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

May 22, 2006

The second 2006 meeting of the Peer Review Advisory Committee (PRAC) convened at 8:30 a.m. on Monday, May 22, 2006, at the Hyatt Regency, Bethesda, Maryland. The entire meeting was held in open session. Drs. Toni Scarpa and Jeremy Berg presided as Co-Chairs.

Members

Jeremy Berg, Ph.D., Co-Chair
Antonio Scarpa, M.D., Ph.D., Co-Chair
Dean E. Brenner, M.D.
Edward N. Pugh, Jr., Ph.D.
Louise Ramm, Ph.D.
Anne P. Sassaman, Ph.D.
Beverly Torok-Storb, Ph.D.
Matt Winkler, Ph.D.

Ad hoc Members

Faye Calhoun, Ph.D.
Leslie A. Leinwand, Ph.D.
R. Lorraine Collins, Ph.D.
Daria Mochly-Rosen, Ph.D.

Dr. Craig J. McClain, M.D., and Dr. Joe. L. Martinez, Jr., Ph.D., were not present at the meeting. Dr. Elias Zerhouni, M.D., and Dr. Norka Ruiz Bravo, Ph.D., attended as Ex Officio members.

Dr. Michael R. Martin, Ph.D., was the Executive Secretary for the meeting.

Introductions, Approval of the January 2006 PRAC Minutes, and Upcoming Meetings

Dr. Berg welcomed participants to the PRAC meeting and asked them to introduce themselves. He then asked for approval of the minutes of the January 2006 meeting. The minutes were unanimously approved. He also asked for approval of the meeting dates for the rest of 2006 and for 2007: August 28 and December 4, 2006, and May 21 and August 27, 2007. The dates were chosen to coordinate with Institute Council meetings.

Update on Electronic Submission of Grant Applications

Ms. Megan Columbus, Program Manager for Electronic Receipt of Grant Applications, at the Office of Extramural Research (OER), reported on progress made on electronic submission for applications received in April 2006. She said that electronic submission has become a reality for the National Institutes of Health (NIH).

Improvements in April 2006
Submission of Small Business Innovation Research (SBIR) applications went much more smoothly in April, when 680 applications went fully through the system by the receipt date, than in December 2005, when only 53 did. The system's processing speed increased dramatically.

Delays in getting assistance from the help desk were reduced by staffing up the desk during "surge" periods, as well as by providing clearer on-screen instructions and by increasing training and outreach to lessen the need to contact the help desk. A new Web ticketing system reduced redundant requests.

Further improvement is needed, and OER is working with the applicant community to refine the process. For example, at the community's request, the timeline for receipt of R01s has been moved to February 2007 so institutions have more time for training and putting electronic systems in place. The deadline has also been changed to 5:00 pm local time, again at the request of institutions. Feedback helped in revising the SF424 Application Guide, released on April 7, 2006. As of May 10, principal investigators (PIs) and business officials are not required to verify application images, although they are strongly encouraged to review them in the NIH Commons.

Grants.gov experienced some software glitches in April, in part because, unlike in December, more agencies shared April 1 deadlines. They have fixed many of the problems and continue to work on those still unresolved. The Citrix workaround developed for electronic submission from Macintosh or UNIX computers held up, although the June deadline for small research grants (R21s, R33s, R03s, and R34s) from academic institutions will provide a better test. In addition, IBM and Grants.gov have reconfirmed that they will have a platform-independent solution in place by November 2006.

Looking Ahead
The June 1 deadline will provide an opportunity for monitoring and refinements. So far, more applicants have submitted in advance of the deadline, many on their first attempt. Plans continue to ensure preparedness for the transition for the R01s, including looking at ways to spread out the workload on peak submission dates. Nine working groups are looking at the challenges posed by different mechanisms. Communication with the applicant community will continue.

Discussion
Dr. Faye Calhoun praised the progress and asked about capturing help-desk data. Ms. Columbus said that every error message is analyzed. The Web ticketing system also helps track data. Dr. Calhoun asked whether callers to the help desk know how long their call will be on hold. Ms. Columbus said the fact that requests take such varied amounts of time might not make that feature possible, but she would check into the possibility.

Dr. Matt Winkler said that he asked for feedback from the grants manager in his organization. The manager reported much improvement between April and December, although he had some frustrations, such as the number of e-mails generated by each problem and the fact that the software does not allow for "pasting" of the same information in different places. Dr. Winkler also reported that his manager praised the application guide, NIH-specific instructions, and help desk. The bottom line is that people understand the need for change and are working with it, but he asked how those with less sophisticated technical capabilities are faring. Ms. Columbus said that some people were having more trouble than others.

Dr. Louise Ramm asked about plans for the R01 round in February 2007. Ms. Columbus said that they are comfortable handling a large number of applications, but are now looking at the impact of file size. In response to a question from Dr. Beverly Torok-Storb, Ms. Columbus said

that a proposal to stagger deadlines has been made, based on the fact that applications stack up, both in the institutions and at NIH, for a few days. However, no real analysis of alternatives has taken place. Dr. Torok-Storb said that such a change would be welcome. Dr. Berg noted that it is not the capacity, but rather the periodic surges, that can cause problems in the system.

CSR: New Challenges and Opportunities

Dr. Scarpa said that "business as usual" is no longer possible at the Center for Scientific Review (CSR) with the number of applications almost doubling and reviewers willing to review half the load as previously. This challenging situation leads brings opportunities for highly desirable changes. Welcome changes are the recruitment of Dr. Cheryl Kitt as the new CSR Deputy Director and Dr. Cheryl Oros as CSR's Director of Planning, Analysis, and Evaluation. Dr. Scarpa then summarized other changes that have occurred, are in progress, or are under discussion, as well as present and future challenges.

Changes in CSR Operations
Communications: CSR has taken steps to increase communications within CSR, elsewhere in NIH, and with the scientific community. Publications are sent out regularly about peer review, and Dr. Scarpa often speaks with different stakeholder communities.

Uniformity: Efforts to increase uniformity in handling applications are under way. For example, 95–97 percent of all summary statements are now posted within one month of a study section, with those of new investigators posted within one week. Summary statement resumes now are more complete and structured. Unscoring is being done uniformly at 50 percent.

Efficiency: Increasing efficiency is essential. Big steps include the upcoming electronic submission of R01s and the use of text-fingerprinting and artificial intelligence software to process applications and recruit reviewers. A major pilot to use knowledge management software to assign study sections will take place in October 2006, with the hope that the system can be operational in February 2007.

