

What We Learned on Our Summer Vacation: NIH reviewers tell us how the 12-page R01 does (and doesn't) change things

Over the past couple of months, the Office of Proposal Development, in collaboration with Naomi Rosenberg from the Sackler School and Linden Hu from Tufts Medical Center, worked to learn as much as possible from the first round of NIH reviews that require the new, shorter format for most proposals. We assembled two focus groups with reviewers, read a number of grants and summary sheets from a variety of fields and with a range of scores, and we put together a discussion of our results. The following document summarizes the major points that we've learned thus far, and we anticipate that we will continue to refine our advice to applicants based on subsequent review cycles.

Major Take-Home Points

The major criterion was the future impact of the proposed work, defined as the significance coupled with feasibility. This applied across the board, whether the grant was new or a competitive renewal. The "so what?" factor is so predominant that solid science proposed by productive investigators did not always get favorable reviews because the work wasn't seen as sufficiently ground breaking. Likewise, the production of numerous publications was not seen as sufficient, particularly if the impact wasn't high – these were described as "cluttering" the journals. On the other hand, lack of relevant publications was seen as a deficit if there were already noted problems with the impact. However, "flawed" proposals with a high impact tended to score well. The impact is that on the field, not necessarily on human health. People who tried to focus very basic science applications as relevant to human health, or who added a weak aim to that effect, did not succeed.

The most successful proposals were by those who had clearly started over with the new format, rather than trying to condense the longer format into 12 pages. Those proposals that had a substantial amount of background material, figures that were non-existent or too small, or that didn't well-integrate preliminary data tended to fare worst.

The corollary to that is there was a general sense that anything proposed had to take the research out of the PI's "comfort zone," and collaborations may be required to push the project beyond the PI's traditional system/approach. For example, if the PI were a geneticist doing experiments in yeast, proposing only continued yeast genetics experiments was not favorably reviewed. The expectation is that the PI will collaborate with someone who can approach the question in a different way (e.g., biochemistry or physiology) or move into different models. In the case of people who have always done cell culture experiments, they need to either move forward into animals, or employ more mechanistic/genetic approaches. Cancer researchers can't just keep using mouse models; people using yeast as a model need to take it into mammalian systems; etc.

The perception is that the bar is higher for renewal applications, and both reviewers and POs are thinking that every grant could/should be a new application. In part this is psychological: as a PI, you'll think about a "new" grant differently than a continuing grant. Additionally, for resubmissions, applicants lose the ability to respond to all critiques because of the one-page limit for the response to reviewers. The general advice is to summarize the critiques, look at the bigger picture, and make reviewers feel as if all critiques were carefully thought through.

While some applicants used Vertebrate Animals and Human Subjects sections to offload details of study design, this will not be allowed in future. The questions the applicant should address in these sections are changing, and while SRO instructed reviewers not to count this tactic against applicants in this round, stricter guidelines will be enforced in the future.

Reviewers liked the shorter format, and felt they were still putting the best grants to the top, even though what constitutes "best" is changing (see first point above). Reviewers noted that it was easier to quickly evaluate the impact of the proposed project, and that they could "hold the entire proposal in [their] heads." They still feel that the review format doesn't help the applicants as much, but are becoming a bit more accepting of NIH's repeated statement that

"the reviewer's job is not to write the applicant's grant for them." The most useful part of the review for those considering resubmission is to separate the responses to Significance from the responses to Approach.

The Applications Themselves

What is the most effective format you've seen that gives you sufficient information to judge the significance and approach? What is the least effective format or most consistent deficiency in presentation?

There was very little consistency in format, ranging from 1 page for Significance and Innovation to 4-5 pages. The worst was an application that left only 2 pages for the Approach, which was not favorably regarded. The highest rated grant in one study section had the progress report at the end, and it included a good bit of repeated material, but the feeling was that the grant rated highly despite this (not necessarily because of it). Reviewers said that it worked to either have a progress report ahead of the Approach, or to have it integrated throughout the Approach.

