing new language for the R21 with greater emphasis on innovation and paradigm-shifting research. Another aspect to consider is whether R21 applications should be evaluated alongside applications for other mechanisms in reviews, or instead should have a focused review locus. Dr. Armstrong said that NIH has an impact on innovation through communications. The Pioneer Awards, workshops, seminars, and other meetings can raise consciousness about innovation extramurally and within NIH. In addition, perhaps a working group should look at scoring individual criteria, as discussed earlier in the meeting, perhaps scoring innovation differently or with a different weight than the others.

Project Innovation
Dr. Armstrong discussed an idea that he called Project Innovation—a trans-NIH initiative to promote funding of high-risk, potentially high-impact grant applications that fail to meet the payline. It would involve collaboration between CSR and all funding ICs, in which nominations of innovative projects would be evaluated and prioritized by Innovation Committees within ICs. He envisions Project Innovation as a way to provide limited support to about 75 highly innovative applications per year that would otherwise go unfunded. This project would also show NIH as willing to take risks and support truly paradigm-shifting research. Dr. Armstrong said that NIMH is currently experimenting with this approach.

Discussion
Dr. Martinez commented that perhaps other ways should be found so that reviewers are more supportive of innovation. He said that second submissions help, but that reviewers sometimes question the likelihood of completion of innovative projects. Dr. McClain pointed out that often younger reviewers are more detail-oriented and conservative. He said that education with SRAs is a critical factor. He also said that older investigators can often be more innovative in their grants, because they are in a better position to take chances on funding.

R21: Dr. Ramm related NCRR's experience with the R21. The applications are looked at in captured study sections, rather than alongside R01s or other grants. No preliminary data are requested, and the expectation is explicit and accepted that only a few will succeed.

Dr. Marcia Steinberg, Chief of the Cell Biology IRG, suggested that perhaps the R21s, which provide limited support for two years, provide "just enough money and time to fail." Perhaps it is time to take a big leap and provide larger amounts of money. To follow up, Dr. Ramm said that NCRR copied an NCI effort to fund innovative projects for three more years at a higher level, without another review, if good data are emerging. Dr. Armstrong said that many ICs echoed Dr. Steinberg's concern. Dr. McClain observed that R21s are sometimes used for the wrong purpose—to try and get preliminary data for an R01, which he attributed to bad mentoring.

New Ideas: Dr. Eileen Bradley, Chief of the Surgical Sciences, Biomedical Imaging, and Bioengineering IRG, suggested going beyond "tweaking around the edges" of innovation. She urged more of an emphasis on ideas; otherwise, reviewers tend to focus on methodology and feasibility. Dr. Armstrong characterized the process as circular—innovative ideas have to be presented so that reviewers have something to consider, but because reviewers tend to prefer more conservative projects, innovative applications may not be as forthcoming.

Drawing from his experience with the Pioneer Awards, Dr. Berg commented that a broadly based committee helped keep attention on the big picture. When there was overlap between a reviewer and the proposed research, the discussion drilled down to a more detailed level. He said that even unsuccessful applicants e-mailed him to say they welcomed the opportunity to develop Pioneer-type proposals. He also noted that Institutes have used the R21 for different purposes, such as a way to fund more grants with the same amount of dollars.

Shortening the Review Cycle

Dr. Hortencia Hornbeak, Associate Director for Scientific Review and Policy, National Institute of Allergy and Infectious Disease, spoke of multi-pronged approaches to shorten the review cycle that preserve the important features of NIH peer review. These approaches included: electronic submissions, knowledge management tools to aid in referral and recruitment, electronic recruitment of SEP reviewers, Internet-assisted review, structured critiques, abbreviated summary statements, and Council approval independent of scheduled meetings. The National Institute of Allergy and Infectious Diseases (NIAID) used "hyperaccelerated reviews" for AIDS initiatives in the 1980s and 1990s and, more recently, for biodefense initiatives, including Project Bioshield. Most of the reviews that took place in FY03 were on a 73-day cycle, with five cycles taking place in one year. She acknowledged that this timeframe was very stressful on staff.

Her IC worked with the Office of Extramural Research (OER) and others to develop mechanisms to secure reviewers with appropriate expertise while managing conflicts of interest. Tools developed to accelerate the review process included: structured reviews, structured critiques, and practical guidance to reviewers in pre-review orientation sessions and at the review meetings. During pre-review teleconferences, SRAs and Program staff explained the objective of the RFA and provided review instructions so that reviewers would receive clear and defined instructions before writing their critiques.

A reviewer support site (RSS) is being developed in NIAID to enable the SRA to post everything a reviewer needs to know and do for a review. The RSS is an Access-based system that will be piloted with chartered committees in January 2006 to further facilitate communication with reviewers. A macro was developed to scan names of reviewers to identify any co-publications with an applicant that a reviewer may have overlooked, which helped avoid last-minute conflicts of interest. Flexible teams consisting of SRAs and Grants Clerks were created and trained in effective teamwork. These teams included 20 newly hired SRAs.

Structured Critiques

Dr. Hornbeak then turned to structured critiques, which she said led to more focused discussions and more concise written evaluations, as well as facilitated preparation of abbreviated summary statements. Structured critiques were used for both simple and complex award mechanisms. Reviewers received a template, specific to each initiative, in which they honed in on an application's strengths and weaknesses within the five review criteria, as well as any initiative-specific criteria. She provided an example, with identifying information removed, and explained

that the reviewer critique information could then go right into the resume of the summary statement.