Monitoring of Integrated Review Groups (IRGs): Dr. Scarpa asked for PRAC assistance in dispelling concern about CSR monitoring of IRGs and study sections. After the reorganization, the intent had been to review each IRG every 5 years. The pace of science, however, has meant that these reviews need to occur more frequently. Thus, each month, one IRG is the subject of an in-house review by the directors, division directors, chiefs, and scientific review administrators (SRAs). This process means that each IRG is reviewed every 2 years, in addition to the 5-year cycle. If small problems are uncovered, the staff, along with the extramural community or NIH program staff, address them. Substantive issues are brought before PRAC. He presented the review schedule for 2006 and 2007. In addition, senior staff visit study sections frequently. Dr. Scarpa debriefs retiring study section chairs by phone to learn about problems and possible improvements. CSR hosts two or three visits weekly from different scientific societies and is considering open houses to invite larger groups of stakeholders.

Possible Change in Current Systems

Review cycle: The pilot to shorten the review cycle is in progress and was discussed later in this meeting. If successful, it will be extended to more applicants. In addition, posting the summary statements earlier, as mentioned above, gives applicants more time to revise.

Clinical research: Dr. Scarpa reviewed data, presented to PRAC by Dr. Michael Martin in January, that showed that, on a percentage basis, clinical researchers (defined as those who use human subjects in their research [HS+]) are more unlikely to apply for a Type 2 A0 or A1 application than researchers who do not use human subjects (HS-); and that funded HS+ new PIs are also more unlikely to submit for another activity than HS- new PIs. These data show that NIH is losing successful HS+ researchers at a greater rate than HS- researchers.

Innovative research: Dr. Scarpa said that assessment of innovative, high-risk/high-reward research must be improved, although difficult cultural changes will be necessary.

Reviewers: Dr. Scarpa called the recruitment and retention of high-quality reviewers a crisis. SRAs report that it is more difficult to obtain reviewers, and many agree only if they can take on a reduced load. Maintaining the best reviewers is key to the success of peer review. The number of special emphasis panels (SEPs) has increased. In addition, the number of applicants and the average number of applications per investigator has essentially doubled the number of applications in the last 3-4 years. At the same time, reviewers review an average 6 grants per person, down from 11 about 10 years ago. Study sections are larger. To lessen the reviewers' burden, CSR has made a commitment to conduct 10 percent of all reviews electronically (via phone, video, or asynchronous discussion) by the end of 2006. This will help recruit clinical reviewers who cannot attend two-day meetings. Physicists and computational biologists, as well as those who use international reviewers, also prefer electronic reviews.

Dr. Scarpa hopes to decrease the number of reviewers and increase their level of experience, as well as increase the number of applications each person reviews without increasing their workload. Possible solutions include (1) replacing many SEPs with parallel study sections, (2) enlarging study section membership and decreasing frequency of participation, (3) convening pre-meetings to streamline applications, (4) using new electronic review platforms, (5) unscoring 40 percent of the postdoctoral fellowships (F32s), (6) shortening applications, and (7) creating more structured applications and reviews.

Pilots are under way to explore these approaches. The closest thing to a "silver bullet" to lessen the workload is a shorter, more structured application for some R01s. This change could increase the number of applications per reviewer and decrease the number of reviewers needed for a study section. There is strong, although not uniform, support for this change. A trans-NIH committee was recently formed to look at whether a shorter application is desirable and to design a pilot.

Discussion

Reviewer workload: Dr. Winkler said that the fact that applications are rising while fewer are funded on a percentage basis means a large body of effort is wasted by applicants and reviewers. Altering the system so people do not review grants with no chance of funding would yield an enormous benefit. He asked whether allowing people to apply for fewer grants would unburden

the system. Dr. Norka Ruiz Bravo said that Institutes and Centers (ICs) would have to decide if they would fund an investigator who already has a grant.

Dr. Lorraine Collins made three suggestions to deal with workload. First, foundations have a two-tier process. Applicants prepare a letter of intent and get feedback about whether they should prepare a complete application. She suggested that NIH consider a similar process. Second, perhaps reviewers should receive a higher daily rate. Third, perhaps there should be shorter terms for review panel members. She also expressed support for restructuring applications so they are easier to review. Dr. Scarpa said that the Pioneer Award uses a two-tier process. He then said that increasing payments would be problematic to a system that depends on reviewers who are essentially committed volunteers, receiving compensation only for the time they attend meetings and not for the many days they spend reviewing applications. An extra $100 would not be much of an incentive, and higher compensation would be difficult budget-wise and also might create unintentional reactions. Finally, he said that he gets feedback that reviewers would sign on for longer terms, perhaps 10 years, if they only had to attend one meeting per year.

Dr. Torok-Storb said that she strongly supports shortening the application and was very encouraged by the appointment of a committee to study the idea. She also suggested a system in which institutions that receive grants would be required to field reviewers on a percentage basis.

Basic and clinical research: Dr. Daria Mochly-Rosen said she was glad to see the progress made and asked how basic research is dealt with in study sections, specifically how much truly basic research is covered and supported by NIH. Dr. Scarpa said that both basic and clinical researchers feel that they are handicapped when funds are tight, but that basic science is doing reasonably. One reason behind more frequent review of IRGs is because the science changes. He noted that Dr. Zerhouni remains strongly committed to basic science. Dr. Ruiz Bravo said that the percentage of basic versus applied research has remained about the same for the past few years.

Dr. Dean Brenner said the financial model of academic centers is increasingly to have clinical and research "pots," each with overhead. The faculty is under pressure to address them both, with stress on clinicians to generate clinical resources. Dr. Scarpa said that clinicians' schedules often require holding a study section telephone meeting at 6:30 or 7 a.m. Dr. Brenner said that he, like other clinicians, is getting more pressure from his institution to spend more time in the clinic. Another issue raised by Dr. Brenner is the reduction in resources that might be addressed by capping the number of grants for individual investigators. He said he did not have an opinion, but wanted to raise the issue. Finally, he said that sharing study section appointments lessens the workload for an individual, but also affects the culture and flow of the section, and, therefore, perhaps also the quality of the reviews.

Dr. Berg noted that the National Institute of General Medical Sciences (NIGMS) has a policy that well-funded investigators ($750,000 in direct costs annually) will not receive additional funding unless staff and council make a strong case otherwise. It is not a cap, but more of a discipline in looking at well-funded investigators. Dr. Brenner noted that institutions are escalating their expectations of their staffs. Dr. Scarpa said that, traditionally, institutions paid the salary of the principal investigator (PI), but that cost has shifted to NIH. Dr. Ruiz Bravo

pointed out, however, that while NIH is paying more of investigators' salaries, the institutions now must spend more to comply with regulatory requirements.

Dr. Mochly-Rosen noted that there are very few NIH investigators with more than four grants. She said it was a good idea to push institutions to pay salaries but noted that the wide disparity in the percentage of salaries that institutions pay makes it difficult to set an absolute cap. Dr. Collins asked about the possibility of an indirect cost cap. Dr. Scarpa said that increasing requirements impose a burden on institutions. At one point, he read that the average institution loses 20 cents on the dollar when someone receives a grant. Dr. Ruiz Bravo said that indirect costs are capped for universities, and the wide range is due to the fact that they are negotiated by institution.

