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

December 3, 2007

The third 2007 meeting of the Peer Review Advisory Committee (PRAC) convened at 8:30 a.m. on Monday, December 3, 2007, in the Natcher Conference Center, National Institutes of Health, Bethesda, Maryland. The entire meeting was held in open session. Dr. Antonio Scarpa and Dr. 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.
R. Lorraine Collins, Ph.D.
Heidi E. Hamm, Ph.D.
Story C. Landis, Ph.D.
Leslie A. Leinwand, Ph.D.
Joseph L. Martinez, Jr., Ph.D.
Craig J. McClain, M.D.
Daria Mochly-Rosen, Ph.D.

Ad Hoc Members

Jill P. Buyon, M.D.
Paulette S. Gray, Ph.D.
Andrew W. Murray, Ph.D.

Dr. Norka Ruiz Bravo, Ph.D., attended ex officio. Dr. Cheryl Kitt, Ph.D., was the Executive Secretary for the meeting.

Welcome, Upcoming Meetings, and Appreciation to Departing PRAC Members

Dr. Antonio Scarpa called the meeting to order and asked PRAC members to introduce themselves. He confirmed dates of the 2008 meetings: April 7 (later rescheduled to April 30), August 4, and December 8. He expressed appreciation and presented plaques to Dr. Jeremy Berg and Dr. Edward Pugh, who are rotating off of PRAC at the end of 2007.

Enhancing Peer Review at NIH—An Update

Dr. Lawrence A. Tabak, Director of the National Institute on Dental and Craniofacial Research and co-chair of the Advisory Committee to the Director (ACD) and the Steering Committee (SC) Working Groups on Peer Review, updated PRAC on the working groups' activities. NIH is conducting this self-study in partnership with the scientific community to strengthen peer review in changing times.

He reviewed the principles behind the study and Dr. Zerhouni's charge to figure out how to "fund the best science, by the best scientists, with the least administrative burden." The groups are coordinating their efforts with Center for Scientific Review (CSR) initiatives such as shortening the review cycle and realigning study sections.

Project Phases

Diagnostic: In the diagnostic phase, which began in July 2007 and will end in January 2008, NIH received more than 2,600 responses to a Request for Information to identify challenges and potential solutions related to the peer review system. Two teleconferences involved about 100 deans, and five regional meetings engaged professional organizations, patient advocacy groups, researchers, and administrators. Scientific liaisons increased outreach with the scientific community. In a parallel process within NIH, more than 200 NIH staff provided input.

A contractor is now developing a final coding schema to organize and analyze the responses. The working groups are also analyzing how other domestic and international agencies conduct peer review. On December 7, the groups were scheduled to present interim findings to the Steering Committee and Advisory Committee. He stressed these findings are interim. Recommendations will be crystallized in January.

Implementation: In February, the NIH leadership will determine the next steps, which will likely include piloting of potential interventions to begin in March. There may also be some "low-hanging fruit" changes that can happen immediately, without the need for pilots and evaluations. Based on the results of the pilot evaluations, new policies may be established.

Emerging Ideas

Dr. Tabak presented ideas emerging from the diagnostic phase. The ideas relate to review criteria, new models of review, maximization of review(er) quality, reviewer mechanisms/mechanics, peer review culture, scoring, mechanisms, and other issues related to the system that supports research. He stressed he presented the ideas in no particular order, and they have been circulated to a number of groups to prompt discussion. He also stressed the ideas are not predicated on a "village vote"; sometimes, the most interesting ideas have come from just one or two people.

Discussion

Dr. Lorraine Collins asked how PRAC could get involved in the process. Dr. Tabak said he made his presentation in order to involve PRAC more fully. In addition, Dr. Scarpa and Dr. Berg are on both working groups, and he expected they would return to PRAC with ideas and suggestions for discussion. Dr. Norka Ruiz Bravo said PRAC recommendations are discussed in various NIH governance committees and go to the Extramural Activities Working Group (EAWG) and the Steering Committee. Dr. Berg urged PRAC members to keep abreast of the process because of its rapid timeframe. A lot will happen before the next PRAC meeting in April, much of it virtually, so he said he will ensure PRAC members are on the appropriate mailing lists.

Dr. Dean Brenner said the scientific community will be interested in the conclusions and analyses and hoped the findings will be widely circulated. Dr. Berg said information has been posted on the Web site (http://enhancing-peer-review.nih.gov) and will be updated after the ACD meeting.

