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

December 4, 2006

The fourth 2006 meeting of the Peer Review Advisory Committee (PRAC) convened at 8:30 a.m. on Monday, December 4, 2006, in the Natcher Conference Center, National Institutes of Health, 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
Toni Scarpa, M.D., Ph.D., Co-Chair
Dean Brenner, M.D.
Joe Martinez, Jr., Ph.D.
Matt Winkler, Ph.D.
Craig McClain, M.D.
Edward Pugh, Jr., Ph.D.
Louise Ramm, Ph.D.
Beverly Torok-Storb, Ph.D.

Ad Hoc Members

R. Lorraine Collins, Ph.D.
Leslie Leinwand, Ph.D.

Dr. Daria Mochly-Rosen, Ph.D., was not present. Dr. Cheryl Kitt, Ph.D., was the Executive Secretary for the meeting.

Welcome, Approval of the August 2006 PRAC Minutes, and Upcoming Meetings

Dr. Scarpa welcomed participants and introduced Executive Secretary Dr. Cheryl Kitt. Dr. Kitt asked for approval of the minutes of the August 2006 meeting. The minutes were unanimously approved. She then announced PRAC's next meeting dates: April 19, August 27, and December 3, 2007, and April 7, August 4, and December 8, 2008. The dates were chosen to avoid overlapping with Institute and Center (IC) Council meetings. She asked PRAC members for confirmation at least for the 2007 proposed meeting dates.

Changes in CSR Operations: An Integrated Vision for Peer Review

Dr. Scarpa prefaced his progress report on CSR by noting recent travel outside the United States confirmed for him the strategic value of peer review to advance science and health. Unlike in other countries, most U.S. Government funding of biomedical research is allocated through peer review. Although it has served very well, however, peer review is now a system under pressure.

Dr. Scarpa receives five major complaints about peer review: (1) the process is too slow; (2) there are not enough senior/experienced reviewers; (3) the process favors predictable over innovative research; (4) clinical research may not fare as well as other research; and (5) the application and review process burdens applicants and reviewers. The increased number of

applications, combined with reviewers busy with their own grants, leads to some of these concerns. Dr. Scarpa discussed current and potential changes in CSR to address them.

Changes in CSR Operations
Increased communication and transparency: CSR is increasing communication with stakeholders to make its operations more transparent.

Increased uniformity: CSR is taking steps so that the approximately 260 scientific review administrators (SRAs) coordinate study sections in a similar way. For example, summary statements are now posted within 1 month of the study section meeting, with new investigators' statements posted within 1 week. An appeals committee, chaired by Dr. Don Schneider, Director of the CSR Division of Molecular and Cellular Mechanisms, was recently formed to deal with applicants' complaints, many of which deal with a lack of consistency in how applications are handled. The committee is working well, starting with the fact that applicants now know that it is in place and will respond to their concerns within a week. In addition, Dr. Kitt chairs a Best Practices Committee with representatives from throughout CSR

Increased efficiency: Electronic submission opens new avenues to increase efficiency. A text fingerprinting and artificial intelligence software pilot is under way, with full implementation scheduled for June 2007. This process will save 2 to 4 weeks in the review process as well as improve it. (Dr. Thomas Tatham, CSR Associate Director for Knowledge Management, reported on the pilot later in the meeting.)

Improved study section alignment and performance: CSR has aggressively responded to complaints that the alignment of study sections, if undertaken too infrequently, can jeopardize some applicants. Each month, one Integrated Review Group (IRG) is reviewed. Minor adjustments are made by the staff; more substantive changes are made with the affected community and brought before PRAC. Starting in early 2007, CSR will host six open house workshops so leaders from the various scientific communities can meet with CSR staff, study section chairs, and others. PRAC input on the workshops will be needed.

CSR is finding ways to deal with a lower travel budget. For example, a pilot to purchase nonrefundable, restricted tickets has saved about $4 million to date. (Ms. Diane Bernal, Director of the Scientific Review and Evaluations Activities (SREA) Management and Service Center, reported on the pilot later in the meeting.)

Vision for Peer Review
Looking at the big picture, NIH must find ways to shorten the review cycle, recruit and retain more experienced and higher-quality reviewers while decreasing the burden on applicants and reviewers, and improve the identification of innovative and high-impact research.

Shortened review cycle: The pilot in which new investigators were able to revise and resubmit an application in the very next review cycle has worked well so far. Dr. Scarpa reviewed the numbers of eligible applicants, the percentage who did so, and how they fared. He hopes to expand the program to all new investigator R01s by the end of 2007 and then to all investigators.

Reviewer recruitment and retention: Currently, about 18,000 reviewers are used, and the number of applications keeps increasing. He singled out the R21s, which increased from 1,000 to 9,000 applications in about 6 years and which respond to 220 program announcements with different requirements. Study sections are getting larger, which takes away from the chemistry of people working together. There is a decrease in the number of applications per reviewer. Part of the reason is that the science may require an ad hoc reviewer to cover just one aspect of an application, but, overall, there are too many ad hoc reviewers.

Near-term solutions to recruit the best reviewers include using electronic modes, such as telephone-enhanced, video-enhanced, and asynchronous discussions, to decrease reviewer travel. He stressed the best review platform varies from science to science. The goal by the end of 2007 is that 10 percent of all reviews are electronic, up from 6 percent now. Other potential solutions include compressing face-to-face meetings into 1 day and shortening the application length.

