Recommendations for Simplifying R01 Review Criteria

Report from the Simplifying Review Criteria Working Groups to the CSR Advisory Council

Working Group 1: Non-Clinical Trials Applications

Co-Chairs: Tonya Palermo, Ph.D., Bruce Reed, Ph.D.

Council Members: Jinming Gao, Ph.D., Alfred George, M.D., Yasmin Hurd, Ph.D., Deanna Kroetz, Ph.D., José López, M.D., Tonya Palermo, Ph.D.

Ad hoc Participants: Kevin Corbett, Ph.D., Michelle Janelsins, Ph.D., Brooks King-Casas, Ph.D.

NIH Staff: Sally Amero, Ph.D., Bruce Reed, Ph.D.

Working Group 2: Clinical Trials Applications

Co-Chairs: Tonya Palermo, Ph.D., Bruce Reed, Ph.D.

Council Members: Alfred George, M.D., Yasmin Hurd, Ph.D., Tonya Palermo, Ph.D.

Ad hoc Participants: Brian Boyd, Ph.D., Matthew Carpenter, Ph.D., Michelle Janelsins, Ph.D., Brooks King-Casas, Ph.D., Pamela Munster, M.D.

NIH staff: Sally Amero, Ph.D., Bruce Reed, Ph.D.

Contents

I. Introduction

The purpose of peer review is to judge the scientific and technical merit of grant applications submitted to the NIH. As currently implemented, review criteria are complex and numerous, so much so that it is difficult for reviewers to keep their focus on judging the science. This happened incrementally and with good intentions, but even so the current criteria are a barrier to obtaining the best review. Restoring a focus on science by simplifying review criteria is a priority of CSR.

In September 2019, the CSR Director raised the issue with CSR Advisory Council, and called for a working group, largely consisting of extramural scientists. The Simplifying Review Criteria Working Group (SRCWG, "WG1") was formed in December 2019. It was charged with recommending changes to RPG peer review criteria that will improve review outcomes and reduce reviewer burden. WG1 was asked to focus on criteria for non-clinical trial R01s, with the intention that modifications recommended for those criteria would inform subsequent efforts to reshape review criteria for other NIH grant mechanisms and for those involving clinical trials. Subsequent actions are outlined below:

- January-March 2020 Meetings and work of WG1
- March 30, 2020 CSRAC meeting. The WG1 interim report is presented. Council endorses the WG recommendations. That report recommended that a working group be formed to evaluate the criteria for clinical trials applications.
- May 2020 A WG to evaluate clinical trials criteria for R01s is formed. (Clinical Trials Criteria Working Group, CTCWG, "WG2").
- June-September 2020 Meetings and work of WG2
- September 29, 2020 Interim report to CSRAC.
- October 2020 – March 2021 Meetings and work of WG2

This report integrates the views and recommendations of both the SRCWG1 and the WG2.

II. Regulatory Context

The Code of Federal Regulations C.F.R 52h.7 states that no research grant award can be made unless the application has been reviewed by a peer review group that makes recommendations "concerning the scientific merit of that application". The CFR addresses the question of review criteria for grants in section 42 C.F.R. Part 52h.8:

"In carrying out its review under §52h.7, the scientific peer review group *shall assess the overall impact that the project could have on the research field involved*, taking into account, among other pertinent factors:

1. The *significance* of the goals of the proposed research, from a scientific or technical standpoint;
2. The a*dequacy of the approach and methodology* proposed to carry out the research;
3. The *innovativeness and originality* of the proposed research;
4. The *qualifications and experience of the principal investigator* and proposed staff;
5. The *scientific environment and reasonable availability of resources* necessary to the research;
6. Inclusion of women, minorities, and individuals across the lifespan
7. The adequacy of the proposed protection for humans, animals, and the environment, to the extent they may be adversely affected by the project proposed in the application.
8. Vertebrate Animals
9. Budget
10. Biohazards (this is NIH's interpretation/implementation of the regulatory requirement to evaluate environmental protections)

The following considerations flow from NIH policies. Peer review of these items is not required by statute. Thus, they could potentially be reviewed administratively, as appropriate:

a) Resubmissions
b) Renewals
c) Revisions
d) Applications from Foreign Organizations
e) Select Agent Research
f) Resource Sharing Plans
g) Data Sharing Plan
h) Sharing Model Organisms
i) Genomic Data Sharing Plan (GDS)
j) Authentication of Key Biological and/or Chemical Resources

III. Issues of Concern

1. The complexity and number of current criteria and additional review considerations unnecessarily burden reviewers and adversely affect the quality of peer review. This was the top issue identified for WG1 and was very important to WG2. In addition, the working groups considered several other important issues, listed below.
2. Misinterpretation/misapplication of the review criteria Significance and Approach. Multiple studies have shown that Significance and Approach are the strongest determinants of overall impact scores. Problems have been observed reviewers' understandings and use of both criteria.
    - Significance is too often evaluated with respect to the disease or problem the application proposes to study rather than the potential impact of the proposed science.
    - Reviewers are too often reluctant to say that the proposed study lacks significance. This reluctance makes it difficult for committees to judge the relative importance of

