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

April 30, 2008

The first 2008 meeting of the Peer Review Advisory Committee (PRAC) convened at 8:30 a.m. on Wednesday, April 30, 2008, at the Hyatt Regency Hotel, Bethesda, Maryland. The entire meeting was held in open session. Dr. Antonio Scarpa and Dr. Story Landis presided as Co-Chairs.

Members

Story C. Landis, Ph.D., Co-Chair
Antonio Scarpa, M.D., Co-Chair
Jill P. Buyon, M.D.
R. Lorraine Collins, Ph.D.
Garret FitzGerald, M.D.
Paulette S. Gray, Ph.D.
Heidi E. Hamm, Ph.D.
Craig J. McClain, M.D.
Daria Mochly-Rosen, Ph.D.
Louise Ramm, Ph.D.
Jane Steinberg, Ph.D.
Beverly Torok-Storb, Ph.D.

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

Welcome, Introductions, and Approval of the Minutes

Dr. Landis called the meeting to order and asked PRAC members to introduce themselves and their research interests. Dr. Landis said, as new PRAC co-chair, she will work with Dr. Scarpa and Dr. Kitt to develop agenda items related to CSR and Institute review.

Dr. Kitt asked PRAC to approve the minutes from the December 2007 PRAC meeting. The motion passed unanimously.

Update on CSR Initiatives

Dr. Scarpa thanked PRAC members for their participation in the meeting. His said his presentation would cover new data about peer review and improvements to peer review through changes in the CSR organization and the realignment of CSR peer review.

New Data

Applications and Review Load: Data drive the business of peer review. Beginning in 2002, the number of grant applications increased significantly, which created problems for reviewers' workloads and for the percentage of applications that receive funding. The numbers dipped in 2007, primarily because of a decrease of several thousand Small Business Innovative Research (SBIR) applications. Because SBIR grants are funded with set-aside funds, this decrease has meant that a higher percentage of SBIR applications are funded than other mechanisms.

The number of R01 applications submitted is steady, while the number of R21s submitted in 2007 held steady, and even declined slightly, after a sharp increase since 2001.

Dr. Scarpa next showed several graphs depicting the number of A0, A1, and A2 Type 1 (new) and Type 2 (renewal) R01 applications. The number of A2 (resubmitted twice) applications rose slightly for both Type 1 and Type 2.

Reviewer load (the average number of applications reviewed) increased slightly from six to seven applications in 2007, versus 11 a decade ago. Even that small increase means a decrease in the overall number of reviewers by several thousand, from a high of 18,000 in 2005. This decrease results in fewer people per study section, which helps section group dynamics, as well as savings in travel and related costs. The number of ad hoc versus permanent study section members, while still high, is improving.

Budget: In 2008, the CSR budget is $63 million, with an additional $39 million in the Scientific Review and Evaluation Activities (SREA) budget, which covers reviewers' expenses. Overall, the review cost, which includes travel for 17,000 reviewers, is less that 0.2 percent of the budget requested in the applications reviewed. Dr. Scarpa called this a great value to insure that outstanding science is funded.

Many changes that improve peer review also save costs. The SREA budget went down from $43 million to $39 million for several reasons: fewer reviewers traveling, more electronic reviews, the purchase of nonrefundable air tickets, and more meetings held on the West Coast, where hotel costs are lower than in the Washington, D.C., area.

Improving Peer Review
Dr. Scarpa presented three matrices on improving peer review, related to changes in CSR operations, changes in peer review, and potential system changes.

Changes in CSR Operations: The reorganization of CSR divisions and Integrated Review Groups (IRGs) is continuing. The goal is to realign the IRGs to reflect changes in science, improve efficiency and effectiveness, and improve consistency. In addition, efforts are underway to recruit Scientific Review Officers (SROs) proactively. The plan for a new Neuroscience, Aging and Development Division was developed last year, with PRAC assistance, and a search is underway to hire a director. (Dr. Don Schneider, Director of the CSR Division on Molecular and Cellular Mechanisms, will present more information about the review divisions later in this meeting.)

Nine review-enabling committees improve CSR operations, covering policy, scientific community outreach, SRO recruitment, reviewer recruitment, knowledge management, electronic review, SRO training, mentoring and staff development, and best practices.

Because the budget is fixed, increasing efficiency is necessary. With electronic submission of applications implemented, most applications can be assigned to review groups through text fingerprinting by October 2008. [

In addition to hiring a director for the new division mentioned above, two new division directors and a sizable number of SROs must be recruited.

Changes in Peer Review: To improve study section alignment and performance, a second round of internal reviews of all IRGs is beginning and will take about two years to complete. Dr. Scarpa identified two major challenges, for which he seeks PRAC input: the percentiling or ranking of applications reviewed in Special Emphasis Panels (SEPs) and the locus of review for "orphan" applications. He suggested forming a PRAC subcommittee to deal with the first issue, and noted Dr. Schneider would present on the second topic later in the meeting.

Progress has been made on shortening the review cycle. The pilot that enabled applicants to resubmit in the very next review cycle is being scaled up. If applicants choose, their application can be reviewed three times within one year.

Recruiting and retaining high-quality reviewers: A review is as good as the reviewers. Contrary to the perception by some applicants that their applications are reviewed by less experienced reviewers, the percentage of ad hoc reviewers at assistant professor level rank decreased from 10 percent to about 5–6 percent, and only 1 percent of permanent study section members are assistant professors. A registry of reviewers, developed with scientific societies, is growing as a valuable resource. The decision to hold meetings on the West Coast has aided in reviewer recruitment, as has expansion of peer review platforms to various types of electronic reviews.

