On My Mind: Overall Impact of Research…How Reviewers Score It

As I noted in the previous *On My Mind* note, *"Enhancing Peer Review at NIH…What Users Think About The Changes"*, one of the changes to the NIH peer review process involves assigning an overall impact score to applications that are discussed at review meetings. The overall impact score represents an assessment of whether the findings of the proposed research will "…***exert a sustained, powerful influence on the research field…***" (from January 20, 2010, "Guidance for Reviewers: http://enhancing-peer-review.nih.gov/docs/Reviewer_Guidance_Application_Changes.pdf"). In other words, if all of the technical issues identified for an application are successfully addressed, reviewers are asked to predict how important outcomes from the proposed research will be to the field in general.

The overall impact score is said to be independent of the five criterion scores (i.e., Significance, Investigators, Innovation, Approach, and Environment) and reviewers are instructed not to simply assign an average of the criterion scores as the overall impact score. ***And, most importantly for applicants, the overall impact score is the one that is used either in its raw form or as a percentile as the 'priority score' that is used for funding decisions.***

Judging from discussions with and feedback from reviewers, applicants and NIH staff, there was some confusion about the difference between overall impact and, for example, the review criterion designated as significance. Concern was expressed and questions were raised as to whether or not overall impact was different from, or could be judged independently of, significance. In fact, a Q and A about the difference between overall impact and other review criteria was posted on the on the NIH web site "Frequently Asked Questions/Enhancing Peer Review" (http://enhancing-peer-review.nih.gov/faqs.html).

"Significance is a stand-alone assessment of the project's goals in the context of the relevant field, and to a large extent assumes that the investigator(s), approach and environment are adequate to allow for successful completion of the aims of the project even if later discussion of each of these review criteria will identify problems. When reviewers assess the Overall Impact of an application they are expected to take into account the scored review criteria (significance, investigator(s), innovation, approach and environment) and the additional review criteria to judge the potential of the project to exert a sustained, powerful influence on the field."

In September 2010 NIH began requiring reviewers to include a narrative statement to explain the overall impact score. As stated in the FAQs about the changes in peer review, all applications whether or not they have been discussed in the review meeting, receive scores for the individual criteria. Not only do reviewers use the criterion scores to help them determine the overall impact or priority score, but the criterion scores are also thought to be useful in providing feedback to applicants about strengths and weaknesses of their application as well as to members of advisory councils and institute staff involved in making funding decisions.

The question that remains for applicants is whether reviewer behavior in assigning overall impact scores is consistent with the intent of the changes in peer review…specifically, whether or not overall impact scores are independent of the other criteria. This information is important for several reasons not the least of which it can help and guide applicants in preparing their applications.

The first attempt to address this question was by Dr. Jeremy Berg, Director of the National Institute of General Medical Sciences (NIGMS). Dr. Berg reported results of an analysis of 360 R01 applications assigned to NIGMS and reviewed at the October 2009 meeting of the institute's national advisory council. The analysis involved calculations of Pearson product moment correlations between overall impact score and each of the standard criteria. Results showed that approach (0.74) was most highly correlated with the overall impact followed by significance (0.63), innovation (0.54), investigator (0.49) and finally environment (0.37). Thus, according to Dr. Berg, given that overall impact was correlated with each of the criteria was an indication that reviewers were indeed using all of the criteria in determining their overall impact score.

However, the fact that the correlations, while all statistically significantly different from no correlation, appeared to each account for different levels of variability in the overall impact scores. The differences among the size of the correlation coefficients suggested that some of the individual review criteria may have been exerting more of an influence on overall impact than others. A principal component analysis showed that a single component or factor accounted for 71% of the variability in the overall impact scores. This means that 71% of the differences among the overall impact scores could be explained by one factor.

What was that factor? It should be no surprise, given the instructions to reviewers to take into account all of the review criteria in determining the overall impact score, that the factor was made up of contributions from all five criterion scores with weights as follows: approach = 0.57; innovation = 0.48; significance = 0.44; investigator = 0.36; and environment = 0.35. However, it is interesting to note that the criterion with the greatest influence (i.e., the greatest weight) was the proposed *approach* (i.e., methods) and not significance.

The second component accounted for an additional 9% of the variability in overall impact score, the major contribution was from *approach* (0.63) and significant negative contributions from investigator (-0.51) and research environment (-0.56). The third component accounted for 8% of the variability in overall impact, with the primary contribution from innovation (0.78). The fourth component accounted for 7% of the variability, mainly influenced by significance (0.85). The final component accounted for 5% of the variability and was made up of contributions from investigator (0.72) and environment (-0.67). No component seemed to have significance as its major contributor suggesting that reviewers were able to decouple it from overall impact.

