Report of the 2020 Committee of Visitors
Division of Mathematical Sciences
National Science Foundation
September 14-17, 2020

COV Members:

Anna Amirdjanova	Federal Reserve
Kelly Bickel	Bucknell University
Pavel Bochev	Sandia National Laboratories
Tara Brendle	University of Glasgow
Russel Caflisch, COV Co-Chair	New York University
John T. Chen	Bowling Green State University
Jessica M. Conway	Pennsylvania State University
Oliver Dasbach	Louisiana State University
Clint Dawson	University of Texas, Austin
Jesus De Loera	University of California, Davis
Sara Y. Del Valle	Los Alamos National Laboratory
Ailana Fraser	University of British Columbia
Wilfrid Gangbo	University of California, Los Angeles
Anette Peko Hosoi	Massachusetts Institute of Technology
Christopher K.R.T. Jones	University of North Carolina, Chapel Hill
Sabine Le Borne	Hamburg University of Technology
Steven L. Lee	DOE Advanced Scientific Computing Research
Po Shen Loh	Carnegie Mellon University
Bernard Mair	Association of Public & Land-Grant Universities
David Manderscheid	University of Tennessee, Knoxville
Mary Meyer	Colorado State University
Jennifer L. Morse	University of Virginia
Deborah Nolan	University of California, Berkeley
Sean T. Paul	University of Wisconsin
Jill Pipher	Brown University
Alexander Polishchuk	University of Oregon
Raquel Prado	University of California, Santa Cruz
Adriana Salerno	Bates College
Deborah Sulsky	University of New Mexico
Simon Thomas	Rutgers University
Tatiana Toro, COV Co-Chair	University of Washington
Rodolfo H Torres	University of California, Riverside
Anthony Várilly-Alvarado	Rice University
Nathaniel Whitaker	University Massachusetts Amherst
Carol S. Wood	Wesleyan University
Tommy Wright	Census Bureau

Executive Summary

The 2020 Division of Mathematical Sciences (DMS) Committee of Visitors (COV) was charged with evaluating and reporting on:

- the integrity and efficacy of processes used to solicit, review, recommend, and document proposal actions;
- the quality and significance of the results of the Division's programmatic investments;
- the relationship between award decisions, program goals, and Foundation-wide programs and strategic goals;
- the Division portfolio's balance, priorities, and future directions;
- the Division's response to the prior COV report of 2016;
- the Division's efforts in broadening participation and application of the Broader Impacts criteria in proposal actions.

This review was for the period of the four Fiscal Years 2016-2019. Because of the ongoing COVID pandemic, all work and meetings were done online. Within the limits of COI constraints, the members of the COV had access to documentation for all DMS review panels from the period, various reports that DMS prepared of their own initiative or at the request of the COV, and relevant publicly available information, such as the reports of previous DMS COVs. In addition, the COV met with leadership of DMS and with many DMS Program Officers (POs). The COV is grateful to DMS personnel for their significant efforts to provide whatever was needed by the Committee, their transparency in answering any questions, and their tremendous dedication to the important role that they play in the mathematical sciences community.

As described in more detail in both the Overview and the Subcommittee Reports that make up this report, the COV found that DMS is fulfilling its mission (described more fully below) of supporting research, training, conferences and mathematical sciences research institutes. It manages a large proposal process with over 3000 proposals per year; its funding leads to outstanding results and supports some of the most important priorities of the NSF; and it promotes diversity and broader participation of underrepresented groups.

DMS also faces some daunting challenges: It is unable to fund many excellent proposals; despite its significant efforts, the participation of women and underrepresented minorities (URMs) is far too low; and it is difficult to recruit members for the review panels.

The COV makes a set of recommendations to address some of these challenges:

1. Encourage virtual participation in future panels.
2. Improve the use of the Broader Impacts criterion.
3. Improve the quality and consistency of reviews and panel summaries.

4. Make efforts to increase the response rate to requests for demographic information.
5. Consider implementing a systematic process for handling resubmitted proposals.
6. Continue to make significant steps to increase diversity at all levels.

More detail on these recommendations are provided in the Overview, and there are many additional suggestions that are included in the subcommittee reports.

In summary, DMS is a superb organization, and its work is vital for the mathematical sciences. It leads the way for mathematicians and statisticians to engage with some of the most important challenges of our time. Its operations are at an excellent level overall, and the 2020 COV presents a set of recommendations in a few significant areas, especially for diversity.

Structure of the COV Review and Report

The Committee of Visitors was organized into 6 subcommittees, and this report reflects that organization. After an Overview which includes a set of overall recommendations, reports from each of the subcommittees are presented, each of which includes the roster of its members.

Subcommittees A, B and C were assigned to disciplinary programs of DMS, and Subcommittees D, E and F were assigned to special programs. Each member of the COV served on two subcommittees: one from the disciplinary programs (A, B, C) and one from the special programs (D, E, F). Each subcommittee had a Lead and a Co-Lead, who organized the work of the subcommittee.

Each of the subcommittee reports is organized around a set of questions that was posed by DMS in the COV Report Template. In addition to the overall recommendations that are presented in the Overview, each of the subcommittee reports includes a number of suggestions from the subcommittee.

Co-Chairs Caflisch and Toro had COIs with the mathematical research institutes, which were included in the purview of Subcommittee F. Accordingly, they did not participate in any of Subcommittee F's activities that were related to the institutes, and in particular they were not involved in writing its report. At the end of the Overview section, there is a subsection on an overview and recommendations of Subcommittee F, which was written by its Lead, without involvement of Caflisch and Toro.

Overview

Russel Caflisch, COV Co-Chair — New York University
Tatiana Toro, COV Co-Chair — University of Washington

Introduction

The DMS Committee of Visitors met under extraordinary circumstances during the third week of September 2020 to conclude the work initiated a few weeks earlier. Due to the COVID crisis all interactions took place virtually. The COV was charged to address and prepare a report on:

- the integrity and efficacy of processes used to solicit, review, recommend, and document proposal actions;
- the quality and significance of the results of the Division's programmatic investments;
- the relationship between award decisions, program goals, and Foundation-wide programs and strategic goals;
- the Division portfolio's balance, priorities, and future directions;
- the Division's response to the prior COV report of 2016;
- the Division's efforts in broadening participation and application of the Broader Impacts criteria in proposal actions.

This evaluation was done through the light of DMS's mission:

DMS supports research in mathematics and statistics, training through research involvement of the next generation of mathematical scientists, conferences and workshops, and a portfolio of national mathematical sciences research institutes.

DMS strengths stem from the combined skills, knowledge and experience of all its stakeholders. The administrative staff and the staff in the office of the Division Director (DD), Program Officers (PO) and Principal Investigators (PI) work together to make DMS a very successful branch of NSF. Some of the most important developments in the mathematical sciences over the past few years have been partially funded by NSF via DMS. DMS receives a huge number of excellent proposals. They are submitted to disciplinary programs, special research programs, infrastructure & workforce programs and research institutes. A number of them are submitted in response to solicitations addressing NSF's 10 Big Ideas. DMS has traditionally been the main source of funding for fundamental mathematics projects while responding to new and emerging trends through innovative programs, despite the fact that their budget has been flat for the last 14 years. This support has greatly contributed to making the US into an international leader in the mathematical sciences.

The overall quality of the proposals in most programs is outstanding and seems to improve over the years. The process is so competitive that it forces PIs, both junior and senior, to develop better and better proposals. The amount of time devoted to this by

PIs should not be underestimated and it is one of the reasons it is so frustrating for PIs to have their proposals declined.

Unfortunately, many outstanding proposals are not funded or are inadequately funded. The success rate oscillates around 30% for the disciplinary programs, where less than 7% of the proposals are submitted by members of under-represented minorities. Flagship programs such as FRG, CAREER and MSPRF were discussed at length in the context of the lack of funding. CAREER grants are very prestigious, they target future leaders in their research area, unfortunately many excellent proposals are not funded. Postdoctoral associates (including those in the MSPRF program) comprise only 6% of the individuals supported by DMS. Under the current COVID crisis the need of postdoctoral support is likely to increase.

The COV commends the POs for their handling of a heavy workload and the efficient overall processing of reviews and awards under time and budget constraints. The management is excellent, with dedicated and talented program officers. The COV also would like to highlight the efforts by the current DD to create a balanced roster of POs. Today 13 out of 28 POs are women. This represents an increase of more than 200% since his arrival in 2018. A quick look at the composition of this COV indicates that DMS is sensitive to the societal shift taking place in the country.

Highlights

PIs supported by DMS received a large number of very prestigious awards and were recognized in different ways, during the time period under review. We will highlight those that stand out with a focus on the junior scientists as they constitute the future of the field.

CAREER holders (FY16-19) have made remarkable contributions both from the scientific point of view as well as in terms of workforce development. According to the information reviewed by the COV, they were recognized with Sloan Fellowships (8), Presidential Early Career Awards for Scientists and Engineers (2), the Doeblin Prize (1), the MacArthur Fellowship (1), the Loeve Prize (1), the New Horizons Prize in Mathematics (3), and the Sadosky Prize (1). Some CAREER awards also have an extraordinary educational component.

We would like to highlight CAREER grant DMS-1554130. The PI Ryan Hynd works closely with educators at an inner-city high school in Philadelphia to introduce mathematical modeling into their curriculum and train teams of their students to compete in a national applied mathematics competition. He aims to give these students a broader appreciation for mathematics and to help prepare them for higher education. He started a Bridge-to-Ph.D. program at the University of Pennsylvania to increase the number of women and members of underrepresented groups earning Ph.D. degrees in mathematics. The program offers two years of fully funded study, culminating in a master's degree and the possibility of continuing on to the Ph.D. program. Each student is paired with a faculty mentor who helps select classes and seminars. In pursuing this

project, Prof. Hind aspires to create a template for other mathematics departments that seek greater diversity at the Ph.D. level.

During the same time period eleven PIs were elected to the National Academy of Sciences, two to the National Academy of Medicine and one to the National Academy of Engineering; three were awarded Breakthrough prizes.

PIs also contributed to enhance the public image of mathematics. Jordan Ellenberg was awarded the 2016 Euler Book Prize for his book *How Not to Be Wrong: The Power of Mathematical Thinking*, Penguin Press, 2014. The Prize is intended to recognize authors of exceptionally well written books with a positive impact on the public's view of mathematics. On June 28, 2017, David Donoho, gave a congressional briefing entitled *Blackboard to bedside: How high-dimensional geometry is transforming the MRI industry* for Members of Congress and staff on Capitol Hill. Donoho explained how federally funded mathematical research transitioned in just 10 years from 'brainiac' math journals to FDA-approved medical devices. His Stanford patents on compressed sensing are licensed by both GE and Siemens in their new generation FDA-approved scanners. The improved technology will save lives, reach new demographic groups, and increase productivity in the use of healthcare resources.

An excellent illustration of how fundamental mathematics can have an impact in other areas is the work of PI Lai-Sang Young. She has proposed to work on increasing the number of fields and examples where dynamical systems theory is applied. In particular, she uses tools from dynamical systems to study problems coming from the biological sciences. In collaboration with her colleagues Robert Shapley and Logan Chariker, they have constructed a new mathematical model to explain how the brain processes visual information; this work was the subject of a recent Quanta Magazine article by Kevin Harnett titled, *A Mathematical Model Unlocks the Secrets of Vision*. In studying mathematical models to prevent an incipient infection from developing into an epidemic, she (together with T. Pereira) used (delay) differential equations to model the disease process; these equations required analysis of an infinite dimensional dynamical system. With her former PhD student, Alex Blumenthal, Young extended the theory of hyperbolic dynamical systems from finite dimensions to infinite dimensions. Young was recently named Distinguished Visiting Professor at the Institute for Advanced Study in Princeton and was elected as a member of the National Academy of Sciences in 2020.

The Role of DMS within the NSF and the Broader Community

DMS has been actively engaging with other divisions and directorates within NSF, as well as with various government and private entities outside of NSF. Within NSF, these interactions can best be viewed through the lens of the Big Ideas. These were first announced in 2016, so that they have influenced DMS during the period covered by this review, even if the first direct funding for topics from Big Ideas only started in 2019.

DMS responded mainly to three of the Big Ideas. The Transdisciplinary Research in Principles of Data Science (TRIPODS), Partnerships between Science and Engineering Fields and the NSF TRIPODS Institutes (TRIPODS+X), and the Computational and Data-Enabled Science and Engineering in Mathematical and Statistical Sciences (CDS&E-MSS) programs, which were joint between DMS, the CISE Directorate and other directorates, are a direct example of the “Data Revolution”; the Designing Materials to Revolutionize and Engineer our Future (DMREF) program, which was joint between DMS and DMR, is largely focused on parts of “Quantum”; and “Rules of Life” is addressed by the NSF-Simons Research Centers for Mathematics of Complex Biological Systems (MathBioSys) program, which is jointly funded by DMS and the Simons Foundation.

Outside of NSF, DMS has worked with a number of federal agencies. These activities included the Algorithms for Threat Detection (ATD) program which was joint between DMS and the National Geospatial Intelligence Agency, the Algorithms for Modern Power Systems (AMPS) program which was joint between DMS and the Department of Energy, and the Joint DMS/NIGMS Initiative to Support Research at the Interface of the Biological and Mathematical Sciences (DMS/NIGMS), Joint NSF/NIH Initiative on Quantitative Approaches to Biomedical Big Data (QuBDD), and Joint DMS/NLM Initiative on Generalizable Data Science Methods for Biomedical Research (DMS/NLM) programs which were joint between DMS and the National Institutes of Health.

Outside of government, DMS has worked with the Simons Foundation on MathBioSys. This was a very successful initiative that started four centers for mathematics and biological systems. DMS also works with the Conference Board in the Mathematical Sciences (CBMS) on the Regional Research Conferences in the Mathematical Sciences.

We also note two recent initiatives by DMS that we believe are influenced by its activities during the review period. First, the NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning (MoDL) was initiated in spring 2020, following the successful collaboration of Simons and DMS on MathBioSys. Second, DMS has strong participation in NSF’s RAPID program that addresses the COVID-19 pandemic, with 20 funded proposals by DMS (out of 960 over all of NSF).

Diversity, Equity and Inclusion

Not surprisingly given the historical context, this COV paid special attention to questions of Diversity, Equity and Inclusion at all levels from the most exclusive scientific programs supported by DMS to initiatives designed to broadening participation. Specific remarks appear in each subcommittee report. Below we summarize the common themes.

The COV had the impression that there is a concerted effort from the DMS leadership to address this delicate issue. Where there are challenges there are also opportunities. A successful approach should include these two components:

1. Incentivize all PIs to become part of a global effort to train and mentor members of under-represented groups.
2. Provide the scientific community with evaluation tools that facilitate the process of identifying excellence in applications coming from PIs with backgrounds very different than one's own.

NSF and DMS already have several programs and evaluation tools in place intended or that could be refined to broaden participation:

- Emphasize Research at Undergraduate Institutions (RUI) and Research Opportunity Awards (ROA). Most members of the COV were unaware of these opportunities and found upon studying the materials under review that both programs were underutilized. The intention and functioning of the ROA were particularly mysterious. Advertising these programs broadly, possibly refining their description, might help attract applications.
- As mentioned in the 2016 COV report, the Broadening Participation Initiative (BPI) was created in response to the 2013 COV report. This process provides additional funding for some regular research grant applications with a particularly strong diversity component. Upon analyzing the data provided to us, the COV wondered whether BPI was truly having the intended effect. The COV recommends that metrics be put in place to evaluate whether this process is accomplishing its goal of broadening participation. Does BPI fund competitive proposals that would not have been funded otherwise?
- The Broader Impact (BI) review criterion was discussed at length in this and other contexts. Its implementation continues to be inconsistent. Some of the issues highlighted on the previous COV reports persist. Well delineated guidelines concerning the possible interpretations and use of BI would be helpful to the PIs, the reviewers/panelists and the POs. An effective and transparent use of the BI criterion will help NSF accomplish some of the goals that go beyond excellence in research. Several members of the COV have effectively used rubrics in the evaluation of proposals or hiring processes. If well done, this systematic assessment technique helps counteract the effects of implicit biases, without limiting evaluator's ability to judge. It was noted that some solicitations phrase their objective in a way that might help to delineate the BI criteria to be applied in evaluating these proposals; e.g. the CAREER solicitation states as goal of the program "to identify junior PIs with the potential to serve as academic role models in research and education and to lead advances in the mission of their department or organization."