Dr. Andrew Murray asked whether ideas relating to an obligation by grant recipients to serve as reviewers and to terms of service had come up. Dr. Tabak said phraseology has been suggested

that recipients "will serve [as reviewers] if asked." In addition, reviewing twice a year resonates better than three times with many people, even if length of service were extended.

Dr. Daria Mochly-Rosen asked for examples of the "low-hanging fruit" that might be quickly implemented. Dr. Tabak said nothing had been decided, but one possibility is providing scores of unscored applications. More generally, these changes would be benign in that they do not cost the system and there are no untoward consequences. More dramatic ideas would need to be piloted on a subset of applications to ensure no untoward effects. A PRAC discussion will need to take place on the working groups' suggestions, via a videoconference or other means, since waiting until April will be too late to have meaningful input.

Improving Peer Review: CSR Initiatives

Dr. Scarpa organized his presentation around four areas: new data, CSR organizational initiatives, recent activities, and realignment of CSR peer review.

New Data
As shown in a series of graphs, the number of applications declined by about 3,500 in 2007, with much of the drop attributed to a decline in R01s. In contrast, R21s have increased from about 4,000 in 2001 to almost 30,000 in 2007, which Dr. Scarpa termed a major concern. CSR reviewed almost 50,000 of the 75,000 applications received in 2007, with the rest reviewed in Institutes and Centers (ICs). He showed data on where different mechanisms are reviewed.

Looking at data from 2000 to 2007, female applicants, on average, scored better than males. In 2006, 11 percent of female applicants scored in the 0 to 10 percentile, and 22 percent scored in the 0 and 20 percentile. These numbers are noteworthy, especially considering researchers in some countries feel the need to remove female names so as not to jeopardize applications.

He next showed data from May 2001 through May 2007 Councils to look at new and established investigators' scores. New applications (Type 1) submitted by new applicants and new applications submitted by established investigators fare about as well as each other. Competing renewal (Type 2) applications receive better scores, but this result appears to have more to do with the difficulty of proposing new research than a review bias against new investigators.

CSR Organization Initiatives
Realignment: CSR divisions and Integrated Review Groups (IRGs) are being realigned to reflect changes in science. PRAC approved the new Emerging Technologies and Training in Neurosciences (ETTN) IRG in August 2007, and it is being organized within the new Division of Neuroscience, Aging, and Development. There are now five CSR divisions, up from four, which will be discussed later in this meeting.

Review-enabling committees: Nine CSR committees have formalized efforts that were taking place on a more ad hoc basis. He reviewed the committees, their chief functions, and their chairs.

Increasing efficiency: Most applications are now submitted electronically, which allows for electronic assignment to IRGs with text fingerprinting. Implementation is scheduled for June 2008.

Recruitment of Scientific Review Officers (SROs): Between 35 and 45 of CSR's 250 Scientific Review Officers (SROs) leave each year. More proactive recruitment has resulted in very senior people with established careers applying for SRO positions.

CSR Recent Activities

Improvement of study section alignment and performance: Since the last complete alignment of IRGs, science has changed. Fourteen IRGs were reviewed in 2006 and nine in 2007, and the cycle will begin again. The six Open Houses held in 2007 provided important input.

Shortened review cycle: The goal is to provide a review and score within 3 months of application submission to permit three reviews within 1 year. He said limited progress has been made through three steps: posting summary statements sooner, a pilot to shorten the review cycle for 2,000 new R01 investigators, and expansion of the pilot to cover all new R01 investigators. Eighty-one new investigators (up from 30) and 124 established investigators (up from 72) resubmitted in the very next Council cycle in October 2007 compared with May 2007.

Reviewer recruitment and retention: Recruiting and retaining high-quality reviewers and decreasing the burden on applicants and reviewers are important challenges. He said there are far too many reviewers (18,000 in 2007), and too many of them are ad hoc. Near-term solutions include remaking peer review a learning experience, maintaining a national registry of volunteer reviewers, moving one-third of the meetings to the West Coast or Chicago, rewarding reviewers such as by abolishing application deadlines for permanent study section members, expanding peer review platforms to require less travel, and shortening applications.

While still high, the number of reviewers has decreased in the past 3 years, and the staff has tried to get reviewers to increase their loads, even slightly. He said a good balance exists among full, associate, and assistant professors. Involvement by scientific societies in suggesting reviewers resulted in slight increases in more senior reviewers.

