SUPPORTING STATEMENT PART B: COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1. RESPONDENT UNIVERSE AND SELECTION METHODS

Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

This data collection will involve a survey of all individuals who have submitted proposals to NSF (applicants) and/or served as reviewers for NSF proposals between fiscal year (FY) 2018 and FY 2020 (for 2021 Merit Review Survey) and between FY 2020 and FY 2022 (for the 2023 Merit Review Survey). Primary objectives of the survey are to assess applicant and reviewer perceptions of the NSF merit review process, which reviews proposals and awards funds to researchers across a variety of fields of study, and to gather information on the experiences of key subpopulations. In addition, the proposed survey will gauge whether satisfaction levels with the merit review process have changed since the most recent survey on this topic conducted in 2019. Results from the survey will be used to improve the merit review process and promote fairness, transparency, effectiveness, and efficiency in decision making.

NSF will create the universe using administrative data to identify individuals who participated in the merit review process as applicants or reviewers during this time frame. This data collection is a census in which all universe members will be considered eligible and invited to complete the survey. It is expected that the 2021 and 2023 Merit Review Survey universes will each comprise approximately 87,000 individuals. The survey will include filters to identify individuals who have served only as applicants, only as reviewers, or as both to reduce burden

and avoid asking respondents nonapplicable questions. No additional screening or sample selection will be required. The survey will be conducted via the Web and will take approximately 20 minutes to complete.

In 2015, 2017, and 2019 NSF conducted the "Assessment of Investigator and Reviewer Experiences Survey" (OMB # 3145-0215), a similar survey of the same population. These surveys garnered 30 percent, 36 percent, and 30 percent response rates, respectively. Based on the enhancement to the outreach strategy planned for 2021 and 2023 Merit Review surveys (described in section B.2), and the response rates of prior iterations of the Merit Review survey, the estimated survey response rate for this study is 40 percent. Outreach to universe members will occur via email using email addresses provided by NSF. Data collection, outreach, and communication strategies are discussed in section B.2 below.

Survey data collection as a whole, and particularly web survey data collection, has experienced a decline in response rates in recent years as the public experiences survey fatigue and is less willing to participate in online surveys.[1] Despite these challenges, lower rates can be mitigated by sample recruitment and outreach strategies such as: 1) capitalization of existing NSF communication platforms to convey legitimacy and importance of the survey; 2) increased advanced confirmation of email address quality by third-party vendors; and 3) use of secondary email addresses, when available, to contact nonresponsive universe members. Additionally, while survey response is an important data quality indicator, research suggests that lower response rates are not necessarily indicative of nonresponse bias. This is in part because bias has been shown to be determined at the survey item response level rather than at the

---

[1] Daikeler, J., Michael Bosnjak, and Katja Lozar Manfreda. 2020. "Web versus other survey modes: an updated and extended meta-analysis comparing response rates". *Journal of Survey Statistics & Methodology*, 8: 513-539.

respondent level, as it is dependent on the relationship between the survey item and the response pattern.[2] Furthermore, this can particularly be the case when there is additional information available about nonrespondents that can help to calculate nonresponse bias analyses that can counter-balance lower response rates.[3] This is the case for the Merit Review Survey for which the universe file, constructed from NSF administrative data, will include data about all participants, such as individuals' participation as an applicant and/or reviewer, associated directorate, and select demographic data. Planned nonresponse bias activities as relates to the proposed survey effort are discussed in greater detail in section B.2 below.

B.2. PROCEDURES FOR THE COLLECTION OF INFORMATION

Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection**

- **Estimation procedure**

- **Degree of accuracy needed for the purpose described in the justification**

- **Unusual problems requiring specialized sampling procedures**

- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden**

The proposed survey instrument will provide updated information on applicant and reviewer perceptions of and experiences with the merit review process. Topics include:

1. Assess applicant and reviewer perceptions of, and satisfaction with, various aspects of the merit review process.

2. Document the time burden the merit review process places on reviewers and applicants.

---

[2] American Association for Public Opinion Research (2016), "Evaluating survey quality in today's complex environment: American Association for Public Opinion Research Report.".
[3] Meterko, M., et al. 2015. "Response rates, nonresponse bias, and data quality". *Public Opinion Quarterly*. Vol 79, No. 1: 130-144.