One year later a similar analysis was completed on 654 R01 applications reviewed at the October meeting of the NIGMS council. Results were similar to those reported earlier. Pearson product moment correlations between overall impact score and each of the standard criteria showed that approach (0.81) was most highly correlated with the overall impact followed by significance (0.71), innovation (0.59), investigator (0.47) and finally environment (0.38). The principal component analysis for these data were in agreement with those from the previous analysis. For example, as stated by Dr. Berg in his blog (weights in parentheses are from a table in the blog), "the first principal component accounts for 72% of the variance, with the largest contributions coming from *approach* (0.59), followed by innovation (0.48), significance (0.45), investigator (0.34) and finally environment (0.33). This agreement between the data sets extends through all five principal components, although there is somewhat more variation for principal components 2 and 3 than for the others."

Dr. Berg repeated these analyses using percentiled scores with essentially the same findings for the principal component analysis. Subsequent to publishing these data on his blog he noted that the Division of Information Services in the Office of Extramural Research, Division of Extramural Research, had completed a correlational analysis on "…32,608 applications (including research project, research center and SBIR/STTR applications) that…received overall impact scores during October, January and May Council rounds in Fiscal Year 2010." Results for each IC and for NIH as a whole are presented in the following table.

IC	Approach	Significance	Innovation	Investigator	Environment	Number of Applications with Impact Scores
FIC	0.78	0.59	0.51	0.45	0.54	125
NCCAM	0.78	0.63	0.60	0.60	0.54	285
NCI	0.80	0.67	0.59	0.53	0.45	5396
NCMHD	0.82	0.69	0.75	0.71	0.57	57
NEI	0.83	0.69	0.62	0.59	0.49	777
NHGRI	0.79	0.69	0.61	0.58	0.52	224
NHLBI	0.82	0.67	0.64	0.56	0.48	3157
NIA	0.84	0.73	0.65	0.58	0.55	1521
NIAAA	0.84	0.71	0.63	0.51	0.41	427
NIAID	0.82	0.67	0.62	0.55	0.47	3809
NIAMS	0.84	0.65	0.65	0.57	0.49	1051
NIBIB	0.77	0.68	0.63	0.54	0.49	894
NICHD	0.83	0.70	0.63	0.54	0.49	2074
NIDA	0.83	0.69	0.60	0.54	0.47	1230
NIDCD	0.82	0.69	0.58	0.51	0.40	443
NIDCR	0.86	0.70	0.68	0.62	0.54	538
NIDDK	0.83	0.69	0.63	0.60	0.50	2271
NIEHS	0.83	0.68	0.64	0.56	0.49	490
NIGMS	0.83	0.72	0.63	0.62	0.53	2856
NIMH	0.80	0.68	0.58	0.50	0.44	1896
NINDS	0.81	0.67	0.60	0.55	0.49	2262
NINR	0.83	0.70	0.66	0.59	0.53	260
NCRR	0.81	0.69	0.65	0.59	0.56	426
NLM	0.88	0.74	0.82	0.71	0.67	139
NIH	**0.82**	**0.69**	**0.62**	**0.56**	**0.49**	**32,608**

The same trends found with the NIGMS applications are also reported for correlations based on this large data set. Once again, the highest correlation was between overall impact score and *approach* (0.82), the next was with significance (0.69), followed by innovation (0.62), investigator (0.56), and finally environment (0.49). Data for a principal component analysis are not available as of the publication of this note.

Keeping in mind that the data analyzed at this point are only relevant to a subset of grant mechanisms reviewed, the answer to our question about what information reviewers are using to arrive at an overall impact score is that it is clear that they use all criteria but to a different degree. The principal component analysis indicates that they rely mostly (i.e., put most weight) on the quality of the proposed approach (i.e., the ability of the overall strategy, methodology, and analyses to accomplish the specific aims of the project) and less so on innovation (i.e., the extent to which the application challenges and seeks to shift current research or clinical practice paradigms by utilizing novel theoretical concepts, approaches or methodologies, instrumentation, or interventions) and then significance (i.e., assessment of the project's goals in the context of the relevant field in the extent to which it addresses an important problem or a critical barrier to progress in the field). The investigator and research environment appear to have the least influence in determining the overall impact.

So what do all these correlations and statistical machinations mean for individuals preparing or revising applications? Beyond the fact that it is imperative that the reviewers be convinced that the

problem being addressed is an important one, the merit of the proposed approach is key and applicants should be careful in addressing this section of the application. Unfortunately, the data do not give any clues as to what part of the approach the reviewers view as being most important. However, it would not be surprising to learn that they are looking at the appropriateness of proposed methods to provide data that specifically address hypotheses generated by the specific aims in conjunction with the capability of the investigator and environment to support the appropriate, reliable and valid use of the proposed methods. But that's a topic for another *On My Mind*.

In the meantime, good luck and remember that there's only one thing that's guaranteed in the world of obtaining support for your research…if you don't apply you cannot get funding.

Norm Braveman
Braveman BioMed Consultants
Rockville, MD
March 28, 2011