The Proposal Review Process

Proposal review is at the heart of the mission for DMS, as well as throughout NSF, and the COV review was accordingly focused on assessment of the proposal review process and its management. This is a large-scale undertaking with approximately 3000 proposals submitted per year, and 30 DMS program officers (POs) who are handling the proposal reviews.

The COV found that the proposal review process is working well for the most part. Proposals are reviewed by individual reviewers then analyzed and ranked by review panels. The low acceptance rate for invited panelists, makes the POs' job very challenging. The COV commends DMS for its hard and successful work to obtain appropriately diverse and qualified panels. Conflicts of interest are handled well and with care. POs try hard to maintain high quality reviews and are successful in the vast majority of cases. The review methods are efficient and appropriate. In general, the COV was able to understand the rationale of decisions in most cases using only the jackets.

Overall the Review Analysis were very impressive, and they contained a detailed explanation of the review process for a given proposal. Those that quoted extensively from individual reviews and panel summaries were not as useful as they often failed to express the rationale for a given decision. The COV members noted that in most cases the Equalization process seemed to be an appropriate tool to balance the portfolio at the end of the deliberation process. Unfortunately, in general, there was no detailed documentation of this process. It was noted that the Computational Mathematics program operates differently than all other disciplinary programs, in that panels do not discuss proposals that do not get at least one V rating, and that panels only consider two categories: competitive/not competitive. Their decision process seems more along the lines of Equalization from the beginning.

The DMS is very responsive to emerging research and education opportunities. There are generally a number of challenges in bringing to life programs that involve agencies outside NSF. We were struck by how well the DMS program officers navigated these tricky waters.

Response to the Previous COV Report

The 2016 COV Report included a number of questions about the following:

- evaluation of programs and of the review process
- the quality of reviews and of the documentation and PI feedback of review decisions
- use of the Broader Impacts (BI) criterion
- the handling or proposals from PUIs and from minority- and women-serving institutions.

The subsequent response from DMS was that, within various constraints, they would strive to improve on these issues. Although the (current) COV continues to find questions about all four of these issues, we also found significant examples of improvement in the first two issues; for example, a workshop for evaluation of RTGs and better documentation of PO decisions. On the other hand, use of the BI criterion continues to be spotty and mechanisms for increasing diversity, in terms of gender demographic and institution type, have been put into place (RUI, ROA, and BPI) but do not seem to be used effectively. This is described in further detail in the subcommittee reports below.

There are two ongoing issues raised in the 2016 COV report that we would like to note here:

The first relates to the Focused Research Groups (FRGs). The 2016 report states: *"*The small number of proposals in FRG awarded each year (4) makes it challenging to create a balanced distribution among fields and to support diversity in this portfolio. …. In 2013, there were no female PIs or co-PIs on any of the awarded FRG grants. In 2015, there were only four female PIs or co-PIs on the awarded FRGs. Minority representation among PIs is low. Our committee felt that this program should make stronger efforts in diversity*."* Our subcommittee found that distribution across fields and diversity in the portfolio continues to be a challenge with the FRGs.

The second ongoing issue that arose in both our analysis and the 2016 COV report concerns the evaluation of proposals for some programs (e.g. DMREF) by two separate panels. In many cases, the two panels arrived at very different conclusions, which is both unsurprising and appropriate given the different areas of expertise and the different lenses through which the panels viewed the proposals. However, the process by which these conflicting panel assessments are resolved was not clear. One suggestion that was made in the 2016 (and 2013) report was: *"We concur with the recommendation made in COV 2013 that mathematical scientists participate in those DMR panels evaluating DMREF proposals."* We continue to recommend that DMS consider joint panels wherever appropriate and feasible.

While the MathBioSys program did not exist at the time of the previous COV assessment, several of its activities indicate responsiveness to previous COV comments. These included setting clear expectations, support of members of underrepresented groups, and detailed assessment plans.

Recommendations

The COV is presenting to DMS the following list of recommendations. Many additional suggestions are included in the Subcommittee reports.

1. Encourage virtual participation in future panels, even after the pandemic ends. Now that many researchers have learned to effectively use Zoom and other tools, this could greatly improve the acceptance rate for panel invitations,

especially for panelists from the west coast. The COV noted that DMS has been successful in making virtual panels effective.

2. Improve the use of the BI criterion. The BI criterion should play a stronger role in funding decisions, at least as a positive factor; i.e. more proposals should be funded, partly on the basis of their Broader Impacts. Moreover, the COV strongly believes that consistency for evaluating Broader Impacts would greatly improve the process. This should include stressing the importance of BIs and directing the attention of PIs and reviewers to the five questions that provide a consistent framework for Broader Impacts (described in the Proposal and Award Policies and Procedures Guide (PAPPG)).

3. Improve the quality and consistency of reviews and panel summaries. More specific guidelines to reviewers would help them to make their reviews detailed and to address all of the criteria for the review. The equalization process would benefit from having a set of more explicit “guiding principles” and more documentation of the reasons for funding decisions. Program officers should consistently include details of the decisions in the Review Analysis, especially for potentially controversial cases. We notice that there has been some grade compression in proposal ratings. Perhaps numerical ratings would be more effective.

4. DMS (and possibly NSF) should work to increase the response rate to requests for demographic information (e.g., gender and race). This is important data for DMS, as well as the mathematical sciences community, for assessing the diversity of the community and the equity of funding decisions. Survey methodologists could possibly provide mechanisms to encourage sharing of this sensitive data; e.g., by including a description of how the data will be used. Moreover, modernizing the survey could be an important step, by including additional options for gender and race; e.g., an "other" option.

5. The committee recommends that DMS consider implementing a systematic process for handling resubmitted proposals. This could include making a concerted effort to provide constructive feedback and encouragement to resubmit, e.g., for proposals that are recommended for funding but do not receive it, as well as for proposals from female and URM PIs. Another possibility is that PIs of a resubmission could be asked to provide a one-page description of the revisions that were made to address comments in the previous reviews (staying within the 15-page limit), and that reviewers would assess how well the previous reviews were addressed.

6. DMS has made significant steps to increase diversity, for example among the POs, and the COV recommends that DMS continue in this direction. Additional steps could include increasing the percentages of reviewers who are women or URMs, funding more research projects at HBCUs and minority serving institutions (MSIs), ensuring that proposals from PUIs would be considered by at

least one reviewer from a PUI, and better publicizing and utilizing the ROA and RUI programs.

Overview and Recommendations of Subcommittee F

David C Manderscheid, Subcommittee F Lead University of Tennessee

Subcommittee F assessed the review process, management, and portfolio of the Mathematical Sciences Institutes, Simons Math Bio Centers, and TRIPODS grants. This portfolio represents a substantial investment by DMS, the Simons Foundation, and other areas within NSF. The subcommittee found that this investment is paying off quite well. Transformational work in the mathematical sciences is being done and the next generations of researchers are gaining valuable education and experience. The emphasis on inter- and multi-disciplinary work is impressive as is the response to the strategic priorities of NSF. This success is due in large part to the excellent management of the institutes and centers by the community in collaboration with DMS staff. The subcommittee found that the DMS staff stewardship of these programs to be outstanding.

Subcommittee F makes the following recommendations concerning the programs they considered. First, DMS should continue the excellent “hands on” management they are providing. This management has made the programs better and addressed the concerns of the previous COV. Second, DMS should work with the programs to better publicize the successes of these programs to the mathematical sciences community and more generally. The subcommittee felt this would garner even greater support for the programs. Third, DMS should work with the programs to get better data on impact of the programs on members of underrepresented groups and mathematical scientists who do not have current NSF support. Finally, DMS should consider the possibility of funding another new institute in the next round of funding, either through reallocation within the institutes program or from other sources. Subcommittee F was quite impressed with the quality of the institute proposals that could not be funded in the most recent competition. In particular a number of the proposals could have contributed substantially to supporting underrepresented groups.

Subcommittee A
Algebra and Number Theory, Combinatorics, Computational Mathematics

Subcommittee Members:

David Manderscheid (Lead)	University of Tennessee, Knoxville
Jesus De Loera (co-Lead)	University of California, Davis
Pavel Bochev	Sandia National Laboratories
Clint Dawson	University of Texas, Austin
Sabine Le Borne	Hamburg University of Technology
Steven Lee	DOE Advanced Scientific Computing Research
Po Shen Loh	Carnegie Mellon University
Jennifer Morse	University of Virginia
Alexander Polishchuk	University of Oregon
Adriana Salerno	Bates College
Tony Várilly-Alvarado	Rice University

I. Questions about the quality and effectiveness of the program's use of merit review process.

1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?

The review methods are efficient and appropriate. The subcommittee was able to understand the rationale of decisions in most cases using only the jackets. The panel system is a valuable way to blend a range of qualified perspectives.

While the existing panel review system is successful, the DMS should continue to fine-tune the process. For example, DMS could consider an experiment where the same set of proposals is evaluated twice by two separate panels or ask for more ad hoc reviewers. Similarly, there are mathematical algorithms to create total orderings based on partial information. These are often used by computer science conferences. Such algorithms might be useful since not all reviewers read all proposals. The review process could integrate a quality control mechanism and continue to improve itself based on data and mathematics.

The panel review system has seen changes due to COVID19. For example, a program reorganized its panels to be greater in number, but with fewer proposals each. This gave panels more time to discuss each proposal, which was a benefit. DMS has been successful in making virtual panels effective. This may have benefits post-COVID. Perhaps virtual participation will become more popular and increase participation.

2. Are both merit review criteria addressed

2a. In individual reviews?

There is variability in the quality of the individual reviews. Some reviewers are extremely conscientious, while other reviewers are not as careful. Broader Impacts is an area for improvement. Intellectual Merit is nearly universally addressed and significantly stressed in reviews, to the degree that some panelists make decisions on the basis of Intellectual Merit even when Broader Impacts are lacking. We observed that different programs had different interpretations for Broader Impacts.

The subcommittee suggests training for first time panelists. One program officer gives examples of good and bad reviews to help the panelists improve their reports. This is an excellent practice. Alternatively, it may be useful to complement the usual review by a rubric that tracks numerical scores by pre-specified criteria, to facilitate a final consistent scoring. The European Union scientific agencies use a rubric style set of scoring guidelines. A rubric could also help eliminate any issues of Broader Impacts evaluations being applied inconsistently. The subcommittee did not have a unanimous opinion on whether a rubric should be used.

2b. In panel summaries?

Panel summaries were generally of higher quality than individual reviews. The panel summaries served as an efficient way for an outside observer (e.g., the COV) to understand the nature of the discussion. This serves as evidence of the benefit of panel discussions to the individual reviews. The program officers can play an important part in steering the panel to write quality summaries. There is an opportunity here to take Broader Impacts to the next level, as suggested above.

2c. In Program Officer review analyses?

The quality of program officer review analyses is generally impressive. They tended to be of even higher quality than panel summaries. As expected, there was variability in style among program officers. The subcommittee encourages all program officers to continue to include details of the decisions behind potentially controversial cases.

3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals?

There is a significant variability among the individual reviews with respect to their level of detail and usefulness. Many reports provide only terse statements. Phrases such as *'the proposed research plan is somewhat vague"* appear in many reports. These statements do not provide meaningful feedback to the PIs. Reviewers should be guided to write focused reports with comments about specific aspects of the proposals (specific sections, figures, theorems, etc.) that can provide a consistent basis for their evaluation. A number of committee members saw a troublesome trend of a declining quality of reviews. The subcommittee suggests that the DMS considers possible steps to mitigate this. Perhaps NSF can investigate successful approaches for reviewing proposals by looking at what other scientific agencies (within the US or in other countries) do. Some

type of scoring system might be tried that forces reviewers to specifically address certain aspects of the proposal.

4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?

The panel summaries can vary by panel. The most useful information was in the review analysis by the program officers. Some panels used mostly boilerplate text across multiple summaries. Program officers should work to improve the uniformity.

5. Does the documentation in the jacket provide the rationale for the award/decline decision?

The jackets provide sufficient information supporting the decision-making process. Access to both the jacket copies and PI copies of reviews was helpful. Decisions regarding budget cuts are often explained by a short reference to limited budgets. The reasons why some budgets are cut more than others are not always transparent.

The minutes for the equalization meetings also proved to be very helpful. However, the decisions of equalization meetings are sometimes less thoroughly documented. Nonetheless, the equalization process does seem to work well and provides an effective mechanism to improve the overall handling of the proposals. The POs are doing a great job, especially when taking into account the large number of proposals that they process.

6. Does the documentation to the PI provide the rationale for the award/decline decision?

For the most part yes. Program officers should continue to strive to provide this in all cases.

7. Additional comments on the quality and effectiveness of the program's use of merit review process:

Overall, the merit review process is good. However, maintaining consistency between the reviews and the outcomes still presents some challenges. For example, a few proposals were described as "very poorly written", having "poor Broader Impacts'', or as having mistakes, and yet these proposals were highly recommended for funding, seemingly solely on the prior record of the PI. This is unfair to junior researchers who do not have the benefit of an extensive prior record as they would have likely received a "Do not recommend" for the exact same proposal. As commented earlier, a more effective mechanism to score and evaluate Broader Impacts may be necessary. Also, for proposals that had been reviewed by more than one panel, some COV members were concerned that this might disadvantage the proposal. Program officers assured us that this was not the case – in fact, if anything, it helped the chance of funding. This should be monitored.

II Questions concerning the selection of reviewers.

1. **Did the program make use of reviewers having appropriate expertise and/or qualifications?**

The subcommittee commends DMS for its hard and successful work to obtain appropriately diverse panels. We recognize the difficulty in recruiting panel members, with many prospective members declining the invitation. In most cases, DMS gets a reasonable mix of senior/mid-career/junior reviewers, from different institutions, different genders, and so on. In some cases, the subcommittee found panels with a lack of diversity of topic expertise. In cases like these program officers should not hesitate to get mail reviews. The subcommittee encourages DMS to continue to expand the pool of reviewers and to stress the importance of diversity, in all dimensions, in panels. The choice of reviewers particularly affects how interdisciplinary proposals or those involving more than one area of mathematics are rated. Breadth of expertise in the panelists should be a priority. NSF Program officers are aware of this issue and the subcommittee trusts they will continue to try to make panels as diverse, welcoming, equitable, and inclusive as possible. This will not only ensure the best outcomes but also make it easier to recruit panelists.

2. **Did the program recognize and resolve conflicts of interest when appropriate?**

Conflicts of interest are handled well and with care. Jacket documentation identifies reviewers with conflicts of interest, but does not mention the nature of the conflict. The subcommittee suggests that DMS consider also documenting the nature of the conflict.

3. **Additional comments on reviewer selection:**

The subcommittee offers the following ideas on how to increase panel participation as well as diversity. First, if a potential PI has tried a number of times to get funding, but has been repeatedly borderline rejected, they should be considered for panel service, to help them better understand what it takes to be successful. Second, consideration should be given to maintaining a core reserve of panelists with earlier invitations to serve, this might increase acceptance of invitations.

Question III. Concerning the management of the program under review. Please comment on the following:

1. **Management of the program.**

Program Officers are commended for their management of the heavy workload (100+ proposals each, on average) and the efficient overall processing of reviews and awards

under time and budget constraints. The extensive jacket documentation is evidence of their conscientiousness and their excellent responsiveness to the demands. The subcommittee observed that the POs worked well together. The training provided to new POs was important and effective in giving them a common understanding of the duties. Proposals received substantive reviews in most cases, but not always. It is clear that the POs try hard to maintain high quality reviews and are successful in the vast majority of cases.