The Trans-NIH Committee to Shorten Applications is focusing on the R01, considering a reduction in the page limit and aligning the application more closely with review criteria. This committee is co-chaired by Dr. Robert Finkelstein, Director of the Division of Extramural Research at the National Institute of Neurological Disorders and Stroke, and Dr. Schneider. It has strong support by most, if not all IC Councils and Directors, the scientific leadership, and PRAC members. Reviewers should be able to read more applications, thus reducing the number of reviewers needed and keeping study sections smaller. Shorter applications also may facilitate reviewer recruitment and sharpen focus on impact and innovation. Dr. Scarpa acknowledged a shorter application is a cultural change because it would change how applications are written and reviewed. Responses received to date from a survey on application size show that almost three-quarters (74 percent) favor shorter applications.

Dr. Scarpa said the great people in both NIH review and program offices make these changes do-able. Peer review is important for scientists but, more importantly, for success in fighting diseases.

Discussion

Appeals Committee: Dr. Joe Martinez praised Dr. Scarpa's comprehensive strategy to speed up and improve reviews. He asked how the appeals process works and the types of issues the committee handles. Dr. Scarpa said the committee has operated for 3 or 4 months and has dealt with procedural issues such as deadlines or study section assignment. It averages one or two complaints a week. Dr. Schneider added recently the committee has dealt with what are termed A3 issues—A2 applications that receive good scores but are not funded. Applicants want to submit them again; however, NIH policy states an application can only be submitted three times. Dr. Scarpa stressed that this committee focuses on administrative matters; if applicants have concerns about the way their application was reviewed, they should contact their IC program officer.

Reviewer recruitment and shortened applications: Dr. Martinez asked how potential reviewers are identified. Dr. Scarpa said many strategies are used. First, there is an attempt to reduce the number needed for a study section. Professional societies recommend potential reviewers. Full-time staff within CSR will be responsible for collecting and following up on suggestions. He said

some people suggest a higher per diem for reviewers. However, a raise would not only have budget implications, but also alter the peer review culture. Another suggestion is giving permanent reviewers an extra year on their grants, but this, too, would have large financial implications. Alternately, some have suggested that grants carry a requirement to serve as a reviewer. The goal is to make the peer review experience rewarding from a scientific point of view.

Dr. Leslie Leinwand asked about results from the survey about whether shorter applications would result in willingness by reviewers to read more applications. The responses have not yet been analyzed. Currently, reviewers spend an average of 7.2 hours to read an application.

Dr. Louise Ramm asked whether supporters of shorter applications identified themselves as reviewers or applicants. Dr. Scarpa said the data he provided was very preliminary, but that both reviewers and applicants seem to support the change. Most favor 15 pages, 10 to 20 percent favor 10 pages, and 5 percent favor 5 pages. She also asked for data about outcomes when applicants submit in the very next cycle versus waiting another round. Dr. Scarpa said program officers have said investigators did somewhat better, but only a small number have gone through the process to date. Dr. Ramm also expressed concern about SRA burnout in this accelerated process. Dr. Scarpa said the total amount of work is the same but distributed differently.

Consistent scoring: Dr. Craig McClain suggested study section chairs update ad hoc reviewers who participate via telephone so their scoring is in line with the rest of the study section. He also said the flattening out funding for R01 applications may be behind the large increase of R21s and, thus, the need for more reviewers. The various criteria of R21s are difficult for reviewers in study sections. Dr. Scarpa said that the proliferation of R21 Program Announcements, combined with the increasing number of applications, is a problem.

Potential drawbacks: Dr. Lorraine Collins said many issues relate to funding. With less funding, applicants submit more new and revised applications. She also warned returning reviews too quickly may mean SRAs cannot do the quality of work required, which will be detrimental to applicants. She said shortening applications does not necessarily mean reviewers will review more of them. Reading grants requires thought, and it is the content, rather than the number of pages, that takes time. She also said the emphasis on interdisciplinarity and collaboration works against reducing the size of study sections. Funding and dealing with these broad issues should be considered, instead of tweaking around the edges.

Dr. Edward Pugh said that most people would favor a shortened application at first blush, but there are some downsides. For instance, investigators may end up submitting two shorter applications, therefore, the workload of reviewers will not decrease. Going in the direction of a shorter application is irreversible, so the decision must be thought out. A shortened application will focus on other criteria, such as innovation. The R21 was supposed to engender innovation, but experience shows how difficult innovation is to evaluate. Two things to consider are whether or not a shorter application will result in a better, deeper review process and whether the system can handle a move to emphasize innovation. Dr. Scarpa said he thought the discussion should center on the criteria of impact rather than innovation. He asked for suggestions about how to structure a pilot to evaluate these questions. Dr. Pugh stressed he is raising these questions

because PRAC should serve as a devil's advocate, especially in the face of such a big change. Dr. Leinwand said from previous meetings and conversations with colleagues, respondents who favor shorter applications did not assume that they would spend a lot less time preparing a 15-page application.

Dr. Dean Brenner asked about follow-up when the comment period closes. Dr. Scarpa said there is a high level of interest in peer review. NIH Director, Dr. Elias Zerhouni, and a small advisory committee he has convened will get the results the survey, but there is no specific plan about next steps. Dr. Pugh said that it would be difficult but necessary to pilot a shorter application. Dr. Scarpa said any plans would be brought to PRAC.