Another incentive for permanent study section members, effective as of February 2008, is the removal of deadlines for their applications. This change means they are not reviewing applications and writing their own at the same time. Reviewers have been very appreciative.

Dr. Scarpa concluded by noting the Enhancing Peer Review process, addressed by Dr. Zerhouni in the next presentation, may point to larger system changes.

Discussion
Dr. Landis praised Dr. Scarpa for the changes made in his relatively short time as CSR director. Dr. Jane Steinberg noted he has brought scientific methodology to an evaluation of CSR to make changes.

In response to a question from Dr. Louise Ramm, Dr. Scarpa said a priority is to calculate the total number of applications received more precisely. Electronic submission, in which a small number are resubmitted because of computer issues, has caused some imprecision.

Discussion on Dr. Scarpa's presentation resumed after the presentation by Dr. Zerhouni.

Update on Peer Review Enhancements
Dr. Scarpa introduced Dr. Zerhouni, noting the NIH Director has been a champion for peer review and for improving it in a participatory way with the scientific community and all of NIH.

Dr. Zerhouni praised Dr. Scarpa as a change agent. He said peer review is a core issue for NIH to address because it is part of a larger ecosystem that relates to programs, extramural scientists,

funding, and other issues. While first-rate peer review is a cornerstone of NIH and is praised and emulated around the world, no system can remain perfect. The increasing breadth, complexity, and interdisciplinary nature of science, coupled with funding trends, are creating new challenges and stresses. The continuing charge is to fund the best science by the best scientists with the least administrative burden, but an added challenge is to recognize that "best" depends on many factors. An important goal is to reduce the administrative burden on scientists and reviewers, so more time is spent doing science than managing its ancillary aspects.

The review of NIH peer review has three phases that will lead to new policies and actions: a diagnostic phase which began in mid-2007, design of an implementation plan, and phased implementation of selected actions to begin during June 2008.

Diagnostic Phase
The diagnostic phase was planned so people would not jump to proposing solutions before understanding the issues. This phase led to a formal report released in February 2008. Through internal and external advisory groups, this phase engaged the scientific community in addition to NIH. The report identifies challenges and recommended actions in seven major categories: (1) reducing the administrative burden of applicants, reviewers, and NIH staff; (2) enhancing the rating system; (3) enhancing review and reviewer quality; (4) optimizing support at different career stages; (5) optimizing support for different types and approaches of science; (6) reducing stress on the support system of science; and (7) meeting the need for the continuous review of peer review. This last category in particular needs PRAC assistance.

Design of Implementation Plan
The report is being used to design an implementation plan in the second phase. Three working groups or clusters are looking at (1) applications, review, and ratings; (2) quality of peer review; and (3) support at different career stages. These clusters are getting input from across NIH, often revealing different perspectives about the same issue. A crosscutting committee will provide an integrated set of recommendations. Dr. Zerhouni said the emphasis first needs to be on tackling big challenges to transform the system, rather than smaller details—the "rocks" versus the pebbles or sand—guided by the principles of doing no harm; maximizing the freedom of scientists to explore; and focusing on changes most likely to add significant value at a reasonable cost/benefit ratio. Three core themes have emerged in the implementation plan: excellence of reviewers, fairness and clarity of peer review, and support of scientists at various stages in their careers.

Excellence of reviewers: Driving any changes is the realization that the excellence of peer review is directly correlated to NIH's ability to recruit, retain, and motivate the most accomplished, broad-minded, and creative scientists to serve in study sections. Doing this requires reducing the burden of review, flexibility, a training strategy, and recognition and reward for distinguished service. Many people have shared their concerns that the review experience has become a chore, and the experience must be improved. Ways to lessen the burden of review and introduce more flexibility are needed. Another point that has come through is that the quality of reviews depends on training of SROs, reviewers, and study section chairs. Tangible reward does not necessarily mean paying people more, but perhaps extending a grant for distinguished service. Again, he stressed these are all preliminary ideas.

Fairness and clarity of peer review: The perception exists that NIH is not consistent in ranking applications by their relative merit and/or public health impact and feasibility. Applicants and program officers say they need clear and purposeful review feedback, including more useful summary statements, as well as a rating system consistent across study sections and fields of science. The statement is often made that peer review is inherently conservative. Many recommendations addressed this perceived tendency, including recommendations aimed at better feedback and summary statements aligned to rating criteria. Reviewers say an area of frustration is the process does not offer the opportunity to step back and rank a block of applications. More in-depth discussion would also restore an element of the meeting reviewers like. He asked for PRAC feedback and said meetings will take place with study section chairs, with several pilots proposed.

A strong recommendation that has emerged is to pilot a shorter application. The "straw man" proposal is 12 pages in the body, with eight pages of appendixes. This change has implications in terms of computer systems and e.gov submission, and it would require approval by the Office of Management and Budget.

Support scientists at different stages of their careers: The message that came through in a recent workshop on innovation in science is the difficulty in assessing innovation prospectively. To insist on preliminary data for early-stage investigators means they will not take chances but rather replicate or expand what they have learned. The Pioneer and Eureka awards and other mechanisms are designed to encourage innovation. Peer review should fairly evaluate proposals from all scientists, regardless of their career stage or discipline, and avoid a bias toward more conservative and proven approaches at the expense of innovation and originality.

Dr. Zerhouni said one of his main concerns as NIH director is the long-term implications to biomedical science of what is happening with early-stage investigators. Data support the assertion of a bias against early-stage investigators' applications. A proposal to develop separate study sections for them did not pass muster on full evaluation, but one possibility is to group these applications together in a study section.