It was generally seen that integrating preliminary data with the appropriate aim was an effective approach. Both too little preliminary data and too much preliminary data were seen as ineffective. "Shortchanging" preliminary data hurt scores, particularly if the data were relevant to the innovation. Even with published data, including enough context is key. The proposal should be able to stand on its own, and the burden is on the applicant to make certain that there is enough information for the reviewers.

The most consistently effective strategy for the Approach was to treat each aim like a story. These proposals integrated necessary background information and preliminary data into the approach for each aim.

Some investigators chose to "save space" by not using any figures. This was considered a major failing. Lack of figures or tables and lack of white space indicated that the grant writer was having difficulty adapting to the new format, and this approach was not viewed favorably.

In this first round of reviews, SRO instructed reviewers not to penalize applicants for not following the format guidelines fully. Lack of a Personal Statement, using the old headers, etc. will start to matter soon. In a study section focused on clinical research, many of the investigators used the old Background, Preliminary Data, Research Design sections. This was less common among basic scientists. Reviewers did not like it when applicants did not follow the new format.

Because there was less discussion allowed on the details of the Approach (see below), reviewers adapted to the diminished detail in the descriptions of experimental design. One major plus for new investigators was to include solid collaborations with investigators who had the necessary expertise, especially when they proposed technical approaches that they didn't have much experience using. Overall, the quality of the Approach section drove the score of the proposal.

There was consensus that 12 pages (plus Specific Aims) were sufficient to convey the important information. Many reviewers liked the shorter format because they could see the project as an integrated whole. Also, some reviewers reported that the grants took just as long to review, but the time was spent in thinking, not "plowing through details."

How did you use the review sheets? Do you expect the grant to guide you by using the proscribed headings?

Most read the grant once, then went back through it with the review sheets. If the applicant had given them solid

sections/statements on innovation and significance, it was a big help. In particular, reviewers report that Innovation is the most difficult criterion to evaluate. They are looking for “flagged” statements, such as “The proposed research is innovative because....”

Most of the reviewers did the "Overall Impact" bullets last. Part of the learning curve for scoring proposals is to learn what each number represents since reviewers are not supposed to calibrate the new scoring system to the old.

While the success of a proposal is largely driven by impact, there was no specific section on the review sheet to discuss what drove the Overall Impact score. [According to OER, NIH will begin requiring reviewers to include a paragraph in their written critiques to explain the factors that informed his or her overall impact score." (August 2010 NIH Nexus)] Reviewers were discouraged from criticizing specific experiments, and some reviewers felt that they were less able to provide concrete advice about what portions of the proposal hurt the applicant's score the most.

The reviewers felt that the most useful feature of the review sheets for the applicant was the separation of Significance and Approach.

When reviewing a grant, to what degree did you consider the personal statement section of the biosketch, and the 'other' and 'ESI' resources sections?

The Personal Statements were considered useful when the reviewers didn't already know the applicant, and generally were read unless they were too long (the optimal length seems to be roughly 1/3-1/2 of a page). The Personal Statements were also useful in buttressing the protocol detail, and were helpful to reviewers as they wrote the Investigator section on the summary sheets. Those proposals that used a "standard" personal statement, rather than tailoring it to the grant, tended to be viewed less favorably. Worse, some applicants treated it as similar to a personal statement for a professional or graduate school application. Words like "sufficient energy and enthusiasm for the project" or "strong desire to undertake this research" were unfortunately regarded as inadvertent comedy. Objective detail regarding experience that does not show up on the training table, or needs to be highlighted from the publication record, is most useful.

The same is true for the intellectual environment sections under Other in Facilities and Resources. These need to be to the point to be effective. For early stage investigators (ESIs), the Personal Statement and discussion of Intellectual Environment was helpful (if they were not too long to be useful to the reviewers) because it overcame the lack of experimental detail in Approach. (E.g., "Considering where they are and the investigators nearby, they'll have a lot of resources to troubleshoot any technical issues.") In this section, ESIs do not need to give dollar amounts for laboratory renovations, start-up packages, etc. Instead, we recommend discussing the square footage of the laboratory that was renovated, that there were funds provided to hire a technician or buy particular pieces of equipment, what the start-up package allowed you to accomplish, etc.