Dr. Beverly Torok-Storb agreed that funding is the "elephant in the room," and a shorter review cycle will not result in more projects funded. She also asked why well-scored A2 applications cannot be resubmitted. Dr. Scarpa said reviewing them again will not make a difference in terms of getting funded. Dr. Torok-Storb said the burden on SRAs to get summary statements posted rapidly is incredible, and the quality of information can suffer if summaries are pushed out too fast. To recruit reviewers, she suggested requiring institutions to field reviewers in some proportion to the amount of funding they receive, as well as videoconferencing from bigger institutions that have good facilities. Dr. Scarpa said security issues would prohibit other institutions from convening review sessions or from having too many reviewers from one location. Dr. Torok-Storb observed it was considered a great privilege to be asked to participate in a review 20 years ago. Creative ways to revive that atmosphere are needed. Dr. Eileen Bradley, Chief of the Surgical Sciences, Biomedical Imaging and Bioengineering IRG, came to the microphone to clarify that regional locations have been used when reviewers come to a central secure site. Reverse site visits have also been used successfully for the National Center for Research Resources and other ICs, in which applicants come to NIH to answer questions posed by an A1 committee, rather than reviewers traveling out to the various sites.

Motions: Dr. Kitt asked for motions on two items. First, a motion to shorten the review cycle for all new investigators applying for R01 grants in 2007 was made and passed. Without the motion, Dr. Kitt explained, the pilot would expire in February before three cycles' worth of data could be collected. Second, a motion to run the open houses described by Dr. Scarpa under the aegis of PRAC was made and passed.

Update on Electronic Submission of Grant Applications

Ms. Megan Columbus, Program Manager for Electronic Receipt of Grant Applications, reported that electronic submissions of R01s are on schedule for February 2007. An expected peak receipt day of 2,500 will represent a 1-day high both for NIH and for grants.gov. She said that systems are in place to handle the R01s. Since December 2005, 18,000 applications have been received electronically, which represents a lot of experience.

The main concerns expressed regarding the February cycle are that system problems will result in late applications, Mac users will have problems with PureEdge forms and other usability issues related to these forms, the help desk will be unable to handle the volume of calls, and principal investigators' (PIs) concern that their sponsored research office, rather than the PIs

themselves, are now responsible for sending in the applications. She summarized how these concerns are being addressed.

System problems: Procedures are in place to ensure that applicants are not penalized for systems-related issues. An applicant will call the eRA help desk to report the problem. In October, four out of 3,000 applications were affected. Additionally, if grants.gov or eRA has problems, the window for submission will be adjusted accordingly.

PureEdge forms: On December 1, the Mac version of PureEdge forms was tested and seems greatly improved over IBM's previous release. A Citrix server solution will also work in case the software is not in place in February. By spring 2007, grants.gov will move to an Adobe forms solution that is platform-independent. As part of this move, she and others at NIH are participating in an application development group to provide user input on form functionality.

Help desk: While the help desk is handling the call volume adequately, additional help will be called in for February to handle any increased demand.

Outreach to institutions: Press releases, listservs, presentations, and other means of communication are being used to ensure that all potential applicants know about electronic submissions. Hands-on training is also conducted.

Discussion
Dr. Brenner said that he receives frequent e-mails about electronic submission. His institution has moved up administrative deadlines in order to ensure a smoother upload to grants.gov and, so far, he has heard no complaints from colleagues. Dr. Torok-Storb agreed NIH has done a great job preparing institutions for electronic submission.

Dr. Collins said some of the electronic submissions she receives to review are formatted strangely and asked whether applications can be submitted through any means to maintain their original formats. Ms. Columbus said using a PDF format results in separate files for every section, so the computer can verify that each required section was submitted. She is looking at the computer validations to see if any improvements can be made. However, a request for a forms change must go through grants.gov, which would take many months and would render previous versions of the application form unusable. Although they are looking at this further, Ms. Columbus did not foresee any short-term changes.

Dr. Pugh asked what happens in the event of a foul-up at a particular institution. Ms. Columbus said the normal late policy has not changed. A request for an exception can sent to the CSR Division of Receipt and Referral—an option that applicants currently have when they have difficulties submitting an application.

Update from the Office of Extramural Research

Dr. Walter T. Schaffer, Acting Director of the Office of Extramural Policy in the Office of Extramural Research, reported on the NIH Multiple PI Initiative. A committee has looked at how

to include multiple PIs on a single application for a few years. He acknowledged the committee's efforts. Applications with more than one PI will be accepted as of February 2007.

Allowing more than one PI recognizes that PIs sometimes work in teams and that many projects depend on collaboration. However, a consensus has grown that team science is discouraged by recognition of only one PI. Other federal agencies, such as the National Science Foundation (NSF) and Department of Defense, have recognized co-PIs for many years. A number of entities, including the NIH Bioengineering Consortium and NIH Roadmap, expressed the need for change at NIH, as did a directive from the Office of Science and Technology Policy. There is widespread community support for the change, as expressed in Requests for Information (RFIs) issued by NIH and the U.S. Office of Science and Technology Policy in July 2005.

During a pilot, nine Requests for Applications (RFAs) and program announcements resulted in 210 applications; more than 60 named multiple PIs. Ultimately, a lower usage rate is expected. About 10–15 percent of applications to the NSF identify multiple PIs.