Another set of recommendations centered on reducing the need for multiple applications for researchers who come near to being funded, as well as focusing on the appropriate level of balance between prospective and retrospective review, and less on technical details.

Beyond peer review lies a systems issue about the aging demographics of researchers. He showed a time-series graphs comparing the age of NIH principal investigators (PIs) and medical school faculty, beginning in 1980. At that time, there was a good match between the time of appointment as assistant professor and the time of receiving one's first grant. Over time, the average age of a first grant has risen to age 41 or 42, up from 35 or 36. Many factors are at play, including increased length of training and complexity of science. But the fact is that PIs are on average five years older than the general population, whereas they reflected the average age in 1980. Looking ahead, the aging of the PI population becomes more pronounced. Action must be taken now to avoid a workforce in 2020 with too few young people.

There were about 1,850 new investigators in 1980, compared to about 1,500 now. Drops in the NIH budget hit them disproportionately. In an extreme case, between 2005 and 2006, the number went from 1,686 to 1,350. A concerted effort raised the number to 1,590. However, 40 to 45 percent were people who, while new to NIH, had been funded elsewhere. He presented data on streamlined and fully reviewed applications of first-time and established investigators. First-time investigators had lower success rates at every submission stage for Type 1 R01s.

Continuous improvement of peer review: Dr. Zerhouni asked for PRAC involvement to establish a continuous quality control and improvement process for peer review, which he said Dr. Scarpa has already begun, rather than an ad hoc process every ten years or so. It will be a work-in-progress to improve rigorous and independent prospective evaluations that favor, rather than discourage, adaptive and innovative approaches to peer review and program management.

Phased Implementation of Selected Actions
The third phase of the process to enhance peer review is to implement a communication plan and select some actions to pilot. Some proposals will not occur, such as eliminating administrative costs in grants or instituting a "not recommend for re-submission" rating. An experiment to offer a "prebuttal" may be considered for some large programs, where there is a need for give and take given the complexity of the science. A requirement for a minimum level of effort by PIs was considered but rejected. However, there is a sense in the community that resources must be carefully distributed to fund complementary science from different PIs.

The analysis and diagnostic phases have taken place; now it's time for "therapy." He asked for PRAC involvement on several issues, including criteria for evaluating reviewers and for determining what makes a reviewer "distinguished," how to continuously measure and maintain quality, and responsiveness of the peer review process. Peer review drives not only science, but also the scientific workforce.

Discussion
Dr. Scarpa asked each PRAC member to pose a question or comment to Dr. Zerhouni.

Reviewers: Dr. Beverly Torok-Storb suggested a nomination process by peers to identify distinguished reviewers. She noted people do not serve as reviewers because of the money and believed that many would forego an honorarium altogether if the money went to fund more research. To further save money, reviews could take place in buildings funded by NIH, after being sure this practice would not constitute a conflict of interest. Reviewers with similar experience could serve as a tag team—if one could not make a certain meeting, the other one could.

Dr. Heidi Hamm suggested CSR call on a few top people in each study section to ask how they feel the section is working and how to improve it. This ongoing review would also give extramural scientists a sense of real interaction with the study sections. She asked about any thinking to increase grant amounts so investigators could submit fewer applications. Dr. Zerhouni said an analysis revealed a lab cannot function on the funds from just one grant, but a balance is needed between providing sufficient resources and serving the greatest number of people. Dr. Landis, who helped develop this analysis, said it costs $750,000 to $900,000 to run a

lab of seven to nine people, which shows the necessity of more than one grant. The implication of awarding larger modular grants, however, is that fewer grants would be funded.

Funding: Dr. Craig McClain said support for young investigators is the biggest concern in his field. With limited funding, it is harder to recommend they go into academic medicine rather than private practice. The gap between joining a faculty and getting the first R01 is a huge issue, and he cautioned against overlapping grants with senior investigators to help junior investigators.

Dr. Garret FitzGerald termed mechanism-based clinical research by physician-scientists an area of deficiency in the NIH portfolio. This research faces increased regulatory hurdles. The conventional time line of the R01 is not accommodating to that reality, and funding strategy is intertwined with the need to revise expectations about promotions. He welcomes the engagement of international reviewers and noted this global view needs to extend to the service U.S. scientists provide to peer review outside the United States. Dr. Zerhouni said these comments illustrate how the components of the system are closely connected. The regulatory and administrative overhead is daunting, especially for early career scientists involved in early translation efforts that involve human populations.

Innovation: Dr. Daria Mochly-Rosen said dramatically disparate scores for the same application might indicate it is an innovative grant that should be separately reviewed, as it might contain a potential breakthrough in thinking. Dr. Zerhouni said several new programs were created to evaluate whether or not the review system is too conservative. The Pioneer Award was created in 2004, and the outcomes of several studies it funded are emerging. At this point, the analysis is based on anecdotes and not data. New Innovator awards support people in the early stages of their careers, as do the Pathway to Independence awards. NIH must be a learning organization, trying new things even if they are not perfect. Another recommendation is a transformative pathway, in which 1 percent of R01s are judged like the Pioneer and Eureka awards.

Aging of PIs: Dr. Mochly-Rosen observed some faculty in their 70s and 80s are still running labs. A possible way to use their learning is to create a distinguished position within study sections, in which they review fewer grants but serve as guides and advisors with a more global look. It would help transition people out of active labs and improve the review process.