2. **Responsiveness of the program to emerging research and education opportunities.**
3. **Program planning and prioritization process (internal and external) that guided the development of the portfolio.**

The DMS has aligned itself well with NSF priorities while maintaining core disciplinary strength. The alignment with NSF priorities can be seen in programs like TRIPODS and the Broadening Participation Initiative. The subcommittee also saw that individual disciplinary programs respond well to trends identified by PIs and reviewers through the process POs use to determine which proposals are funded. For example, this can be seen in the equalization process. The POs also get ideas as to emerging trends from going to conferences and in their working with the mathematical sciences institutes. It was less clear to the subcommittee to what extent POs take a more top down approach if they feel a particular trend should be championed more. The DMS should consider how they encourage POs to be responsive to emerging challenges

4. **Responsiveness of program to previous COV comments and recommendations**

Successful efforts have been made to address many prior concerns. Unfortunately, demographic data continues to be limited. Although the information is given only on a voluntary basis and can consequently be hard to obtain, serious measures to improve this situation should be taken. A body of knowledge of best practices in voluntarily obtaining sensitive data is part of the field of survey methodology and can provide suggestions to improve response rates. For example, ICERM saw an improved response rate to their demographic questions when they sent a follow-up request with a short explanation about why the information is important. Another idea is to ask the program officers to discuss providing the data and the reasons for doing so during panel meetings.

IV Questions about Portfolio. Please answer the following about the portfolio of awards made by the program under review.

1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity?

The subcommittee was impressed with the balance of the portfolio across major areas of the program. The NSF should continue to have an active dialogue with the mathematical sciences community about emerging areas of active research and future directions.

2. **Are awards appropriate in size and duration for the scope of the projects?**

The POs are doing a good job of distributing the scarce resources available to them. Often though there are budget reductions, which can be discouraging to PIs. The subcommittee was glad to learn about a recent influx of money to help support graduate students -- about $10 million in a recent cycle. The subcommittee hopes that such support can continue and possibly be increased. Funding of postdocs in programs considered by this subcommittee is not common. An increase in funding for postdocs similar to the increase for graduate students should be a top priority, especially in light of the current job market.

3. **Does the program portfolio include awards for projects that are innovative or potentially transformative?**

DMS has funded a number of high-risk, high-reward proposals that have led to spectacular results. The PO's seem to be striking a good balance on funding innovative research and more incremental but necessary efforts.

Examples of transformative research done by Algebra and Number Theory PIs, include the work of Bhargav Bhatt and collaborators on p-adic geometry, the work of Dennis Gaitsgory on geometric Langlands correspondence, and the work of Yunqing Tang and collaborators proving deep results on the arithmetic of Shimura varieties of K3 type. Moreover, Alena Pirutka and collaborators have developed new tools to study rationality of algebraic varieties. The success of the ANT program is also reflected in many awards given to the PIs over the last 3 years: a Fields Medal, a Wolf Prize, a Shaw Prize, and five New Horizons Prizes.

Examples of transformative research in Computational Mathematics include Yifei Lou's CAREER project on *Mathematical Modeling of Data to Insights and Beyond*, which has application in computed tomography scanning; Andrew Christlieb's project *Method of Lines Transport for Balance Laws, an Application to Plasma Physics* that focuses on magnetohydrodynamics modeling with applications to space weather; and junior scientist Soledad Villar's stellar work on *Optimization Techniques for Geometrizing Real-World Data* that has applications in computational biology. Computational Math researchers were highly recognized through prestigious prizes, including new members of the National Academy of Sciences and the election of many PIs as SIAM Fellows.

Examples of transformative contributions in Combinatorics include Jozsef Balogh's new results in Ramsey Theory, an area of mathematics that uncovers order within large chaotic systems; Igor Pak and Greta Panova's work at the crossroads of combinatorics and computational complexity theory; and Isabella Novik's breakthroughs in counting faces of polyhedral manifolds, which extends classical work from 1700's by Euler. The researchers were also recognized with awards, for example at least two PIs funded by

this program were invited to speak at the 2018 International Congress of Mathematicians in Brazil.

4. **Does the program portfolio include inter- and multi-disciplinary projects?**

The portfolios do include inter- and multi-disciplinary projects. Some of the most interesting developments in the mathematical sciences, leading to surprising advances, occur at the interstices between two fields, or at intersections with other fields of knowledge. The subcommittee applauds DMS staff for their recognition that as the mathematical sciences evolve the barriers between pure, applied, statistical and computational mathematics must be broken. The program officers are attentive to this issue.

5. **Does the program portfolio have an appropriate geographical distribution of Principal Investigators?**

The subcommittee found that DMS makes an excellent effort to distribute funds equitably across geographical areas. Precise information is difficult to decipher from the data provided, however.

6. **Does the program portfolio have an appropriate balance of awards to different types of institutions?**

The subcommittee would like to see more PI's from departments which serve large minority populations but which do not have a Ph.D program. The DMS has attempted to address this through joint proposals between PIs at R1 departments and co-PIs at non-Ph.D departments. More efforts in this direction would be desirable. Further, it is not clear what criteria are used to evaluate proposals from RUIs nor how these are evaluated within a panel. The subcommittee found that some panels are unable to weigh successfully the extra teaching demands of RUIs.

7. **Does the program portfolio have an appropriate balance of awards to new and early-career investigators?**

The PO's fund a good number of early-career investigators. Equalization panels pay particular attention to this point, with commendable results.

8. **Does the program portfolio include projects that integrate research and education?**

The majority of proposals are mostly focused on research, but a large number of grants request funds for graduate students. However, with limited funds, it is difficult to support students so the educational benefits of the projects are diminished. Many PIs do outreach activities aimed at the broader mathematical community, although these activities are not necessarily financially supported as line-budget items.

9. **Does the program portfolio have appropriate participation of underrepresented groups[1]?**

The number of submissions from underrepresented groups is too small, certainly well-behind demographic numbers and even from states with growing minority populations. This systemic problem cannot be addressed by DMS alone. The subcommittee found proposals by underrepresented minority PIs that have not had success in obtaining NSF funding after trying for several years. The subcommittee recommend that PO's make a concerted effort to provide constructive feedback and suggestions for improvement that could lead to successful proposals. The COV was pleased to hear from the Assistant Director of Mathematical and Physical Sciences of a potential new initiative that could help broaden participation.

10. **Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.**

The programs we reviewed all respond to the agency's stated mission to promote the progress of science. They also have strong intersection with the ten big ideas initiative.

11. **Additional comments on the quality of the projects or the balance of the portfolio:**

V. Other Topics.

1. **Please comment on any program areas in need of improvement or gaps (if any) within program areas.**
2. **Please provide comments as appropriate on the program's performance in meeting program-specific goals and objectives that are not covered by the above questions.**
3. **Please identify agency-wide issues that should be addressed by NSF to help improve the program's performance.**
4. **Please provide comments on any other issues the COV feels are relevant.**
5. **NSF would appreciate your comments on how to improve the COV review process, format, and report template.**

The subcommittee makes the following suggestion: Invite to each panel two dedicated scribes. The scribes would be junior researchers who have not received NSF funding as a PI. Since these scribes would not contribute to the panel discussion, they could take turns taking notes during alternate panel discussions and then refining their notes

[1] NSF does not have the legal authority to require principal investigators or reviewers to provide demographic data. Since provision of such data is voluntary, the demographic data available are incomplete. This may make it difficult to answer this question for small programs. However, experience suggests that even with the limited data available, COVs are able to provide a meaningful response to this question for most programs.

during the next discussion. This would contribute substantially to the record quality of panel discussions. It would increase the efficiency of the entire process, because much of the panel summaries would effectively be written after each discussion. Moreover, it would provide a valuable development opportunity to the scribes, and so recruiting high caliber scribes would not be difficult. Scribes would likely gain perspective on the proposal writing process, enhancing their chance of getting funding.

SUBCOMMITTEE B
Applied Math, Foundations, Probability, Geometric Analysis, Topology

Subcommittee Members:

Tara Brendle (Lead)	University of Glasgow
Peko Hosoi (Co-lead)	Massachusetts Institute of Technology
Anna Amirdjanova	Sandia National Laboratories
Oliver Dasbach	Louisiana State University
Aliana Fraser	University British Columbia
Chris Jones	University of North Carolina, Chapel Hill
Bernard Mair	Association of Public & Land-Grant Universities
Deborah Nolan	University of California, Berkeley
Sean Paul	University of Wisconsin
Deborah Sulsky	University of New Mexico
Simon Thomas	Rutgers University
Carol Wood	Wesleyan University

I. Quality and effectiveness of the program's use of merit review process.

I.(1) Subcommittee B found the review methods were effective in obtaining valuable information for making award decisions. Although panel reviews are the main mechanism for reviewing individual grant proposals, the ad hoc reviews are used to augment them in cases of conflicts of interest, lack of expertise on the panel, or other considerations.

The panel review system seems especially effective in equalizing ratings between individual reviewers; the individual scoring may not be consistent across reviewers, but the panel often reconciles the variations in grading. The panels usually involve a large number of experts (10 – 20) so the POs have to be quite adept in obtaining a meaningful summary evaluation from many, often disparate, views. We encourage DMS to pursue strategies for ensuring that the panels are kept on topic and focus on Intellectual Merit and Broader Impact. We are pleased that NSF provides training in this area and support enhancing the training provided to the POs.

I.(2) Both merit criteria are technically addressed, but the quality of reviews is highly variable. Overall, the reviews contain more substantive comments on Intellectual Merit than on Broader Impacts; in addition, the evaluation of Intellectual Merit often takes into account the PI's record of achievement, which is closely linked to their status in the community. This status is sometimes used to compensate for weaknesses in the proposal.

The criterion of Broader Impacts is applied less consistently in the panel ratings and among those proposals in the highly recommended for funding category there were cases with only modest Broader Impacts. In applied mathematics panels, Broader Impacts are too often assessed by the project's importance in applications which is inconsistent with the National Science Board report on Merit Review Criteria. The Program Officers are aware of this issue and expressed frustration during their interview with our subcommittee. One suggestion that may improve the understanding and application of the Broader Impacts criterion is to require reviewers and panelists to specifically address the five questions, most recently described in the NSF PAPPG (nsf20001), that elaborate on Broader Impacts (see footnote[2]).

I.(3) Our subcommittee found considerable inconsistency in the review analyses. Some are well done, detailing the rationale for the decision, whereas in other cases, boilerplate language is overused, sometimes giving little indication as to why a proposal was awarded or declined.

I.(4) The subcommittee found a breadth of quality in the individual reviews. While many are informative and well-argued, sometimes reviewers merely summarize the project summary then give a rating without much commentary. In other cases, the criticism is so detailed that it seems as if the reviewer expected a proof of a proposed conjecture which may cause PIs to “play it safe” and shy away from potentially transformative work. Neither extreme is helpful in supporting the case being made for the funding decision, and POs need to play an important role in guiding the reviews in a productive direction.

I.(5) Our subcommittee found that some of the panel summaries are inadequate, whereas other panels provide substantive comments that are useful to inform the PI as to why the proposal received its ranking. We would encourage the PO leading the panel to ensure that the summaries reflect the give-and-take of the discussion. The summary should reflect not just the reviews but the reaction of other panelists and highlight which comments in the reviews were taken as most significant by the other panelists in reaching their decision.

I.(6) This documentation in the jacket varied with some cases being carefully documented, but others with only very cursory explanations. The inclusion of boilerplate statements about the overall context of NSF and the challenges faced generally in reviewing NSF proposals should not be included in every jacket as they only serve to obscure the case.

[2] 1. What is the potential for the proposed activity to: (a) Advance knowledge and understanding within its own field or across different fields (Intellectual Merit); and (b) Benefit society or advance desired societal outcomes (Broader Impacts)? 2. To what extent do the proposed activities suggest and explore creative, original, or potentially transformative concepts? 3. Is the plan for carrying out the proposed activities well-reasoned, well-organized, and based on a sound rationale? Does the plan incorporate a mechanism to assess success? 4. How well qualified is the individual, team, or organization to conduct the proposed activities? 5. Are there adequate resources available to the PI (either at the home organization or through collaborations) to carry out the proposed activities?

The quality of individual reviews, panel summaries and PO comments varies. Often specific suggestions are made by individual reviewers that provide advice to the PIs about where improvements could be made. Generic language, such as the research 'not being compelling compared to other proposals' without further detail is common and is unhelpful to a PI.

I.(7) In general, the panel process seems to work quite well in determining Intellectual Merit. However, as brought up in the 2016 COV report, the assessment of Broader Impacts is not carried out uniformly across programs, and this has not been adequately addressed (see above suggestion and footnote 1). Consistency in the quality of reviews could be improved, perhaps by more specific guidelines being provided to reviewers. It is useful to list strengths and weaknesses in a review, but an indication of how much weight was given to the specific issues raised would be helpful as well.

II. **Selection of reviewers.**

II.(1) Overall, the program uses reviewers with appropriate expertise and qualification. We found a few instances where the panelists' opinions diverged and additional expert advice was not obtained. For these cases, we suggest a more extensive use of ad hoc reviewers.

II.(2) The program carefully documents potential conflicts of interest and has a process to remove conflicted reviewers. In rare instances, all of the experts in a sub-discipline were in conflict with a proposal in that sub-discipline; and, in these cases, it seems desirable that at least one mail review should be requested. If such conflicts of interest are identified before the panel meeting, it would be helpful to obtain the mail review in time for the panel to see it.

II.(3) In general, the panel members' gender and racial composition (as self-identified by the panelists) appeared to be well-balanced. However, over half of the panel reviewers did not provide their gender. It was not clear whether these reviewers declined to state their gender, an appropriate option was not identified, or they simply did not bother to respond to the request. It is recommended that options for gender should be modernized; e.g., an "Other" option could be included. The importance of this information should be stressed and additional opportunities for panelists to provide this information should be offered during or after the panel.

III. Management of the program under review.

III.(1) Overall, the Subcommittee felt that programs were well managed. We commend the collaborative approach to management taken by the division. The process of selecting and appointing reviewers for various panels is crucial to the successful management of DMS programs. The program officers devote an incredible amount of time and energy to this important task and would benefit from more support in carrying this out. We would ask the wider mathematical community to support DMS in this

regard by responding promptly and, whenever possible, positively to requests to serve on panels.

The subcommittee liked the Equalization process, because it provides a way to bring to bear additional insights and perspectives on difficult funding decisions. We particularly liked the fact that this included the possibility of co-funding with other programs.

We would encourage the DMS to consider whether the practice of having a single panel review two or more subfields is beneficial, as panels seem to find it challenging in such cases to rank proposals across different areas.

Finally, the subcommittee encourages the DMS to continue and even expand the training of the panelists in areas such as evaluating Broader Impacts and implicit bias.

III.(2) We were pleased with the philosophy put forward by program officers that allows the community to determine trends and areas of interest in research and education opportunities. This bottom-up approach seems to capture emerging research very well. Committee members generally found that funded proposals are in exciting, bold areas of research and were representative of important themes and topics. The subcommittee also felt that the balance between top-down (e.g. solicited) and bottom-up (e.g. unsolicited proposals) is very good and notes that maintaining some top-down influence to encourage new areas is beneficial.
In many panels, Program Officers clearly paid careful attention to awarding innovative and transformative proposals, to having an appropriate balance among subfields in the program, balancing by geographic location and institution type, PI diversity, including awards to high quality RUI proposals, and to contributing to priority areas or to interdisciplinary efforts. Revised versions of proposals declined in one year sometimes received awards in the following year, and research that had been supported for multiple funding cycles was often declined as other strong proposals that had not yet enjoyed NSF support entered the competition.

On the other hand, in some cases, the program planning and prioritization process was not evident from the documentation provided. The COV found discussions with the program officers helpful in clarifying their procedures in these cases, and the lack of documentation did not necessarily imply absence of implementation.

Success in balancing between new investigators and those previously supported seemed a bit mixed. In some panels this balance was felt to be achieved, in others less so. When panels did not achieve this balance, it could go either way: in some cases the vast majority of awards went to early career mathematicians, and in other cases to more established mathematicians.

We discussed whether there might be unintended consequences of disproportionate support for one PhD-age group over others. For example, early career mathematicians might be discouraged from continuing their careers if they perceive that very few senior mathematicians are successful in obtaining funding.

III.(3) The 2013 and 2016 COV reports both flagged issues with Broader Impacts evaluation, particularly in terms of consistency. One subcommittee member felt that progress had been made in the panels they looked at, while most others did not see evidence of progress in this regard in other panels. We note that the NSF previously declined the COV 2016's suggestion to provide panels with 'good' sample reviews due to concerns about reviewers adopting a cookie-cutter approach in proposal evaluation. Nevertheless, this committee strongly believes that consistency for evaluating Broader Impacts would greatly improve the process; see **I.(2)** above.