It was noted that in this first round, some investigators used the Other section of Facilities and Resources to discuss information that should have been in the Approach. Although applicants were not penalized at this point, this was perceived as an attempt to circumvent the page limit.

Some applicants discussed publications from previous award periods in the Personal Statement. Some annotated the publications in the list in the Biosketch. This approach did not serve them well, because reviewers did not read it. However, annotating publications in the Progress Report citation list was seen as potentially helpful because it could tell the reviewers why each publication generated during the previous funding period was important.

The Study Section

Did you see a real change in how reviewers discussed applications, or are they trying to shoehorn the old expectation into the new system, and penalizing applications that don't give enough information?

The cultures are generally the same for long-standing study sections, but in terms of how the grants are judged, reviewers seem to be adjusting their reviews and expectations to the new, shortened format. The change in order of review (starting with the best, and reviewing in descending order) was reported to have had a bigger impact on study section culture than the shortened format. Furthermore, both chairs and reviewers are receiving additional training to help them adapt to the new review format. How grants are discussed is different:

- There was much less "bean counting" in the discussion of approaches. When reviewers seemed to get into detailed discussion/criticism, the SROs or session chairs would rein them back in. The discussion focused more on whether an application was conceptually flawed.
- There was less "bean counting" in terms of productivity, unless it was remarkable (high or low).
- SROs and Chairs would not let traditional introductions continue. Reviewers were asked to summarize in two sentences, rather than the longer narrative.
- SROs would not allow criticisms for "lack of detail" to continue for long, reminding reviewers of shortened format.
- Discussions of particular strengths were sometimes interrupted with, "But you haven't told me why this project is significant."
- In this round there was some self-correction with respect to scoring. Sometimes other panelists would push for justification of reviewer scores. A reviewer discussing many weaknesses would sometimes be met with, "That sounds like a 4 or 5. Why did you give this grant a 3?" from other panel members.

Innovation is not being stressed, although it can lift the discussion of an application if there is a good and credible case for innovation. "What's the new idea/twist?" was a constant point of discussion, however. Using the Innovation section effectively will help reviewers answer that question (see note under "How did you use the review sheets..." above). Innovation can be seen either as methodological or conceptual. Renewal applications tended to have the most trouble with Innovation, especially in human subjects grants where the PIs proposed to continue following a cohort without adding something new and different to the study. In basic science grants, aims that were highly innovative seemed to have a higher bar for preliminary data, however.

Study sections varied in terms of how ESIs were reviewed. Some reviewers noted that ESIs were shown more leniency, while others did not.

In some study sections the third reviewer is as outside the field as possible and still within a study section. This makes it more important to be sure to write a comprehensible application.

Serving on a review panel was seen as essential because review panels have their own "personalities" and unwritten rules. Additionally, finding the appropriate study section for one's grant is increasingly crucial. One piece of advice is to use "key words" from a proposal to search RePORTer for funded applications with a similar focus. From RePORTer, you can find the study section that positively reviewed similar grants, and you can use your cover letter to specify the expertise needed for your proposal that's represented in that study section.

In your experience, do outlying reviews drive an overall score?

Yes. In general, the culture of how panelists interact is the same. Trusted reviewers are still trusted. However, it seems that SROs in general are trying to make it easier to vote "outside the range," although this happens progressively as the panels continue discussion. The top grants tend to have good consensus among reviewers, but sometime grants discussed later (e.g. because initial Impact scores were 1, 1, 4) can move up after discussion. However, while most reviewers changed their scores after discussion, it was noted that there was a tendency for scores to get worse the longer the study section discussed a proposal.

Did your review group give any 1s? Were there many tie scores?

One definition of a 1 was for a "1 in 100" proposal. Very few 1s were given in the basic science sections. More were seen in the clinical section. Overall, the scores tend to spread, but there are still many tied scores.

Other critical notes:

In many study sections, competitive renewals were seen as a liability. Some of the panels attended by the reviewers in our focus groups had no competitive renewals in the entire batch of applications. The bar for what was considered "new," or a vertical step in the field, was higher for these than for new applications, "...and if it's that new, then why isn't this a new application instead?" Even some POs are suggesting that every application be a new application.

Focus the application on a single, key idea of high significance.