Four core principles emerged from the pilot and discussions. First, "PIs are PIs": in other words, the same level of authority and accountability applies whether there are one or many PIs on an application. To avoid confusion, NIH will not use the term "co-PI." Second, multiple PIs are an option, not a requirement, as decided by the investigators and institutions. The decision must be consistent with the goals of the project. Third, the use of multiple PIs does not imply more than one project. If PIs are at multiple institutions, then the project is managed using a consortium agreement or a subaward. In the future, Dr. Schaffer said, linked awards may be possible. Fourth, all PIs will be recognized in NIH reports including the NIH Commons, summary statements, notice of grant awards, and other documents and databases.

A leadership plan is required. The plan covers the rationale for using more than one PI, how decisions will be made and funds will allocated if desired, the process for conflict resolution, and each PI's roles and responsibilities in the project. Minimal modification of review criteria is required. A contact PI will be the first person listed on the application, similar to the corresponding author on a multi-author publication.

Dr. Schaffer noted the impacts on new investigators. The most significant is that if a new investigator and an established one are multiple PIs, the "new investigator" box is not checked on the application.

Some awards, such as fellowships and career awards, will not accommodate multiple PIs, but most will have the option available. A Guide Notice was published on November 20, 2006.

In the future, the plan is to develop the ability to manage research projects spanning more than one institution using linked awards, in addition to subcontracts, and to develop the capacity to recognize non-PI key contributors. They will be assessing the desirability of permitting formal apportionment of funds to identified components of a project or the associated PIs.

Discussion

Institutional issues: Dr. Ramm called the initiative a great step forward. She expressed hope that institutions will seriously consider including women as PIs who are in leadership roles on a project. She also asked about the effect of multiple PIs in the tenure process at institutions. Dr. Schaffer noted some institutions are better than others in recognizing the many collaborators on a project. The effect will be variable, but this was one reason behind using the multiple PI, rather than co-PI, designation. Dr. Scarpa said institutions should develop a mechanism to recognize multiple PIs. Dr. Schaffer agreed but did not know of any way that tenure and promotion could be controlled outside an institution.

Dr. Torok-Storb said designating multiple PIs would help investigators' biosketches, but the issue of subcontracts and who actually submits the grant could be a problem. That is why, said Dr. Schaffer, apportionment and linked awards are still being explored. Dr. Torok-Storb asked how budget pages could be presented so that individuals can show the dollar amount they bring in on their biosketches. She said she advocated separate budgets for PIs within the same institution. Dr. Schaffer said the allocation could be indicated in the leadership plan. The disadvantage to separate budgets is if they limit communication among the PIs. Feedback from the community will help determine how this plays out.

Dr. Brenner agreed with Dr. Torok-Storb that departments within institutions fight over allocation of credit because of the implication for overhead dollars. On the other hand, separate budgets cast doubt on whether two PIs are truly on the same project. The management plan becomes important, which might mean an administrative review beyond peer review.

In response to a question from Dr. McClain, Dr. Schaffer said a great grant idea but a terrible leadership plan could result in an application not being funded. Dr. McClain said the process for resolving conflict should be highlighted in the leadership plan. He also asked whether departmental rankings are being eliminated. Dr. Schaffer said NIH has wanted to eliminate them for some time, as they are often used for unintended purposes. Instead, NIH will provide spreadsheets or some other listing of all the awards made and the institutions can rank themselves. A new Web site at http://grants1.nih.gov/grants/award/trends/FindOrg.cfm is under development that will allow users to examine all awards made to a particular organization.

New investigators: Dr. Martinez asked about treatment of new investigators in a multiple PI application if the goal is to get new PIs into the system. Dr. Schaffer stressed they are being treated as new investigators, but the application cannot be assigned new PI status. He agreed the solution was not perfect.

Evaluation: Dr. Collins asked about data collection. Dr. Schaffer said an evaluation will look at some of the cultural changes that might ensue. In addition to surveying investigators, Dr. Collins suggested finding out how institutions are handling allocation of responsibility and other issues. That feedback could help NIH make better decisions about how to proceed.

Alternative Peer Review Meeting Formats

Presentations on asynchronous electronic discussion (AED) and video enhanced discussion (VED) were followed by PRAC questions and comments.

Asynchronous Electronic Discussion
Dr. George Chacko, Chief of the Bioengineering Sciences and Technology IRG, described AED as a way to bring in the very best reviewers to CSR. He explained AED works like a threaded message board in which reviewers log in at different times during a defined discussion period. Since the method is asynchronous, reviewers are separated by both time and space. This separation fosters more thoughtful discussions because reviewers have more time to formulate responses to comments. The separation also serves to reduce confrontation between participants and enables soft spoken or diffident reviewers. A meeting transcript is available for review, and conflicts of interest are easily managed. The asynchronous method also overcomes the problem of participating remotely from different time zones and eliminates travel costs.

He briefly summarized how the process works. After applications are assigned to study sections and then to specific reviewers, the reviewers write up preliminary critiques and post them on Internet Assisted Review. Before the meeting, the site becomes read-only, and all members can read the critiques and scores. The meeting occurs online and comprises three phases: streamline, discussion, and final scoring. Unlike a face-to-face meeting, however, these phases may overlap in AED. He showed two screen shots of a typical AED meeting.

Eleven reviews, looking at R21, R03, R15, and R01 applications, took place in an AED pilot between July 2005 and January 2006. This part of the pilot showed improvements were needed in the software; new software is now in prototype form and was used in 38 Special Emphasis Panels (SEPs) between July 2006 and December 2006. The Office of Planning and Evaluation is conducting an evaluation of AED, but the results are not yet available. A technical evaluation is also taking place.