Expanded reviewer platforms: The goal was to have 10 percent of all reviews take place electronically in 2007; actual numbers are closer to 13 percent. The bottom line in selecting which platform to use is to determine which one will result in the best review possible.

Shorter applications: A limited pilot of shorter applications has begun, as discussed later in the agenda.

He closed by thanking CSR staff for their work in making these changes to improve peer review.

Discussion

Comparisons: Dr. Leslie Leinwand asked about the data on scoring of applications submitted by women. She asked for comparison data with male applicants. The total number of applications submitted by women is about 27 percent, and study sections are about 33 per cent female.

Dr. Heidi Hamm asked whether the scoring percentages for new investigators will increase given the infusion of new monies to support them. Dr. Scarpa clarified his data referred to peer review percentages, rather than funding outcomes.

Application numbers: Dr. Craig McClain asked about any NIH plans to address the growth in R21s. Dr. Scarpa said beyond the numbers, more than 200 program announcements call for R21s, which is difficult for reviewers. Some ICs are looking at how to reduce the proliferation.

Dr. Mochly-Rosen asked about the increase in the number of grants submitted per person. Dr. Scarpa and Dr. Berg said the issue is being discussed in different NIH committees. She also asked about the outcomes when applicants resubmit more rapidly. Dr. Scarpa said throughout the pilot, about 13 percent of resubmitted in the very next cycle.

New reviewers and investigators: Dr. Jill Bunyon said a perception in the community is that new reviewers are harsher than more senior members. She asked if any way existed to test the validity of this perception. Dr. Scarpa said data would not be available, but he and Dr. Kitt do not observe this tendency at section meetings. Newer members often focus more on detail. Dr. Bunyon said first-time reviewers on the journal she edits often focus on specifics as well.

Dr. Brenner said people often talk about the fear new investigators have of the review process. One benefit of involving assistant professors in study sections is they see the process is fair and what occurs at a meeting. Institutions could identify committed, younger investigators who could then have limited involvement with a study section, perhaps on a one-time basis or with a reduced load to learn the system. Dr. Scarpa said this idea has been discussed, but could conflict with the goal of having the best reviews. Reviewer training is definitely needed, but the issue of how best to do it remains. Dr. Leinwand said her institution sets up internal study sections, which assistant professors have said are invaluable.

Dr. Joe Martinez suggested a solution may be to assign younger people as tertiary reviewers initially, then move them up as they gain experience. Dr. Scarpa said SROs usually do this.

CSR Best Practices Committee

Dr. Cheryl Kitt, CSR Deputy Director, said the Best Practices Committee, which she chairs, was formed to ensure that the peer review process continues to operate at a high level. Study sections are quite variable in terms of training and how they conduct themselves. As she said, "if you've seen one study section, you've seen one study section."

She reviewed the four goals of the committee: (1) review existing policies and procedures, (2) ensure alignment with the Federal Advisory Committee Act and other regulations, (3) ensure consistency, (4) promote sharing of best practices across CSR.

Best Practices Guidelines
To help achieve these goals, the committee authored peer review best practice guidelines in five areas, which serve as companion pieces to the SRO Handbook.

Criteria for selecting reviewers and assembling rosters: The quality of reviewers affects the quality of the scientific evaluation of a grant application. SROs must ensure that the study section remains responsive to emerging areas of science and is balanced in all aspects. Reviewer selection is based on many considerations—including accomplishments, honors, publications, and respect by peers, with specialized considerations for those serving on fellowship and small business committees—as well as geographic, gender, ethnic, and other representation.

With some exceptions, the goal for reviewers attending a meeting is seven assignments per reviewer, which gives the SRO flexibility to bring in a limited number of reviewers with light loads balanced by the rest of the study section who are assigned a more substantial number of applications to review.

Conduct of study section meetings: An effective peer review process requires an SRO to be active during and in the weeks before a study section meeting. SROs must establish early and frequent contact with the reviewers, and they must help foster a team environment within CSR.

Summary statement production: The most visible output of peer review is the summary statement. How these are produced varies across study sections. High-quality statements depend on the reviewer and SRO. Because SROs assemble what reviewers say, reviewers must communicate their concerns clearly and in a way that is consistent with the score. The SRO should provide orientation and oversight, as well as monitor critiques in Internet-Assisted Review (IAR) to detect and correct problems before a meeting. In some study sections, a reviewer, rather than the SRO, drafts the summary statement. Complaints, while few in number, are partly a result of the pressure to get the summary statements completed quickly.