3.   Examine applicant and reviewer perceptions of the quality of reviews and of proposals.

4.   Assess the changes in applicant and reviewer perceptions of burden, satisfaction, and quality between the 2019 and 2021 surveys and then 2021 and 2023 surveys.

5.   Examine the variation of applicant and reviewer perception of satisfaction, burden, and quality by key population subgroups, including race/ethnicity, gender, and disability.

6.   Describe the extent to which NSF's reviewer orientation video is correlated with awareness of different types of cognitive biases and the use of strategies to reduce cognitive bias and to provide constructive feedback.

7.   Describe the extent to which the elimination of annual proposal deadlines affected reviewer and applicant burden, perceptions of proposal and review quality, and satisfaction with the merit review process.

8.   Describe applicants and reviewers experiences with student support programs as well as what NSF application and funding support is associated with the receipt of financial support from NSF as an undergraduate or graduate student.

The proposed data collection includes two separate survey rounds, the first scheduled

to begin in late August 2021 and the second in August 2023.

The outreach plan for both survey rounds will include the following email types:

- **Prenotification email from NSF**: Three days before the survey's distribution, NSF will

    send a prenotification email from its servers to the intended survey recipients as

    identified by the universe file. The prenotification email will introduce the study

    contractor, describe the survey effort, and underscore the importance of responding to

    the survey. This email will also notify participants that they should expect an invitation

    to participate in the survey from the contractor to arrive within the week.

- **Initial survey invitation email**: Universe members will receive an email invitation

    containing a unique link to the web survey and requesting that they complete the

    survey.

- **First nonresponse follow-up email**: Within two weeks after the initial invitation is sent, nonrespondents will be sent a follow-up email from the web-survey platform. Emails will only be sent to individuals who have not yet begun the survey (nonrespondents) and individuals who have partially completed the survey (partial completes).

- **NSF reminder email**: NSF will send a reminder email approximately two weeks after the nonresponse follow-up email to all individuals who were invited to participate, reminding those who have not yet finished the survey to do so and thanking those who have already completed it.

- **Last-chance email**: The study contractor will send a final reminder email less than one week before the end of the data collection period, making a final request for their participation. Unique emails will go to two groups: 1) individuals who have not yet begun the survey (nonrespondents) and 2) individuals who began but did not complete the survey (breakoff).

Incorporating an adaptive design outreach approach, the 2021 data collection effort will include an experiment of the initial survey invitation and first follow-up emails wherein the control group receives the standard email (based largely on the 2019 survey invitation email) and the treatment group receives a much shorter email. The purpose of the experiment is to understand if the shorter or longer email formats yield higher completion rates. Past literature has examined the effects of a variety of email contacting strategies on web survey response rates,[4] such as if a respondent is notified in the initial email invitation that they will receive a

[4] Klofstad CA, Boulianne S, Basson D. Matching the Message to the Medium: Results from an Experiment on Internet Survey Email Contacts. Social Science Computer Review. 2008;26(4):498-509. doi:10.1177/0894439308314145

reminder or if a web survey question is embedded within the web survey link included in the email.[5] While there has been limited research conducted on how the length of email invitations affect web survey response rates, some literature has shown the length of the email survey invitations can affect overall response rates, with longer, more detailed invitations yielding higher response rates.[6] However, while brevity of the email invitation text may be helpful in encouraging respondents to read the entire email, the text should not be shortened at the expense of critical information that may provide legitimacy to the survey effort. This experiment is intended to determine the appropriate balance of email length to yield the most survey responses.