Challenges include developing a culture of usage, since the review is less immersive than a face-to-face meeting. Reviewers have to log in three or four times a day to keep current with the meeting and must look at multiple parallel threads of discussion. The layout of the Web site needs to be easier to understand. An expert in Web design is being brought in to improve the layout based on reviewer feedback.

Future directions include scaling up to use AED with more panels, more applications, and different mechanism, but without losing quality. It is possible that some communities may be more attuned to this method than others. He concluded by acknowledging the efforts of the many people involved in the project.

Video Enhanced Discussion
Dr. Eileen Bradley, Chief of the Surgical Sciences, Biomedical Imaging and Bioengineering (SBIB) IRG, reported on the use of video enhanced discussion (VED) in her IRG. Dr. Xiang-Ning Li, an SRA in SBIB, has taken the lead in a project that began about 2 years ago.

VED is an important tool to involve reviewers who cannot travel for review meetings, especially clinicians, and a potential tool in the event of a catastrophe that precludes travel. Simultaneously with their efforts, the NIH Center for Information Technology (CIT) convened a committee, chaired by Rich McKay, to investigate video conferencing software that CIT could support, and Breeze was chosen. Dr. Scarpa provided resources to help with the pilot that began with six SBIB reviews. She acknowledged the work of SRAs, Technical Services Branch, and CIT.

Dr. Bradley explained how the technology works and noted Dr. Li trained reviewers in how to use it, a time-intensive task that would not be possible if the system were deployed throughout CSR. CIT has taken over the training and technical support. In addition, when a VED takes place, a technical person is now on hand to provide assistance.

The pilot showed recruitment can work for this type of meeting. A load of 20 to 40 proposals can be handled, and as many as 30 participants included successfully. The maximum time limit for a VED seems to be four hours. SBIB is now ramping up the number of VED reviews, adding about six in each cycle. Overall, CSR's goal is to provide additional ways to conduct reviews and to include VED as part of the SRA toolbox. Dr. Bradley concluded it is reasonable to consider using VED or other alternate formats for reviews of 20 applications.

Discussion
Dr. Ramm asked if clinical participation in reviews rose as a result of these alternate formats. Dr. Bradley said, although they have no data, they were able to recruit the people they wanted.

Electronic versus face to face: Dr. Matt Winkler expressed reservations about losing the cultural component of physical reviews, even while it is inevitable and advantageous to move in this direction. He suggested a hybrid in which reviewers rotate between on-site and remote participation. Dr. Scarpa said the first meeting of a study section would be face to face. The goal is not to replace face-to-face meetings but to provide a different platform and get the best reviewers.

Dr. Kenneth Dill, Associate Dean for Research at the University of California, San Francisco, said he participated in an AED. Two downsides were the inability to read body language during a meeting and the opportunity to meet with colleagues, but the great advantage was saving 10 to 15 hours of travel time and the associated cost savings.

Scoring: Dr. Torok-Storb questioned the assumption that a lack of confrontation is a good thing as it resolves disparate scores. Dr. Chacko said AED reviewers have been diligent in resolving scoring differences, and their text tends to be highly focused. Additional data will help in addressing any problems. Dr. Scarpa said an AED is an equalizer among reviewers, especially if a study section has someone who tends to dominate a discussion. He said there might be a gentle migration to AEDs, as people feel more comfortable with the technology.

Dr. Schneider called himself a skeptic but said he was impressed when he used AED in a recent round. The discussions were more lengthy, in-depth, and thoughtful than in traditional settings. He plans to use it again, but noted his small study section is well suited to this type of review.

Maintaining records: In response to a question from Dr. Ramm, Dr. Scarpa said CSR had legal advice to destroy records of the meeting to avoid confidentiality problems. Records are erased immediately after the meeting.

Dr. Martinez suggested attaching voices to the AED when people are simultaneously online, or using software to translate voices into text. Dr. Torok-Storb asked if reviewers can call each other offline to reconcile a difference or clarify a point. Dr. Chacko said, if needed, an SRA can schedule a teleconference and assemble everyone, which has happened in a few cases.

Deep Innovation at NIH

Dr. Kitt introduced Dr. Kenneth Dill, Associate Dean for Research at the University of California, San Francisco. Dr. Dill thanked Dr. Scarpa for the invitation to speak and praised NIH for the science it has funded and its capacity to be user-friendly and flexible.

Deep Innovation and Current Constraints
When he speaks of deep innovation, Dr. Dill explained he refers to developing a program to go after the next "big thing" as opposed to the next "little thing." He shared several anecdotes about ideas that have, in retrospect, advanced science but that did not fare well in study sections. R01s are an outstanding mechanism for funding small innovation, but he urged NIH to consider a smaller program for deep innovation. Supporting deep innovation would require moving from the faultfinding that tends to take place in study sections to a system based on advocacy, reducing the tendency to over-review, and welcoming the allied disciplines.

Communal reviews tend to focus on faults, problematic when the goal is deep innovation. Reviewers who are also competitors can be particularly critical. The reasons that an idea won't work often dominate the discussion.

Dr. Dill praised the Pioneer Award Program for handling the issue of proposal length. Lengthy proposals discourage deep innovation. In the first five to eight pages, a good PI should be able to communicate his or her intent. People make many important decisions, such as on faculty hiring, with information not nearly as long and detailed as a 25-page proposal.