NIH Director's Pioneer Award Program

Dr. Berg explained that the Pioneer Award Program is one component of the NIH Roadmap and falls under the "high-risk research" implementation group area. It supports individuals, rather than projects, with demonstrated ability to solve important problems. The application process is very different than other NIH mechanisms. Applicants submit a five-page essay, three letters of reference, and a single representative work. In 2004, the first year of the program, a multi-tiered review process resulted in nine awards being made from a pool of 1,300 nominations, of which 20 were finalists. Each award is for $500,000 in direct costs per year for 5 years. Winners must commit at least 51 percent level of effort to their Pioneer Project.

The first awardees in 2004 were an outstanding group of scientists spanning a range of areas, but all were male. Also, as a group, they were more well-established in their careers than Dr. Berg thought was intended by the program. NIGMS was assigned to run the Pioneer Program in 2005 and instituted some changes to broaden the applicant pool. Only self-nominations were allowed, and outreach was stepped up to encourage women and underrepresented minorities in their early to mid-careers to apply. In addition, a section was added in the nomination form for applicants to address why the Pioneer Award was appropriate to their goals. The review process was also tweaked to bring in outside reviewers in the initial stage and to include a reviewer from outside an applicant's field of expertise.

Dr. Berg said that most, including a high percentage of the final winners, applied in the last few days before the deadline. He shared data about the 20 finalists that correlated their applications to the different award criteria. The correlation between impact and overall score was nearly complete. Twenty finalists, out of a pool of 800, were interviewed, and 13 were funded through a combination of Roadmap and Institutes' funds. He highlighted a few of the recipients.

The 2006 program was announced with some minor modifications. Rather than a two-tiered process, applications were submitted at one time through Grants.gov. More than 450 applications were received and a single round of outside reviews performed. Interviews for finalists will take place in August 2006.

Dr. Berg concluded with some personal views about the program. First, the shorter application can better focus on the investigator, the problem selected, and the evidence for innovation.

Second, there is no attempt made to match the expertise of the reviewers and the applicants, which has the advantage that the impact of the problem selected is looked at carefully. Finally, the interviews provide an evaluation of the applicants' credibility, which allows the reviewers to take more risks. He invited PRAC members to visit the website at nihroadmap.nih.gov.

Discussion

Dr. Leslie Leinwand asked about correlations between the awardees and their seniority and any previous NIH funding. Dr. Berg noted that a few applicants were assistant professors, while others were more senior but wanted to move in a new direction. The requirement that applicants address the appropriateness of the mechanism to their goals helps the reviewers.

Dr. Collins asked about the breakdown of clinical versus basic research among the awardees. Dr. Berg said the number of applications for clinical research has increased each year, but is still relatively low. About 10 percent were clinical applications in 2005, but none were funded.

Dr. Torok-Storb asked if any elements of the Pioneer Program could be incorporated into other reviews, such as the use of reviewers who were experienced but without expertise in the specific field, or the shorter application. Dr. Berg said that the Institutes are looking at these and similar issues now. Dr. Ann Sassaman said that the National Institute of Environmental Health Sciences (NIEHS) is looking at some of the concepts for its newly created Outstanding New Environmental Scientists Award. Interviews will be held for this award shortly.

Dr. Winkler asked about the cost on a per-awardee basis of this type of review. Dr. Berg said that the first few years had high set-up costs, particularly in setting up information systems. Review costs have been lower. Dr. Winkler suggested maybe this program should be expanded, given its lower cost and ability to address innovation. Dr. Berg said there is interest in expanding it.

Dr. Brenner asked about the role of individual salesmanship and charisma in choosing awardees through an interview. Dr. Berg said that even though the program was conceived as funding individuals and not projects, reviewers still weigh the potential impact of the proposed project heavily. Feedback from both successful and unsuccessful applicants was that they enjoyed writing the application and coming up with new ideas. He said this review process would not work for many other NIH mechanisms.

Dr. Ramm said that one of the Roadmap programs funded by the National Center for Research Resources (NCRR) asks applicants first to prepare a short paper, reviewed via Internet-assisted review; the subset who fare best then prepare full-blown applications. She also asked how the Pioneer Award would be evaluated. Dr. Berg said that a thorough process evaluation took place for 2004 and is under way for 2005. They are thinking about how to do an outcome evaluation.

In response to a question from Dr. Mochly-Rosen about the decrease in the number of applicants, Dr. Berg said that there was initial confusion about the criteria, with some people interpreting the Award as a prize for past accomplishment. About three-quarters of the 400-plus applications received in 2006 were new. He said that the program may be reaching steady state. Dr. Mochly-Rosen said that neither 2004, with no women receiving awards, nor 2005, when 6 out of the 13 were women, seemed an accurate reflection of the field. Dr. Berg said that in 2004, in the

scramble to find reviewers, it turned out that 60 out of 64 reviewers were men. In 2005, the reviewers were more proactively recruited, and the panels were more balanced. Beyond encouraging women and minorities in early to mid-career to apply, no additional coaching or assistance was given. The women finalists did very well in their interviews. Dr. Berg also clarified that the 51 percent level of effort was of research time, not overall time.

Dr. Collins asked if there were any plans to address the imbalance between basic and clinical research. Dr. Berg said that the clinical and behavioral sciences were explicitly mentioned in the 2006 announcement, and there is directed outreach to clinical groups. There was already a good balance of reviewers with clinical experience.

Peer Review Outcomes for the R03 and R21 Mechanisms

Two presentations compared review outcomes of the R21 (designed for developmental research) and the R03 (designed for smaller research projects) mechanisms with R01s reviewed in CSR, while a third looked at how the R03 is used in an Institute.

Review of R21s in CSR
Dr. Elaine Sierra-Rivera, SRA in the Oncological Sciences IRG, explained the features of the R21 and how it differs from other mechanisms. CSR reviews about 70 percent of all R21s submitted to NIH. SRAs emphasize to reviewers the indicators to focus on when reviewing R21s, such as the proposal's conceptual framework and significance to the field.

In the October 2005–May 2006 councils, a total of 8,579 R21 applications were received in response to various announcements. Of them, 73 percent were new, 21 percent were in their first revision, and 4 percent were in their second revision. Dr. Sierra-Rivera discussed how they fared in review as compared to the 23,445 R01s reviewed during the same time. She showed data that summarized the priority scores for Type 1 applications in study sections that typically review different mixes of applications. In both the study sections that primarily review R01s and the study sections that primarily review R21s and R03s, R21s and R01s fared about the same. In the study sections that primarily review small business applications, both the R21s and R01s fared a little better than the SBIR applications. There also did not seem to be a difference in scoring between R21 and R01 applications that involved human subjects compared to those that did not.