Dr. Hornbeak said that structured critiques increase the time for discussion of competitive or challenging applications in the review meetings and allow committees to manage a larger number of applications. She then presented data on the numbers of applications, reviewers, and other information for a number of mechanisms, including R21s and U01s, for which structured critiques were used. She noted that the summary statements were sent only to the funded applications within the accelerated timeframe, with the summary statements of unfunded applications prepared and sent later. This practice needed buy-in from the Institute and the scientific community.

Project Bioshield used a different process than the usual NIH peer review. For example, summary statements were not in the usual NIH format. The guidance provided to reviewers was critical to meeting the required timeframe for review and award. Reviewers received a specific instruction manual and participated in a pre-review, orientation teleconference. Necessary resources to decrease time included a sufficient number of well-trained staff, effective work teams, flexible staff assignments, and versatile, state-of-the-art IT support.

In summary, structured critiques can shorten the review cycle and preserve the features of NIH peer review, as long as the reviewers are properly oriented. In addition, shorter critiques decrease the time for summary statement preparation and actually result in a more useful product for Program staff and other stakeholders.

Discussion
Dr. Brenner asked Dr. Hornbeak about any applicant feedback. She said that no complaints were received. There was some informal feedback from reviewers, mostly about the challenge in only listing strengths and weaknesses in their critiques.

Dr. Calhoun praised the accomplishment of five rounds in one year, which she noted took place without adequate staff. She asked Dr. Hornbeak whether she thought the process could be applied in other situations, perhaps doing four rounds per year. While there are significant time-saving possibilities, particularly with electronic submission, Dr. Hornbeak warned against speeding up the time allocated for reviewer's assessment of the applications and deliberations.

In response to a question from Dr. Martinez, Dr. Hornbeak explained that discussion leaders facilitated consideration of complex applications. Dr. Torok-Storb said she thought that listing strengths and weaknesses in a direct fashion would be useful for all concerned. Dr. Hornbeak noted that besides listing the strengths and weaknesses, reviewers had space to write any other comments they wished. Dr. Torok-Storb called this a happy compromise between what currently exists and the use of checkboxes. She called for anything that cuts down on the narrative so the communication is more deliberate and accurate between reviewer and applicant.

An audience participant observed that most SRAs would like to have some of the flexibility that Dr. Hornbeak described. Dr. Hornbeak noted some of NIH's procedures and guidance are sometimes self-imposed and can be changed. The challenge in the biodefense instance, according

to Dr. Hornbeak, was to identify the best science through peer review within the available time to ensure that awards could be made within the fiscal year. The deadline meant getting the reviewers to focus on the strengths and weaknesses that support their score and deliver their evaluation concisely.

New Study Sections

Two short presentations were added to the agenda to inform PRAC about the creation of new and reorganized study sections.

Brain Disorders and Clinical Neuroscience
Dr. Anita Miller Sostek, Director of the CSR Division of Clinical and Population-Based Studies, told the committee that she had made a presentation to the previous CSR Advisory Committee in May 2004 outlining basic principles for modifying study sections. The new process starts with evaluating the stability of the workload over time and brings in input from many stakeholders.

She used this process with the Cell Death and Injury in Neurodegeneration study section. Over the past few years, the scientific material in this study section has broadened and the number of applications steadily increased to 145 in January 2005. After some examination and consultation with stakeholders, it was proposed that one study section would focus on chronic neurodegenerative diseases, or Cell Death in Neurodegeneration (CDIN). Another would focus on acute brain injury, or Brain Injury and Neurovascular Pathologies (BINP). She exhibited the guidelines for the two sections and noted that overlap statements are being developed.

Dr. Pugh commented that the presentation serves as a good example of how a new study section should be created under the reorganization. In response to a question from Dr. Calhoun, Dr. Sostek said the study sections are not particularly captive. A number of different Institutes might have applications under review in these study sections.

Diabetes and Obesity
Dr. Elliot Postow, Director of the CSR Division of Biologic Basis of Disease, presented information about creating a new study section related to diabetes and obesity. Given the rise in applications in these areas, IRG staff developed plans for creating three study sections out of two. Detailed discussions of the proposal were held with nine extramural scientists at two professional meetings. Their feedback was that the proposed study sections were too heterogeneous and suggested some changes. The plan was revised and, during a second round of discussions, the consultants supported the plan. Dr. Postow explained that the three study sections now track the continuum from basic to clinical research: Cellular Aspects of Diabetes and Obesity (CADO), Integrative Physiology of Obesity and Diabetes (IPOD), and Clinical and Integrative Aspects of Diabetes and Obesity (CIDO).

There was some discussion about whether there were enough applications to support a mostly clinical study section, but there was general support for the new study sections.

Final Business Items

Dr. Sassaman asked whether the minutes, which she said are used in preparing for PRAC and other discussions, could be distributed earlier and whether the presentations were available online. (They can be found and downloaded by searching for "PRAC" in the search function of the NIH website.)

Dr. Scarpa closed by thanking all participants and reminding them that the next PRAC meeting is scheduled for January 23, 2006. He adjourned the meeting at 3:40 p.m.

We do hereby certify that, to the best of our knowledge, the foregoing minutes of the third meeting of PRAC are accurate and complete. The minutes will be considered at the fourth meeting of the Committee, and any corrections or comments will be made at that meeting.

Karl Malik, Ph.D.
Executive Secretary
Peer Review Advisory Committee

Antonio Scarpa, M.D., Ph.D.
Co-Chair
Peer Review Advisory Committee

Jeremy Berg, Ph.D.
Co-Chair
Peer Review Advisory Committee