applications across neighboring scientific fields. It also encourages technical revisions of applications that have little prospect of scoring well.
    - Reviews of approach too often focus on technical minutia, overly weighting problems that skilled scientific teams could overcome, while neglecting important issues of rigor and reproducibility.
    - An over-emphasis on approach contributes to a risk-averse bias in study sections.
3. Confusion over Innovation. WG1 received many comments from reviewers who have noted confusion and inconsistencies in the application of the criterion Innovation.
    - There are different types of scientific innovation, for example innovation in design, technical innovation, and conceptual innovation. Reviewers interpret the criterion in different ways.
    - The importance of innovation is quite variable. Ground-breaking innovation may be a major strength of an application but studies that definitively answer an important question with established methods are also critical to scientific progress.
4. Concerns about the quality of science and need for improved stewardship of public funds were the basis for the current NIH policies on clinical trials.
    - Aspects of approach that specifically improve rigor and reproducibility, including sex as a biological variable, are too often inadequately evaluated.
    - The diversity (e.g., race, ethnicity) of subjects recruited into clinical trials is too often inadequately evaluated.
5. Persistent racial disparities in NIH funding raise the question of whether review criteria in any way perpetuate an unfair advantage or disadvantage.

IV. Practical limits on change

SRCWG sought clarification from the NIH Office of General Council (OGC) comments on the role of NIH in defining review criteria. The committee wanted to better understand what was required by statute and regulation versus what could be changed as matters of internal NIH policy. Listed below are some points from that discussion that guided both WG's. (This list should not be interpreted as a formal opinion from OGC regarding any specific recommendation of this Working Group).

1. NIH decides how to interpret the language of the criteria stated in the peer review regulation. Within reason, those definitions are up to NIH.
2. All matters of scoring are matters of NIH policy. Which criteria are scored, how they are scored, whether and how criteria should be weighted, whether they are all scored on the same scale or using the same system, as examples, are policy matters for NIH.
3. Superordinate factors based on existing criteria in the peer review regulation, and that NIH interpreted as falling within those criteria could be implemented as matters of interpretation and scoring.
4. The *peer review group* needs to consider all criteria.

V. Differences in NIH clinical trials peer review compared to non-clinical trials RPGs:

This is a summary of differences in R01 review criteria for clinical trial and non-clinical trial R01s. For additional detail and background, visit the OER pages on clinical trials. Definitions of each of the 5 core criteria (Significance, Innovation…) are modified by the addition of questions intended to specifically frame the criterion to enhance relevance to clinical trials applications. The modifications add substantial additional material to all criteria. Because clinical trial and non-clinical trial applications use different review criteria, NIH must issue multiple versions of funding opportunity announcements to cover clinical trial and non-clinical trial applications.

Applicants complete a study timeline document that contributes to the evaluation of an additional criterion "Timeline" for clinical trial applications.

Additional Human Subject and Clinical Trial Information forms are required and are subject to peer review. The forms collect a) title & registration, b) information on the study focus and inclusions, c) human subjects protections and safety monitoring, d) a protocol synopsis, including a detailed description, outcome measures, statistical power. In addition, other clinical trials documents are required including a recruitment and retention plan, inclusion enrollment report, data and safety monitoring plan, overall structure of the study team, statistical design and power, and dissemination plan. One set of forms is required for each distinct clinical trial proposed in the application. It is not unusual for an application to propose multiple studies that qualify as clinical trials. Thus, the additional material can be voluminous.

VI. Key ideas from the Clinical Trials Working Group Discussions

Below is a list of influential ideas that emerged from multiple WG meetings and online discussions. Each enjoyed broad support, but not necessarily perfect consensus.

- Review criteria have become incrementally more complex for all applications. This is especially so for clinical trials applications, which now require submission of substantial additional material for applicants to prepare and for reviewers to review. These complexities and additions were perceived as reducing the quality of peer review and contributing to reviewer burden.
- Criteria should ideally be broadly applicable to research project applications (R01s, R21s, R03s, etc.) across the full range of NIH science, including fundamental investigations, bioengineering, hypothesis-free explorations, clinical, translational, implementation and population studies.
- Criteria definitions should acknowledge the intelligence and adaptability of reviewers and should encourage high-level scientific judgment and critical evaluative thinking on the part of reviewers.
- Much of the additional clinical trial criteria-defining language is essentially duplicative of standard review criteria.
- The group identified several considerations as especially important for review of clinical trials applications.

    - Innovation should include reference to innovations in trial design.
    - Investigators should reference the need for clinical-trials specific expertise.
    - Environment should reference the institutional capabilities with respect to clinical trials.
    - The most important additional considerations for clinical trials, compared to non-clinical trials, concern approach. Feasibility is critical; well defined and justified endpoints and analysis plans, and a strong empirical foundation for the study should all influence review outcomes.
- The Human Subjects/Clinical Trials Information forms and additional documents that are required for each clinical trial proposed are generally not useful in peer review. Applicants complete the forms differently and they are often not thoroughly integrated into review meetings. Preparation is a major burden to applicants and reading the additional material—dozens, even more than 100 pages-- adds substantial burden to reviewers. The additional material is often not informative and rarely drives review outcomes.
- Rigor and reproducibility are critical to clinical trials, but not uniquely so. Rigor and reproducibility are important considerations in all research.
- Additional language is needed to direct reviewers' attention to feasibility and rigor considerations that are especially important for clinical trials applications. The WG considered, but rejected, creating different rigor language for different kinds of clinical trials, or different kinds of science. The critical considerations for a field always have some degree of field or method specificity and trying to spell these out and keep pace with them as they evolve was deemed futile. It was also seen as not likely to be helpful, and for this reason, the WG rejected a checklist approach, instead favoring uniform, conceptual level language that could guide reviewers in evaluating a wide range of research.
- It is both possible and desirable to use a single set of review criteria for clinical trials and non-clinical trials applications. The NIH definition of clinical trial spans an enormous range and variety of science, joined only by the common features of being experimental and involving human participants. In this context, there are no criteria or review considerations unique to clinical trials. There are criteria that apply to all clinical trials; these are basic questions including how important the proposed science is, and whether it is rigorous, and is feasible. These fundamental questions should be applied to all research project grant applications received by the NIH.