In summary, R21s are being evaluated fairly in CSR, and the study section environment does not affect the score distribution. Study sections are following the review criteria specific to the R21s. The SRAs play an important role in making sure this happens.

Review Outcomes of R03s in CSR
Dr. Valerie Durrant, SRA in the Health of the Population IRG, focused on review outcomes of the R03 mechanism within CSR. She explained the features of the R03, noting that applications increased from about 2,500 in 2001 to 4,000 in 2005. CSR currently reviews about 44 percent of them, so R03s are a small part of all CSR-reviewed applications. They mostly concentrate in a few IRGs in the Division of Clinical and Population-Based Studies. She noted that R03s are more likely than R01s to have a new investigator as the PI and less likely to be resubmitted.

The R03 guidelines instruct reviewers to focus on the conceptual framework and overall approach. Most are reviewed in standing study sections. Challenges include avoiding "R01 expectations," keeping budget considerations out of the review, ensuring a fair review when there are so few R03s, and finding the "peers" with R03-type experience who can review them.

The two main questions are: Does the score distribution of R03s differ from that of R01s? Does the score distribution of R03s differ when they are reviewed in different types of review groups? As with the R21s discussed above, Dr. Durrant focused on raw score distributions of Type 1 applications. She found very few differences in distribution of priority score between R03s and R01s, whether they were considered in standing study sections or in small mechanism SEPs. Reviewers are keeping the R03 guidelines in mind in their reviews.

NIDCD Small Grant (R03) Program
Dr. Craig Jordan, Director of the Division of Extramural Activities, National Institute on Deafness and Other Communication Disorders (NIDCD), spoke about the R03 from an Institute perspective. The program began in NIDCD in 1990. The current emphasis is to support scientists with no NIH or federal research support in the early stages of establishing an independent research career and transitioning from postdoctoral status to their first independent position.

Dr. Jordan explained the details of the R03 mechanism as used at NIDCD. Reviews occur in-house, three times a year, within a SEP that looks only at R03s. In FY 2005, there were 108 awards for $8.1 million. Nineteen percent of new applications are funded, but those numbers jump to 46 percent for A1s and 62 percent for A2s, for an average of 29 percent overall. Over the past 6 years, 58 percent of all applications have used human subjects, and about 51 percent of the successful applications used human subjects.

The goal of the program is to help unestablished investigators to then go on to compete for R01s and other higher-level awards. Dr. Jordan shared an analysis made in 2002 of the funding history of NIDCD-funded new investigators to obtain R01s. Looking at awardees from 1993 to 2002, 40 to 50 percent received R01s in subsequent years, often with a lag of a few years.

Dr. Jordan shared some conclusions from an FY 2002 evaluation. The funding comes at a critical career point and can serve as an important bridge to R01 support, although most successful applicants experienced a lapse in support between the two. A single, dedicated SEP allows for careful selection of reviewers and common orientation, goals, and review criteria. There has also been a high degree of continuity of reviewers, although NIDCD plans to try breaking out the applications into smaller, more scientifically focused panels.

Discussion
Further data analysis: Dr. Leinwand asked Dr. Sierra-Rivera and Dr. Durrant if they had outcomes for the mechanisms they discussed of the percentage of researchers who later obtained R01 support. Dr. Sierra-Rivera noted that the R21 is open to all investigators, not just new ones, so the correlation would be difficult. Dr. Torok-Storb asked whether data showed if some study sections do a better job than others in reviewing R21s and whether reviewers deal with them properly in deciding which ones to triage. Guidelines are given to reviewers when they are triaging applications. Dr. Torok-Storb said she has heard feedback that a lot of R21s get triaged.

Dr. Sierra-Rivera said that she could provide data that shows scoring history in more detail.

Scoring innovation: Dr. Mochly-Rosen said that the innovative nature of the R21 should mean, in fact, that there is not consensus among reviewers. Dr. Berg noted that when NIGMS had an R21 program, scoring was affected by the presence or absence of preliminary data, leading to concern about whether the mechanism was fulfilling its intent. Dr. Collins said that human nature makes it hard to change hats in reviewing different mechanisms. She also commended NIDCD, which extends support for up to 3 years.

Dr. Brenner said that the clinical study sections have struggled with some of the issues raised and the balance between innovation and the feasibility of an idea. There is a lot of variability in how to review the R21 mechanism, with a clustering phenomenon sometimes resulting. Dr. Sassaman suggested analysis go beyond the numbers presented, as she hears concerns from program staff that R21 applications are at a disadvantage. Dr. Sierra-Rivera said that she checked with National Cancer Institute (NCI) offices that have a lot of R21 reviews in their groups. They reported no difference in score distributions between R21s reviewed in study sections with more or fewer R21s in comparison to R01s.

Dr. Ruiz Bravo suggested that the R21s and R03s might fill a niche for some communities more than others. Dr. Mochly-Rosen agreed that the R21 is an important tool, but is not used to its full capacity. A spread of scores, rather than unanimity, should characterize an innovative grant.

Dr. Leinwand acknowledged the difficulty in tracking outcomes, but said that a critical issue is whether the R21s and R03s are used for their intended purpose. Dr. Sierra-Rivera said that she would discuss how to track the data with CSR analysts.

Interim Report on Evaluation of Shortening the Review Cycle

Dr. Bettie J. Graham, Associate Director of Extramural Research of the National Human Genome Research Institute and co-chair of the Shortening the Review Cycle Evaluation Design Subcommittee, briefed PRAC on behalf of the committee. Her presentation described the pilot, progress to date, and future plans, with the long-term goal being to allow for submission of amended applications in the very next review cycle.

Implementation of the Pilot
An initial study recommended a pilot in a few study sections of R01 applications from new investigators over three cycles, after which the NIH leadership could decide when and whether to move to a next phase to cover more study sections. The overarching principle governing the pilot was to maintain the core values of NIH peer review. About 600 applications were included in the pilot in 40 study sections from 10 IRGs. The committee will collect qualitative data from surveys and quantitative data from IMPAC II in order to conduct an evaluation for NIH decision makers.

In the pilot, applications were referred to study sections earlier than usual, and reviewers had four weeks to review them. Pilot study section meetings will be held earlier than other study sections. Summary statements were written within one week for R01s and within 30 days for other applications. Applicants can then discuss the feasibility of an early resubmission with

program staff. Dr. Graham stressed that an early resubmission was most suitable for an application that was not seriously flawed. Amended applications could be submitted 20 days after the receipt date for Type 2/amended applications.