To maximize the benefit of this experiment on the 2021 survey effort, we will conduct data collection in two sample waves, applying the outcomes we learn in wave 1 to the email outreach strategy for wave 2. Wave 1 will include a smaller portion of the universe split equally into treatment and control groups on which the initial invitation and follow-up emails would be tested. Estimating a total universe size of approximately 87,000 applicants and reviewers, we expect that wave 1 sample will include approximately 10,000 applicants and reviewers, which will be equally split into two subgroups. Power analyses indicate that a wave 1 sample size of 10,000 would allow for the detection of a 3 percent difference in response rates. Analysis of wave 1 survey completion rates for each email type would then inform which type of each of the two emails would be sent to wave 2. The wave 2 sample will include the remaining universe

[5] Liu, Mingnan, and Nick Inchausti. 2017. "Improving Survey Response Rates: The Effect of Embedded Questions in Web Survey Email Invitations." *Survey Practice* 10 (1). https://doi.org/10.29115/SP-2017-0005.
[6] Kaplowitz MD, Lupi F, Couper MP, Thorp L. The Effect of Invitation Design on Web Survey Response Rates. Social Science Computer Review. 2012;30(3):339-349. doi:10.1177/0894439311419084

(approximately 77,000 applicants and reviewers) and will include the same email to all wave 2 sample members.

The 2021 data collection schedule will run approximately eleven weeks. Table B.2.1 describes the sample wave assignment and the timing of each email. Templates for each email type can be found in the references appendix.

Table B.2.1
Sample Wave Assignment and Email Timing

Data collection week	Wave	Group	Email type	Appendix
0	1	All	NSF pre-invitation	B.1
1	1	Experimental	Initial survey invitation	B.2
		Control		B.3
2				
3	1	Experimental	First nonresponse follow-up	B.4
		Control		B.5
4				
5	1	All	NSF reminder email	B.6
	2	All	NSF pre-invitation	B.1
6	1	All	Last chance (nonrespondents)	B.7
		All	Last chance (breakoff)	B.8
	2	All	Initial survey invitation	B.2 *or* B.3
7				
8	2	All	First nonresponse follow-up	B.4 *or* B.5
9				
10	2	All	NSF reminder email	B.6
11	2	All	Last chance (nonrespondents)	B.7
		All	Last chance (breakoff)	B.8

The lessons learned from the 2021 experiment design will be applied to the 2023 survey data collection schedule. The schedule of outreach activities for the 2023 Merit Review survey data collection will following a similar approach to that described above for 2021 with all

universe members contacted in a single wave using the email templates that were demonstrated to be most effective in increasing completion rates in 2021.

Throughout the survey field period, the contractor will monitor a project-specific email address and toll-free number included in communications to respondents to provide answers to survey-related questions or troubleshoot technical issues. Potential respondents may choose to opt out of the data collection at any time by clicking the "unsubscribe" link embedded at the bottom of all survey invitation and reminder emails. A survey management system will track completed cases, partially completed cases, and nonresponse cases throughout the data collection period. These data will inform who will receive follow-up email prompts and will be used to calculate response rates.

B.2.1 Statistical Methodology for Stratification and Sample Selection. The Merit Review Survey will be conducted with a census of applicants who submitted proposals to NSF or served as reviewers between FY 2018 and FY 2020 (for the 2021 effort) and between FY 2020 and FY 2022 (for the 2023 effort). Therefore, there will be no stratification or sample selection. NSF will provide the full list of the target population for the survey using administrative records. Survey invitations will be sent to all individuals included in the final universe file.

Given the breadth of NSF programs and granting opportunities and the number and nature of subgroup analyses that are planned, NSF has determined that a census of applicants and reviewers is most appropriate for this survey. This is the same approach taken for the administration of the 2015, 2017, and 2019 iterations of NSF's "Assessment of Investigator and Reviewer Experiences Survey" (OMB # 3145-0215). NSF includes eight separate directorates and program offices; within these are over 30 divisions that house over 600 distinct programs.