Distribution of funds: Dr. Lorraine Collins said she appreciated the list of things that NIH will not be doing, because there is concern in the field about some of them. Beyond review issues, she asked about the balance of money between direct research and indirect costs. Dr. Zerhouni said a study by the Rand Corporation showed institutions contribute about 20 percent to the cost of research. It would not be fair not to compensate them for developing the environment in which scientists work. He noted the Department of Defense is engaged in this discussion as well and has come up with a rule that no more than 35 percent of a grant can go to indirect costs. The doubling of the NIH budget stimulated the infrastructure that scientists benefit from, although institutions, donors, states, and other sources paid for it. NIH is in partnership with institutions and care must be taken not to dis-equilibriate the balance. Dr. Collins clarified her suggestion was not to remove indirect costs, but perhaps set upper limits. Dr. Zerhouni described the A21 federal-wide government approach to cost recovery, in which an independent agency sets indirect costs. Other entities, such as land-grant universities and Veterans Administration

facilities, receive lower indirect costs for research but they receive other federal funding. The reality of sustaining a viable scientific enterprise requires these indirect costs. Dr. Jill Buyon said some of the indirect money also goes to fund new investigators as a type of seed money.

Resubmission: Dr. Buyon described unevenness in reviewing resubmissions. In many study sections, reviewers look at how the applicant responded to earlier feedback; in others, reviewers start anew. Guidelines should be consistent across study sections. Dr. Zerhouni said this is a common complaint. An initial recommendation was made to allow only new applications. However, looking in real terms at not giving people an opportunity to correct an application resulted in huge pushback from many scientific societies. A pilot is intended that may look at methods of journal editorial boards as a model.

Dr. Buyon also suggested study section chairs have a role in selecting members of their section.

Dr. Hamm said science on science, is important rather than relying on anecdotes. Dr. Zerhouni termed it a need for science on science management. The best scientists are artists who use a scientific method, and they cannot be judged based on the number of papers produced or other measures. In contrast, science on science management would focus on distribution of resources, appropriate grant size, and other issues.

Dr. Torok-Storb said her colleagues asked her to emphasize the need for more feedback from study sections for streamlined applications. Dr. Zerhouni said many people have asked for more clarity to have a sense about whether to resubmit an application. This issue has a divergence of views. Many senior scientists favor a "not recommended for resubmission (NRR) designation," while younger scientists say they would not want to be stamped in this way. Experiments with clearer feedback are necessary.

From the floor, Dr. Alan Krensky, Director of the Office of Portfolio Analysis and Strategic Initiatives, asked PRAC's views about a slide in Dr. Zerhouni's presentation related to development of a queue of A0, A1, and A2 applications. Dr. Zerhouni said a sense of obligation in the community has developed with an inherent commitment to fund A1s or A2s over A0s.

Update on CSR Initiatives: Continued Discussion
Discussion on Dr. Scarpa's earlier presentation resumed.

Early investigators: Dr. Torok-Storb asked whether delay in funding of early investigators could be attributed to a disproportionate number of foreign investigators, who may experience a lag time until they are in the system. Dr. Landis and Dr. Ruiz Bravo have looked at the funding issue, although data have not been analyzed about the impact of foreign investigators. Dr. Landis said only 50 percent of new investigators are in their early stage: that is, no more than 10 years out of clinical training or a Ph.D. Defining early stage investigators and providing them an advantage is under discussion but is complicated. The Pathway to Independence award is a way to move people through training and into jobs to gain independence. Dr. Scarpa said one reason new investigators may be established elsewhere is that NIH is funding new areas, such as bioimaging.

Small Business Applications: Dr. Mochly-Rosen asked about the decline in SBIR applications. Dr. Scarpa said Dr. Joanne Goodnight, from the Office of Extramural Research, points to several factors, including electronic submissions and the deadline cycle, both of which might discourage inexperienced applicants. There is a need to increase the pipeline of these grants.

Global perspective: Dr. Mochly-Rosen praised NIH's leadership in integrating the review system and making it more global. Dr. FitzGerald said there is global competition for the best young people, with some going outside the United States. A high hurdle for foreigners who train in the United States and would like to stay is restriction on training grants. Dr. Landis said the decision to make the K99/R00 Awards open to non-U.S. citizens, which Dr. Zerhouni favors, generated much discussion. About half of these awards go to non-U.S. citizens. Fewer K08 applications from physician-scientists in neurology to the National Institute of Neurological Disorders and Stroke may be because few U.S. residents specialize in this field, but the K08 requires citizenship. Other training grants are also only open to U.S. citizens by law. Dr. FitzGerald said it is in the U.S. national interest to broaden them.

Paused mechanisms: Dr. FitzGerald also noted many institutions take account of major life events with a pause mechanism in terms of a timed promotion. He suggested funding agencies act in concert with this. Dr. Landis said a committee on career progression for women has been looking at related issues, but she is not sure if this particular concern has arisen.

Dr. Torok-Storb said perhaps there should be grants for valued senior mentors, just as grants exist for trainees, which would free up funds for research by junior people. Dr. Buyon said the K24 could be recognized as a true mentorship award. Dr. Landis said a committee is looking at many workforce issues. The magnitude of the aging issue had not been recognized until Dr. Zerhouni requested the data he presented earlier. Support for the next generation of scientists is one of Dr. Zerhouni's highest priorities.