VII. Recommendations

Recommendation 1. Reorganize the five core review criteria into three factors, 1) Importance of the Science, 2) Feasibility and Rigor, 3) Investigator and Environment.

Review should be simplified by focusing reviewers' attention on what matters most in judging the scientific merit of an application. Arguably, there are three basic questions: Should it be

done? Can it be done well? Will it be done? The WGs propose that review criteria be restructured so that reviewers are guided in evaluating those three questions. Instead of scoring 5 criteria, reviewers would score 3 factors, each of which maps on to the basic question of merit:

Factor 1) Importance of the Science (Should it be done?)

Factor 2) Feasibility and Rigor (Can it be done well?)

Factor 3) Investigators and Environment (Will it be done?)

Factors 1 and 2 pertain only to the science that is proposed and are intended to capture judgements regarding conceptually distinct aspects of scientific merit. By titling Factor 2 "Feasibility and Rigor" the WG intended to highlight the critical importance of those two aspects of approach. The third factor, Investigators and Environment, is intended to capture judgments about how the investigators and environment shape the likelihood that the project will succeed, that the science will be implemented well, and the project will be productive.

Each factor score would derive from consideration of one or more of the current 5 required review criteria (Significance, Innovation, Investigator, Approach, Environment). Factor 1, Importance of the Science derives from the consideration of "Significance". Factor 2, Feasibility and Rigor derives from the criteria "Approach" and "Innovation"; Factor 3 derives from the criteria "Investigators" and "Environment". Thus, the factors are not new criteria, but rather new interpretations of the existing criteria required by CFR 57.h. In addition, reviewers would provide an overall Impact score that should reflect the reviewers' overall judgement of scientific merit based on an integration of the three factor scores. As with existing scoring methods, reviewers will rightfully weigh these factors separately and uniquely, depending on their view of score-driving strengths/weaknesses. As is currently the case, the final overall impact score would be used to prioritize applications.

Recommendation 2. Define each criterion and factor conceptually.

Review criteria can be simplified and strengthened by using conceptual definitions rather than lists of questions. A good definition can be applied across a wide range of science and methods. Extensive sets of specific questions or checklist criteria tend to encourage checklist thinking — quick yes-no's, rather than thoughtful consideration—or reviewing by counting pluses and minuses rather than by thoughtful integration. To get the best review, reviewers should be encouraged to make intelligent, informed high level scientific judgement of well-defined questions.

Recommendation 3. Recommendations regarding the criterion Innovation

NIH in general and review specifically is frequently criticized as being risk averse, as favoring sure and established ideas over potential high impact but unproven ideas. The WGs considered how review criteria might be modified to reduce this tendency. A common observation of WG members was that an over emphasis on minor weaknesses in approach, methodological

"weeds", technical/methodological minutia, too often hurts otherwise promising grant applications. The problem is amplified when committees pay too little attention to the potential importance of the proposed science. This was one of the driving concerns that led the groups to propose a major restructuring of review criteria, emphasizing the three big questions (Recommendation 1). The concern also drove Recommendation 9, to revise the review template to discourage reviewers from identifying methodological problems that skilled scientists are likely capable of overcoming. The WGs also gave considerable attention to the review criterion Innovation.

Innovation is a multidimensional concept. For example, innovation may pertain to the theoretical or conceptual frameworks, model systems, technical advances, computational advances, design and analytic innovations, etc. The relationship of innovation to overall scientific merit is complex and context-dependent. Innovation, perhaps conceptual, perhaps technical, other or both, is central to Significance and Impact. The promise of a scientific breakthrough is often appropriately judged to be highly significant whereas derivative, incremental ideas are judged less so. At the same time, very important studies such as critical phase III trials may lack innovation. In fact, their impact may derive from testing a narrow, highly defined, amply supported hypothesis using well-established (e.g. conventional) methods. The concept of innovation is also relevant to evaluating Approach, and Rigor/Feasibility. Innovations in technology, study design, computational methods, etc. may provide the foundation for making a study rigorous (by fixing previous problems), or feasible (by overcoming prior obstacles). Because it can be variably defined, is relevant to different questions, and is variably important, many reviewers struggle with the concept of Innovation, and the WG felt it important to provide a better framework. After extensive discussion the WG settled on the following.