Each program announces their own solicitations and receives and awards proposals. The diversity in fields of study and funding opportunities necessitates a census of applicants and reviewers to ensure all the different fields, ways in which funds are received, and applicants' and reviewers' satisfaction with and thoughts about the merit review process are reflected in the responses. NSF is interested in learning about the experiences among participants within each directorate so a census has been determined as the best choice. In addition, beginning with the 2021 Merit Review Survey, the study contractor will conduct analyses by population subgroups of particular interest to NSF, breaking down findings by race/ethnicity, gender, and disability status in order to better understand variation in the respondent experience in the merit review process. These new subgroup analyses will give NSF a clearer understanding of the diversity of participants and their experiences within the merit review process in order to promote more equal participation.

B.2.2   Measurement/Estimation Procedures. Based on prior NSF surveys of the same population combined with an enhanced data collection outreach design, we anticipate that the response rate will be 40 percent. Because we estimate the response rate will be lower than 80 percent, we plan to conduct a non-response bias analysis and implement a non-response weight adjustment, if appropriate, to compensate for missing data and reduce the potential for bias.

A key to understanding the potential for nonresponse bias is to understand patterns of nonresponse and how those patterns relate to survey characteristics. For instance, in surveys with persons as sampling units, response propensity is often related to age and/or gender. If age or gender is then a correlate of the survey characteristic of interest, the potential for

nonresponse bias is high. If such a potential is high, survey researchers must seek methods to mitigate this potential. The survey contractor will conduct a nonresponse bias analysis to examine any known differences between respondents and nonrespondents and to illuminate any potential bias introduced by nonresponse. In addition, expanded analyses for this effort will consider nonresponse when examining outcomes associated with key population subgroups, which have not been conducted for previous iterations of the survey. Some of the components of this study will be to:

- ► Understand the levels of nonresponse, both unit and item

- ► Understand differences between respondents and nonrespondents

- ► Understand the impact that those differences may have on survey estimates and adjust accordingly

- ► Consider information that could be used to mitigate the potential for nonresponse bias

Results of this analysis will be included in the final report.

Although the Merit Review Survey will be a census, and thus not subject to potential bias due to sampling error, non-response error, or the inability to obtain responses to survey items from the entire sample, is a potential concern. Because nonresponse is typically not proportionately spread across sub-groups of interest to the survey, there is a potential for nonresponse bias if survey characteristics and responses to survey items differ by sub-group. Nonresponse adjusted weights will be used to compensate for this type of disproportionality. Characteristics that are related to both survey estimates and to response propensity may be used as weighting variables to support a nonresponse weighting design that mitigates the

potential for nonresponse bias. The study contractor will formulate an appropriate nonresponse weighting approach for the 2021 Merit Review Survey based on the final response rate, item nonresponse, and universe member characteristics.

To further address item nonresponse, the study contractor may use existing administrative data provided by NSF as part of the universe file to populate nonresponse survey items for respondents. This will be conducted for key demographic variables only (i.e., race/ethnicity, gender). These data may also be used for nonresponse bias analyses.

Furthermore, the study contractor will conduct imputation on survey response data in order to populate missing items. Determination of imputation for each survey item will be based on the percent of missing values and the accuracy of the imputation method. The study contractor will begin by calculating a response rate to each survey item by reviewer and applicant characteristic and use these data to determine survey items for which imputation might be most critical. Then the study contractor will apply statistical modeling, such as logistic regression models, to predict likelihood of response from reviewer and applicant characteristics. Those analyses will be expanded to quantify the accuracy with which these models can predict survey items and response and determine if the statistical models predict well enough to be considered for imputation. Other forms of imputation for items with missing values may also be considered, such as "borrowing" reported values from respondents with similar characteristics.