This accelerated process requires bidirectional communication among the applicant, reviewer, and program staff. Committee co-chair Dr. Eileen Bradley, Chief of the CSR Surgical Sciences, Biomedical Imaging and Bioengineering IRG, is meeting with SRAs in CSR, while Dr. Philip Smith, Deputy Director of the Division of Diabetes, Endocrinology, and Metabolic Diseases and Co-Director, Office of Obesity Research of the National Institutes of Diabetes and Digestive and Kidney Diseases, is chairing the Triangulation Committee to facilitate communication between review and IC program staff. It is critical for program staff to know that they will receive calls from applicants earlier than usual about their summary statements and that applicants may ask them for advice about resubmission. Notices about the pilot appeared in the NIH Guide and as an addendum to the affected applicants' summary statements.

Evaluation
Evaluation instruments for the various stakeholders are being developed on a rolling basis, starting with referral staff, to be followed by surveys with reviewers, SRAs, and program administrators. Dr. Graham reviewed some of the questions asked in each instrument. Responses from referral staff have not been examined carefully, but Dr. Graham said that a preliminary look reveals that the results are confounded because IMPAC II had technical problems at the time of the pilot. Overall, the referral officers seemed to think that accelerated review was do-able. The SRAs will ask reviewers involved in the pilot to complete a survey at the end of their meetings.

Comparison groups among applicants will be pilot new PIs who submit early, pilot new PIs who do not submit early, and non-pilot PIs. In collecting applicant data, the committee will look at when applicants had access to their summary statements and how many resubmitted early, later, or not at all. They will also look at any differences in priority scores, funding rates, and time between the first review and any funding. She stressed that early resubmission must be separated from the funding situation in the evaluation.

The committee's next steps include data analysis, development of additional survey instruments, feedback to NIH leadership, and keeping others in NIH informed about the pilot and the evaluation. She closed by noting that all members are very engaged in the committee task.

Discussion
Scalability: Dr. Edward Pugh praised the goal and execution of the pilot. He questioned the speed with which the evaluation is being done and the scalability of the process, and asked that the PRAC have an opportunity to discuss the evaluation in open forum. He questioned whether the resources are available if the next phase is to expand to all R01s, particularly as electronic submissions of R01s is also getting underway. He also noted that the values not only of external stakeholders, but also within NIH must be taken into account. He said that accelerated review seems like a good idea, but recommended considering changes slowly and carefully. Dr. Graham agreed that electronic submission is a confounding factor that must be taken into account.

Dr. Scarpa said that electronic submission would help with faster assignments of applications to IRGs. Dr. Ramm said that she shared Dr. Pugh's concerns and philosophy. She underscored Dr. Graham's comment about separating out ICs' funding decisions from the accelerated process as applicants get in a fundable range. She also expressed concern about stress, especially on SRAs.

Dr. Torok-Storb asked about the analysis tool and specifically who is included in its denominator. Dr. Graham answered that it would include all the people in the pilot and the goal is to see any effect of coming back quickly on their scores. Dr. Ruiz Bravo stressed that a careful analysis would be done before moving forward. She suggested that PRAC receive copies of the evaluation forms and instruments to provide feedback. Dr. Sassaman also asked about how program staff would be surveyed, stressing the importance of keeping them involved. Dr. Graham said that Dr. Smith is talking to program staff and a survey is being planned.

Dr. Mochly-Rosen said that reviewing the impact of the program by comparing against a historical group was important. Dr. Collins asked about applications that are triaged. Dr. Graham said these applicants could also have a dialogue with a program director about resubmission.

Resubmission considerations: Dr. Calhoun said she had two suggestions, given the importance of the pilot. First, she asked if data could be collected on the outcome of resubmitted applications reviewed by the original reviewer. Second, she suggested that program staff might need some scripting so they are not blamed for misguiding applicants about accelerated resubmission.

Dr. Olivia Bartlett, Chief of the NCI Research Programs Review Branch and member of the Shortening the Review Cycle committee, noted that the policy is to purge reviewer information, so it would be difficult to implement Dr. Calhoun's first suggestion. There may be new reviewers looking at the application when it is resubmitted. Reviewers are told not to tell applicants how to fix a problem.

Reorganization of Study Sections in the Risk, Prevention and Health Behavior IRG

Dr. Anita Miller Sostek, Director of the CSR Division of Clinical and Population-Based Studies, spoke about creating a new study section entitled Risk, Prevention and Intervention for Addictions (RPIA). She briefly reviewed the principles for modifying study sections. In this case, a SEP was first formed to deal with workload issues in two existing study sections in the IRG, and it considered applications related to psychosocial risk, such as substance use and abuse. After several review rounds, it was clear that these applications formed a coherent, well-integrated area of science.

By the May 2006 review round, the SEP had a full review load. At the same time, content for a possible standing study section was discussed with experienced reviewers, program officials, and leaders in the field, and many stakeholders helped with planning and guidelines. Dr. Sostek reviewed the topics covered and characteristics of the reviewers in the SEP. For the October 2006 round, the SEP has 73 applications to consider (mostly R01s), and the largest from the National Institute on Drug Abuse. Draft guidelines for the study section have been written and are being circulated. During a telephone conference, support for the need for the study section

was unanimous. Comments from the conference were used to update the guidelines.

Discussion
Dr. Collins, who has participated in reviews in the SEP, concurred about the importance of creating the standing study section. She said other colleagues are similarly enthusiastic. She asked that research around the acute effects of alcohol on behavior be included.

Dr. Brenner asked about the interaction between the genetic and behavioral scientists. Dr. Sostek said that collaboration tends to work better with behavioral geneticists familiar with these types of studies. Dr. Calhoun praised the formation of the study section and suggested that the councils of affected Institutes be briefed on its formation.

Dr. Pugh praised the formation of the study section as a wonderful example of the evolution of the peer review process. He also suggested that this study section might benefit from some kind of involvement of advocacy or other nontraditional groups.

Dr. Scarpa asked for a motion to approve formation of the new study section and it was passed unanimously.

NIH at the Crossroads: Myths, Realities, and Strategies for the Future

NIH Director Dr. Elias Zerhouni thanked PRAC members for their ongoing contributions to NIH peer review. Over 31,000 individuals serve each year on NIH councils, committees, workshops, and peer review panels; together they make NIH the best Government agency. He then focused on the difficult budget environment at NIH. Conditions for a "perfect storm" exists: Federal and trade deficits have gone up as the country seeks to meet needs related to defense, homeland security, the aftermath of Hurricane Katrina, the threat of pandemic flu, and the need to better fund the physical sciences. The pressure on the NIH budget is compounded by the fact that the inflation rate for biomedical research is 3 to 5 percent higher than the general inflation rate.

Myths
Dr. Zerhouni shared feedback he hears from scientists. Many believe NIH favors applied over basic research, solicited over investigator-initiated research; and its new NIH Roadmap initiative over regular grants. He then addressed each issue.