Telephone reviewers at face-to-face meetings: Until recently, SROs handled the phone calls that came into a regular meeting, which was very disruptive. A new system, called Meeting One, makes this process more seamless and also allows program staff to listen in. Best practice guidelines include limiting telephone reviewers to initially 10 percent and no more than 20 percent of participants. They should have experience as reviewers and not lead off the discussion if they are not regular members. SROs should not have to handle the logistics of the calls, and reviewers should submit scores in a secure manner, preferably in IAR or via a secure Meeting One phone connection.

Mail reviews: The use of mail reviews should be consistent. In balancing expert coverage, an SRO may determine that an application requires special expertise not covered by existing panel members. A mail review is appropriate when the size of the need for this expertise does not warrant adding another person to the meeting.

Parking lot issues: Additional issues raised that the committee will address in the future relate to conflict of interest, streamlining, appeals/rebuttals, deferral for re-review, selection of study section chairs, training of reviewers and chairs, and role of program staff at meeting.

Dr. Kitt listed and thanked the committee members, which included division directors, IRG chiefs, and SROs.

Discussion
Supplemental materials: Dr. Bunyon said applicants and even reviewers are confused about such issues as sending in supplemental material and debriefing. Dr. Kitt clarified the policy about supplemental materials. Dr. Scarpa said inconsistencies exist in practice in terms of how close to a meeting a study section accepts materials. Another NIH committee is looking at this issue and will draw up guidelines.

Study Section membership: Dr. Murray observed study sections vary in how well they can recruit high-quality members. There needs to be a level of oversight to ensure that study sections do not decline in quality. Dr. Scarpa said the IRG chief and division director review what is going on in each study section, especially so that section membership is not perpetuated forever. Particular oversight is needed for close affinity groups.

Ad hocs: Dr. Brenner said the status of ad hocs was an issue when he chaired a study section. People who had served as ad hocs for a long time were frustrated because they only get credit at their institutions if they are regular members of chartered sections.

Communication after the review: Dr. Bunyon and Dr. Brenner pointed out information transfer to applicants after a study section meeting is particularly difficult when the program officer was not at the meeting. The program officer, not the SRO, must communicate with an applicant, as part of the firewall between review and programs. This requirement applies even if the program officer did not attend. Dr. Paulette Gray suggested a program officer contact the SRO, if needed, and then get back to the applicant.

The summary statement may not give the applicant enough information about whether to fix sections. Dr. Collins said there is often a disconnect between the summary statement and the score, which makes it difficult for an applicant to decipher what to do next.

Trans-NIH Short Application Design Team

Dr. Kitt updated PRAC on efforts toward a shorter R01application. Internal and external stakeholders have provided much input on this issue. Program announcements that have requested R01 applications shorter than 25 pages can serve as a type of pilot. Shorter applications have both operational and cultural advantages.

The first trans-NIH application committee, as discussed at the April 2007 PRAC meeting, issued a Request for Information to the scientific community and developed recommendations for shortening applications in an integrated way with other peer review initiatives. Almost three-quarters of the respondents to the RFI (73 percent) favored shorter applications.

For the March review cycle, NIH has about ten program announcements or requests for applications that ask for fewer than 25 pages; they vary in terms of the information requested of applicants. Two surveys were designed as a follow-up with reviewers and NIH staff to ask about their ability to assess the science with these applications and about issues related to the length of the research plan and the format of the application. Questions on the survey for reviewers include the number of applications they were able to read as compared to longer applications, and whether the reviews took more or less effort. In a meeting Drs. Kitt and Scarpa participated by teleconference, Dr. Kitt noted that the reaction to reviewing shorter applications was favorable.

The current team's goal is to design a short R01 application template, review criteria in concert with the template, and a format/template for written critiques. An additional critical piece is the design of an evaluation instrument.

The members of the design team come from many ICs and include people who have worked with shorter applications. The team is gathering examples of shorter applications from within and outside of NIH and welcomes suggestions about the design of the application and review criteria.

Discussion
There was no discussion after this presentation. Dr. Kitt said the team hopes to report back to PRAC with a proposed template at a future PRAC meeting.