He said the current system is too reliant on preliminary data, and the three-digit scoring system is not meaningful in discriminating one good proposal from another or predicting which will result in better science. Even looking at his own best two or three papers, already published, he could not distinguish within three decimal places which one is better than the others. In some cases, the quality of a piece of work does not emerge until many years later, such as the mathematics behind CAT scans.

The allied disciplines, including physics, math, and chemistry, should be harnessed to encourage deep innovation. NIH funds these disciplines only when applied to biological systems. If proposals in these disciplines have any biomedical relevance, they are considered outside the mission of NSF, Department of Energy, and the Defense Advanced Research Agency (DARPA). But, as with x-ray crystallography, ideas can take many years for their biological significance to

emerge. In the study section of which he is a member, technology development grant proposals are ranked on whether or not there is an immediate biological payoff.

We are not good at predicting big advances, noted Dr. Dill. Deep innovation looks for what he termed disruptive advances. With deep innovation, evaluations should focus less on guessing payoffs and more on the PI, the science, and whether the person has the drive to explore.

Proposal for Deep Innovation Grants
Dr. Dill proposed a program that he said the Pioneer Award Program partly embodies. Proposals would be short (5 to 8 pages) and reviewed at arm's length so there is not as much attention to the details in the approach. The focus should be on people and not payoff. Relevance should be separated from the review, and proposals should be ranked, rather than scored. Many of the best investigators have deeply innovative ideas they would like to explore. Opportunities for deep innovation would allow them to develop new ideas, as well as attract new scientists to biomedical research. A deep innovation program would take some investigators out of the 25-page R01 proposal machine.

Ranking versus scoring proposals would eliminate the need for reviewers to meet, since the main function is to score proposals. Ranking, rather than scoring, proposals would better leverage reviewers' insights. It eliminates the clout of one loud reviewer who is strongly opposed to a particular application, which works against deep innovation. Reviewers would be able to advocate for proposals they find particularly strong. He showed an example of how ranking would work with four hypothetical applications.

A mechanism like the R21 is not addressing deep innovation, because the same reviewers are looking at the same long proposals using the same review mechanism. Reviewers asked to look for "high risk, high reward," often don't realize that you often cannot see "high reward" at all; the biggest payoffs are often the most unexpected. In addition, deep innovation is not necessarily high risk. Deep innovation should be about charging people to be explorers rather than developers.

The Pioneer Awards come close to a deep innovation program. The program would need to be scaled up from 10 or so to hundreds of investigators. It could be a flexible program based on needs. It should be untargeted and disconnected from direct mission payoffs, as the great industrial research labs were able to do in the 1950s and 1960s. The private companies were unable to harness the advances for their needs, which led to the end of such research support, but this should not be the concern of a public institution.

In summary, NIH is the only place to deepen biomedical research. Google's founders claim that 70 percent of time should be spent on core business, 20 percent on peripheral business closely related to the core, and 10 percent on deep innovation. At least a small part of the NIH budget should go for deep innovation, a very different beast than more incremental innovation.

Discussion
Dr. Berg said Dr. Zerhouni has used a calculation to determine how much support should be given to innovation, depending on how quickly one's field changes. For example, if it changes every 10 years, 10 percent of support should go for the truly innovative.

Dr. Collins said NIH should consider Dr. Dill's ideas about ranking versus scoring more broadly. She said finding good people and letting them explore must go beyond the old boys' network. Dr. Dill said some support should go to those who are easy to identify, but a goal of deep innovation should be to find those who are not as easy to identify. A key factor is what he termed an investigator's fire in the belly, which could come out in a five to eight page proposal.

Dr. Pugh said he found the presentation provocative and interesting, and perhaps a pilot could be designed to look at ranking versus scoring. He asked Dr. Dill for ideas to bring in people from other disciplines, noting the creation of the National Institute of Biomedical Imaging and Bioengineering (NIBIB). He said he is not opposed to the idea of deep innovation, but his role as PRAC member is to be skeptical. For example, we know in hindsight what has received a Nobel Prize, but not how many other ideas failed miserably. He pointed to some potential conflicts that would have to be worked out. Fairness is a core value of peer review, but this program might result in the perception that fairness is breaking down. It also conflicts a bit with the idea of impact. NIH is pushing for more translational research, yet deep innovation looks at a long horizon. He reiterated his support for Dr. Dill's ideas, but stressed not racing ahead.

In response, Dr. Dill noted NIBIB does fund research in the physical sciences, but in a very focused way that misses huge parts of the physics, math, and computer science communities. He said a deep innovation program would bet on people, and some gambles must be taken. It is not possible to predict what will be transformative. Dr. Winkler said no one can identify innovative science, but most people know who within their institutions are exceptionally creative. It might be interesting to experiment with the concept of a nominative process in which institutions identify their most innovative people. Dr. Dill said a colleague involved with the Sandler Family Foundation told him that the foundation at first received a huge number of applications, but the numbers were manageable.

Dr. Leinwand said some study sections rank as a matter of course. She agreed a pilot would be very informative. She also expressed strong support for a deep innovation program that could be worked out to incorporate fairness and other factors. Dr. Martinez suggested an idea to place grants in quality groups so that a tier one group would be those that should be funded, a tier two would perhaps merit funding, and a tier three would be triaged. Program staff could make decisions from within the groups.