Basic research funding remains strong: Dr. Zerhouni reviewed basic and applied research allocations from FY 1998 to FY 2005. Except for a dip in basic research after 9-11 for biodefense investments, the balance between basic vs. applied research has remained relatively constant—about 54 to 56 percent basic and about 40 percent applied.

Funding for solicited research has declined relative to investigator-initiated research: In FY 1995, 91 percent of NIH grants were unsolicited, and 9 percent were solicited, such as through program announcements and requests for applications. In FY 2005, the ratio was 93 percent to 7 percent.

Roadmap funding: Roadmap funding represents a very small part of the total NIH: 0.8 percent in FY 2005, 1 percent in FY 2006, and 1.2 percent in 2007. The Roadmap is funding 400 separate grants to 350 new investigators—40 percent in basic science, 40 percent in translational research, and 20 percent in high-risk science. The Roadmap was created to address community concerns that NIH is too conservative by finding new ways to address roadblocks to science.

Realities
Dr. Zerhouni then discussed three realities will influence the ability of investigators to secure NIH funding.

Increased capacity in U.S. research institutions: Tremendous capacity increases at U.S. research institutions—new labs and more tenure-track faculty—in the last few years led to a significant increase in applications just as the doubling of the NIH budget ended.
Dr. Zerhouni noted that there were as many new applicants in FY 2003–2005 as in the previous 5 years.

Budget: While NIH appropriations have increased, they have not kept up with inflation in recent years; and across-the-board cuts to respond to Katrina led to a flat FY 2006 NIH budget.
Dr. Zerhouni noted after 9-11, about $2 billion or 20 percent of the funds devoted to doubling the NIH budget were redirected to biodefense.

Budget cycling: In any given year, the funds available for new grants come from uncommitted funds, or the money that comes from grants that are ending, as well as any budget increase. The funds available in 2006 come from grants that ended 4 or 5 years ago—which was before the doubling. But in 2007–2009, more funds will be recirculated back into the system from grants that began during the doubling.

Educating Stakeholders
Dr. Zerhouni said NIH is concerned and is seeking to educate the public about the need for sustainability in biomedical research. In his recent Congressional testimony, he stressed the long-term nature of medical research and showed examples of the impact of research on human health.

Returning to the applicant concerns about funding levels, Dr. Zerhouni noted that success rates are higher than pay lines. The success rate per application in 2006 is currently about 19.8 percent. Many applicants submit more than one application, and many revise and resubmit applications; hence, the expected success rate per applicant is greater—about 25 percent.

Future Strategies
Dr. Zerhouni said that the situation like the current ones have occurred before. He had four ways to deal with them: (1) know the facts; (2) develop adaptive strategies that are true to the NIH mission, such as increasing the number of competing grants as much as possible, and supporting new investigators so they do not become discouraged and leave the field; (3) convey a unified message about the positive impacts NIH has had in helping to save lives and to spur economic development; (4) articulate an exciting vision for NIH that shows its impact on the national interest.

Dr. Zerhouni acknowledged that increases in number of applications and applicants complicate the task of peer review. He urged PRAC to explore ways to make the process more efficient for reviewers, applicants, and NIH. The NIH goal is clear: to transform medicine through discovery. Basic research lies at the base of a pyramid that supports translational and clinical research, which will ultimately advance the delivery of health care. Dr. Zerhouni explained how a major paradigm shift is occurring as medicine moves beyond seeking to cure disease and to preventing it by focusing on new predictive, personalized, and preemptive measures.

Discussion
Dr. Mochly-Rosen agreed with Dr. Zerhouni about the need to educate the public. She identified two key issues: the long timetable of return on biomedical investments, and the lack of hard data to show the ultimate success of basic research. Dr. Zerhouni said he receives very positive reactions when he is able to show the direct impact of research on people's lives.

Dr. Torok-Storb suggested that it might be useful to show how NIH funds have a positive impact on the state level. Dr. Zerhouni agreed, saying that he tells policy makers that, from 1998 to 2004, almost 4,100 new technologies were licensed from institutions that received NIH funding, and thousands of new companies were created.

Dr. Brenner said that the important message he took away from Dr. Zerhouni's presentation was to encourage new investigators—particularly those doing translational research—to "buck it out" until better times so we do not loose a generation of researchers. Dr. Zerhouni said the Clinical and Translational Science Awards, partly funded through the Roadmap, will create a home for the next generation of clinician-scientists and translational scientists.

Adding a Study Section in the Oncological Sciences IRG

Dr. Elliot Postow, Director of the CSR Division of Biological Basis of Disease, spoke about creating a new study section in the Oncological Sciences (ONC) IRG, as the number of applications referred to three existing study sections has increased. A new SEP was formed in June 2005 as the number of applications per study section reached 100 or more. A panel from the extramural community advised CSR and helped formulate guidelines. If approved by PRAC, the SEP would become a new study section, entitled Molecular Oncogenesis (MONC), and would hold its first official meeting in October 2006. Dr. Postow explained the new study section's guidelines, as well as resulting adjustments in the three existing study sections. He reviewed the distribution of applications and said that the new study section would help balance the workload.

Discussion
Dr. Calhoun said that this change benefits a number of Institutes, but that the workload, which is still large, bears watching. She also asked about the number of R01 applications within the Cancer Etiology study section, since, according to Dr. Postow, it still reviewed 129 applications even with the new addition. Dr. Postow said it had a mixture of types of applications to review. Dr. Brenner said the reorganization and addition of the new study section reflects the state of the art of the field.

The motion to approve formation of the new study section passed unanimously.

Principles and Philosophy of Evaluation and Evolution of Study Sections and IRGs in CSR

Dr. Don Schneider, Director of the CSR Division of Molecular and Cellular Mechanisms, said that his presentation would provide context on how study sections and IRGs are organized. The system is set up to further the core values of peer review. Organizationally, IRGs seem to be the right-sized work unit, each with about six study sections that can cluster related science.

From 1945 to 1998, new study sections were formed on a case-by-case basis. Broad changes took place in the late 1990s with the merger of the Alcohol, Drug Abuse and Mental Health Administration (ADAMHA) into NIH, the formation of the AIDS IRG, and the formation of the Panel on Scientific Boundaries for Review (PSBR). The PSBR was designed to look broadly at how to organize study sections and to involve the scientific community in the process.

The current organization is based on the PSBR work. CSR has 17 IRGs that focus on an organ system or disease, three that focus on basic scientific discovery, and three that focus on the development of methods and crosscutting science. The organization is not expected to be static and is continually monitored to ensure that the core values are maintained, the organization reflects changing scientific opportunities, and communication and transparency are promoted.