a) <u>Incorporate the concept of Innovation in both Factors 1 and 2.</u>

The WG noted that Innovation is appropriately a driving consideration in scoring Factor 1 (Importance) and Factor 2 (Feasibility and Rigor). An innovative model of pathophysiology, for example, could promise new therapeutics (relevant to Factor 1), while an innovation in technology could make it feasible to test a hypothesis more rigorously than previously possible (relevant to Factor 2). The WG entertained a proposal to explicitly reference the criterion Innovation under both factors. This idea was ultimately rejected because it was likely to confuse reviewers, and because of concern that including Innovation twice, explicitly, might lead to an over-weighting of innovation. Instead, the WG proposes to explicitly reference Innovation under Factor 2, and to use language in defining Significance that captures the contributions of innovation to significance. Thus, the value of creativity, conceptual and technical advances would be captured in Factor 1, by asking reviewers to consider whether the grant would "create a valuable conceptual or technical advance". Innovation would be included as a criterion only for evaluating Factor 2 (Feasibility/Rigor) where it would be further defined.

b) <u>Incorporate types of innovation that are specifically relevant to interventional clinical trials.</u>

Phased clinical trials of drugs, devices and other interventions derive value from using standard methods and well-established approaches. Yet, there is room for innovation in some aspects of

these clinical trials. For example, innovations in trial design or recruitment approaches may improve the feasibility of a trial or improve its generalizability and thus its rigor. The WG incorporated these ideas into the definition of innovation.

c) Remind reviewers that innovation may appropriately carry different weights for different types of science.

The WG felt it was important to explicitly state that innovation can appropriately be given different weights in different scientific contexts. The goal was to avoid having otherwise highly meritorious scientific applications be inappropriately downgraded by lack of innovation, while not diminishing the general value of scientific innovation.

Recommendation 4. Recommendations for incorporating NIH policies on Rigor and Reproducibility into review criteria.

How to shape the evaluation of Factor 2- Feasibility and Rigor- to obtain high quality review of clinical trials was a major focus of discussions. The group noted that otherwise well-designed trials often fail because of feasibility problems, and that the importance of feasibility should be highlighted. A major consideration was how to promote good review of applications with respect to rigor and reproducibility. The WG noted that rigor and reproducibility are critical to clinical trials, but not uniquely so. Problems with clinical trials and preclinical science have been prominently documented, but it is abundantly clear that other types of science face the same challenges. Rigor and reproducibility are critically important in all research and should always be important in review.

The WG considered, but rejected, creating different rigor language for different kinds of clinical trials (e.g. BESH, mechanistic, treatment trials) and for different kinds of science. They noted that critical rigor considerations for a field always have a degree of field specificity and that standards evolve as methods and science itself evolves. They concluded that to spell these out and keep pace with them as they evolve by listing them out as particulars in review criteria is futile. A high degree of specificity was also seen as unlikely to be helpful. That is, checklists, while appropriate in some review contexts (administrative perhaps), were thought inappropriate for NIH peer review. Thus, the WG rejected a checklist approach, instead favoring uniform, conceptual level language that would guide reviewers in evaluating a wide range of research. The WG recommends the following:

a) Factor 1 should include language directing attention to the empirical foundation of the proposed science, the “rigor of the prior research”.

There was broad consensus that evaluation of the premise, or scientific foundation of the proposed work, needs to be assessed across the board for all science; it is not a consideration for clinical trials only. Wording should capture major clinical trials considerations, and general considerations applicable to all studies.

b) Highlight attention to rigor and reproducibility in Factor 2.

Obviously, one step to highlight attention to rigor was to include "rigor" in the name of the factor. In addition, the importance of rigor would be emphasized by giving it a distinct bullet in the definitional statement. In explaining what reviewers should consider, there will be direct language on rigor and reproducibility considerations, including sex as a biological variable (SABV). The WG identified the following as considerations that are critical for clinical trials: appropriateness/rigor of the control group, appropriateness/rigor of methods for recruitment and retention (feasibility), appropriateness/rigor of the study population (including representation by sex/gender, race and ethnicity), rationale for the estimated sample size. Because these considerations apply broadly, the WG thought they should be listed even though they are not relevant to all areas of science.

c) Do not require additional pages of material from applicants.
The WG entertained a proposal to add an additional page to the existing 12, one dedicated to rigor. However, the panel did not support this idea. There were concerns about increasing the burden for applicants and reviewers. It was also noted that a well written approach section of the R01 application incorporates rigor/reproducibility considerations as it is.

Recommendation 5. Recommendations regarding issues of bias and diversity in the scientific workforce.

Multiple members expressed strong concern about persistent racial and ethnic funding disparities at NIH, about the severe underrepresentation of some minority groups in science, particularly Black Americans, and voiced strong support for urgent action. Multiple members also expressed the view that peer review is about evaluating scientific merit, and voiced concern that using peer review to try to accomplish other goals, however worthwhile, would detract from that. The role of peer review at the NIH is to evaluate the scientific and technical merit of applications. That is a unique and valuable role. Thus, the WG felt that efforts to modify peer review criteria to address equity issues need to be framed in the context of improving the evaluation of scientific merit.