The 2013 COV also suggested that the NSF consider new ways to support mid-career researchers, and we did not see how this had been taken forward. As noted above, it did appear that in many panels, few mid-career researchers had been awarded grants.

IV. Questions about Portfolio.

IV.(1) While the balance among subdisciplines may vary from year to year, the subcommittee found that in general the program portfolio was well-balanced by disciplines and subdisciplines. The awards were appropriate, and the broadening of the portfolio, especially through the influx from junior researchers, is encouraging. As an outlier, a committee member noticed that in one particular panel the funded proposals were largely in one subarea, while other important subareas received little support.

IV.(2) The subcommittee overall found that the size and duration of awards was appropriate for the projects. However, questions were raised about whether revised budgets left sufficient funding for graduate students, while acknowledging that there are field differences in expectations for such support. In one case it was noted, perhaps as a consequence of flat budgets over a decade, that the budget revisions were substantial and that the PIs were expected to do the proposed work with two-thirds of the requested funds.

IV.(3) The subcommittee found the overall performance of the program highly successful, including awards for innovative and potentially transformative projects. As an example, over the review period, PIs from one program won a Wolf Prize, a Bôcher Memorial Prize, two Oscar Veblen Prizes and a New Horizons Prize. The portfolio includes innovative and transformative projects, for example, awardees made spectacular progress on min-max minimal surface theory, which led to the resolution of Yau's Conjecture and the proof of the Multiplicity One Conjecture.

IV.(4) Approximately half of our panels were described as inter- or multi-disciplinary and half were not. This was not deemed to be a problem as in some cases, the focus of the panel was on strictly disciplinary projects. However, it was noted that areas that are too tightly focused may limit access to underrepresented groups. As we describe in the diversity section below, one significant hurdle that needs to be overcome is the ability to attract members of underrepresented groups to apply. It is well-known that, e.g. in hiring, open-area calls will typically lead to a more diverse pool of applicants. It was

also noted that these broad strategies to increase diversity should be pursued concurrently with individualized efforts to encourage underrepresented researchers to engage, which can be a successful strategy in both multi-disciplinary or strictly disciplinary settings.

IV.(5) Across the board, the subcommittee found the panels presented a well-balanced geographical distribution of PIs both in terms of proposers and awards.

IV.(6) The subcommittee recognizes that most research awards will go to faculty at research-intensive universities. However, it was noted that faculty from departments with small or fairly recent Ph.D. programs were represented in the program portfolio, and that RUI projects and projects from Minority Serving Institutions were supported. The subcommittee also felt that more efforts could be made towards funding research projects at HBCUs.

IV.(7) Subcommittee members found substantial variety from one panel to another, with mixed views on whether there were enough new and early-career investigators. In one panel, 5 of the 9 awards went to early-career investigators. Another panel had a high number of new and early career awards, which seemed appropriate, in light of the recent infusion of generational talent in the area. In contrast, one subcommittee member had concerns about this sort of imbalance, noting that among the 20 top-ranked proposals in two panels, the majority had a PhD age of 10 years or less, and only 4 or 5 had a PhD age of 15 years or more. This may send the message (to mathematicians at all career stages) that there is little hope of being funded at the mid-to-late career stages, which could serve to discourage people from applying.

IV.(8) The views of the subcommittee on the integration of research and education were also mixed. A few members of the committee spoke highly of the program officers' efforts: "POs paid close attention to including high quality RUI proposals." However, others felt that this was not a priority in panels they looked at.

IV.(9) Despite the good efforts of DMS, we found the portfolio did not have appropriate representation from underrepresented groups. Subcommittee members cited low numbers in several instances and also found some cases where highly-rated proposals submitted by PIs from underrepresented groups were not recommended for funding. This is a very difficult matter and one for which we have no easy solutions.

There are at least two issues: the number of PIs from underrepresented groups who apply, and the number that are funded. In the proposals considered by this subcommittee, the first number was disappointingly low. The second number varied; in some -- but not all -- programs the success rate for underrepresented groups was encouragingly higher than the average.

Regarding the second issue, the subcommittee found that overall the panels are doing a reasonable job engaging underrepresented groups given the proposals at hand. (There are certainly examples where PIs from underrepresented groups were under-rated by

reviewers but this was not the norm and these instances were sometimes corrected by a savvy program officer in equalization). This leaves the vexing question of the first issue: how to get more proposals from underrepresented groups. Self-selecting in the mathematics community definitely occurs and is not optimal. The subcommittee felt that it is important to get the message out to underrepresented groups that "This is for me" and to create a culture of "It's ok to apply", possibly building on visible successes.

The committee suggests that a proposal from a URM which is not funded in one year despite good ranking be given strong encouragement to apply again. This sort of explicit attention might be the encouragement needed to persist. In addition, there may be a mentoring role for senior colleagues to encourage proposals from URMs and also to advise PIs on how to present proposed work. The committee also discussed what role the NSF should play in encouraging applications from underrepresented groups; Program Officers later clarified that there is already considerable activity in this regard.

IV.(10) Although this is typically not explicit in the proposals, the committee largely felt that the panels we reviewed were relevant to agency mission, relevant fields, and other constituent needs. In general, the panels are not targeted directly at the national priorities[3]; however, mathematics sets the foundations in most of these areas.

V. Other Topics. The COV is grateful for the enormous amount of data provided to us and the access we were granted during this process, though it would have been helpful to receive it sooner in order to facilitate our work. In addition we suggest that the NSF develop more rigorous and informative evaluations of the review process for COV consideration. For example, it would have been helpful to see the data in the context of appropriate benchmarks to assess diversity (e.g. for geographic diversity, awards by state could be assessed relative to state population). Presumably the NSF is tracking progress relative to these benchmarks and it would be helpful for the COV to have access to that information. Other examples of additional information that would provide a better understanding of the review process are: more nuanced analysis of panelists' scores of projects; observations of panels in situ with formal rubrics; and documentation of pre- and post-review processes.

Our subcommittee also spent a considerable amount of time discussing the challenge of attracting more PIs from underrepresented groups to apply. Program Officers are largely doing a reasonable job of funding underrepresented PIs given the diversity of the applicant pool, but this does not mean that the diversity of the portfolio is appropriate. Our subcommittee suggests that DMS institute mechanisms to increase diversity in the pool of applicants. This is not an easy task; some possible strategies include:

- Invite strong applicants who have been declined to serve on a panel and learn from the process.

[3] National science priorities are available at: https://www.whitehouse.gov/wp-content/uploads/2020/08/M-20-29.pdf

- Encourage unsuccessful applicants of highly-rated proposals to reapply. Inevitably some applicants who get declined stop applying, and only a tiny fraction of research active mathematicians apply and are supported.
- In some years there was a high success rate for applications for conference funding, but in other years the success rate was lower. Funding conferences is an effective way to support a large number of junior researchers with a relatively minimal investment and can have a significant impact. We encourage DMS to continue to support high quality conference applications.
- Automatically turn over strong unsuccessful proposals to be considered in the next year. This should not be done for all proposals; merely for those that are identified as having unique qualifications that make them strong contenders for the next round (e.g. exceptional Broader Impacts might be a necessary but not sufficient condition).

Subcommittee C

Analysis, Mathematical Biology, and Statistics

Subcommittee Members:

Rodolfo H Torres (Lead)	University of California, Riverside
Sara Y. Del Valle (Co-lead)	Los Alamos National Laboratory
Kelly Bickel	Bucknell University
John T. Chen	Bowling Green State University
Jessica M. Conway	Pennsylvania State University
Wilfrid Gangbo	University of California, Los Angeles
Mary Meyer	Colorado State University
Jill Pipher	Brown University
Raquel Prado	University of California, Santa Cruz
Nathaniel Whitaker	University Massachusetts Amherst
Tommy Wright	Census Bureau

The subcommittee was charged with reviewing a sample (790 of 2,512 projects with ~25% success rate) of the Analysis, Mathematical Biology, and Statistics portfolio between 2016 and 2019. The results from this review, as well as suggestions, are included below.

I. Questions about the quality and effectiveness of the program's use of merit review process.

1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?

In the panels reviewed, the process includes at least three individual reviews with Poor to Excellent rankings, subsequent panel discussions where proposals were ranked against one another and recommended for funding, and program officers' equalization meetings (in the Analysis and Statistics programs) where proposals are discussed in the broader context of the specific disciplinary programs, DMS priorities and funding decisions. The subcommittee finds these methods, which include both significant community input and NSF oversight, both appropriate and quite effective. Overall, there is agreement amongst reviewers and panelists as to which top proposals were most worthy of funding, and which have resulted in significant impacts in the analysis, mathematical biology, and statistics research communities. However, the subcommittee was dismayed by the number of very strong proposals that could not be funded due to significant budgetary constraints. Given this, it seems possible that subjectivity in the review process could affect funding decisions for proposals "near the margins." Thus, some subcommittee members suggest that DMS considers providing reviewers and panelists with a more detailed and quantitative or objective rubric.

2. Are both merit review criteria addressed by individual reviews, panel summaries, and in Program Officer review analyses?

The application of the Intellectual Merit criterion appeared to be quite consistent; however, subcommittee members noted that some reviewers merely summarized the proposed research while others used language such as “innovative” and “incremental” without further elaboration to justify their ratings. The application of the Broader Impacts criteria was less consistent. There seemed to be some confusion amongst reviewers and PIs about what constituted broader impacts and how to assess those. For example, sometimes proposed broader impacts did not go beyond the standard responsibilities of research faculty, but were viewed by reviewers or panels as sufficiently strong. Thus, the subcommittee suggests reviewer training and more objective metrics to address this shortcoming.

3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals?

The majority of the written reviews contains expert’s comments including the impact of the proposal in the field, significance of the proposed work, and chance of success. Most of the comments are very constructive, encouraging, and substantive. These comments may greatly help the PIs prepare for the next round of the competition. However, the subcommittee also noticed a few cases where the written reviews were very brief, especially for proposals that were not selected for funding.

With the current system, a revised proposal will be assigned to a new set of reviewers, and the new reviewers may interpret the criteria differently. The committee believes that it would be more helpful if NSF could implement a process to treat revised or resubmitted proposals systematically.

4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?

Most of the panel summaries contain panel consensus and rationales for the final decision after panel discussion. The panel decisions are supposed to be unanimous and indeed the subcommittee did not see any proposal funding decision that contained any dispute on the final panel decision. According to the Program Officers and documents in file, the panel summaries were read to all the panelists, consulted for different opinions and suggestions, and finally uploaded to the system after reaching the panel consensus. The committee was impressed by the extremely well-organized structure of DMS.

5. Does the documentation in the jacket provide the rationale for the award/decline decision?

The jacket provides informative documentation for the award/decline decision on each proposal and most of the documents contain detailed rationales, reviewer information, and panel summaries.

6. Does the documentation to the PI provide the rationale for the award/decline decision?

With few exceptions, the individual reports and the panel summaries provide sufficient rationale for the decisions made on the proposals. The rankings given by the reviewers normally agree with the narrative provided in the reports. It was observed that more information about weaknesses and strengths of the proposals were offered in the panel summaries than in the individual reports by the reviewers. Reviewers should be encouraged to be more specific about weaknesses and strengths for the benefit of the PIs in future submission. Perhaps more details should be provided when a proposal is declined. It was observed that sometimes proposals with very high rankings of multiple Excellent or Very Good ratings still get declined. The panel summaries should provide more context in those cases to the PIs and better explain that it was not the quality of the proposal per se, but rather the comparison with other proposals that determined not to fund given the budgetary constraints.

7. Additional comments on the quality and effectiveness of the program's use of merit review process:

The committee was very impressed by the amount of proposals that DMS has successfully evaluated each year. The great efforts of the Program Officers should be applauded. However, NSF might explore a more quantitative assessment of the merits of the proposals similar to what some funding agencies in other countries do. Likewise, a more detailed template for the reviews similar to what NIH uses could be provided, facilitating consistency in reviewers' assessments.

The current ranking Poor to Excellent system seems to have been inherited from the time most reviews were done via mail. It would be useful to have some expert analysis of whether this is the best possible method to rank the proposals. There appears to be grade compression taking place, which results in many highly ranked proposals in the individual merit reviews still getting declined.

A recurrent concern among the members of the subcommittee was the need for more clear guidelines about the Broader Impacts criteria used in the merit review proposes.

II. Questions concerning the selection of reviewers.

1. Did the program make use of reviewers having appropriate expertise and/or qualifications?

Expertise is defined as general knowledge and publications in the broad area of the proposed research. From a scientific standpoint, the Program Officers do an excellent

job in recruiting relevant experts for review panels. In some cases, proposals are inter- or multi-disciplinary, and therefore expertise in multiple areas is required to effectively gauge a proposal's merits. To meet this challenge, Program Officers take different strategies, including (1) review and evaluation of the proposal from a second panel specializing in the second discipline area, (2) obtaining an ad-hoc, "mail-in" review from an outside expert, or (3) consultation with a different program or directorate. We found all three used effectively in the difficult process of evaluating and ranking quality proposals.

The subcommittee found that proposals receiving reviews from multiple panels seemed less likely to receive funding. After discussion with Program Officers, we realized that typically this is the result of referral of multi-disciplinary panels receiving a low score from a first panel and referred to a second panel for potential "rescue".

The subcommittee noted a concern about the 2016 COV recommendation that "*DMS should pay special attention to make sure proposals from Primarily Undergraduate Institutions (PUIs) have at least one reviewer from a PUI."* Within the sample of panels that this committee reviewed, many PUI proposals did not appear to have a reviewer from a PUI. DMS should increase its efforts to fulfill this recommendation. The perspective of a PUI reviewer seems critical in evaluating the potential of research and broader impacts activities proposed at an undergraduate institution.

2. Did the program recognize and resolve conflicts of interest when appropriate?

A major concern in the review process is the potential for conflicts of interest (COI), both professional and personal. DMS Program Officers take great care to identify the potential for professional conflicts and to educate reviewers so that they may identify personal conflicts. During the review process, any reviewer with a COI for a particular proposal is asked to leave the room, effectively "quarantining" the reviewer, for the duration of the discussion of that proposal. The existence, or lack thereof, of COIs per proposal, and the ensuing mitigation described above, is clearly documented.

3. Additional comments on reviewer selection:

We appreciate that reviewer selection is a major challenge given the lack of gender balance and demographic diversity in the research mathematics profession, and we were pleased to learn of the significant efforts that Program Officers make in order to achieve appropriate representation. We encourage the Program Officers to continue in their efforts, and hope that as the mathematical profession becomes more inclusive, so too will the panels.

The subcommittee noted that the percentage of self-identified minority reviewers in DMS panels ranged from 4%-6%. Thus, the subcommittee strongly urges DMS to consider mechanisms for increasing the percentages of reviewers from minorities and underrepresented groups.

III. Questions concerning the management of the program under review.

1. Management of the program.

Management needs of each program vary within DMS and can occur at various levels: within and among programs in the division, with other divisions in MPS and NSF, and with other federal funding agencies. While the Program Officers have independence in forming and managing panels, all the POs contribute to the discussions in equalization of the awards. The more interdisciplinary programs sometimes require interactions even with private funding agencies such as the Simons Foundation, as in the case of the Mathematical Biology Program.

When expertise is needed outside of the DMS, the Program Officers seek co-reviews, and may have a proposal reviewed in multiple panels. It may be that jointly-convened panels, or panels with experts from within and without DMS, could be explored. The added benefit is the discussion across disciplines, where appropriate, that could identify new research directions as well as inform reviews.

The subcommittee finds that the management is excellent, with dedicated and talented Program Officers in all of the programs we evaluated. We were particularly impressed that the Mathematical Biology program is managed excellently by only two Program Officers. We would like to highlight coordination with other agencies, guidance, and communication to panel reviewers as areas where all of the Program Officers excel.