Dr. Schneider outlined the procedures for systematic evaluation of study sections: identifying issues, concerns, and problems; collecting and analyzing data; and involving applicants, reviewers, relevant advocacy groups and scientific societies, and NIH staff in suggesting any changes. Working groups' assessments of IRGs are on a 5-year cycle, and CSR is now on the second 5-year cycle, starting with those related to neuroscience and to small business. The others will follow in the order of reorganization. The results of each assessment are presented to PRAC, as required by the Federal Advisory Committee Act. Working groups may also look at trans-IRG issues or broader principles or practices related to the CSR structure.

In addition, as Dr. Scarpa described earlier, the IRGs are monitored more frequently. Workload is reviewed every cycle, and IRG chiefs and other senior staff attend as many study section meetings as possible. In addition, CSR staff focus on one IRG per month, spending as much as a half-day looking at data and hearing from the SRAs. These monthly meetings might point out the need for minor modifications. For more substantive issues, such as the new study sections discussed by Dr. Sostek and Dr. Postow, a working group is formed.

In addition, Dr. Scarpa contacts retiring study section chairs. CSR staff attend scientific society meetings and host visits by members of these societies. CSR is considering inviting representatives from about 30 groups at a time to explain CSR in a more organized way and have an open discussion. Dr. Schneider concluded by stressing the importance of communication and transparency with all stakeholders.

Discussion

Dr. Leinwand asked whether a working group has been formed to focus on how to obtain and retain the top reviewers. Dr. Scarpa said that he thought PRAC, or perhaps a working group that included some PRAC members, would be the way to handle that issue. Dr. Leinwand suggested

putting the topic on the agenda of the next PRAC meeting.

IRG Review: Dr. Pugh said that the system described by Dr. Schneider shows an amazing process of government self-examination. He asked for clarification about reviewing the review process itself. Dr. Schneider said an example of an issue to look at is the fact that study sections have become quite broad and involve more reviewers. Dr. Scarpa said that looking at more study sections will help point to any concerns related to workload or overlapping areas. In response to a question from Dr. Pugh about whether the IRG system itself is at stake, Dr. Scarpa said he does have this view, but the system could be examined if PRAC wished. Dr. Pugh stressed that this was not his recommendation because, as shown in Dr. Postow's presentation, the formation of a new study section from parts of others shows how the IRG system can work.

Dr. Scarpa asked for PRAC feedback on the idea of open houses for scientific societies as a proactive way to reach out to them. Dr. Pugh expressed support for the idea.

Dr. Sassaman said that NIEHS has been exploring the possibility of a new IRG with a segment of its constituency. Rather than rely on anecdotal information to make decisions, they have worked with CSR to look at data to analyze any problems and potential solutions. She suggested developing a process when new IRGs are proposed, because looking at a new IRG, rather than new study sections, means going back to PSBR and the whole process of community engagement. Dr. Ruiz Bravo said that the issue becomes one of evaluation: How is "success" defined? She suggested developing with some definitions and criteria to determine when a new study section, or perhaps an entire IRG, should be considered. The structure should serve a function.

Dr. Schneider said that CSR thinks of success in terms of review outcome. Dr. Scarpa asked PRAC to think about criteria for evaluating an IRG. Dr. Pugh said people would need forewarning of any change. The IRG structure has provided a great deal of stability to the organization and has allowed change without overhauling everything.

Dr. Torok-Storb asked what staff looks for when they visit study sections. Dr. Scarpa said that he looks at the dynamics of the meeting, especially those with large (60 or more) numbers of participants. Staff is not there to intervene. Feedback is presented in a general way. He meets with SRAs, but he also does not want to micro-manage.

Dr. Brenner said that it seemed like two approaches are used to look critically at the function of the organization in real time: an internal evaluation system and a "customer survey" by talking with scientific societies. He asked about how IC feedback is solicited. Dr. Scarpa said that the ICs are involved in all decisions. Dr. Brenner said he liked the idea of CSR staff attending society meetings, but he was not sure of the value of inviting their representatives to CSR. Dr. Scarpa said that open houses would be a way of involving all the societies. Dr. Collins said she thought the discussion would be informative. Dr. Brenner asked more about how the discussions would be set up. Dr. Scarpa said that details are being worked out, but acknowledged that preparation is very important as group meetings can be more or less successful.

Dr. Ruiz Bravo noted that policies about peer review are done in an analytical way at NIH and go beyond CSR. The Extramural Activities Working Group (EAWG) and NIH Steering

Committee would be involved in any changes, as would ICs. She agreed that developing criteria by which to make decisions about change would be helpful.

Statistical Parameters in Peer Review

For the final presentation of the day, Dr. Scarpa introduced Dr. David Kaplan, professor of pathology at Case Western University. Dr. Scarpa noted that he and Dr. Kaplan were at Case Western together, but that it is Dr. Kaplan's ideas that he has published about peer review that elicited the invitation to speak to PRAC.

Dr. Kaplan began by stating his assumptions: (1) CSR is the gatekeeper for NIH and an important player in affecting NIH policies; (2) NIH has difficulty in recognizing innovation in part because there is not a good measure to do so; (3) to promote innovation, it is most reasonable to make changes in CSR procedures. He said this syllogism comes from the minutes of the 1997 Peer Review Oversight Group. In addition, he said that peer review should utilize statistics in the most robust or powerful way possible and that peer review should reflect the peer group as broadly as possible.

Sampling for NIH Peer Review
The current system in which two or three reviewers read through each application utilizes only very small sample sizes. Peer review uses quota sampling, rather than random sampling, which makes it subject to bias. Small sample sizes are constrained by the size of the grant applications, yet even four or five hours spent reading a grant that was literally years in the making means that peer review is a low-precision exercise. Because peer review involves discussions among reviewers that may result in altered scores, the scores are not independently derived.

Peer review as currently practiced uses an arithmetic mean. The scores produce ordinal evaluations (rank ordering), but not assessments on a parametric scale. As an alternative, other statistical calculations known to be important for nonparametric evaluation should be considered.

In order to use statistical parameters other than the arithmetic mean, larger samples would need to be collected in a more random manner with independence among the opinions offered. The implications of this alternative would be shorter applications, a different selection scheme for reviewers, no meetings, and scores set in a low-precision scale.

Dr. Kaplan stressed that these changes could be complementary to the current system, and not a replacement, as a means to identify innovation. The current system has done well in identifying excellent proposals. But innovativeness and excellence are not the same thing, and there is no robust measure for innovativeness at present. Dr. Kaplan said that perhaps statistical measures other than the arithmetic mean could better identify innovative grant proposals.

Hypothesis: Dr. Kaplan said that his hypothesis, which he said was a first approximation hypothesis that could be modified with data, is that variance (the scatter of distribution) and/or kurtosis (the peakedness of the distribution) would be a robust indicator of innovativeness. Innovation should elicit controversy because the ideas are new and unusual, while proposals close to what is already generally accepted tend to engender consensus. Variance and kurtosis are statistically valid measures that could indicate the degree of controversy or consensus

associated with a proposal. These measures would require a statistically robust system of sampling and scoring other than what is currently used.