a) <u>Modify the criterion definitions for Investigator and Environment to reduce positive bias</u>. Bias distorts judgments of merit; conversely, stronger judgments of merit reduce the impact of bias. While instances of overt bias against women or minorities are rare, reviewers too often introduce an application with generalities about the scientific reputation of the PI and/or their institution. This sort of reputational bias, connected to scientific networks and pedigree is common and tends to favor the most senior PIs, who are disproportionately White and male. Thus, to reduce bias, the WG proposed language under Investigators and under Environment intended to diminish halo effects—inappropriate, positive bias that results from non-specific consideration of investigator or institutional reputations.
b) It was noted that there have been proposals to include evaluation of diversity within the research team as part of the Investigators criterion. While the value of demographic diversity on research teams was acknowledged, a variety of concerns were expressed about incorporating team diversity as a review criterion. Points in discussion included

that while team diversity is always desirable, it may not be directly connected to the merit of the proposed science. Institutions are in a much better position to promote diversity in science through training, hiring and retention programs. Review may be an ineffective point of action. Awards like the NCI K award for junior minority faculty help train more scientists in a strong manner and give extensive opportunities to minorities. It was noted that NIH could create additional awards like this to develop and support mentorship and diversity within the scientific workforce.

c) It was pointed out that when the number of URM investigators is as small as it is now any quasi-requirement for team diversity may have the perverse effect of increasing burden and barriers for URM scientists. The "diversity tax" on URM scientists is well recognized, and building a diversity criterion into review criteria could result in URM scientists being asked to serve on numerous grants and scientific organizations in order to add diversity to those teams, to the point that their own science might suffer.

Recommendation 6. Add clinical trials considerations under Investigators and Environment

This recommendation flowed from the goal of improving review of clinical trials applications, and the specific point that clinical trials failures are often failures of feasibility based in investigator inexperience. The knowledge and skills of the investigators and the resources of the institution where the work would be conducted were identified as critical determinants of clinical trials success. Therefore, special mention of clinical trials was recommended for the definitions for the criteria under Factor 3.

Recommendation 7. Use the same set of factor and criterion definitions for all clinical trials.

The Working Group clearly recognized the wide range of clinical trials supported by NIH, from BESH to phase III interventional clinical trials. The CTWG discussed whether a single set of criteria could serve all clinical trials or whether separate or additional criteria should be developed for therapeutic/interventional clinical trials. A variety of approaches were considered.

An influential idea was "less is more". All members believe that it is best to keep reviewers focused on fundamental ideas for evaluating applications and let them apply those ideas specifically, in an informed, intelligent way as appropriate to the science. This steered the group away from separate criteria sets for different types of clinical trials. It also steered them away from checklist approaches.

Recommendation 8. Drop the additional clinical trials criterion "Timeline".

This criterion was perceived as not adding anything to the quality of the review. It is often an afterthought of reviewers and frequently is not critically evaluated. The WG notes that "timeline" should not be confused with "Milestone plans" which can be very appropriate but might better be developed for applications likely to be funded, working with the funding institute/center.

The WG acknowledged that "Timeline" may have been an attempt to have reviewers evaluate the feasibility of recruitment and retention targets. The panel noted that feasibility is a critical aspect of the evaluation of clinical trials, as many trials fail because of predictable problems with feasibility. The WG proposed to emphasize the importance of clinical trials feasibility in Factor 2 "Feasibility and Rigor", and by including additional references to recruitment and retention in the definitions.

Recommendation 9. Alter templates to focus reviewer attention on score-driving factors.

The current template asks reviewers to list bulleted strengths and weaknesses. The observation of the WG is that the perceived necessity of listing weaknesses may contribute to an unnecessary focus on minor technical aspects of the proposal. The WG sought to encourage reviewers to explain the salient points that drive their score. They developed language to encourage reviewers to think about major strengths, major weaknesses, and to not focus on minor problems that competent scientific teams are likely to overcome.

The recommendation is to remove headers for "Strengths" and "Weaknesses" below each scored factor, and instead provide headers for "Major Score-Driving Factors" and "Minor Points (optional)".

Rather than specify a bulleted or narrative format, reviewers would be instructed as follows:

For factors 1, 2 and 3: "Using sentences or short narratives, explain the points that determine your score, clearly and concisely. Identify and weigh the most important strengths and weaknesses of the application with respect to [factor X]."

For Overall Impact: "Write a clear, concise paragraph that explains the basis for your score. Identify and weigh the most important strengths and weaknesses of the application."

Recommendation 10. Simplify reviewer responsibility for evaluating the budget.

A detailed, line-item level analysis of the proposed budget is unreasonable to expect of reviewers. To accomplish it well would require more time, attention and information than they can give in the context of evaluating the scientific merit of the proposals. Based on their scientific evaluation of the proposal reviewers will have a good sense of the scope of work required. Based on their experience as practicing scientists most reviewers will have a general sense of the resources, time and money required to accomplish the work. It is reasonable to ask reviewers to weigh the scope of the work against the resources described in the budget. Thus, reviewers should only be asked to choose between these options:

> Budget is appropriate to support the scientific activities proposed.
>
> Budget appears excessive. Further justification is needed.
>
> Budget appears inadequate and raises concerns about project feasibility.

Recommendation 11. Relieve peer review burden by not requiring peer review of select "additional considerations".