2. Responsiveness of the program to emerging research and education opportunities.

The DMS is very responsive to emerging research and education opportunities. The call for proposals is very broad, unlike other funding agencies that choose the research topics. Hence the proposal topics are driven by the experts in the community and reflect the cutting-edge research trends. We commend Program Officers in the programs reviewed for responding to new cross-disciplinaries activities by securing co-funding, thereby expanding the scientific reach, support, and products of DMS-funded research. For instance, research in the Analysis program has pushed the boundaries of the discipline and created analytic tools of wide application in other areas; the Statistics program has significantly moved forward the research fields in Big-Data, statistical learning, and deep-learning methodologies; and programs in Mathematical Biology have been partially funded by CIF21 Software Reuse Venture who support computational and data-enabled science and engineering.

3. Program planning and prioritization process (internal and external) that guided the development of the portfolio.

From conversation with the Program Officers, it was clear that programs were committed to creating portfolios that include early career researchers, have demographic and geographical diversity, are balanced across subdisciplines, and have various of types of institutions

represented in them. Nevertheless, the committee noticed that the process of equalization would benefit from having a set of more explicit “guiding principles” and more documentation of the decision process in support of these guiding principles.

4. Responsiveness of program to previous COV comments and recommendations.

While the subcommittee found evidence that DMS has been responsive to prior COV reports, we identified four issues that still persist:

a) Understanding and evaluating the Broader Impacts criterion is an issue. Reviewers should be provided with examples of what represent meritorious broader impacts as well as examples of what may seem to be broader impacts but are not. We suggest that DMS (and all of NSF) consider providing a more detailed guided training session for reviewers and panelists.

b) While all reviews and documentation address the criteria, there is significant variability from review to review and panel summaries as well as subjective judgments. Thus, we suggest exploring the use of a template to ensure that reviewers address all the criteria objectively and in detail, as well as training to avoid subjective language.

c) The 2016 COV suggested that DMS might "extract information from its own data, information that could be used internally, as well as for future COV meetings". It is encouraging to see that DMS has provided additional demographic data on proposers, panelists, and reviewers in the Summary Data. To improve the data analysis, we suggest that all PIs could be asked, after grants are funded, to provide gender/minority information as well as to remind reviewers and submitters to update their FastLane profile.

d) There is insufficient internal documentation in the equalization processes regarding specific targets that need to be met (this issue was already mentioned in the 2013 COV report).

e) Involvement of reviewers from PUI in proposals from such institutions as mentioned in Q.II.1.

IV. Questions about Portfolio.

1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity?

For the panels that we reviewed in Analysis, Mathematical Biology, and Statistics, there seems to be a good representation in the disciplines and an excellent portfolio quality.

2. Are awards appropriate in size and duration for the scope of the projects?

The average award duration across the different disciplinary panels seem to be similar, around 3 years. For the time period reviewed by the COV, there was some variability across disciplines with respect to dollar amount requested and dollar amount awarded. We noticed that Statistics has the lowest median rate of dollar amount awarded / dollar amount requested (50% for Statistics and 62%-75% for other disciplinary programs), while not being the program with the largest mean requested dollar amount (309K for Statistics and 247K-345K for other disciplinary programs). After consultation with Program Officers in the discipline, it was noted that this seems related to efforts made to bring the rate of funded proposals in Statistics to the level of other DMS disciplines.

The limited funds in DMS means that award sizes are too small to fully support graduate students and postdocs in some of the programs in mathematical sciences. The lack of support for postdocs has a negative impact on progress in mathematics and on the development of the workforce.

3. Does the program portfolio include awards for projects that are innovative or potentially transformative?

All awards that we reviewed proposed innovative approaches. The DMS annual reports provided highlights of funded proposals in the programs we reviewed. These highlights were evidence that DMS is funding cutting-edge research and accelerating breakthroughs in these fields.

4. Does the program portfolio include inter- and multi-disciplinary projects?

The program's portfolio includes inter- and multi-disciplinary projects, especially in the Mathematical Biology and Statistics programs.

5. Does the program portfolio have an appropriate geographical distribution of Principal Investigators?

There was geographical distribution among states of awarded proposals. However, we were not clear how NSF wants to measure "appropriate" geographic distribution, that is, specific targets or guiding principles to quantify "appropriate."

6. Does the program portfolio have an appropriate balance of awards to different types of institutions?

Regarding the balance of awards to different types of institutions (e.g., R1, R2, PUI), an appropriate response seems to require a listing of the categories of institutions, the number of institutions in each category, the number of institutions that applied in each category, and the number of institutions awarded. We did not have, or requested, this data at the program level.

7. Does the program portfolio have an appropriate balance of awards to new and early-career investigators?

The sample of programs reviewed in Analysis, Mathematical Biology, and Statistics do have an appropriate balance of awards to new and early-career investigators.

8. Does the program portfolio include projects that integrate research and education?

The program portfolio does include projects that integrate research and education.

9. Does the program portfolio have appropriate participation of underrepresented groups?

Regarding appropriate participation of underrepresented groups, this is a difficult question to answer, partly because race is often unreported. The percentage of PIs that did not report minority status in the years reviewed is increasing over time: 2016 –10.43 %; 2017 – 12.78%; 2018 – 14.68%; 2019 – 18.71%. We suggest that DMS consult with survey methodologists to help define experiments to look at effective ways of eliciting answers to these questions.

Among the panels we reviewed, we also noted differences in the rate of funding (i.e., ratio of funded proposals to submitted proposals) for women, with Analysis having a higher rate than Statistics and Mathematical Biology.

We note a point that should be investigated further: a review of a subset of panels suggests that while male and female PIs submit collaborative and sole-PI proposals at the same rate, female PIs seem more likely to receive funding from collaborative proposals while male PIs seem to receive more sole-PI proposals. If this observed trend within our sample is statistically significant across DMS programs, it might be an indication that female PIs are less entrusted with scientific leadership than reports of PI awards by gender implies.

10. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.

DMS projects are very relevant to national priorities, agency mission, relevant fields and constituents needs. NSF plays a unique and fundamental role driving basic science in our country and moving us forward. The fundamental research supported by DMS is key for a lot of innovation and further technology development that takes place in the US and makes the nation a leader in many areas.

11. Additional comments on the quality of the projects or the balance of the portfolio:

The quality of the proposals is really outstanding, in fact, there are many excellent proposals that cannot be funded.

V. Other Topics.

Below are additional suggestions to address some of the issues identified in this report:

Increase diversity

- Consider developing a database where researchers can submit their CVs and areas of expertise for POs to use as a resource to recruit a more diverse pool of reviewers.
- Consider providing incentives to reviewers.
- Encourage the participation of early career researchers including postdocs in review panels.
- Recruit more reviewers from PUI institutions in order to increase the potential pool of proposals coming from these institutions and ensure participation of at least one reviewer for proposals from PUIs.
- Consider rotating the panel meetings throughout the U.S. in order to increase the number of reviewers and encourage more potential reviewers to attend.
- In certain fields, consider recruiting reviewers working in non-academic institutions such as National Labs or industry research labs.

Increase continuity in the reviewing process

- Consider creating a membership program for reviewers (similar to NIH) where reviewers agree to serve for multiple years but still have ad hoc reviewers join to address specific expertise.
- Consider requesting that all “resubmissions” provide a one-pager addressing the reviews/suggestions provided by the review panel (also similar to NIH). As part of the review criteria, reviewers could be asked to score whether or not the previous reviews have been appropriately addressed.

Increase objectivity of the reviewing process

- Ensure transparency of number of reviews required for proposals, especially for multi-disciplinary programs.
- Consider providing reviewers and panelists with a more detailed and quantitative or objective rubric.
- In order to address ambiguity in the interpretation of Broader Impacts, we suggest refining this criterion as: general contributions to society, the development of human resources, and contributing to diversity and inclusion.
- Provide and require reviewer and panelist training that address unconscious bias and reviewing guidelines.

Subcommittee D
Special Research Programs

Subcommittee Members:

Tara Brendle (Lead)	University of Glasgow
Peko Hosoi (Co-lead)	Massachusetts Institute of Technology
Pavel Bochev	Sandia National Laboratories
John Chen	Bowling Green State University
Jessica M. Conway	Pennsylvania State University
Clint Dawson	University of Texas, Austin
Chris Jones	University of North Carolina, Chapel Hill
Mary Meyer	Colorado State University
Deborah Nolan	University of California, Berkeley
Jill Pipher	Brown University
Alexander Polishchuk	University of Oregon
Adriano Salerno	Bates College

I. Quality and effectiveness of the program's use of merit review process.

I.(1) Subcommittee D found the merit review process to be clear and efficient overall. We were pleased to discover a lot of "eyes on the process". We initially had concerns about how conflicts between panels are resolved in making funding decisions, but discussion with the program officers clarified this issue for the most part. We recommend DMS consider joint panels wherever appropriate and feasible. In addition, there was some initial confusion about how panels are conducted when reviewing proposals that are joint with another agency, and further explanation in the review analysis would have been helpful. For example, in the AMPS panel, there seemed to be an initial screening by DOE, but this was not clear in the documentation sent to us. In at least one panel, there was a full panel, ad-hoc reviews, and small panels, and further explanation in the review analysis of how these influenced the final decisions would have also been helpful.

I.(2) The merit review criteria are addressed in individual reviews, in panel summaries, and in Program Officer review analyses. However, there was a lot of variability in the quality and detail of the written reviews. Also there was a lot of variability in the reviews in the analysis of Broader Impacts; although we now understand that these are considered in aggregate across the program. We found in some of the reviews activities that were deemed as having Broader Impacts may not in fact have such impacts (e.g. placing code in a repository at the end of a project, or applications to another area of mathematics). If there is a rubric for broader impact tailored to the specifics of a particular panel, or subject area, it would be good to see this in the review analysis. This is especially true in these specialized panels that are meant to address more focused topics. The review analysis is generally useful in determining how the merit review criteria weigh into the final decision.

I.(3) The individual reviewers giving written reviews sometimes provide substantive comments to explain their assessment of the proposals, but this is not uniform. Sometimes the feedback is very detailed and sometimes it is very terse. This is true across most panels we reviewed. Also some comments can be somewhat inappropriate (or even offensive). We know that NSF tries to redact these types of statements and the POs receive bias training. We would encourage the POs to ensure that reviewers back up negative statements with facts. We found that reviewers who made comments like “this proposal ranked in my top third” helped us understand how to calibrate scores from different panelists.

I.(4) The panel summaries overall do a good job of providing the rationale for the panel consensus (or reasons consensus was not reached). However, their quality varies across different panels. We encourage the POs to continue to ensure that panel summaries explain the consensus process for reaching a decision on the ranking of the proposal, especially in cases where it is not self-evident from the individual reviews.

I.(5) The review analysis and other documents in the jacket were generally thorough and
helpful in understanding how decisions were reached. More detailed minutes of meetings where decisions are made would also be helpful. Also if more than one panel reviewed a proposal, more explanation of how the panels’ recommendations were reconciled would be helpful.

I.(6) The Subcommittee found that overall, the documentation provided to the PI containing the rationale for the award/decline decision was of good quality.

I.(7) Additionally, we commend NSF for their thorough review process. As with any analysis, one can second-guess how specific panels were handled or how decisions were made but, overall, we find the process to be fair and thoughtful. Some big themes that came up in multiple panels are further clarity on what counts and what doesn't count as Broader Impacts, and the decision process when there is more than one panel involved in reviewing the proposal.

II. Questions concerning the selection of reviewers.

II.(1) The program made use of reviewers having appropriate expertise and/or qualifications. In the special programs we looked at, the panels were composed of reviewers with adequate background. Some proposals were sent to external experts as necessary. The NSF gets a good mix of senior, mid-career, and junior reviewers from different institutions and demographics. As one example, one of the AMPS panels in this program consisted of 12 people (5 female) with a range of expertise across applied mathematics and statistics, including one panelist with experience in biomolecular engineering. As another example, in an NIGMS panel there were 5 experts from life sciences in addition to mathematics and statistics. Statistics in particular is well-represented and for good reason: a majority of proposals are aimed at developing methods for big biomedical data for different applications.

We appreciate the challenge, given the fields, scope and multiple institutions involved, of convening a panel of experts for FRGs. Great effort is made to get expert review of content and also feedback on the FRG-specific criteria.

II.(2) The program recognized and resolved conflicts of interest when appropriate. NSF gives a good deal of training to aid in self-identification of COIs. Review analyses show anyone with a COI left the room while relevant proposals were being discussed. We commend the DMS for their care and their transparency in addressing COIs. In some cases, however, COIs seem unnecessarily restrictive, which drastically reduces the pool of potential reviewers and panelists. This seems to be an issue for panels that consider multi-institutional proposals such as FRGs and possibly for panels in smaller research areas. See **V.(3)** below for suggestions and recommendations.

II.(3) Additionally, we note that the short timescale from proposal submission to panel meeting may make it difficult for panelists to write detailed reviews. The Subcommittee suggests that either the submission dates could be moved earlier, or that letters of intent could be used so that program officers can get started on forming panels earlier. The latter could be used to identify potential conflicts of interests and areas of expertise that are required, in order to facilitate the panel selection process.

III. Management of the program under review.

III.(1) The subcommittee agreed that the programs under review are very well-managed in almost all cases. These are highly impactful initiatives. The reviewed panels represented a broad spectrum of proposals, running the gamut from theoretical analysis research to specific application-driven research. The review analyses indicate that the program officers are deliberative in their decisions and take into account many factors. We were very impressed at the efficient and effective program management from reviewers' reports to panel discussion, panel consensus, and decision delivery. It was disappointing, however, to see the lack of funds impacting both the size of project budgets and the ability of DMS to fund more worthy projects.

The FRG program is acknowledged as a flagship program of DMS and, as such, it rightly commands considerable effort and focus of program officers across the division. It is funded differently from other initiative programs in that it has apparently no intrinsic budget but each project is funded by a combination of funds from a central pot at DMS and funds assigned to the relevant specific disciplinary program; the latter taking away from the funds that programs would otherwise use for individual investigator awards and other initiatives such as RTGs. The individual programs seem to react to this funding structure in different ways with resulting imbalances between numbers of FRGs across areas of mathematics. The central FRG management team should work to ensure a greater uniformity across areas in both funding levels and reporting documentation. On the latter point, some disciplinary programs kept minutes or notes about the decisions made, while others merely reported award and declinations in the individual jackets.

The FRG program presents a massive opportunity for DMS to make progress in diversity. We suggest adding an explicit, mandated element of mentorship in FRG proposals, perhaps with an analogy to ROAs in mind; see also **IV.(6).** The mentorship could bring in diversity on all axes: for instance, early-career researchers at the institution from one of the lead PIs, early-career female co-PIs and/or PIs from minority-serving institutions (HSIs or HBCUs). The FRG projects encompass exciting new directions that often combine different areas of mathematics in exciting ways. The idea has always been to fund short-time scale projects with specific ambitious goals to resolve an outstanding challenge. As such, the teams have mostly been made up of well-known and established researchers at top institutions. These teams should find the inclusion of early-career researchers as a boon, not a drag, on the project. To bring in such broader participation within these projects will take a deliberate act on behalf of DMS. We suggest that DMS provide incentives in the form of additional subsidies to disciplinary programs to fund projects that add this kind of mentorship.

There are generally a number of challenges in bringing to life programs that involve agencies outside NSF. We were struck by how well the DMS program officers navigated these tricky waters. There are, of course, many advantages to working with other agencies, not the least of which is the increased funding for mathematical work that is opened up. But many compromises are required of the DMS program officers in the way proposals are reviewed and also attributed to the programs, as instances were noted whereby proposals were funded directly by the partner agency midstream and thus withdrawn from the joint pool. Despite these many challenges, the DMS program officers are to be commended for developing so many opportunities and for their flexibility in program organization that facilitates these many initiatives.

The AMPS program was a particular challenge as this represented an almost completely new area to DMS. The first year of this program was 2017, and it is evident that the program officers were finding their way. We noted significant improvement in the handling of AMPS proposals in 2019 with a use of the panel that adhered more closely to standard DMS practices.

The CDS&E program suffered from serious underfunding in 2016 but seemed to improve in 2018. In one panel in 2016 most of the funded projects received only $100K for three years. This would be insufficient to even cover a student at some major universities.

Finally we note that DMS program officers are not always given their due! The Joint DMS-NIGMS program, a joint NSF/NIH initiative attracts impressive proposals, some of which are ultimately funded by the NIH only. Any funding acknowledgements of the resulting work acknowledges the NIH only and not the excellent joint initiative with NSF. We suggest adopting a mechanism for acknowledging the DMS program in addition to the funding source in order to credit and further advertise this important program.