Dr. Hamm asked about the first renewal by young investigators. Dr. Scarpa agreed the first competing renewal is a critical time. The data need to be analyzed. Dr. Landis said Councils often deal with this. There are bridge funds of about $90 million going to A0s and A1s that are within 10 percentile points of the payline, although there are twice as many requests as funds. She said many Institutes use these preferentially for first competing renewals. Dr. Ruiz Bravo said the NIH directors will discuss the workforce in a meeting in the next several months, addressing these kinds of questions. Dr. Mochly-Rosen noted the time of a first competitive renewal is also when tenure decisions are made. It is important for people to get tenure, but quality control is also important because tenure means they are permanently in the system. She suggested reducing the amount of money or time for a grant to provide the opportunity for investigators to improve their projects. Dr. Landis said Dr. Zerhouni is concerned because he does not want universities in effect to allow NIH to make their tenure decisions.

Dr. Buyon said a differential persists between ad hoc and permanent reviewers. Ad hocs may be asked to come back repeatedly, yet remain unable to gain academic promotions within the community based on service. Dr. Scarpa said reviewers are learning during the first few times as reviewers, and some may find they do not want to continue. In England, frequent ad hoc reviewers are given a special classification, which gives them more recognition.

A0s: Dr. Landis returned to the issue of the application queue raised by Dr. Krensky. Dr. Krensky said the data show a decline in funding of A0s, although they are eventually funded as A1s or A2s. Some people say NIH should fund more A0s to avoid disrupting labs, while others say doing so would have a big impact because it goes against expectations. Dr. McClain said ultimately there can only be so many set-asides, and the system has to be left alone to work. Dr. Steinberg said an option may be to phase out A2s, so it reduces the burden of review and releases more money for A0s. Dr. Collins said applications are usually better after revision and resubmission, and agreed the system has to be left to work. Reducing the maximum of resubmissions to three helped the process, but going down to two might do harm. Dr. Landis said the question NINDS has been discussing is whether the science or the grantsmanship improves in resubmissions. Dr. Torok-Storb said everyone in her institution is guaranteed one cycle of interim support, but the support is dependent on the rating of the grant, which is why feedback on an application is important.

Update on Electronic Submissions

Ms. Megan Columbus, NIH Program Manager for Electronic Receipt of Grant Applications, reported that electronic submission is now business as usual, with about 100,000 applications received electronically since 2005. Applicants are now familiar with the system. As of January 8, the error correction window went from five to two days, which means SROs get applications quicker.

Grants.gov now has an Adobe form, and the next transition will be to these forms from PureEdge. Benefits include platform independence and the need only for Adobe Reader and no special software. The transition to Adobe is tentatively set for December, although Grants.gov is working through some issues around different versions of Adobe. Changes in forms will be coordinated with the Adobe transition to minimize the impact on the applicant community.

Once the Adobe transition occurs, the Career Development (K), Fellowship (F), and Training (T) grant programs will transition to electronic submission, with a very tentative timeline of February 2009 for K awards, April 2009 for F awards, and September 2009 for T awards. Her office is exploring how to use Grants.gov for complex applications, which it currently cannot handle. The discussion is in very preliminary stages, with no timeline yet set.

Discussion

Dr. Landis noted a side benefit of electronic submissions is the freeing up of space once used to receive and process paper grants. Dr. Steinberg and Dr. Torok-Storb praised the process.

Update on CSR Realignments

Dr. Don Schneider reported on the effort to realign CSR divisions and IRGs to respond to changes in science and workloads. In 1999, three divisions handled about 10,500 applications each. In 2008, three divisions would handle almost 17,000 applications each. He outlined the process of realignment to date, beginning in April 2007. Since then, in a series of PRAC discussions and consultations with internal and external communities, five review divisions have been proposed. He presented a series of slides that detailed the proposed changes.

Divisions: The five divisions would encompass (A) neuroscience, development, and aging; (B) AIDS and behavioral and population sciences; (C) basic and integrative biological sciences; (D) physiological and pathological sciences; and (E) translational and clinical sciences. PRAC approved the first two at previous meetings. The specific names of the divisions would be determined later, and three new division directors would be hired.

IRGs: The division changes mean a different clustering for the IRGs, including the creation of nine new or reorganized IRGs. Dr. Schneider outlined the study sections that will be within these IRGs, as well as under which division they would fall.

Study sections: Increasing workloads are also behind the need to split three study sections: Neural Basis of Psychopathology, Addictions, & Sleep Disorders (NPAS); Clinical Neuroscience & Disease (CND); and Medical Imaging (MEDI). In each case, working groups looked at how to split the study sections from scientific and workload perspectives.

Dr. Schneider summarized what was before PRAC: approval of three new CSR divisions, for a total of five; creation of six new IRGs along with the three new IRGs approved previously; and splitting of the three study sections.

Discussion
Dr. Ramm asked how working groups were involved in the recommendations. Dr. Schneider said once PRAC approved the changes conceptually, feedback was solicited from affected ICs. Dr. Landis thanked Dr. Schneider for the opportunity for PRAC to be involved in the process. She reflected the sense of the committee that the realignment is a good effort.

Clustering of Grant Applications for Review
Dr. Schneider's next topic on the agenda focused on clustering of applications for review, particularly so-called orphan applications, so they are treated to achieve optimal fairness.

Peer review implies clustering of similar applications and review by people by equal rank. The nub of the matter is whether clusters should be broadly composed of any research scientist who can understand the scientific problem proposed or more narrowly composed only of scientists with direct experience in the area of proposed research. The Panel on Scientific Boundaries for Review (PSBR) and other recent trends altered clustering, but with somewhat conflicting messages. On the one hand, PSBR set a high bar for clustering, with a goal of at least 30 percent of applications in a given study section similar, but it also de-clustered some communities. Some areas of science are diverse and bridge several study sections, while research in other areas is diminishing. Extreme clustering essentially establishes an entitlement and is counter to broad study sections. The question becomes how to balance these trends.