The following considerations were identified as suitable for administrative review. Each is an important issue for the NIH and deserves careful attention if projects are to be funded. In peer review the focus should be on scientific merit. Evaluating these topics adds a half dozen additional questions that are often, not always, tangential to scientific merit. They thus add additional burden to review—unless it is treated trivially—and divide the reviewers' focus. The WG acknowledges that there are times when these issues do bear on merit; for example, inadequately validated key resources can undermine a project. However, they also note that applicants always need to demonstrate methodological rigor in the body of the application. Administrative review would give the ICs the opportunity to require uniform information and to impose uniform agency standards as a condition of award.

1. Biohazards
2. Foreign components
3. Select Agent Research
4. Resource Sharing Plans
5. Authentication of Key Biological and/or Chemical Resources

Recommendation 12. NIH should drop the requirement of an additional Human Subjects/Clinical Trial information form for clinical trial studies.

The WG charge was to reduce reviewer burden and improve the quality of review. There was strong consensus on the Working Group that the current HS/CT forms increase burden and do not improve peer review. Because an additional set of forms is required for each study that qualifies as a clinical trial, and because some form fields require extensive information, these forms can add dozens of pages to the basic 12-page application. Panelists noted that applicants complete the forms differently, which makes it difficult for reviewers to know how to approach the additional information. Sometimes it is entirely redundant with the 12-page application, while other times it includes substantial new information. This creates both confusion and frustration.

It was the consensus of the Working Group that the forms rarely add value from the perspective of evaluating the application for scientific merit. Members voiced concerns that reviewers currently treat the HS/CT information quite variably. Some evaluate it carefully, while others neglect it. Because the HS/CT forms create confusion, frustration, and high burden without adding clear value, a change is needed. There was strong consensus that the HS/CT forms should be dropped in their entirety, and that the critically relevant material instead be presented within

the 12- page research narrative. There was discussion about whether the 12-page limit was too short for certain clinical trials, but the panel came to a consensus that it was not.

The panel acknowledged that the CT/HS forms are used to collect information required for registration on clinicaltrials.gov in a standardized way. However, they argued that registration requirements could be met outside of the peer-review process. The consensus of the group was that peer review should not be saddled with materials deriving from clinical trials registration requirements. Materials needed to satisfy requirements for the registration of clinical trials should be outside of peer review.

Recommendation 13. Use a single set of criteria for all R01s, clinical trials and non-clinical trials applications alike.

The WG believes that by directing reviewers to evaluate a small number of basic questions and by defining the considerations that should shape those judgments conceptually it is possible to create criteria that support high quality review across the entire range of applications supported by the NIH. Acknowledging that there are always considerations particular to different levels of science, populations, techniques, and designs, the WG believes that the scientists who are expert in those areas are best suited to define the specific implementation of general review principles.

Recommendation 14. Proposed simplified criteria for the review of NIH R01 applications (attachment 1).

Recommendation 15. Extend the simplified criteria, on a modified basis, to all NIH RPGs, Training, and Career Development Grants.

There are obvious benefits to NIH using a unified, consistent set of review criteria across all RPGs. While acknowledging that differences in the scale, purpose, and structure of different award mechanisms would necessitate some modifications of the proposed R01 criteria, the WG recommends the proposed simplified criteria for R01s serve as the template for a unified RPG criterion set for NIH. An important next step will be to consider how to modify these criteria to serve the specific needs of other RPGs, R21s, R03, R15s SBIR/STTR grants, multi-component awards, etc. The emphasis on Innovation might be different in R21s, for example. Small business applications (SBIR/STTR) require attention to considerations such as commercialization that are not generally relevant to other RPGs, and so need special attention. The WG recognizes that training and career development grants are reviewed according to a different set of core criteria than are RPGs. It will take the attention of an additional group to apply the principles of this report to those criteria.

Recommendation 16. Implementation.

The WG acknowledges that implementation of these criteria would require a major effort from NIH and feels that the benefit warrants the effort. Implementation would require changes in FOAs, in FOAM, and in eRA. SROs would need to be trained and their training of reviewers would be key. Training of committee chairs would be vital. In addition, NIH would need to make multiple and repeated efforts to directly reach the scientific community through multiple channels including blogs, conference presentations, webinars etc. Training and review materials, including new review templates and positive examples of good reviews, would be needed. Websites would need to be scrubbed of old materials and refreshed with new.

Attachment 1. Proposed criteria for R01 grant applications

FACTOR 1. IMPORTANCE OF THE SCIENCE

Significance: Assess how important it is to accomplish the proposed science. Try to evaluate the importance of this application not simply with respect to other very similar applications, but rather in the broad context of current scientific challenges and opportunities.

Judge the scientific value of the knowledge likely to be gained through the proposed work. Consider the importance of new facts it may establish, the value of the methods, models, and concepts that it may create, develop, or enable and how these may shape future science. NIH supports highly significant work across the scientific spectrum including basic biology, behavior, bioengineering, physiology, pathophysiology, disease and its treatment or prevention. For clinical trials that aim to evaluate an intervention to improve health, evaluate whether the study makes a scientific advance that may ultimately modify disease biology, health, or therapeutic outcomes. Evaluate the rationale for undertaking the study. An empirical foundation is generally important and is critical for clinical trials. Evaluate the rigor of the scientific background.

Significance of the science must be distinguished from the significance of the disease or general scientific challenge that frames it. Choosing to study a significant problem does not necessarily make a proposal significant. Rather, judge whether the application addresses an important gap in knowledge, would solve a critical problem, or create a valuable conceptual or technical advance. Significance of the scientific knowledge to be gained from the proposed research is what matters.