III.(2) It is the very nature of the initiatives in this group of programs (or panels reviewed) that they are highly responsive to emerging research and education opportunities. The

statistics program is particularly notable for its response to the Big-Data research and education initiatives, and these have resulted in many new inroads of mathematical sciences into this hot area. The DMS-NIGMS initiative has also shown the responsiveness of DMS to the significance of big data. The DMREF program is designed to boost the emerging intersection of materials and data. We commend the excellent work being done by the DMS program officers in ensuring a key role for mathematics in this rapidly growing area. The Computational Mathematics program is very responsive to emerging research opportunities and at encouraging high-risk, high-reward research. This is precisely the role computational math should play as it serves as a bridge to other disciplines in science and engineering, and thus as a conduit for carrying mathematical ideas into mainstream application areas.

III.(3) The subcommittee found that the program planning and prioritization process (internal and external) that guided the development of the portfolio met a high standard overall. The programs this Subcommittee looked are cross-cutting. They involve partnerships across the NSF directorates as well as partnerships with other federal agencies, such as Office of Electricity Delivery and Energy Reliability, National Geospatial Intelligence Agency, National Library of Medicine, Nations Institute of General Medical Sciences, and their parent organizations the Department of Energy and the National Institutes of Health. These interdisciplinary integrative programs create excellent opportunities for researchers in mathematical sciences to contribute to transformative research, and they are excellent examples of NSF's long-range agenda to support pioneering research through their 10 Big Ideas campaign. Overall, the outcomes from these programs have been excellent. The final list of funded projects meets several goals such as supporting cutting-edge research, encouraging junior faculty, encouraging gender and ethnic diversity, supporting geographic diversity, diversity among universities, etc. To that extent, the portfolios meet these goals.

The program planning and prioritization process was somewhat opaque from the documentation provided. However, we attribute this lack of clarity to the complexity and variety of these programs. Extensive discussion with the Program Officers revealed the nuances of running these collaborative, cross-cutting programs. DMS Program Officers closely communicate with the relevant partner organizations and directorates. We have ascertained that they are well run and that sound decisions are being made in program planning and prioritization. The flexibility of the decision-making process allows for funding potentially transformative proposals.

Focused Research Groups in the Mathematical Sciences (FRG) projects are entirely within disciplines in DMS, and we found that they are run differently in the different disciplines. For example, Statistics manages their FRG program separately and sees the program as an opportunity to identify exciting innovative projects. Alternatively, Probability and Foundations cooperate on the selection of projects to these disciplines, and Applied Math rarely chooses to fund projects through this program. We understand that each discipline views the purpose of the program slightly differently, and appreciate that it serves to fill funding gaps in various ways.

III.(4) The previous COV report regarded these special research programs as strong, successful and efficient, and we concur. Although we did not necessarily have access to material directly related to changes made in response to the previous report, the continued stellar outcomes from these programs suggest that challenges are being successfully navigated.

There are two ongoing issues raised in the previous report that our subcommittee would like to note here. The first relates to the FRGs. The 2016 report states: *"The small number of proposals in FRG awarded each year (4) makes it challenging to create a balanced distribution among fields and to support diversity in this portfolio. …. In 2013, there were no female PIs or co-PIs on any of the awarded FRG grants. In 2015, there were only four female PIs or co-PIs on the awarded FRGs. Minority representation among PIs is low. Our committee felt that this program should make stronger efforts in diversity."*

Our subcommittee found that distribution across fields and diversity in the portfolio continues to be a challenge with the FRGs. We also believe that NSF is aware of this issue and, while progress may be slow, it is not through lack of will or effort on the part of the leadership; this is a difficult problem to solve and one that most institutions struggle to address. Despite the magnitude of these challenges however, we believe that NSF may have a golden opportunity to offer a small amount of guidance and incentives that could nudge the FRGs to become a cornerstone in the diversity pipeline (as described in **III.(1)** above) while maintaining their status as a flagship research initiative.

The second ongoing issue that arose in both our analysis and the 2016 COV report concerns the evaluation of e.g. the DMREF proposals by two separate panels. In many cases, the two panels arrived at very different conclusions, which is both unsurprising and appropriate given the different areas of expertise and the different lenses through which the panels viewed the proposals. However, as discussed above in sub-question 3, the process by which these conflicting panel assessments are resolved was not clear. One suggestion that was made in the 2016 (and 2013) report was: *"We concur with the recommendation made in COV 2013 that mathematical scientists participate in those DMR panels evaluating DMREF proposals."*

IV. Questions about Portfolio.

IV.(1) The overall distribution of awards in the panels reviewed by Subcommittee D covers well a diverse range of topics and applications in data science, computational topology, optimization, material science, health sciences, and national security. For example, the DMS-NIGMS panels (P180321, P170450) show a reasonable breadth of topics in applications and in awards (e.g. awarded in P180321 include research on the genome, molecular dynamics, and stem cell homeostasis in plants). In the discussion with the POs we learned that the partner agencies for the ATD were very satisfied with the high quality of the proposals and in particular highlighted results that they learned at a PI workshop.

IV.(2) Awards were not always appropriate in size and duration for the scope of the projects. In some panels that we sampled, very few proposals got funded at the requested amount. This is discouraging and seems to aim at funding as many proposals as possible. However, significant budget reductions for interdisciplinary and collaborative proposals can have detrimental effect on the ability of the teams to perform the work as they are structurally very different from single PI projects. The latter were also subject to some dramatic reductions sometimes greater than 50%.

IV.(3) This committee feels that most of the awarded projects are innovative or potentially transformative. Examining the panel summaries and the contents of the awarded proposals, innovation and potential for transformation are two key components frequently addressed in the intellectual merit, which is a one of the two critical parts in the evaluation of NSF proposals. The committee was impressed by the quality of the funded proposals across different fields in the portfolio.

IV.(4) A majority of the panels reviewed by this subcommittee are for programs that are designed to be interdisciplinary and to fund crosscutting work and collaborations across disciplines. For example, the DMS-NIGMS panels are evaluated for impact on both biology and mathematics. This drives a lot of Computational Mathematics research and the number of collaborative teams is encouraging.

IV.(5) In its capacity as a National foundation, the presence and impact of NSF should be felt nationwide. Our members found that the individual programs showed reasonable geographic distributions, for the small numbers of awards per panel. We also examined the aggregated geographic distributions of awards for the Special Programs under our purview collectively to get a clearer picture of geographic distribution. Awards went to all the major regions of the United States (Northeast, Midwest, South, West). Examination of distribution at the state level revealed that states representing 93% of the population of the United States received awards, which we deem very good. No state or region seemed to inappropriately dwarf the funding levels of other states or regions when viewed in terms of population and number of academic institutions. We suggest outreach to schools in the slightly under-served states to increase participations in these Special Programs.

IV.(6) It is desirable for DMS to have appropriate distribution across different types of institutions, and DMS Program Directors put in considerable effort to ensure portfolio balance. Subcommittee members noted that a vast majority of awards went to R1 institutions and that, in the subset of panels where funding rates by institution was examined, PIs from R1 institutions were funded at a significantly higher rate than other institution types (R2, PUIs). Examination of aggregate data across all Special Programs bears out these trends: The vast majority of funding went to R1 schools (2/3 public, 1/3 private). However, committee members noted that projects associated with most of the panels within our purview require a lot of institutional infrastructure to execute and coordinate. In those cases, awards made to big institutions capable of providing that support were more appropriate. We therefore interpret the observation that PIs from

R2s and PUIs were more often funded though collaboration as encouraging, since DMS is managing to reach PIs at smaller institutions anyway through collaboration. One avenue to further increase participation from smaller institutions in these Special Programs are the under-utilized Research Opportunity Awards (ROAs). Thus we find that the skew is understandable and the program portfolio does have an appropriate balance of awards. We suggest that the NSF develop programs that are analogous to the ROAs for HSI and HBCUs. (See **V.(3)** below).

IV.(7) In panels where there were early-career PI applications, there was a good balance of awards between early-career, mid-career, and senior scientists. However we note that there was some variability: in some cases, we found early-career PIs were funded at a slightly higher rate than the average, while in others, the awards seemed tilted towards senior scientists. Therefore, while we feel that the Program Officers are doing a generally good job in maintaining this balance when they can, we suggest vigilance.

In order to advance mathematics it is imperative to support early-career scientists so that they may establish themselves and become productive, starting a long career making contributions. However, it is often seen as "safer" to fund senior scientists with an established record of productivity**.** The balance is a challenge to maintain, a challenge that was generally well-met by the "Special Program" panels in our purview. The bigger challenge to maintaining a good balance of awards is the application rate from early career mathematicians, which we found represents a smaller fraction than expected for these programs. One Program Office has reported success with a strategy of inviting early career mathematicians to serve on review panels, to familiarize themselves with the process**.** We suggest that Program Officers extend this commendable outreach effort to more programs.

IV.(8) The committee found that the program portfolio includes projects that integrate research and education. Although most of the research projects are technically involved, the requirement of broader impact in the evaluation process drives most proposals to include graduate education. A substantial portion of the proposals we reviewed even discuss the integration of high school education in their project, including summer camps for high school students. Some of the committee members feel that the balance is still weighted heavily towards research. With drastic budget cuts from the requested amounts, funding for students and undergraduates is a common target for reduction. Thus it becomes difficult to support students and undergraduates, so sometimes the educational benefits are unclear.

IV.(9) The Subcommittee was asked: does the program portfolio have appropriate participation of underrepresented groups? In our answer, we wish to separately address two stages in the process: application and award[4]. Based on currently

[4] See for example 'Gender Differences in Grant Submissions across Science and Engineering Fields at the NSF' in Professional Biologist, LESLIE J. RISSLER, KATHERINE L. HALE, NINA R. JOFFE, AND NICHOLAS M. CARUSO.

available information, the funding rate for proposals with women involvement increases from 32% (2016) to 40% (2019). The funding rate of proposals with underrepresented minority involvement increases from 27% to 35%. This indicates that the funding rates of these underrepresented groups are improving over the years. Out of the total of 12,974 proposals under review, the overall funding rate is 30% over the four years. The MSI funding rate is 24%. The female PI funding rate is 34%, and the funding rate for minority PI is 29%. This strongly indicates that the program portfolio has appropriate representation of underrepresented group in terms of the funded research.

The participation rates for proposals with women involvement maintain at around 20% over the four years under review. For minority participations, the percentage of proposals with underrepresented minority involvement maintains at 7% over the four years. There are 887 proposals submitted by minority PIs, the participation rate is 7%, with funding rates increase from 27% (2016) to 35% (2019). Program Officers described a variety of activity aimed at increasing the diversity of their applicant pool. We suggest that Program Officers enhance and extend this commendable outreach to underrepresented groups, for example targeting mid-career mathematicians from underrepresented groups who may not have held a grant in the past 10 years (or ever), as this is a stage where people often drop out of the grant pipeline, contributing to the so-called `Matthew effect'.[5]

IV.(10) The programs reviewed are highly relevant to national priorities, agency mission, relevant fields and other constituent needs. Special programs such as CDS&E, DMREF, and AMPS emphasize applications of mathematics and computational science to scientific and engineering applications with strong connections to real-world problems. AMPS is a new area, highly relevant to the energy security of the nation, while ATD is a partnership between DMS and NGA with relevance to national security applications. Similarly, the FRG seeks to fund projects with innovative approaches to major problems. All these programs foster better interactions between mathematicians and application scientists and build strong interdisciplinary teams. They are well aligned with the mission of DMS/NSF since such interdisciplinary efforts also tend to strengthen, rejuvenate and improve research in multiple fields of mathematics.

These programs and their success are also important for arguing for increased funding of basic research to congress and the public. See, for example, "Fueling innovation and discovery: The mathematical sciences in the 21st Century," and "Mathematics in 2025" published by the National Academies.

IV.(11) Additional comments on the quality of the projects or the balance of the portfolio: We found that projects submitted to the programs were in general of very high quality, and the portfolio is well balanced. However in analyzing the Special Programs panels under our purview we have some final concerns:

(a) The FRG program is extremely competitive. It seeks to fund projects with innovative approaches to major problems. In addition, the NSF requires a well-designed logistics of

[5] 'The Matthew Effect in Science Funding' PNAS Thijs Bol, Mathijs de Vaan, and Arnout van de Rijtd

the proposed collaboration between the members of the group. Some members were concerned about the priorities of the funding process.

(b) In terms of the balance of portfolio, members noticed that in some panels, although male and female PIs on average submit the same ratio of collaborative:sole PI proposals, female PIs awards ratio of collaborative:sole PI seems to be higher than the male ratio. The same observation was made for early-career PIs, and may also be true for under-represented groups. Since we could not do a program-wide analysis, we could not make a determination, but we feel this trend merits investigation, as it gives nuance to statistics on funding rates for these groups. For example, known statistics do not address the important question, "Are female PIs being entrusted to lead projects at the same rate as male PIs?"

V. Other Topics. In evaluating all our suggestions and recommendations, we urge the NSF to follow two basic guidelines: (1) Make evidence-based decisions wherever possible; and (2) Avoid assumptions that lead to one-size-fits-all solutions – ask the community what they need in order to avoid unforeseen consequences.

V.(3) We suggest the NSF as a whole consider the following suggestions.

- Explore double-blind reviews (remove name, institution/department). The AMS's new initiative to do this for their journals could be a useful source of information.
- Involve postdocs in the review process as a way to encourage more participation in the grant proposal process, perhaps particularly from underrepresented groups.
- Make it as easy as possible for people to participate in panels, so that program officers are better able to select the most appropriate and diverse set of reviewers. For example, it might be beneficial for panelists to be able to claim extraordinary childcare expenses to be able to serve.
- Reduce COI constraints. One idea is to reduce the timespan for conflicts to apply. For example, collaborators could be considered within a much shorter window, say, two years or even only active collaborations, and a similar reduction could be considered for co-editor conflicts. We also encourage the NSF to continue to explore the possibility of blind refereeing, which could open the possibility of reviewers looking at content rather than person-specific information such as publications, which could reduce the number of COIs. Program Officers raised this issue as well, and are no doubt well placed to make further suggestions.
- Use multi-investigator awards to bring in new people and increase diversity in people and institutions by incentivizing mentorship in flagship programs. Specifically, incentivize senior researchers to include on their team junior researchers, researchers from diverse institutions, and from underrepresented groups, by including the PI composition as part of the assessment of a project's broader impact and/or by offering supplements to projects to expand their PIs and senior personnel to include a more diverse set of researchers.

V.(5). We ask the NSF to provide benchmarks (e.g. census data) to help determine answers to, say, question IV above. Also, we learned a great deal from the program officers about the process for handling proposals in the special research program, how they were co-reviewed and even co-funded. The COV suggests that it would have been helpful to get this summary of processes, history, and measures of success in advance of reading the proposals.

Subcommittee E
Infrastructure and Workforce Programs

Subcommittee Members:

Rodolfo Torres (Lead)	University of California, Riverside
Jesus De Loera (co-Lead)	University of California, Davis
Oliver Dasbach	Louisiana State University
Ailana Fraser	University of British Columbia
Wilfrid Gangbo	University of California, Los Angeles
Sabine Le Borne	Hamburg University of Technology
Po Shen Loh	Carnegie Mellon University
Sean T. Paul	University Wisconsin
Raquel Prado	University of California, Santa Cruz
Deborah Sulsky	University of New Mexico
Tony Várilly-Alvarado	Rice University
Nathaniel Whitaker	University of Massachusetts, Amherst

This subcommittee was charged to evaluate the MSPRF, CAREER, RTG, infrastructure, CBMS, REU, EDT, MSGI, and BPI programs. The report and suggestions of the subcommittee are described in the answers bellow.

I. Questions about the quality and effectiveness of the program's use of merit review process.

1. Are the review methods (for example, panel, ad hoc, site visits) appropriate?

For the most part yes but the answer**,** like most in this evaluation, often depends on the program.