Once these steps have been completed, data management and analyses will be conducted using SAS. Variations in output, per type of analysis, will depend on what measures are appropriate for the variable and the measurement level for each defined variable. For

example, we will conduct descriptive analyses including the calculation of frequencies for categorical variables and means and medians for continuous and discrete variables. To assess whether burden varies by subgroup (e.g., directorate, race/ethnicity, gender, disability status, early career status, or reviewer type), we will conduct a variety of statistical tests depending on the type of variable (i.e., z-tests for each continuous variable, chi-square tests for each binary or nominal variable, and Mann-Whitney tests for each ordinal variable). To assess whether any changes since the 2019 survey varies by subgroup, we will conduct a variety of regression analyses depending on the type of variable (i.e., linear regressions for each continuous variable, logistic regressions for each binary variable, multinomial logistic regressions for each nominal variable, and ordinal logistic regressions for each ordinal variable).

B.2.3. Degree of accuracy needed

We anticipate a final sample size of 34,800 respondents biennially. For many descriptive analyses we will be reporting on the entire sample of respondents (i.e., applicant and reviewer burden and overall satisfaction with the merit review process). However, research questions also call for analyses by key population subgroups (such as directorate, race/ethnicity, gender, disability status, early career status, and reviewer type). Because of this expansion, some subgroup analyses may be based on very small sample sizes. To account for this, the study contractor will conduct a preliminary power analysis to determine whether subgroups have sufficient sample sizes to support reporting and comparisons. Disclosure review will also be conducted on any data tables to ensure that respondent identification cannot occur within the smallest cells.

B.2.4. Unusual problems requiring specialized sampling procedures

There are no unusual problems requiring specialized sampling procedures.

B.2.5. Any use of periodic (less frequent than annual) data collection cycles to reduce burden

The survey will be conducted biennially with one data collection cycle in 2021 and a second separate effort in 2023.

B.3. METHODS TO MAXIMIZE RESPONSE RATES AND THE ISSUE OF NONRESPONSE

Describe methods to maximize response rates and to deal with issues of nonresponse. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

The data collection methodology described above incorporates strategies intended to maximize response rates for the target population of applicants and reviewers. The following strategies will be used to help achieve the estimated response rate:

- ***Convenience of Web survey instrument:*** Web has been selected as survey mode to reduce respondent burden and improve response. Email addresses are the most reliable contact information available for universe members and will allow for quick, cost-effective communications between the survey team and potential respondents. In addition, the web survey will allow respondents to complete the survey at their convenience, which is important for a population that is busy with additional work responsibilities. Respondents will automatically be routed to questions that apply to them, and they can breakoff at any time and resume the survey, as needed.

- ***Quality of email addresses:*** The survey contractor will contact universe members using email addresses from NSF administrative records. The email addresses will be cleaned, deduplicated, and validated as part of universe file creation to ensure that most recent

email addresses are used in outreach. Beginning in 2021, the study contractor will validate email addresses with more than one third-party vendor in order to potentially identify more invalid email addresses than in the past, in advance of data collection. This will allow for the immediate use of alternate email addresses, when available. In addition, secondary email addresses provided with the NSF administrative data will be incorporated into the universe file so that additional outreach can be made to nonrespondents if primary emails prove inactive or outdated.

- ***Participation of NSF sponsorship in outreach activities:*** Universe members are invested the merit review process. NSF is a funding source for approximately 27 percent of all federally supported basic research conducted by America's colleges and universities and has developed an excellent relationship with researchers and reviewers all of which have an interest in ensuring an effective and efficient merit review process. Given the saliency of the topic and existing relationship with NSF, NSF sponsorship of the study will aid in gaining cooperation. NSF staff will serve as a bridge between the survey contractor and potential respondents by sending out prenotification emails that include a letter from a high-ranking NSF official encouraging participation and emphasizing the importance of the survey. In addition, beginning in 2021, NSF will share information about the upcoming data collection as part of their general communications with the broader research community. These communications will provide legitimacy to the study and emphasize its importance to the continual improvement of the NSF merit review process.

- **Targeted outreach to nonrespondents:** Survey prompting will be scheduled using a series of targeted emails to nonresondents which are designed to preempt anticipated concerns and questions that universe members may have that could prevent their participation. Following the survey prenotification email from NSF, the study contractor will be sent an initial survey invitation, including the study purpose and the personalized web survey link. A follow-up email will then be sent to nonrespondents within two weeks stressing the importance of participation, explaining how the survey results will be used to improve the merit review process, and encouraging participation. This will be followed by another prompt from an NSF official and then a final "last chance" email for universe members who have not yet completed or have partially completed the survey.