PRAC Working Group Report on CSR IRG Re-Alignments

Dr. Anita Miller Sostek, Director of the Division of Clinical and Population-based Studies in the Center for Scientific Review, reported on the second stage of the reorganization of CSR's review divisions to better cluster science and manage workloads. The scientific community supports the effort, as expressed at the Open Houses.

At the August PRAC meeting, the new Neuroscience, Development and Aging Division was discussed and is now CSR's fifth division. Neuroscience-related IRGs were pulled from the four existing divisions. The focus now is on better organizing these other four divisions. At this meeting, she said she would present the proposed Division of Healthcare, Population and Behavioral Sciences (DHPB). Tentatively, the remaining three divisions will center on translational and clinical science, integrated biology and clinical science, and basic biomedical sciences. Each will be studied for possible changes in turn.

Dr. Scarpa and division directors developed a draft DHPB plan, and a PRAC subcommittee participated in a November telephone conference. ICs and the rest of CSR were also invited to comment, and their comments were incorporated into the plan presented at this meeting.

DHPB would encompass five IRGs. Dr. Sostek presented several charts with the proposed changes. One change to note is that the Health and Population IRG would split into two, each with 10 SROs: (1) Epidemiology and Population Sciences, and (2) Healthcare Delivery and Methodologies. Other IRGs proposed to become part of DHPB would include AIDS and Related Research (AARR), Biobehavioral and Behavioral Processes, and Risk Prevention and Health

Behavior. The Surgical Sciences, Biomedical and Bioengineering IRG would temporarily be in this division but move in a future phase of the reorganization.

The PRAC subcommittee, during the teleconference, questioned whether the AARR IRG should be part of DHPB, since it covers basic as well as clinical science. Dr. Sostek said study sections in this IRG have always been together because they interact regularly with a fairly strong overlap from epidemiology to healthcare delivery. The proposed transfer of the Biomedical Computing and Health Informatics (BCHI) study section was raised, and Dr. Sostek said guidelines would be developed to ensure an appropriate transfer. PRAC members also had concerns about the composition of some individual study sections, which Dr. Sostek said would be addressed separately.

IC and CSR staff generally support the proposed division. NCI sought a broader distribution of professional disciplines in the areas of rehabilitation, which Dr. Sostek said would be addressed, as would realignment of the BCHI guidelines.

Discussion
In response to a question from Dr. Martinez about Special Emphasis Panels (SEPs), Dr. Sostek said fellowship, small business, and some non-R01 mechanisms have standing SEPs, rather than chartered study sections. Dr. Norka Ruiz Bravo added the Federal Advisory Committee Act limits the number of FACA committees within an agency. Dr. Kitt said from a practical matter, the IRGs are chartered, with study sections serving as IRG subcommittees.

Dr. Brenner said the reorganization reminds him of the two models of cancer centers: organ-oriented or cross-disciplinary. In this case, the shifting structure includes more discipline-oriented and more population-based approaches. He said he sees a struggle, which might require "casting the die" toward one approach or another, with neither being right or wrong: for example, in which of the two newer divisions would applications from neuroscientists in the population sciences be reviewed? Dr. Sostek said CSR took over a structure that includes both broader and organ-based groups. This challenge will be even greater with the remaining divisions, such as the Division of Biological Basis of Disease and Division of Physiology and Pathology. She said the IRG chiefs have some creative ideas.

General Principles of Application Clustering

Dr. Don Schneider, Director of the Division of Molecular and Cellular Mechanisms in the Center for Scientific Review, discussed how applications are clustered and asked for PRAC feedback about how to improve the system, especially for applications on diverse topics with relatively low numbers.

Spread of IRG Review Assignments
Reviews must strike a balance between involving scientists with direct experience in the proposed research and those with broader understanding. Clustering of similar applications lends itself to peer review in the narrow sense: that is, by scientists with direct experience.

Some factors to keep in mind in a discussion on clustering are that the Panel on Scientific Boundaries for Review (PSBR) recommended clustering as a way to minimize scattering; in 2000, the goal of 30 percent was set. In some cases, this is not a practical goal for several reasons. PSBR de-clustered some communities, and some areas of science bridge existing structures, such as the genetics of human behavior. Other areas, such as complications of diabetes, are inherently diverse, which defies clustering. At the same time, research in some areas of science is diminishing, so there are fewer applications. Extreme clustering essentially establishes an entitlement, which is counter to identifying the best science.