From the information in the jacket, the review process for the MSPRF program was not very transparent to this subcommittee. Fortunately, we later received in conversations with the Program Officers a very detailed account of recent changes. The changes made the process faster, but a remaining concern is the lack of explanation to the PIs about the funding decisions. Applicants for MSPRF receive only three scores from reviewers. We realize that the large volume of submissions is the main reason for this review process, but we encourage DMS to provide meaningful feedback to help junior researchers improve their proposals. We agree that for MSPRF the change to group applicants by areas, as opposed to having all in the same panel, is an improvement, but since there are no reports it is hard to evaluate if the process is fair or appropriate.

Overall CAREER proposal reviews are quite detailed, even more than regular discipline specific proposals, but the program has become so competitive that the differences between awarded and rejected proposals are harder to justify.

We learned from conversations with Program Officers that some infrastructure programs are not using panels. This seems appropriate for this type of proposals as it expedites the review process.

The MSGI is an exciting new program that addresses the growing need of graduate students to train for careers in business, industry, and government. Its evaluation process is new and interesting as there is a joint review of applicants with Oak Ridge National Laboratories. It replaces a similar program, EDT. The changes from EDT to MSGI will hopefully increase applicant placement in the internships.

2. Are both merit review criteria addressed by individual reviews, panel summaries and in Program Officer review analyses?

Except for MSPRF, where there was insufficient data, the merit criteria are explicitly addressed at all levels and all programs. However, more guidelines from NSF may be appropriate regarding Broader Impacts, as sometimes PIs are praised for activities which are a standard part of the profession for any PI. The weight given to both merit review criteria differs among reviewers. The Program Officer's review analysis generally addressed both criteria well.

3. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals?

The consensus in the subcommittee about this question is that there is nonuniformity across programs. There was variation in the quality of individual reviews with some being substantial and others not. It was suggested that Program Officers give more specific instructions to the reviewers in order to encourage information that would help investigators and maintain a more uniform level of reviews.

4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?

Overall the summaries are good but our impressions are program dependent. For example, for REU proposals the information was sometimes minimal. It seems difficult to compare proposals with entirely different populations to serve (e.g., top ranked R1 school versus a minority institution in a rural community).

5. Does the documentation in the jacket provide the rationale for the award/decline decision?

REU – For some subcommittee members the documentation in the jackets gives a very clear overview of the review process and decision rationale, while for others the jacket was not well-documented.

MSPRF – As explained, there was hardly any documentation in the jackets about this program, but convincing information about the rationale used was provided by the Program Officers.

CBMS, Infrastructure, and CAREER – The jackets were carefully documented.

6. Does the documentation to the PI provide the rationale for the award/decline decision?

REU – Most of the individual reviews for this program gave sufficient justifications. The Team Meeting Minutes in the Diary are sometimes a little terse but the more details can often be found in the Review Analyses of the individual proposals. However, not all of this information is of course provided to the PIs.

MSPRF – The rationale for the decisions is only given in very general terms; it would be better if some additional feedback could be provided to these junior applicants. It is hard for PIs to compare what they are told with the letter grade they receive and the final outcome for their proposals.

For the CAREER, CBMS and RTG programs the rationale provided is sufficient and clear, but in CAREER the decisions to pick from excellent applicants are so minute they become somewhat arbitrary.

In general, the subcommittee thought that most panels were providing the PI adequate information, but in some cases the decisions were not sufficiently justified. While this does not seem to be a general problem, we encourage Program Officers to refine their process of quality control on the production of reviews and summaries.

II. Questions concerning the selection of reviewers.

1. Did the program make use of reviewers having appropriate expertise and/or qualifications?

The reviewers in the panels appeared qualified, knowledgeable, and diverse. The reports included enough technical details, which indicates they had some level of understanding and familiarity with the proposals. REU panels had an impressive pool of reviewers, but it was suggested that the panels include representation from a diverse range of fields of research to help increase the mathematical diversity of the award portfolio.

2. Did the program recognize and resolve conflicts of interest when appropriate?

The possible conflict of interests and their resolution are usually well-explained in the “Review Analysis” documents. DMS appears to seriously and carefully handle these situations.

3. Additional comments on reviewer selection:

More information about the solicitation of panelists would be useful for future COVs. It would be interesting to know the statistical information of who is asked to serve in the panels (e.g. requests to non-Ph.D. institution faculty). Overall, the pool of reviewers in some of the programs analyzed by Subcommittee E in the Infrastructure programs seems to be more diverse in terms of the type of institutions represented than in the disciplinary panels, however we did not have enough data to make any conclusions. On the other hand, the subcommittee was glad to see the presence of junior researchers in several of the panels reviewed.

III. Questions concerning the management of the program under review. Please comment on the following:

1. Management of the program

DMS is to be commended for being very effective despite its limited resources. Although the different DMS programs conduct their business in a variety of ways, they all seem to function with equal efficiency. In particular, the CAREER, REU, RTG and MSPRF programs are extremely important for advancing research and training for the next generation of researchers and worthy of more resources. Some specific comments about them follow.

CAREER –This is one of the top programs to identify highly excelling and promising leaders across all of NSF, and the Program Officers are doing a very good job at managing the resources for this program. There is, however, no formal assessment of the program's impact. It is not clear that researchers supported by career awards are more successful in the long run than researchers supported by regular awards at potentially lower cost. The current assessment of this program is somewhat based on anecdotal information rather than specific metrics. We suggest a systematic study about the impact of the program.

REU – There were a number of proposals from minority serving institutions in the pool of applications. Some of these proposals that did not receive funding seem to have good potential. We suggest DMS to engage in active mentoring PIs from minority serving institutions to increase their future competitiveness.

RTG – We found this program to be successful and well managed.

MSPRF – As mentioned above there are concerns about the feedback junior researchers receive. We stress that the program was already underfunded before COVID-19 and it should be a top priority to better fund it in order to help junior researchers.

2. Responsiveness of the program to emerging research and education opportunities.

All programs seem to be doing a very good job in terms of their responsiveness to emerging research areas and education opportunities.

3. Program planning and prioritization process (internal and external) that guided the development of the portfolio.

The portfolio is driven in part by the community through the proposals submitted, as neither NSF nor DMS specify the research topics or problems in their solicitations. This approach to developing science is almost unique among federal funding agencies in the US and should be preserved. We commend the Program Officers for working jointly as a team in the planning and prioritization process of the portfolio.

4. Responsiveness of program to previous COV comments and recommendations.

We noticed that similar concerns we have expressed about REUs above were raised in the 2013 and 2016 reports. We did not find any longitudinal study of their performance. DMS should double their efforts to address them. We applaud the creation of BPI to increase participation from minorities in response to earlier calls to do so.

IV. Questions about Portfolio.

1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity?

Overall, Program Officers do an admirable job trying to balance awards across disciplines, but DMS could also explore new mechanisms to find emerging topics and not leave it entirely to the programs. The REU portfolio is not very mathematically diverse, likely due to favoring topics accessible to undergraduates. (Reviewers often raised concerns about the accessibility of the proposed research.) The awards in other programs, such as CAREER, are too few in number and provide an extra challenge in balancing the portfolio.

2. Are awards appropriate in size and duration for the scope of the projects?

REU - An overview of requested and granted award size would be helpful. However, it seems that many funded projects have been funded without cuts.

MSPRF – Though the salary is still competitive, it does not appear to have changed during the period the COV reviewed. The duration of the award is good. DMS (or NSF) might consider gradual cost-of-living adjustments to handle the eroding effect of inflation. A big problem is that postdoctoral awards do not cover benefits nor salary increases mandated by union contracts.

CAREER - We learned that the $400K minimum amount awarded to successful proposals and their 5-year duration are dictated by NSF and not DMS. This translates to $80K/year on average, which is a reasonable amount for the length of the award. In conversation with the Program Officers, they indicated that the number of CAREER awards per year has increased over the years from 20 to 40, in response to the very large number of proposals they are receiving (about 240 last year). This suggests Program Officers are already doing what they can to spread these highly coveted awards within the prescriptions of the program. Some subcommittee members expressed their concern that the CAREER awards have become true "winner-takes-all situations". The rejection of so many extraordinary proposals can be highly disheartening for future generations. It is urgent to increase the number of awards through an overall budgetary increase, without sacrificing other DMS programs.

3. Does the program portfolio include awards for projects that are innovative or potentially transformative?

The CAREER awards have many examples of innovative ideas.

4. Does the program portfolio include inter- and multi-disciplinary projects?

The majority of REU sites are discipline-specific, perhaps due to the fact that undergraduate students have not yet developed a breadth of expertise. The committee observed that it could be valuable to have more REUs in applied interdisciplinary and computational domains. On the other hand, most RTG proposals were substantially interdisciplinary (at least across mathematical areas), as could be expected because they typically originated from departments as multi-faculty collaborations. MSPRF and CAREER projects were narrower in scope, perhaps because the highly competitive nature of these programs makes it more difficult for junior researchers to significantly address multiple research areas.

5. Does the program portfolio have an appropriate geographical distribution of Principal Investigators?

The NSF maintains a website showing the distribution of awards by state, with views in terms of number of awards or the dollar amount of the awards. While there is coverage across the country, there are definitely concentrations in some geographical areas. Each disciplinary program has coverage in most states and so do conferences and workshops in the mathematical sciences. In 2020, there are active CAREER awards in 33 states. Geographical coverage of REU awards is excellent, while RTG awards are active in only 24 states. Similarly, MSPRF active awards are in 24 states, with concentrations in California, Massachusetts and New Jersey. The majority of active Infrastructure awards are in California and Rhode Island.

While Program Officers are cognizant of the need to have geographical diversity, and often work diligently to try to distribute awards to achieve that, some programs are more

successful than others. There does not appear to be an explicit assessment in place at DMS of Program Officers' decisions in these matters.

6. Does the program portfolio have an appropriate balance of awards to different types of institutions?

REU – The subcommittee was glad to see that many awards went to institutions that traditionally do not receive support in the form of disciplinary grants. Some subcommittee members felt that some sites successful in terms of undergraduate publications bring together students from mostly elite institutions, but lack significant broader participation. However, the overall portfolio of REU sites is fairly well-distributed across different types of institutions.

CAREER – The geographical distribution of these awards is commendable, especially considering the rather small number of awards that are made in any given year. The vast majority of awards, however, are made to R1 institutions. This may suggest that DMS should strive for some balance.

MSPRF – Conversation with Program Officers made it clear that a balance of types of institutions and geographical considerations play a role in equalizing situations.

CBMS - There was a mix of institutions that will host a CBMS, but 86% of all awards went to Ph.D. granting institutions that lack minority representation. Data seems to indicate that submissions by minority institutions, minority investigators, or female investigators have slightly lower acceptance rate than overall proposals. This, coupled with a smaller number of submissions, translates in very small numbers of minority groups being supported. It would be desirable to see more non-Ph.D. granting programs, which serve large minority populations, receive some form of support.

Infrastructure - The portfolio seems strong and diverse. It includes some high profile, long-standing, and very successful programs like the IAS/Park City Mathematics Institute and the AMS MRC program. Established infrastructure seemed to be a factor in getting funded: among the funded proposals in a panel reviewed by this subcommittee, three had "established infrastructure", IAS, AIM, and MAA. Hence, it could be beneficial if award recipients are asked to produce some documentation that might help share with others know-how on establishing infrastructure.

7. Does the program portfolio have an appropriate balance of awards to new and early-career investigators?

REU –There seem to be hardly any junior PIs. This is not very surprising since running of these programs required a lot of effort taken away from individuals' research time.

RTG – The PIs are mostly senior, but a lot of Ph.D. students benefit from the support. So this type of award cuts across many different career stages and provides benefits to all.

MSPRF and CAREER – They are of course prescribed by design to early-career investigators.

Across DMS – The data available (Graph IV.7 – Proposals & Awards by PhD Age of PI) indicates that, on average, the success rate for very junior PIs is lower than for more senior ones. It would be desirable that the 30+% success rate for investigators who are 15 to 45 years passed their Ph.D. could be paralleled for new and early-career ones.

8. Does the program portfolio include projects that integrate research and education?

By design, REU and RTG programs are excellent projects to integrate research and education. Other programs, such as the Park City Math Institute and the IAS Women in Math program, have strong direct educational components and expose participants to cutting edge research as well.

The MSPRF projects typically do not stress education but rather focus on research. Indeed, the awards are often used to reduce the amount of educational work for the postdocs. Since this is the design of the program, the committee has no particular criticism.

The CAREER program is unique in its combination of education and research. The prestige of the award produces a very positive effect of celebrating contributions to education in an environment that typically stresses research over education. The committee asked specific questions about the intent of the CAREER program and the review procedures. In their response, Program Officers emphasized that the educational component and, in particular, the PI's proposed efforts to promote mathematics outside the university classroom are of crucial relevance. It is unfortunate that there are so few such awards. It would be beneficial for DMS and NSF to find more ways to support the communication of mathematics to a general audience.

9. Does the program portfolio have appropriate participation of underrepresented groups?

The data provided shows that the number of proposals from PIs with underrepresented backgrounds is small. This reflects an overall problem in the profession. We encourage DMS to continue efforts to increase participation of underrepresented groups. For example, NSF should consider sponsoring grant proposal writing workshops for new PIs.

REU – Overall, the committee felt that REUs do have good participation by underrepresented minorities. This might be attributable to distribution of funds among different types of institutions, and should be further encouraged.

MSPRF – Funding rates by gender and minority status match their percentage representation in the pool of applicants. However, the demographics of the applicant pool do not reflect those of the US population, either in its current state or in its trends. This is a serious source of concern. While this circumstance is a reflection of the lack of diversity in the profession, we encourage DMS to double their efforts to achieve broader participation in the pool of applicants.

CAREER – As mentioned, the number of these awards is too small. Too many excellent proposals are rejected and the difference in quality with those which are funded is often insignificant. We did see good representation of women among awardees, which is encouraging.

CBMS – Only two of the 18 CBMS proposals reviewed featured a female lecturer. However, among the seven funded proposals, there were 5 men and 2 women as lecturers, which was viewed positively by the committee.

10. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.

The United States has a shortage of advanced STEM talent and the programs reviewed by this subcommittee are of fundamental importance to national priorities, agency mission, and constituent needs in this regard.

Transformative innovation often originates from fundamental research. Therefore, it is important to develop a robust pipeline of talent, not only to follow in the technical footsteps of others, but to forge their own new research paths. To this end, the training programs are essential. Through the REU and RTG experiences, students immerse themselves into research, sometimes contributing to the knowledge discovery in substantial ways. The track record of REU program includes students who went on to gain significant undergraduate research awards such as the Schafer Prize and Morgan Prize.

The MSPRF program also represents an important stage in the pipeline, and provides postdocs with the funding, time, and recognition to launch their research careers. The postdoctoral period is short and intense, and it is valuable to provide postdocs with the space to focus.

The CAREER program, through its unique solicitation, is particularly suited to highlight the broader relevance of mathematics and science in general. The committee strongly supported the phrasing of its objective: “to identify junior PIs with the potential to serve as academic role models in research and education and to lead advances in the mission of their department or organization.”

11. Additional comments on the quality of the projects or the balance of the portfolio:

The overall quality of the proposals in most programs is outstanding and seems to improve over the years. The process is so competitive that it forces PIs, both junior and senior, to develop better and better proposals. The amount of time devoted to this by PIs should not be underestimated and it is one of the reasons it is so frustrating for PIs to have their proposals declined. Resources should be sought so more institutions and more individuals benefit from NSF awards.

MSPRF and CAREER set the awardees apart and provide a track record that establishes an advantage (often rightfully deserved) in future competitions of all type. But, an equally big disadvantage accrues to those who are not funded, even if they have very competitive projects. The fact that these awards tend to go to individuals from elite R1 universities can further make the playing field uneven for individuals from diverse backgrounds. Perhaps this is a problem that the profession should address, but NSF should not compound it. The committee stresses the importance of making every possible effort to ensure diversity of gender, minority status, geography, and type of institution; and to advocate for more funding to increase the number of awards. We stress that the MSPRF postdoctoral awards are truly crucial for the profession and, as a result of the pandemic, junior researchers would truly benefit from an increase of funding in this special moment of crisis. In particular, it would be encouraging if emergency action equivalent to those back in the 2008 economic crisis would be implemented.

We encourage DMS to conduct a systematic analysis to compare the trajectory of the CAREER award recipients vs. those who only received regular NSF awards, and to determine the impact of these awards on the profession as a whole.