Motion: Dr. Kitt suggested a motion to develop a pilot on ranking versus scoring. The motion was made and passed.

Automated Referral Workflow System

Dr. Thomas Tatham, CSR Associate Director for Knowledge Management, reported on a pilot to facilitate referral of applications through text mining. CSR is the NIH portal for all incoming

competing applications, with more than 70,000 received by the Division of Receipt and Referral in fiscal year 2006. About 20,000 were referred to ICs for review, and the rest went to IRGs in CSR. Written guidelines and PI requests help determine assignments. It takes more than 2 months in any given review cycle to complete the referrals. Speeding up the process through an Automated Referral Workflow System (ARWS) could help shorten the review cycle. Other advantages to an ARWS include greater efficiency, more transparency, and more consistency.

Electronic submission makes automation more feasible because of the ability to pull out specific sections of an application and make assignment predictions based on those sections. One of the primary ways is through automated mining of PIs' cover letters. About 47 percent of PIs request a specific Scientific Review Group. CSR usually initially refers the application to the requested group; most of the time (89 percent), the application remains there. Algorithms developed for automated mining of cover letters have proven to have meaningful levels of accuracy. Machine learning algorithms are being developed to assign applications that come in without a cover letter request, and referral by experts would continue for difficult cases.

Machine Learning Experiment

Dr. Tatham explained how the machine learning algorithm works. The software looks at a target application and then at the assignments of 15 other applications that are most similar to predict the assignment of the target application. A series of experiments was conducted on 4,500 R01s submitted without a specific PI request that were referred to standing CSR review groups. Predictions were made on study section assignment by extracting the abstract and specific aims of the applications and processing the text with software that can compare pairs of documents based on their shared usage of words that are relatively rare in the overall set of applications. Two dependent variables were (1) the agreement, or concurrence between the automated referral prediction and the historical assignment made by human experts and (2) the yield, or the percentage of applications for which prediction may be made at a given accuracy level. The automated system concurred with the human-made decision about 75 percent of the time (agreement) in 100 percent of the cases (yield). If a frequency of seven matches or better is required of the automated system, then the system matched human assignment about 82 percent of the time, with a yield of about 78 percent.

He differentiated between concurrence with historical human judgments and ultimate IRG acceptance of the assignment. For example, a concurrence of 82 percent would not mean that the other 18 percent of the time an IRG would reject the application. An appreciable percentage would still be accepted by the receiving IRG, since the design of the IRGs is such that any given application may be appropriate for one or more IRG venue.

The model developed estimates that 47 percent of applications could be referred based on cover letter requests, 39 percent with machine-learning algorithms, and 14 percent by human referral experts. Eventually, the model could be applied to IC referrals as well, although more coordination with ICs would be required. Considerations include how to resolve programmatic areas of overlap or disagreement among them, how to consider prior grants or support in making an assignment, and how to protect the interests of smaller ICs. He said stakeholders will be consulted, and a steering committee would include IC representatives.

Next steps
CSR has taken the lead in developing and maintaining ARWS and will use it for some referrals in February. Over the longer term, it is critical to have a data path between the system and the IMPAC II database, as assignment predictions cannot now be passed into IMPAC II. This interface will hopefully be in place by June. The savings gained from implementing the interface include fewer referral officers and review meetings held two to three weeks earlier.

Dr. Tatham closed by acknowledging the support of the Office of the Director and the Extramural Affairs Working Group, as well as the work of the ARWS Project Team, which includes staff from CSR and other ICs, as well as contractors from DiscoveryLogic.

Discussion
Assignment requests: Dr. Martinez praised the system but asked why PIs are not routinely asked to request assignments. He also asked about the use of key words in making assignments and about any other ways to reduce the number of applications handled by humans. Dr. Tatham said requiring all PIs to request an assignment would burden those less familiar with the system. In addition, NIH rules require going back to PIs if their requests cannot be honored, which could become time-consuming if those less familiar with the system made inappropriate requests. Dr. Martinez suggested a pilot on the issue.

Dr. McClain asked what would occur if a PI were dissatisfied with an ARWS assignment. Dr. Tatham said PIs can monitor the routing of their grants through the NIH Commons and contact NIH to request an alternate assignment. This occurs even under the current system and no new procedures would be needed to accommodate PI requests for an alternate assignment. He said a statistician will be brought onto the project team to create a multivariate model for predicting assignments to enhance the acceptance level. In response to a question from Dr. Torok-Storb, Dr. Tatham said incidences of applicants requesting an application not go to a specific section are very infrequent, but ARWS will have a way to handle this situation. In any event, IRG chiefs would still look at cover letters and provide a level of quality control.

Alternatives: Dr. Tatham said assignments based on key words would be difficult to implement because applicants are not consistent in how they use them. Dr. Leinwand asked whether a dropdown list could help PIs request a study section. Dr. Tatham called that "Plan B," and it might be developed if mining cover letters is not as accurate as anticipated. It would place an additional burden on reviewers: they would have to make their request through the NIH Commons, as it would be difficult to the NIH-specific dropdown list within grants.gov. Many people objected to the requirement with electronic submission that PIs and signing officials go into the Commons within a certain time limit, and they might also find the dropdown list on the Commons a burden.

Dr. Torok-Storb said many applicants might peruse the membership roster of study sections to make their requests. Dr. Tatham said quality control helps ensure a request is appropriate for the project. If the request made by a PI is not within the top two or three ARWS choices, the application could be routed to a referral expert for further analysis.