Dr. Schneider reminded the group that CSR reviews more than 15,000 applications per cycle. Study sections average about 80 applications each cycle; thus, the PSBR goal of 30-percent clustering would result in a cluster of about 24 applications. A "low cluster" might be five or fewer related applications per cycle.

Review and program staff have several options at their disposal to promote clustering. CSR can cluster better within IRGs, so that one study section reviews all of one subset, or at least among fewer IRGs. More aggressively, a SEP could handle that cluster, or a working group could be formed that might come to PRAC to propose a new study section to deal with the topic. Program staff could give low-represented areas a high program priority or write program announcements or RFAs to attract applications that would in effect be clustered.

CSR formed a working group to examine clustering and develop recommendations for PRAC and the CSR Director. The group, which includes CSR and IC staff, as well as one PRAC member, met at the end of November.

Review Outcomes in Low and High Clustered IRGs

Overall analysis: The group analyzed scoring review outcomes for three subsets of applications, which each contained about 200 applications per cycle, or about 1.3 percent of the 15,000 reviewed by CSR each cycle. The analysis was designed to compare overall review outcomes in low and high clustered IRGs, with a focus on outcomes for new investigators.

Subset A had low clustering in that over a 3-year period, applications were scattered across 19 IRGs. Good clustering would have meant that only six or seven IRGs would have been involved. A series of scatter plot graphs shows a trend that the applications that were better clustered seemed to receive better scores. The trend for scores in low-representation IRGs were more scattered and, generally, not as favorable. A committee is developing a statistical tool to measure the significance of such differences in scatter plot graphs.

Subset B applications were well clustered, with most of the applications going to one IRG. This subset had less of a difference in review outcomes between the higher and lower clustered applications, although the applications in the high-represented IRGs fared slightly better.

In Subset C, four IRGs had most of the applications. Here, too, applications that were less clustered scored below average, and those that were well clustered scored at least as well as the average.

This analysis indicates that clustering may have an effect on scores. At the same time, however, any changes must not disadvantage applicants, particularly more vulnerable segments of the community such as new investigators.

New investigator focus: Looking at how new investigators fared in these three subsets led to some surprising findings. In all three, as Dr. Schneider showed in a series of graphs, new investigators fared better when their applications were not well clustered, bucking the overall trend.

Clustering seems to be advantageous to the majority of applicants, but low clustering has advantages for new investigators and perhaps others. Tools for assessing clustering and determining the statistical significance of review outcomes are under development by a CSR Scatter Plot Committee.

Dr. Schneider asked PRAC for feedback on a series of questions posed by the working group. What should the clustering goal be? Is low clustering generally a disadvantage? Is a more comprehensive study needed? Should CSR provide interim realignment solutions, such as reversible SEPs, where low clustering has already been shown to be a problem?

Discussion

IFCN: Dr. Martinez observed the Integrative, Functional, and Cognitive Neuroscience (IFCN) IRG was a part of all three subsets and wondered if this was random or somehow related to that IRG's structure. IRG Chief Chris Melchior said the group has several totally captive study sections to ICs, which perhaps explains some of the scoring outcomes.

Scoring outcomes for different groups: Dr. Hamm commented that some PSBR principles, such as more integration of clinical and basic research and bringing basic science proposals into clinical study sections, may result in anti-clustering. She asked about follow-through to ensure expertise in these study sections is balanced. Dr. Scarpa said many complaints from applicants are from the clinical and surgical communities, who feel they are at a disadvantage when reviewed together with basic science.

Dr. Hamm said new investigators could be dealt with in ways independent of clustering, such as being reviewed as a group. Dr. Schneider said that solution could be put in place fairly quickly.

Dr. Murray said one interpretation of the data is that clustering supports the old boys' and girls' clubs. It works if you are established and a "member of the club," whereas in a declustered study section, the club does not exist and new investigators get a bump in their scores.

Dr. Mochly-Rosen said when physician scientists or clinicians propose basic research, their proposals are often not up to par with those of basic researchers. Similarly, basic researchers often do not do as well with clinical applications or those looking at populations or trends. Perhaps a hybrid will answer the needs of applicants who are doing translation research that do not do well in a class mode.

Dr. Brenner said he has been on both sides of the fence. The success rate of MDs has not been so bad, based on data presented to date. The success rate of research with humans as the primary research endpoint has not been as clear, but the evidence is anecdotal. Review environments are so heterogeneous that it is hard to make sweeping conclusions. He said having served on both basic and clinical study sections, the environment and expectations are different, if that is what clustering means.