A clustering working group was established, made up of CSR and IC staff, as well as PRAC member Leslie Leinwand. When the group first reported to PRAC, the committee recommended a comprehensive study, including development of a statistical tool to assess differences in review outcomes and a means to identify orphan applications.

Statistical Tool
Earlier analysis showed applications reviewed in high-representation study sections score better than applications that are more scattered across study sections, but a tool to determine any significant difference did not exist. CSR formed a "scatter plot" committee to develop a tool to assess differences in review outcomes depending on whether an application is in a study section with a high or low number of similar applications. The approach is to calculate a ratio of odds and the 95 percent confidence limits on the odds ratio using multivariate logistic regression.

Dr. Schneider showed an example of how the tool works. In a group of 1,200 applications, 127 in the high-representation group scored well (in the 15th percentile or better) compared to 21 in the low-representation group. The odds ratio is 2.2: that is, an application in the high group has twice the odds of scoring in the top 15th percentile as one in the low group. Other variables are also factors, such as whether the application is amended, is a renewal, or is a new submission, but the tool can take this into account. He explained how it can calculate the odds from raw data.

While the tool confirmed the significance of differences associated with high- versus low-representation applications, he warned the findings must be interpreted with caution. They should only serve as a starting point for further investigation. He presented an example from the Skeletal Muscle and Exercise Physiology (SMEP) study section. Before 2004, applications in this area were scattered, and the community pushed for its own section. Comparing outcomes from 2001-2003 and 2005-2007 shows scores improved when applications were less scattered.

One limitation in using the odds ratio is that 100 data points, in the form of separate applications, are necessary to reach a 95 percent confidence limit. But if there are that many applications on a topic, then they can no longer be considered "orphans."

Definition of Orphan Applications
The SROs and study sections were polled about what constitutes an orphan application. The definition varied, with general agreement that one or two applications per cycle should be considered an orphan. He presented a list of examples of sample orphan topics received.

The interim conclusion is that the analysis of orphan applications is challenging. The odds ratio is useful, but the weakness is the number needed to reach 95 percent confidence. To accumulate the necessary numbers would take an unrealistic amount of time. CSR continues to look at ways to identify and group orphan applications.

Discussion
Definition: Dr. Buyon suggested an orphan application may be an application that requires an ad hoc reviewer. The SROs could be asked how often they need to go outside of the study section to locate appropriate expertise. The definition helps determine how to address the problem.

Entitlement: Dr. Hamm asked why clustering equals entitlement. Using a hypothetical example, Dr. Schneider said if there were a left toenail study section that receives 50 applications, at least 10 percent would get higher scores. The potential entitlement comes from the fact, using his example, other parts of the foot would not have what amounts to a set-aside. Dr. Landis suggested an example from her IC relates to muscular dystrophies: the most common topic for

research is Duchenne's, but researchers who work in other areas feel their grant applications get short shrift. However, it is a more complicated issue than the numbers. Dr. Ruiz Bravo said the enhancement of peer review process has been looking at the scoring system, which may address this problem in a less direct way.

Review versus program: Dr. Collins said with funding made at the Institute level, funding decisions are based on more than peer review scores. To her, the issue seems more a program than a review issue. Dr. Landis said Councils serve as a second-level review and do make decisions about funding. She noted a group of about 40 or 50 applications funded by the National Institute of Environmental Health Sciences is now the subject of scrutiny about whether they were funded out of order based on other priorities. There are different views about how easy it is to override scores at the Council level. Dr. Ruiz Bravo said it is not easy, but is something ICs can do. Dr. Landis said if her Institute's Council wants to fund something out of order, it must show strong justification in discussion and in writing. Dr. Paulette Gray noted CSR should be reviewing the science in grant applications, while ICs should make funding decisions.

Additional data: Dr. FitzGerald suggested a complementary way to get quantitative information: Reviewers could self-rate how well they feel they know the science and the potential public health impact of an application, while applicants could be asked to review the section roster and rate how well they think the membership is equipped to judge their application. Applicants' real concern is whether reviewers are well equipped to review their applications, which goes beyond if an application is an orphan. Dr. Hamm agreed balance is needed, even if ICs make funding decisions, to make study sections function as well as possible to deal with potential problems.

Dr. Buyon said another data point is how often applicants choose a study section versus those that are assigned. She said many people do not recognize the importance of a choice and suggested CSR should reinforce the benefit, especially to newer investigators, of applicant choice. Another checkpoint would be if all the reviewers marked 5 on an application in terms of their expertise to review it. In these cases, an application should be sent to another study section to see if it belongs elsewhere. Dr. Scarpa said applicants' own selection of a study section, combined with the text-fingerprinting tool, places most applications where they need to be.

Peer Review in Germany: A Look at the DFG Model

Dr. Marion Mueller, Director of the North America Office of the German Research Foundation (in German, Deutsche Forschungsgemeinschaft, or DFG), reported on how her agency undertakes peer review.

Facts and Figures about DFG

The DFG is the central, self-governing funding organization that promotes research at universities and other publicly financed research institutions in Germany. Its current budget is approx. 2 billion Euros, about 58 percent from the German federal government and the rest from the states. It is not a government agency, but rather a self-governing membership association under private law. In addition to funding research, the DFG fosters scientific excellence through competition, advises parliaments and public authorities on questions relating to science and research, encourages international collaboration in science and the humanities, and supports the advancement and education of young researchers.