Subcommittee F
Mathematical Sciences Institutes, Simons Research Centers for Mathematics of Complex Biological Systems, and Partnerships between Science and Engineering Fields and the Theoretical Foundations of Data Science (TRIPODS) Institutes

Subcommittee Members:

David Manderscheid (Lead)	University Tennessee, Knoxville
Sara Y. Del Valle (Co-lead)	Los Alamos National Laboratory
Anna Amirdjanova	Federal Reserve
Kelly Bickel	Bucknell University
Steven Lee	DOE Advanced Scientific Computing Research
Bernard Mair	Association of Public and Land-Grant Universities
Jennifer Morse	University of Virginia
Simon Thomas	Rutgers University
Carol Wood	Wesleyan University
Tommy Wright	Census Bureau

The subcommittee was charged with assessing the review process, management, and portfolio of the Institutes, Simons Math Bio, and TRIPODS research centers between 2016 and 2019. We organize this subcommittee report by the five major questions. We answer each question and provide recommendations for each of the three grant programs the panel examined separately, as the programs are quite different in number of grants, size of grants, and maturity of the program.

I. Questions about the quality and effectiveness of the program's use of merit review process.

6. Are the review methods (for example, panel, ad hoc, site visits) appropriate?
7. Are both merit review criteria addressed
   a) In individual reviews?
   b) In panel summaries?
   c) In Program Officer review analyses?
8. Do the individual reviewers giving written reviews provide substantive comments to explain their assessment of the proposals?
9. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?
10. Does the documentation in the jacket provide the rationale for the award/decline decision?
11. Does the documentation to the PI provide the rationale for the award/decline decision?
12. Additional comments on the quality and effectiveness of the program's use of merit review process:

Institutes: The quality of the documentation in the proposal jackets and the overall effectiveness of the review process in collecting substantive input from the reviewers was very good. The panel merit reviews and subsequent decisions for reverse site visits and institute management review were appropriate. Intellectual merit and broader impact criteria were addressed in the individual reviews, panel summaries and Program Officer review analyses. Individual reviewers provided substantive comments to explain the strengths, weaknesses, and overall rating for the proposals. The panel summaries were well written and appropriately synthesized the individual reviewers' evaluations, even when there was a diversity of ratings. The review process and basis for award/decline decisions by the Program Officer was well documented in the proposal jackets.

SIMONS: The review methods for MathBioSys were appropriately extensive. In the first round of the review, a panel of experts in mathematics and biology reviewed the proposals. Then MathBioSys Working Group (comprised of program officers and staff from the NSF Divisions of Mathematical Sciences, Molecular and Cellular Biosciences, and Integrative and Organismal Systems, and from the Simons Foundation) decided which projects would be invited for a reverse site visit at the Simons Headquarters. After the reverse site visits, the Working Group recommended that four of the six finalists be funded.

The individual reviews, the panel summaries, and the review analyses addressed intellectual merit and broader impacts explicitly. Compared to other disciplinary programs, the individual reviews and the panel summaries in this program paid more attention to the Broader Impacts criterion, which is natural given the interdisciplinary nature of the program. The review analyses are very effective at putting the funding decision in the context of the reviews, panel discussion, and additional considerations. For each of the six High Priority projects that were not selected for a reverse site visit, the review analysis gives a clear explanation for the Working Group's decision. These explanations tended to involve "big picture" considerations. The documentation to PIs whose proposals were highly rated by the panel but did not receive funding was inadequate. The DMS is to be commended however for the outstanding review process overall.

TRIPODS: The subcommittee found the quality and effectiveness of the program's use of merit process was good including the review criteria, reviews, and summaries. However, there was variability in both the length and quality of the reviews. Suggestions to PIs whose projects were not selected for funding did not always provide substantive information regarding why those proposals were declined. Additionally, although most of the reviewers address the merit criteria the information is not always presented in a uniform way, making it difficult to find strengths and weaknesses.

DMS should consider providing more direction to the reviewers regarding what needs to be addressed in order to make the evaluation process more uniform and equitable. The NIH and EU could be looked at for examples. Consideration should be given to ensuring that every proposal gets a summary review, including those that are relatively weak.

II. Questions concerning the selection of reviewers.

1. Did the program make use of reviewers having appropriate expertise and/or qualifications?
2. Did the program recognize and resolve conflicts of interest when appropriate?
3. Additional comments on reviewer selection:

Institutes: The reviewers' range of expertise and research qualifications were well suited for evaluating institute proposals. The reviewer selection for panels must be hard due to conflicts of interest, given the number of mathematicians who participate in the institutes and the number of host universities. The members of the panels and the site visits team came from a good range of institutions and subdisciplines. They are geographically and gender diverse. We cannot comment on race and ethnicity as that data (panelists, and RSV) was not provided. Reviewers were conscientious in addressing the merit review criteria and providing substantive comments.

The jacket documentation clearly identifies reviewers with conflicts of interest. There are detailed explanations when a potential COI did not rise to the level of an actual COI.

SIMONS: The reviewers had the appropriate expertise and qualifications. For example, the panel seemed to have a careful balance of math-biologists (about 50%), biologists (about 25%), and mathematicians with expertise in tools that would be used in these centers (about 25%). We did notice that several proposals were reviewed only by mathematicians and suggest that, given the interdisciplinary nature of these proposals, DMS pay special attention to disciplinary representation at both the review and panel level.

The subcommittee applauds the detailed treatment of COIs and finds that the DMS takes the highest level of care with respect to any and all COIs. We did note that of the 16 panelists, three were female, which seems low.

TRIPODS: The reviewers for all proposals were well qualified with more than sufficient expertise to serve as reviewers for the TRIPODS program. In some cases, more reviewers from application areas or education/pedagogy backgrounds might have helped to improve the quality of the reviews. COIs were handled well.

Representation of members of underrepresented groups should be given further consideration to achieve panels that are more balanced. For example, on one panel only one of the seven reviewers was a woman.

III. Questions concerning the management of the program under review.

1. Management of the program.
2. Responsiveness of the program to emerging research and education opportunities

3. Program planning and prioritization process (internal and external) that guided the development of the portfolio.
4. Responsiveness of program to previous COV comments and recommendations.

Institutes: The current portfolio is consistent with the goal of focusing the best mathematical minds on problems of particular importance and timeliness. Broader outreach to the mathematical community and public is also achieved, along with greater impact on major impact on workforce development. This impact is stronger than that which can be achieved by individual investigator awards. Attention to emerging research and education opportunities is a prominent point of discussion in the panel reviews and apparent in the breadth of institutes.

We found almost all of the Program Officer's reports to be clear and concise, giving a good idea of the rationale for decisions. The review analysis for institute awards is thorough in explaining the balancing and prioritization process that guides the Institute portfolio development. The several step reviewing processes are well documented and especially so when the reviews differ from the site visit reports.

There has been a successful effort on the part of the NSF to improve their communication with the institutes; the POs attend institute board meetings, annual directors' meetings, and regular discussions with the PIs about new initiatives that arise during the grant period. This pays off in excellent management of this portion of DMS's portfolio. In particular, the POs set clear expectations for the institutes. Communication efforts have further enhanced during the COVID-19 period; NSF is working closely with institute directors to discuss the reallocation of funds to support appropriate activities during the pandemic and consistent with the mission. DMS is to be commended on its plan to monitor closely the progress of the newest institute. Finally, the move from non-competitive renewals to competitive renewals with competitions each five years has greatly improved the program by providing more oversight.

SIMONS: The subcommittee commends the DMS on the management of this program and finds it fair, effective, and transparent. The post-award management of this program is strong. DMS program officers had frequent communication with center leadership to ensure that the centers are fulfilling their missions and specific proposed activities. Examples include a kick-off meeting with the leadership from all four centers, site visits (with one exception due to COVID-19) with a follow-up report, recommendations, and annual reports from the centers that are checked against proposed activities and site visit recommendations.

The MathBioSys program demonstrates the responsiveness of the DMS to emerging research and education opportunities. Prior to the MathBioSys program, an NSF workshop related to the "Rules of Life," NSF priority identified the need for focused support for mathematical approaches as necessary. As a result, NSF collaborated with the Simons Foundation to fund partially four centers "*to enable innovative collaborative research at the intersection of mathematics and molecular, cellular and organismal biology, to establish new connections between these two disciplines, and to promote*

interdisciplinary education and workforce training." This private-public sponsorship as well as the subsequent funding of a fourth center, indicates DMS's flexibility and responsiveness to the needs of, and opportunities in, the greater science community. It is remarkable how DMS and the Simons Foundation, two organizations with quite different cultures, have worked together so successfully. DMS should consider replicating this successful partnership.

The program planning and prioritization process was exemplary. The subcommittee observed special care paid to the development of new collaborations, the impact of centers outside of their home institutions, and the creation of dual innovations in mathematics and biology. Similarly, strong educational and training components for undergraduates, graduate students, and postdoctoral fellows as well as center outreach activities were heavily valued. Special attention was also paid to research areas that were not already being heavily funded in order to maximize the impact of the centers. All of these priorities seem very appropriate and are quite in line with the advertised mission of the MathBioSys program.

While the MathBioSys program did not exist at the time of the previous COV assessment, several of its activities indicate responsiveness to previous COV comments about the Institutes program as a whole. These included setting clear expectations, support of members of underrepresented groups, and detailed assessment plans. However, the types of evaluation metrics used in assessments varied across the centers. To make comparisons across centers and with the overall mission of the program easier, DMS might consider requesting a summary document in the annual report that includes metrics consistent across the centers.

TRIPODS: TRIPODS +X has separate educational and collaborative research subprograms, which aim to support a very wide range of different types of projects. Thus, in addition to the typical components of program management, there are additional challenges related to assembling of the panels with expertise in very diverse areas and avoiding conflict of interest, when proposals involve large collaborative teams. Overall, the program's management has been efficient and the program is likely to continue to generate enthusiastic responses from the scientific and educational communities.

TRIPODS +X is designed to be very responsive to emerging research and educational opportunities since the program encourages an in-depth exchange of ideas, methods, and information across disciplines and finding new dimensions in "*big data*." The program has already funded a number of innovative projects that promise to solve data-intensive problems across different domains of science and engineering and contribute to the development of the emerging field of data science.

The TRIPODS program seeks balance across sub-areas, aiming to provide support to a wide variety of research and educational projects involving data science. Although the program broadly plans to contribute to harnessing the data revolution and growing

converging research, the program's portfolio is fully shaped by the research communities and their educational needs and has no additional prescribed components.

IV. Questions about Portfolio.

1. Does the program portfolio have an appropriate balance of awards across disciplines and sub-disciplines of the activity?
2. Are awards appropriate in size and duration for the scope of the projects?
3. Does the program portfolio include awards for projects that are innovative or potentially transformative?
4. Does the program portfolio include inter- and multi-disciplinary projects?
5. Does the program portfolio have an appropriate geographical distribution of Principal Investigators?
6. Does the program portfolio have an appropriate balance of awards to different types of institutions?
7. Does the program portfolio have an appropriate balance of awards to new and early-career investigators?
8. Does the program portfolio include projects that integrate research and education?
9. Does the program portfolio have appropriate participation of underrepresented groups[6]?
10. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.
11. Additional comments on the quality of the projects or the balance of the portfolio:

INSTITUTES: The portfolio of awards is in line with DMS's stated purpose. It is clearly hard to break into the elite company of the institute program, and the requirement to serve nationally is taken seriously. One institute did break through in the open competition we reviewed, as another was ramped down with a final year.

The current portfolio has several institutes that are familiar to the mathematical community and most were renewed. Continuity of successful institutes within the overall portfolio is commendable. The Institute for Mathematical and Statistical Innovation is a welcome addition and provides an additional focus on statistics in the mathematical sciences. The overall institute portfolio appears to be well balanced and the jacket documentation is thorough in the justification for each of the individual selections.

The geographic balance of the institutes is appropriate and the new institute at the University of Chicago improves the situation. The institutes draw participants from across the United States and the world. The Program Officers did an excellent job of adjusting budgets to reflect the relative merits of the proposals. Participation of

[6] NSF does not have the legal authority to require principal investigators or reviewers to provide demographic data. Since provision of such data is voluntary, the demographic data available are incomplete. This may make it difficult to answer this question for small programs. However, experience suggests that even with the limited data available, COVs are able to provide a meaningful response to this question for most programs.

underrepresented groups has been a significant concern and the Program Officers are working with the institutes to address it. The subcommittee recommends that their actual success be monitored closely. The subcommittee was encouraged to see how many of the proposals for new institutes integrated diversity, equity, and inclusion into their mission.

SIMONS: There is a good balance of awards. The four funded centers each have a different theme/mission; and, under these general umbrellas, include a wide variety of innovative research projects at the interface of mathematics and molecular, cellular, and organismal biology. These centers are appropriately funded and work on innovative and potentially transformative research. This program is in support of the NSF 10 Big Ideas: Understanding the Rules of Life and is thus aligned well with the NSF mission.

The four centers are very inter-disciplinary and provide excellent geographic distribution. While the centers are all at research institutions, this seems entirely appropriate. It is noteworthy that one of these institutions, University of California Irvine, is a Hispanic Serving Institution.

While one Co-PI and a handful of senior personnel associated to the centers are early-career researchers, most of the PIs and Co-PIs are researchers that are more senior. Given the amount of experience required to run a center, this is entirely appropriate.

The training of inter-disciplinary researchers and various types of outreach appear to be priorities for all of the centers. The subcommittee did not have data to determine whether there was appropriate participation of underrepresented groups.

TRIPODS: The program has a reasonable balance of awards across a variety of disciplines and sub-disciplines in both its research and its educational components. The award sizes and duration are in line with expectations. The program supports a good number of innovative projects and has a potential to identify and fund transformative research All the projects in the program portfolio are inter- and multi-disciplinary. The program is aligned with national and NSF priorities around harnessing big data.

The geographical distribution of the Principal Investigators did seem appropriate as did the balance investigators and of types of institutions, although see the comments in the next section for a further discussion. The TRIPODS + X RESEARCH awarded 24 proposals that focused on research, though educational benefits for students and postdocs were mentioned. Similarly, the TRIPODS + X EDU/VIS awarded proposals focused on a variety of educational activities**.** There is appropriate participation of women and underrepresented groups for the TRIPODS + X EDU/VIS. Specifically, 5 out of the 9 proposals awarded were led by female PIs and at least one self-identified underrepresented minority, However, the TRIPODS + X RESEARCH did not appear to have a proportional representation of women or underrepresented groups.

V. Other Topics.

1. Please comment on any program areas in need of improvement or gaps (if any) within program areas.
2. Please provide comments as appropriate on the program's performance in meeting program-specific goals and objectives that are not covered by the above questions.
3. Please identify agency-wide issues that should be addressed by NSF to help improve the program's performance.
4. Please provide comments on any other issues the COV feels are relevant.
5. NSF would appreciate your comments on how to improve the COV review process, format, and report template.

The subcommittee did not have data on race/ethnicity, and limited gender identities of PI's and co-PI's for proposals submitted and awarded for each program. This information would be helpful. The subcommittee also note the increasingly high rate of nonresponse to these questions, thus, the subcommittee recommends working with survey methodologists trained in the social and behavioral sciences study these very problems of dealing with the asking of sensitive questions. DMS might consider engaging one or two such experts to help in improving response rates. Additionally, DMS may encourage PIs to update their FastLane profile on a regular basis.

Some members of the subcommittee struggled with the notion of what the word "*appropriate*" means in the questions from Section IV, "Questions about Portfolio" above. They suggested providing more detail as to what is desired and specific metrics on how that should be evaluated.

Finally, a member of the subcommittee suggested making use of probability sampling methods in future COVs. In particular, a COV report might benefit from being more quantitative and databased. This is not to minimize the value of qualitative features of the COV report. Quantification can help strengthen the results expressed qualitatively. To this end, NSF may want to consider making more use of (probability) sampling methods as part of the COV review process to add a little more structure to what each member of COV does to carry out the COV work. This could include, for example,

- how to select panels reviewed,
- how many panels to select,
- how to distribute the panels among the members of the COV subcommittees,
- how each COV subcommittee member determines the number of proposal processes he/she will study,
- what each COV subcommittee member measures for each proposal process he/she will study,
- what targets should the assessment may be evaluated against.