Restructuring of Study Sections in Molecular, Cellular, and Developmental Neuroscience

Dr. Don Schneider introduced the potential restructuring of several study sections by highlighting the core values of peer review and summarizing the process to review IRGs. In late 2005, a review spotlighted that the Molecular, Cellular, and Developmental Neuroscience (MDCN) IRG had one oversubscribed study section and one spiraling down. A pre-working group was convened to recommend some changes, which were presented to PRAC in January 2006. However, one of the newly structured study sections—Neurodegeneration and Glial Biology (NDGB)—is already too large. A CSR internal review group and a working group were also involved in assessing the IRG and what to do to improve it. The working group involved at least one representative of each regular study section in the IRG. The members were surveyed about issues that needed attention, which they then discussed in a telephone conference in November. In addition to some generic IRG-wide issues pertaining to reviewers and workloads, the working group identified the large size of NDGB and the small sizes of a number of Special Emphasis Panels (SEPs) as concerns.

The consensus was to split NDGB along degenerative and glial biology lines; develop guidelines for two standing SEPs in neuroengineering and in neuroinformatics/imaging; and continue the neurogenetics SEP as needed but without developing guidelines as yet. The group felt this clustering would result in reviews that would better serve the neuroscience community. Dr. Schneider reviewed the possible guidelines for what would be the Cellular and Molecular Biology of Neurodegeneration (CMND) and Cellular and Molecular Biology of Glia (CMBG) study sections, as well as for the two SEPs. He noted many study sections also deal with bioengineering and that applications would go to MDCN when they have a neuroscience context.

In summary, he asked for PRAC approval to (1) divide NDBG into CMND (neurodegeneration) and CMBG (glial) study sections, (2) form the Neuroengineering and the Neuroinformatics SEPs on a recurring basis, and (3) continue the Neurogenetics SEP as needed.

Discussion

SEPs: Dr. Leinwand said Dr. Schneider had satisfied a concern she had about a SEP specific to neuroengineering when other bioengineering groups are already operating, but warned against too much subspecialization.

Dr. Pugh praised the process for restructuring of study sections and the broad stakeholder representation. He said the stakeholder comments on various aspects of how the study sections function are also valuable. He asked how PIs know about the SEPs. Dr. Schneider said some, but not all, are posted on the Web. Dr. Pugh urged they all be identified in the public domain, rather than only a small group of insiders know about them. Dr. Scarpa said the open houses will clarify issues like this. Dr. McClain warned against SEPs becoming cliques.

Motions: Dr. Torok-Storb suggested separating the issues requested for approval. A motion to approve the two new study sections passed unanimously. In terms of the two recurring SEPs, Dr. Torok-Storb questioned whether they were merited or should be merged with informatics and bioengineering SEPs. Dr. Schneider said the working group felt strongly that the neuroscience context was important. Dr. Leinwand said that to vote she did not have enough information about

the scope of the SEPs and the predicted numbers of applications. Dr. Scarpa proposed going back for some additional thinking, enlarging the working group, and returning to PRAC in 3 months. Dr. Torok-Storb suggested getting feedback from other affected groups. A motion was made and passed for a working group to look at the three SEPs.

Scientific Review and Evaluation (SREA) Operations

Ms. Diane Bernal, Director of the SREA Management and Service Center, reported on two efforts to save money: a pilot on purchasing nonrefundable airline tickets for reviewers and broad purchase agreements (BPAs) for hotels and meeting rooms for committees.

She said the Office of the NIH Director asked CSR to establish SREA for the management and processing of all NIH peer review activities. SREA processes 2,500 hotel contracts and provides travel arrangements and processes reimbursements to more than 65,000 consultants per year.

Air Travel Pilot
CSR conducted a pilot to issue nonrefundable air tickets. Data from June through November 2006, conducted with 11 ICs, showed a total fare savings of more than $4.9 million. Reviewers are allowed to change their tickets one time with no questions asked, and 785 such changes were made. The pilot will end in December 2006, after which an evaluation of cost savings and impact on reviewers will be carried out. The analysis will be presented to the Extramural Activities Working Group for consideration as NIH-wide policy.

Hotel Blank Purchase Agreements
The second pilot involves establishing BPAs with hotels. Terms and conditions will be negotiated, and, as a government mechanism, BPAs are exempt NIH from state and local taxes. SRAs could select hotels without competitive sourcing. Requests for proposals were sent to 140 area hotels. The responses will be reviewed by a working group of SRAs, acquisitions staff, and lawyers. A preferred provider list will be the result.

Discussion
Dr. Kitt noted that BPAs will free up SRAs from spending time calling and negotiating with hotels. Dr. Brenner said the quality of the hotels must be considered. Amenities matter to reviewers who have spent 40 to 60 hours prepping for review meetings. It is easier to recruit reviewers when they know they will go to places they will enjoy. Dr, Kitt said the hotels are those that NIH has used and that reviewers say they prefer.

Conclusion

Dr. Kitt asked for other questions or comments. None were received, so she moved to adjourn the meeting. PRAC adjourned the meeting at 2:30 p.m.

We do hereby certify that, to the best of our knowledge, the foregoing minutes of the December 2006 meeting of PRAC are accurate and complete. The minutes will be considered at the April 2007 meeting of the Advisory Committee, and any corrections or comments will be made at that meeting.

Cheryl Kitt, PhD
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