The DFG serves all fields of science and the humanities. She presented an organizational chart, focusing on the divisions that specialize in life science research. Review boards are not responsible for the reviews themselves but for assessing the peer review process and making funding recommendations to the DFG Joint Committee. Review board members are elected representatives of their communities and the boards comprise several related subject areas. As laid down in the framework policy, the 48 review boards can choose between various working methods and can thus structure the review and evaluation of proposals according to their needs.

The DFG budget has increased over the past few years, with life sciences getting the largest share of funding in 2006 (38.7 percent), followed by natural sciences (26.0 percent), engineering (21.0 percent), and humanities and social sciences (14.3 percent). Within life sciences, medical sciences receive the largest amount. Between 2002 and 2004, about 20 percent of the entire DFG budget went to medical sciences, with most of that amount going to 68 institutions of higher learning and 108 non-university research institutions

The Individual Grants Program is the DFG program most similar to R01s. The Individual Grants Program is the DFG's central form of the research funding. There are no specific thematic solicitations and the research grant can be used to fund staff, scientific instrumentation, consumables, travel as well as most of the other financial requirements of a research project. The DFG receives about 10,000 proposals of this type per year and funds about 5,100 of them.

Mechanics of the DFG Peer Review System
Dr. Mueller discussed how decisions are made about which proposals to fund. She acknowledged that peer review in general has its limitations, but that there is no alternative. She then explained peer review as practiced within DFG and showed the various steps a proposal takes from submission to decision.

There are about 10,000 active peer reviewers. The DFG program directors try to find the closest competitors of the applicant who can be trusted to produce neutral and fair comments to the head office. They are usually distinguished in the field through publications, honors, and other factors. The review boards ensure the quality of the review process. Review board members are elected every four years by scientific and academic representatives of the member institutions and serve in an honorary capacity.

There are two basic review modes: a written process used for proposals submitted in the DFG's Individual Grants Program, and panels for the competitive review of individual grants as well as for proposals in the DFG's Coordinated Program scheme. The Head Office is responsible for the effective management and administration of the whole review process. The activities range from the election of the reviewers to preparing the relevant drafts and recommendations. On the basis of these reviews, the review boards develop their funding recommendations that in turn are forwarded to the DFG Joint Committee, which usually follows the review boards' recommendations, but is not required to. From 2002–2004, 10,833 reviewers evaluated 24,419 proposals, with about a total of 65,665 written reviews taking place. About 12 percent were in programs to promote work of young researchers (research fellowships, Heisenberg Program and

Emmy Noether Program). Panels are used to review applications orally and, in the case of applications in the DFG's Coordinated Programs, will involve on-site visits.

Other Information
Dr. Mueller presented data of interest in comparison to NIH based on a survey of the years 2005 – 2007. More than 75 percent of reviewers worked for DFG once a year and produced an average of 3.3 written or oral reviews over a period of three years. The proportion of written reviews by researchers outside of Germany has increased and now stands at 17.7 percent (or 4930), mostly from Switzerland, the United States, the United Kingdom, Austria, and the Netherlands. The overall participation of peers from North America in the DFG peer review process (2005 to 2007) across disciplines amounts to 3.4 percent. 19 percent of all international reviews come from North America. Again an above-average percentage of reviewers working abroad is documented for the life sciences, the overall participation is 4.8 percent and the percentage of reviews from North America within the group of international peers is 23.8 percent. The percentage of proposals submitted in English is also above average in the life sciences, at about 19.2 percent in 2007.

Two review board elections have taken place since a 2003 reform. The 2007 election was one of the largest and most complex Internet elections in Germany to date. Average elected reviewer age has declined slightly, from 51.9 to 51.6; the proportion of women increased from 12 to 16.8 percent.

The Institute for Research Information and Quality Assurance has carried out a survey of review board members to summarize their experience and identify potential problem areas. The overall feedback was positive. Results are available in German online (http://www.forschungsinfo.de/Publikationen/Download/iFQ_Working%20Paper%20No%202_FK.pdf).

Discussion
Review boards: Dr. Torok-Storb noted, in one of Dr. Mueller's slides, that review board members value compensation through reputation higher than the financial remuneration. Dr. Mueller said it is an honor people include on their resumes. Dr. Torok-Storb, who has reviewed for DFG, noted the agency funds post-docs to work in U.S. labs. According to Dr. Mueller, about 120 German post-docs are working at NIH. Dr. Mueller clarified there is no age limit for review board members, but they can only be re-elected once. In answer to a question about how review board members oversee the review process, Dr. Mueller said it varies by board.

Funding: Dr. FitzGerald asked how the state and federal governments determine research funding. Dr. Mueller said that institutional funding of research organizations is jointly financed by the German Federal Government and the states (*Laender*). There is a formula called the "Koenigsteiner Schluessel" (Koenigstein Key) that determines the share of funding of federal and state funds. Although the DFG receives its funding from the Federal Government and the states, it is not a federal agency. The principle of self-government and self-organization of science and research form the core of the DFG's mission. Whereas federal and state ministries will set priorities in specific research areas, the DFG's focus is on a bottom-up approach. The distribution of the DFG budget is entirely science driven with the life sciences' share currently at

38.7 percent. She added that the DFG is not the only entity funding the life sciences and mentioned among others the Max Planck Society, the Helmholtz Association, the Leibniz Association, and the federal and state ministries. Dr. Mueller was not aware of a comprehensive statistical compilation of all federal and state sources that provide funding for life and medical sciences, but mentioned the recently published "Report of the Federal Government on Research and Innovation 2008" as the best available overview.