Dr. Kaplan reiterated that innovation and new ideas naturally collide with more established ideas, hence leading to controversy. However, innovation can sometime lead to large advances in understanding, which makes it important and valuable. He then showed four examples of different scoring distributions by 30 hypothetical reviewers. Each had different mean, variance, and kurtosis measures. The applications with high means, low variances and positive kurtosis would be the excellent, more traditional grants that tend to score well under the current selection process. The applications with negative means, higher variances, and lower kurtosis would identify the more controversial and more innovative ideas.

Proposed Test: Dr. Kaplan proposed a test that would utilize various statistical parameters to identify innovativeness and to determine the number and types of reviewers that would be needed to provide stable values for these parameters. The experiment might use one-page grant applications with varying degrees of innovativeness and as assessed by an independent panel of between 20 and 100 reviewers. The reviewers would be asked to score the proposals, as well as provide information on the length of time needed to review each proposal, their seniority, and the relative closeness of the proposal their area of expertise. The scores could then be analyzed to ascertain the number of independent evaluations needed to obtain stable statistical values, and the role of seniority and relative closeness to the reviewer's area of expertise.

Dr. Kaplan said that this system would be appropriate when looking for innovation. For established ideas that require further development, the traditional system would be more appropriate. The proposed new paradigm would also enable evaluation of the system itself.

Dr. Kaplan concluded by listing potential benefits of a system with more robust statistics: it would minimize bias, provide greater satisfaction to scientists, provide greater control for administration, identify innovativeness along with excellence, and solidify CSR as a flexible and intelligent regulator of NIH granting activities.

Discussion

Dr. Mochly-Rosen said she was happy to hear Dr. Kaplan's thoughts in more detail. She said that she had found the negotiation of scores at her first study section meeting surprising and agreed that breakthrough ideas might suffer in that process. She proposed a focus on the R21s, as this mechanism already exists to support innovative ideas. She also asked whether resources were available to fund the type of test that Dr. Kaplan proposed.

Retrospective analysis: Dr. Leinwand asked whether existing data from R21 reviews could be analyzed. Dr. Kaplan replied that the sampling would be problematic, as the analysis would start with a contaminated dataset. Dr. Leinwand also asked how an applicant's identity would be factored into a test. Dr. Kaplan said that he did not think that blinding the identity would be required, as it is just one part of an application. Moreover, he said, because reviewers would score on a low-precision scale (1 to 5), they would be giving a general impression. In terms of re-examining completed reviews, Dr. Mochly-Rosen said that scores are already negotiated with the 41-point scale, which leads to consensus and does not give "out of the box" proposals a chance.

Test design: Dr. Torok-Storb agreed it would be a great experiment and could be done inexpensively over the Internet. She said it would be interesting to see what would happen if R21 proposals went out in a brief form as part of an electronic survey. Dr. Collins also supported the experiment, but said that innovation is not the only concern of NIH or any other program of research. She wondered how feasibility enters into the equation. Dr. Kaplan said that feasibility comes into play in considering the potential for advancement of the idea. He said that while he is pegging this concept to innovation, perhaps there is another term for a proposal that skews to the high end with a positive kurtosis and high variance. It would be useful to collect data to find out what kind of scores should be reflected in proposals that have desirable characteristics for a variety of different granting programs. Such a tool could expand the way of looking at grant applications as opposed to the linear way used currently. The linear scale has been proven successful over the long term to identify excellence and should not be eliminated.

Defining innovation: Dr. Pugh said there have been many discussions over the past few years about the small sample size of reviewers. In terms of reviewers' independence, one important change made by CSR has been to put scores into electronic storage before study sections meet. But, he said, this issue is distinct from the issue of innovation. He said there might be a number of alternative hypotheses than Dr. Kaplan's hypothesis about the relationship between innovation and measure of variability. Rather than look for data to confirm Dr. Kaplan's hypothesis, he said he would find ways to test a variety of hypotheses that might explain variance in people's scoring behavior. He also brought up the issue of practicality. To get a robust or reliable measure of variance would also mean a huge sampling problem. His main concern, he said, is the need for an independent, agreed-upon definition of innovation before testing a hypothesis to measure it. Dr. Kaplan agreed that his hypothesis might not win out in the end.

Dr. Winkler said that he was intrigued by Dr. Kaplan's hypothesis. He agreed with Dr. Leinwand that there might be a way to look retrospectively at existing data, perhaps by writing one-page summaries of last year's applications and sending them to different reviewers. Dr. Ruiz Bravo said that if PRAC agreed with the concept of the test, members should not worry about how to do it. PRAC's recommendation would go to the EAWG, who would explore how to get it done or if it is not possible to do.

Dr. Winkler agreed that a determination of how to measure innovativeness is important. As a crude measure, he suggested it could be how many previous grants that scored well in the conventional versus alternative mechanism end up as articles in Science or Nature. Dr. Kaplan agreed with him and Dr. Pugh about the necessity of a gold standard to identify innovation.

Dr. Ruiz Bravo said that the focus here is on peer review, but that program people who make funding decisions in ICs are also an important component in recognizing innovation. Review is part of the equation. Dr. Pugh said he found the discussions about innovation extremely important.

Dr. Scarpa said that an alternative mechanism to test could be discussed at the next meeting.

Action Items and Final Discussion

Dr. Martin read through the action items for the future: (1) the next PRAC meeting will have a report on the shortening of the application; (2) Dr. Sassaman will present about the NIEHS program that mirrors the Pioneer Award; (3) An attempt will be made to obtain more outcome data on the R21 and R03 mechanisms; (4) CSR will send out forms and ask for feedback on shortening the evaluation; (5) Dr. Zerhouni's presentation slides will be sent out to PRAC members; (6) Staff will develop a document on SEP principles that will go back to PRAC; (7) The concept about pursuing a test of the hypothesis of an alternative mechanism on innovation will go to the EAWG.

To this list, Dr. Leinwand added setting aside time at the next PRAC meeting to discuss recruiting and retaining high-quality reviewers. Dr. Torok-Storb suggested a discussion on allowing post-doctoral fellows on T32s to apply for an additional year of lab support. Dr. Mochly-Rosen agreed with the need. Dr. Scarpa agreed on the importance, but said that the issue is not within the purview of peer review. Dr. Calhoun suggested following up to see if there are alternative hypotheses for innovation that might be tested. Dr. Scarpa requested that PRAC members contact CSR with thoughts on that issue and thanked the members for their participation. PRAC adjourned the meeting at 3:54 p.m.

We do hereby certify that, to the best of our knowledge, the foregoing minutes of the May 2006 meeting of PRAC are accurate and complete. The minutes will be considered at the August 2006 meeting of the Advisory Committee, and any corrections or comments will be made at that meeting.

Michael R. Martin, 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