Clarification of concept of clustering: In response to a question from Dr. Bunyon, Dr. Schneider said clustering, as used in this context, refers to subject area, rather than popularity of the topic. Some popular subject areas, in fact, do not have well-clustered applications. The ultimate cluster is the group of applications in response to an RFA. The examples Dr. Schneider used in the subset analysis represented a broad subset of R01 applications, not those in response to an RFA.

Drs. Buyon and Murray expressed difficulty in understanding the concern without knowing more about the subject areas of the three subsets. Dr. Scarpa said some areas, such as toxicology or alcohol, have certain study sections where it is clear they will be reviewed. Others may need many people with diverse backgrounds to gather the necessary expertise.

Assigning applications: Dr. Martinez suggested assigning applications to declustered study sections if that is where they will get a better outcome. Dr. Scarpa said doing so would complicate the review guidelines, in that considerations beyond the science would come into play. Dr. Mochly-Rosen suggested an experiment in which the same application is assigned to two different study sections to see how it fares. Scientific expertise and merit are two separate issues, but perhaps they are getting confused. One concern expressed by clinicians is that they are less likely to be reviewed by their peers because clinician scientists cannot travel to reviews.

Dr. Berg said he favored a larger study. One of the implications of clustering at the program officer level is that applications going to many different study sections are more difficult to track. The better the match between program and review clustering, the easier it is for program staff to provide good service to applicants. In terms of new investigators, the focus throughout NIH on funding at the program level means less dependence on peer review as long as the applications score within broad limits. The system needs to fund new investigators to keep itself intact as established investigators start to retire.

Translational research: Dr. McClain said it is difficult to get a good review for translational research, when basic and human-related research is in the same grant. He also said perception is important in terms of the significance of a topic. For example, research on the effects of lead on the liver may be perceived differently by lead researchers than by liver researchers. Dr. Scarpa said this example shows the challenges with clustering. He also gets many comments about how study sections review translational research with many distinct components.

Dr. Collins noted ICs do their own clustering and asked how well IC and CSR clustering match so as not to confuse applicants. Dr. Schneider said although most R01s are not IC-specific, there is cross-interaction. Any changes by CSR should not take place in isolation.

Dr. Leinwand, who serves on the clustering committee, stressed it is still not clear whether the differences in review outcomes for other than new investigators are statistically significant. This information is needed before going ahead. She agreed a more comprehensive study is needed to get more information on the statistics and also to look further at clustering itself.

Dr. Bunyon asked why reviewers are not brought in to focus on the parts of an application for which they have expertise to avoid the problem in reviewing translational research, as is done with program projects. Dr. Scarpa said the downside is too many reviewers would then need to be involved. As another downside, Dr. McClain said in some translational research where, for example, blood is drawn from a patient as a test of a mechanism, it might actually become more complicated if a clinical investigator was brought in to review that small part of the proposal. Dr. Bunyon pointed out many applications need a special statistical review. Dr. Schneider said many SROs bring in specific expertise successfully, although it is important to look at the minority of study sections where things are not going well.

Interim measures: In the meantime, said Dr. Scarpa, the Open Houses and other feedback show at least 12 groups feel that unclustering is not good for them. He asked PRAC whether to address their concerns on a piecemeal basis or await a broader solution. Dr. Hamm said if data show they are not getting a good review, it is better to address the issue earlier rather than later. Dr. Brenner asked how to act if the statistical significance of the data is still unknown.

Dr. Mochly-Rosen asked whether the Eureka grant review groups could serve as an experiment, since these applications are in effect not reviewed as a cluster. Dr. Berg said data collected could be analyzed, but he was not sure whether the lessons could be applied to this issue.

General Discussion and Future Agenda Items

Dr. Scarpa asked for input on priority topics for PRAC to consider in 2008. Already proposed is a presentation from the National Science Foundation, which will be scheduled for April 2008. In addition, Dr. Gray will have a member of her staff make a presentation on an experiment conducted at the National Cancer Institute.

Locus of review: Dr. Hamm suggested a discussion on locus of review and comparisons across ICs and with CSR about how reviews are done. Dr. Story Landis asked for specific issues on which to focus so ICs would know who on their staffs should participate. She said one issue may be the firewall between program and review. Some people think it is too impermeable, while others do not think it is permeable enough. A panel on this topic, especially if it involved Institutes with different strategies on whether they handle reviews in-house or through CSR, would be interesting. Dr. Scarpa said Dr. Bravo could help identify which ICs to invite.