FACTOR 2. FEASIBILITY AND RIGOR OF THE METHODS

Assess the feasibility and rigor of the study. Evaluate the scientific quality of the approach, judging the likelihood that compelling, reproducible findings will result (rigor). Assess whether the proposed studies can be done well and within the timeframes proposed (feasibility). Base your judgment of Feasibility and Rigor on your evaluation of the application's Approach and Innovation. Projects need not be strong on both to justify a strong score.

Approach:

- Evaluate how well the proposal demonstrates technical competence and feasibility, and whether it demonstrates the capacity to adjust methods appropriately to address problems and new challenges that emerge in the work.

- Evaluate rigor and reproducibility. Assess whether the approach will produce unbiased, robust data, whether the design is well-controlled, and if the plans for analysis, interpretation and reporting of results is technically appropriate and scientifically strong. Judge whether the sample size is sufficient and well-justified. Evaluate whether biological sex is appropriately considered in the design, analysis, and reporting. Additional considerations, especially for clinical trials, are the rigor of the intervention or study manipulation; the appropriateness of control or comparison group(s); whether outcome variables are justified; whether results will be generalizable; whether the sample will contain sufficient demographic diversity to address the proposed question(s), including adequate representation by race/ethnicity; whether implementation is feasible (for example, can recruitment, retention, and timeline goals be met?).

- Focus on major issues that strengthen or detract from the feasibility and quality of the work. Less important technical flaws or omissions, especially those that a capable team of scientists could correct, should carry less weight.

Innovation: Evaluate whether innovations in the grant application contribute to or detract from the feasibility and rigor of the methods. Consider technical, methodological, experimental, and trial design innovations. Innovations may enable more precise observations, make studies more efficient, more rigorous, and overcome existing scientific limitations. However, poorly understood or difficult to implement innovations may detract from rigor and feasibility.

Innovation is often critical and is generally desired, yet studies can be highly meritorious without being highly innovative.

FACTOR 3. INVESTIGATORS AND ENVIRONMENT

Evaluate the application's Investigators and Environment from the perspective of what they contribute to the likelihood that the project will be executed well, the aims met, that the project will be rigorous, productive, and that scientifically valuable outcomes will result. Rather than general reputation, consider the strengths or weaknesses of the investigators and environment with respect to the specific science proposed.

Investigators: Evaluate the scientific background, expertise, skills of the PI and team of investigators with respect to the proposed science. Assess their intellectual and scientific capabilities, using evidence. Judge qualifications, not reputation. Strong investigators will creatively overcome obstacles and flexibly adapt to challenges thus improving the likelihood that the proposed project will be accomplished and will produce important new scientific and health knowledge, tools, or resources. **Evaluate investigators in the context of their career stage**; different indicators and standards are appropriate for early, mid-career, and senior investigators.

Environment: Evaluate the extent to which the scientific environment, institutional support and capabilities, equipment, facilities, the community setting, and other resources available to the investigators will contribute to successful execution of the proposed project. For clinical trials, consider if the capacity exists at the site(s) to conduct clinical trial research of the type proposed.

4. OVERALL IMPACT SCORE

Judge the overall scientific and technical merit of the application. Considering the importance of the science, the feasibility and rigor of the proposed approach, and the capabilities of the scientists involved, assess the likely contribution of the project to advancing fundamental knowledge about the nature and behavior of living systems, or the application of that knowledge to enhancing human health. Write a clear, concise paragraph that explains the basis for your score. Identify and weigh the most important strengths and weaknesses of the application.

Attachment 2. Proposed Scoring Table

Factor 1. Importance of the Science: Significance	1-9
Factor 2. Feasibility and Rigor of the Methods	1-9
Factor 3. Investigators and Environment	1-9
Overall Impact Score	1-9

Attachment 3. Rosters

CSR Advisory Council Review Criteria Working Group Roster (WG1)

CO-CHAIRPERSON(S)

PALERMO, TONYA M., PHD
PROFESSOR AND ASSOCIATE DIRECTOR
CENTER FOR CHILD HEALTH, BEHAVIOR AND DEVELOPMENT
SEATTLE CHILDREN'S RESEARCH INSTITUTE
SEATTLE, WA 98145

REED, BRUCE, PHD
DEPUTY DIRECTOR
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

MEMBERS

GAO, JINMING, PHD
PROFESSOR OF ONCOLOGY, PHARMACOLOGY, AND OTOLARYNGOLOGY
SIMMONS COMPREHENSIVE CANCER CENTER
DEPRARTMENT OF PHARMACOLOGY
UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER
DALLAS, TX 75390

GEORGE, ALFRED L. JR, MD
MAGERSTADT PROFESSOR AND CHAIR
DEPARTMENT OF PHARMACOLOGY
DIRECTOR, CENTER FOR PHARMACOGENOMICS
FEINBERG SCHOOL OF MEDICINE
NORTHWESTERN UNIVERSITY
CHICAGO, IL 60611

HURD, YASMIN L., PHD
PROFESSOR
DEPARTMENTS OF PSYCHIATRY, NEUROECIENCE AND PHARMACOLOGICAL SCIENCES
MOUNT SINAI SCHOOL OF MEDICINE
NEW YORK, NY 10029