In answer to a question from Dr. McClain, Dr. Mueller said clinical and translational research undergoes the same review process as other kinds of research.

Dr. Collins asked if salaries for researchers are covered in the grant funds. Dr. Mueller said salaries for young researchers are included, but not the salary of established and tenured PIs.

Dr. Hamm asked about any changes in the trend of funding to large groups versus individual investigators. Dr. Mueller said this has been the topic of much discussion, with the perception by some that larger groups are favored. However, the core funding instrument remains the Individual Grants Program (total DFG budget 2007: 2,167.6 million EUR; Individual Grants Program: 635.4 million EUR).

Dr. Ruiz Bravo asked about funding for universities. Because most are publicly funded, salaries for researchers come from public entities. Retirement, with some exceptions, is mandatory, although many people would like this to be changed. Dr. Landis noted a point from Dr. Zerhouni's presentation earlier in the meeting that showed that the U.S. system may be creating a system with more older than younger researchers.

Dr. FitzGerald asked whether DFG is addressing how to help translational science gain better access to patients. Dr. Mueller answered this question by referring to a joint program by the DFG and the German Federal Ministry of Education and Research (BMBF). This program funds clinical trials within the framework of two complementary, coordinated funding measures. Funding by the BMBF focuses on interventional trials on pharmacological therapeutic measures, meta-analyses and reviews of clinical trials. The DFG primarily funds interventional clinical trials on non-pharmacological therapeutic procedures, in addition to prognostic trials and controlled trials on secondary prevention.

Dr. Hamm asked about young people entering the system. Dr. Mueller said there have been deep changes over the past five or six years in the German university system, which have opened up opportunities. She drew particular attention to the Excellence Initiative, a nation-wide competition for 1.9 billion Euros aimed to strengthen the German university system and thereby creating thousands of new positions for young researchers.

Dr. Mueller closed by inviting PRAC members and others to contact her by email or in person at the DFG offices in New York and Washington (marion.mueller@dfg.de).

Instantaneous Electronic Scoring of Multi-component Applications
Dr. Olivia Bartlett and Dr. Shamala Srinivas of the Research Programs Review Branch, Division of Extramural Activities, National Cancer Institute (NCI), reported on an innovative way to review large, multi-component applications.

A large portion of the NCI review load involves large multi-component applications such as program projects (P01), centers (P50), cooperative agreements (U19 and U54), and other mechanisms. In 2006, two SROs in the NCI Division of Extramural Activities adopted the Interwrite PRS System, a simple, inexpensive system used by elementary school students to take tests, to score the components (sub-projects) of multicomponent applications in real time during the review meeting. Panel Members use a small hand-held device to transmit a score or adjectival rating after the discussion of each component. During the meeting, the mean scores and score ranges or adjectival ratings for each project/core are recorded on large laminated posters. When it is time to score the overall application, the reviewers have their own vote sheets as well as the "group think" tallies shown on the poster. The system promotes consistency in final scoring (priority scores) because scores for all the components are displayed.

The PRS system consists of a wireless radio frequency receiver that plugs into laptop computer and individual wireless radio frequency transmitters ("clickers") about the size of a TV remote control for each reviewer. Each receiver operates on its own frequency, so neighboring review meetings do not interfere with each other. The clickers have numeric, alphabetic and true/false buttons, which allow reviewers to enter numeric (1.0 to 5.0), adjectival (outstanding, excellent, very good, good, and acceptable) or yes/no scores/ratings for each application component. Each user is assigned a unique clicker ID so answers can be tracked. NCI contractors developed interface software to download review meetings and application information from IMPAC II and to display the means and ranges. Each component of an application becomes a "question" in the PRS system. Some set-up is required before each review meeting. Division of Extramural Activities Support (DEAS) staff were trained to operate the system during the review meeting.

After the meeting, Excel spreadsheets with each reviewer's scores are downloaded from the PRS system. DEAS staff correct the spreadsheet to reflect the scores recorded on each reviewer's paper vote sheet, which remains the "gold standard" for scores. Although telephone reviewers do not have clickers, analysis of the scores showed that the means generated from the clicker results during the review meeting were within 0.1 of the final means calculated from the vote sheets, giving the committee members accurate information on which to base their final scores.

Feedback from reviewers indicates that the PRS system is easy to learn and use, and that having the group means and ranges is useful in discussing and scoring the overall application and in promoting consistent scoring. The PRS system has been used in various review settings and saves DEAS staff time in calculating the final mean scores of individual project/cores after the review meeting. The system costs about $3,000 for the software, transmitter, and set of 32 clickers.

Dr. Srinivas then led a hands-on demonstration of the PRS system with PRAC members.

Discussion
Dr. Buyon asked how the range and average for the various components change before and after a meeting. This format does not specifically affect voting, but Dr. Srinivas said it is important to vote at the end of the discussion. Dr. Landis said the system would be very useful for other ICs. Dr. Kitt said CSR has experimented with live Internet-Assisted Review (IAR) scoring, but noted the NCI system also provides a way to post mean scores.

General Discussion/Future Agenda Items
Dr. Scarpa said PRAC members will be contacted about scheduling the next PRAC meeting in late September or early October. Before then, a video or teleconference, or a hybrid of the two, will be held to discuss scoring in SEPs. Dr. Mochly-Rosen asked that slides be sent in advance. A motion was made and passed to adjourn the meeting.

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

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

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

Story Landis, Ph.D.
Co-Chair
Peer Review Advisory Committee