Electronic referral: Dr. Gray asked for a demonstration of the text fingerprinting system. Dr. Scarpa reviewed how the system works.

Minority investigators and R01s: Dr. Martinez asked for an update on how minority applicants are faring with R01s.

Productivity: Dr. Murray suggested an assessment on how productivity per dollar changes with the size of a research group or of funding. In his experience on study sections, reviewers have a hard time assessing productivity, which seems to work against people in smaller research groups. It is important to get reviewers to realize, for example, that one paper every 2 years from a very small lab represents reasonable productivity compared to a large, well-funded group that publishes more frequently. Dr. Scarpa said he attended a meeting in Sweden in which this topic was discussed and noted salary in Europe and China is often dictated by paper output. He welcomed any ideas to measure outcome. Dr. Bravo said the Science of Science Committee of the Office of Portfolio Analysis and Strategic Initiatives (OPASI) is asking similar questions, and its director, Alan Koretsky, may have enough information to address PRAC in 2008. Dr. Murray said productivity outcome depends on the metric used; however, a broader array of metrics with a similar answer may reveal an underlying truth. Dr. Bravo said the simple measure of how many papers are produced per dollar is not adequate.

Dr. Landis said a colleague has observed that a paper published in a high-profile journal probably means the research is not at the leading front of the wave. Productivity is a more complex issue than counting papers or even a scientist's self-selected five best papers. Dr. Bunyon said training also has plays a role—the influence a scientist has on perpetuating science as their mentees develop.

Dr. Berg suggested a discussion on existing metrics, such as the Hirsch index, as a PRAC agenda item. Dr. Murray said he was not arguing that a metric should be used to decide funding but was relating a concern particularly by those in smaller groups. Dr. Landis said the Bridge awards target smaller groups to give them a way to regroup and resubmit in a funding emergency. Dr. Berg explained how bridge funding works at the National Institute of General Medical Sciences, as well as how well-funded investigators' applications are handled in making funding decisions. He offered to present more data on these topics at a future PRAC meeting.

Review outcomes: Dr. McClain suggested three topics for discussion: whether electronic review platforms are increasing participation by reviewers who cannot attend face-to-face meetings; how non-clinical trial research involving humans is reviewed; and how some of the ICs handle clustering, particularly when innovative science is at issue.

Ad hoc reviewers: Dr. Bunyon said she would like further discussion on the issue of ad hoc reviewers becoming permanent study section members. The reward of converting reviewers from ad hoc to permanent status is very meaningful for them, as ad hoc membership does not help in promotions. She also raised concerns the system might become bogged down if reviewers serve on panels twice, rather than three times, a year. A1s faring worse than first-time submissions might reveal problems that work against applicants. Dr. Hamm said the issue of continuity could be dealt with by asking reviewers to weigh in on an application even if they do not attend the meeting where the resubmitted application is discussed. Dr. Scarpa said the Advisory Committee has discussed this issue.

Reorganization: Dr. Brenner said CSR is changing as science changes, as shown with the new divisions discussed. He said he has been thinking how the organization can best grapple with the science of the future. Dr. Scarpa said the effort is ongoing, and CSR receives much input before

coming to PRAC. He noted some of the reorganization of divisions is to improve workloads and quality control. The Open Houses provided answers from the scientific community to very specific questions. He suggested a PRAC meeting by phone or videoconference before the next in-person meeting about reorganization of the third division and about input to the working groups co-chaired by Dr. Tabak.

Journal manuscript decisions: Dr. Mochly-Rosen suggested asking some editors of general and more specialized journals to discuss how they deal with reviews when they know some of the authors submitting papers better than others. She said they may be dealing with issues analogous to the clustering concerns discussed earlier.

Dr. Scarpa said he welcomed other topic suggestions by e-mail and he would schedule a phone or other conference in the next few months. With no other topics raised, a motion was made and passed to adjourn the meeting at 1:51 p.m.

We do hereby certify that, to the best of our knowledge, the foregoing minutes of the December 2007 meeting of PRAC are accurate and complete. The minutes will be considered at the April 2008 meeting of the Advisory Committee, and any corrections or comments will be made at that meeting.

05/02/08

Cheryl Kitt, Ph.D.
Executive Secretary
Peer Review Advisory Committee

05/02/08

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

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