KROETZ, DEANNA L., PHD
PROFESSOR

AD-HOC

CORBETT, KEVIN, PHD,
ASSOCIATE PROFESSOR
DEPEARTMENT OF CELLULAR AND MOLECULAR MEDICINE
UNIVERSITY OF CALIFORNIA, SAN DIEGO SCHOOL OF MEDICINE
LA JOLLA, CA 92093

JANELSINS, MICHELLE C, PHD
ASSOCIATE PROFESSOR
DIVISION OF SUPPORTIVE CARE IN CANCER
DEPARTMENT OF SURGERY
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14642

KING-CASAS, BROOKS, PHD
ASSOCIATE PROFESSOR
VIRGINIA TECH CARILION RESEARCH INSTITUTE
PSYCHOLOGY, COLLEGE OF SCIENCE
PSYCHIATRY, VT CARILION SCHOOL OF MEDICINE
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIV
BLACKSBURG, VA 24061

NIH STAFF

AMERO, SALLY ANN, PHD, BS
REVIEW POLICY OFFICER
DIVISION OF SCIENTIFIC PROGRAMS
OFFICE OF EXTRAMURAL RESEARCH
NATIONAL INSITUTES OF HEALTH
BETHESDA, MD 20814

EXTRAMURAL SUPPORT ASSISTANT

JAIN, ADITI
REVIEW ANALYST
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH

DEPARTMENT OF BIOENGINEERING AND THERAPEUTIC SCIENCES
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
SAN FRANCISCO, CA 94143

LÓPEZ, JOSÉ A., MD
PROFESSOR
DIVISION OF HEMATOLOGY
UNIVERISTY OF WASHINGTON SCHOOL OF MEDICINE
BLOODWORDS NORTHWEST RESEARCH INSTITUTE
SEATTLE, WA 98102

PALERMO, TONYA M., PHD
PROFESSOR AND ASSOCIATE DIRECTOR
CENTER FOR CHILD HEALTH, BEHAVIOR AND DEVELOPMENT
SEATTLE CHILDREN'S RESEARCH INSTITUTE
SEATTLE, WA 98145

CSR Advisory Council Clinical Trials Criteria Working Group Roster (WG2)

CO-CHAIRPERSON(S)

PALERMO, TONYA M., PHD
PROFESSOR AND ASSOCIATE DIRECTOR
CENTER FOR CHILD HEALTH, BEHAVIOR AND DEVELOPMENT
SEATTLE CHILDREN'S RESEARCH INSTITUTE
SEATTLE, WA 98145

REED, BRUCE, PHD
DEPUTY DIRECTOR
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

MEMBERS

GEORGE, ALFRED L. JR, MD
MAGERSTADT PROFESSOR AND CHAIR
DEPARTMENT OF PHARMACOLOGY
DIRECTOR, CENTER FOR PHARMACOGENOMICS
FEINBERG SCHOOL OF MEDICINE
NORTHWESTERN UNIVERSITY
CHICAGO, IL 60611

KING-CASAS, BROOKS, PHD
ASSOCIATE PROFESSOR
VIRGINIA TECH CARILION RESEARCH INSTITUTE
PSYCHOLOGY, COLLEGE OF SCIENCE
PSYCHIATRY, VT CARILION SCHOOL OF MEDICINE
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIV
BLACKSBURG, VA 24061

MUNSTER, PAMELA N., MD
PROFESSOR
DEPARTMENT OF MEDICINE
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
SAN FRANCISCO, CA 94143

NIH STAFF

AMERO, SALLY ANN, PHD, BS
REVIEW POLICY OFFICER
DIVISION OF SCIENTIFIC PROGRAMS
OFFICE OF EXTRAMURAL RESEARCH
NATIONAL INSITUTES OF HEALTH
BETHESDA, MD 20814

HURD, YASMIN L., PHD
PROFESSOR
DEPARTMENTS OF PSYCHIATRY, NEUROECIENCE AND PHARMACOLOGICAL SCIENCES
MOUNT SINAI SCHOOL OF MEDICINE
NEW YORK, NY 10029

PALERMO, TONYA M., PHD
PROFESSOR AND ASSOCIATE DIRECTOR
CENTER FOR CHILD HEALTH, BEHAVIOR AND DEVELOPMENT
SEATTLE CHILDREN'S RESEARCH INSTITUTE
SEATTLE, WA 98145

AD-HOC

BOYD, BRIAN, Ph.D.
ASSOCIATE PROFESSOR
DEPARTMENT OF APPLIED BEHAVIORAL SCIENCE
UNIVERSITY OF KANSAS
KANSAS CITY, KS 66101

CARPENTER, MATTHEW J, PHD
PROFESSOR
DEPARTMENT OF PSYCHIATRY AND BEHAVIORAL SCIENCES
MEDICAL UNIVERSITY OF SOUTH CAROLINA
CHARLESTON, SC 29425

JANELSINS, MICHELLE C, PHD
ASSOCIATE PROFESSOR
DIVISION OF SUPPORTIVE CARE IN CANCER
DEPARTMENT OF SURGERY
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14642

EXTRAMURAL SUPPORT ASSISTANT

JAIN, ADITI
REVIEW ANALYST
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892