- **Survey support from project (for technical issues and survey content questions).** A helpdesk email will be provided to respondents on all outreach emails and the survey so that they may contact a representative with questions about the survey. We have successfully used this approach in the past to maximize the chances a potential respondent will proceed with the survey rather than dropping out as soon as they encounter a technical issue or question about the survey content. In addition, the help desk can answer questions about study legitimacy, purpose, and content.

- **Survey revisions:** The 2021 survey instrument has been shortened from prior iterations to drop items that are not necessary for analyses to streamline questions. This is expected to reduce respondent burden and potentially increase response rates. In addition, to increase item response for demographic questions, revised survey items will

be included in the 2021 survey instrument, including those asking about race/ethnicity, gender, and disability. These items follow NSF-wide guidance on question stem wording and more inclusive response categories. Increased item response rates for these survey questions will aid in critical population subgroup analyses as well as nonresponse bias analyses.

As discussed in section B.2. above, the study contractor will conduct non-response bias analysis and implement a non-response weight adjustment, if appropriate, to compensate for missing data and reduce the potential for bias.

B.4. TESTS OF PROCEDURES

Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

The 2021 survey instrument is draws heavily on the instrument used for the 2019 Merit Review Survey. No new survey topics will be introduced that would require external pre-testing or validation. All items having appeared in previous iterations of the Merit Review Survey, dating back to 2015. Revisions to survey items have been limited so that comparisons can be made to the prior survey. They include:

1. Updates to the survey reference period and other relevant time frame adjustments;

2. Removal of redundant question text to improve clarity and reduce burden;

3. Removal of survey items not required for 2021 analyses in order to reduce burden;

4. Revision to demographic question to align with NSF best practices in survey methodology as defined by the National Center for Science and Engineering Statistics (NCSES).

The survey instrument will be programmed using Qualtrics web survey software and tested thoroughly by the study contractor prior to launch. Testing steps will confirm: 1) accuracy of survey question text, skip patterns, and range checks; 2) functionality of the survey across multiple web browsers and platforms (including desktop, tablet, and mobile devices); and 3) review of test data to confirm variable format and structure are accurately captured.

The 2021 Merit Review survey data collection will include a test of outreach emails as indicated in section B.2 above.

B.5. CONSULTANTS

Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

NSF has contracted with RIVA Solutions and Insight Policy Research to conduct this survey. Table B.5.1 identifies specific individuals who will consulted on the design and will be responsible for collecting and analyzing the data. The Contracting Officer Representative (COR) for the contract providing funding for the evaluation, Bernice Anderson, will be responsible for receiving and approving all contract deliverables. Her contact information is also included in table B.5.1.

Table B.5.1
Individuals Responsible for Statistical Aspects and Data Collection and Analysis

Name	Title (Project Role)	Organizational Affiliation and Address	Phone Number
Kaye Burton	Program Manager (Project Manager)	RIVA Solutions 8000 Westpark Dr #450 McLean, VA 22102	703-509-1135

Amanda Hare	Senior Study Director (Subcontract Manager)	Insight Policy Research 1901 N. Moore Street, Suite 1100 Arlington, VA 22209	703-758-5009
Marietta Bowman	Senior Survey Researcher (Data Collection Lead)	Insight Policy Research 1901 N. Moore Street, Suite 1100 Arlington, VA 22209	571-200-7932
Richard Griffiths	Statistician (Data Analysis Team)	Insight Policy Research 1901 N. Moore Street, Suite 1100 Arlington, VA 22209	410-302-6303
Zoe Jacobson	Research Analyst (Data Analysis Team)	Insight Policy Research 1901 N. Moore Street, Suite 1100 Arlington, VA 22209	571.200.7933
Bernice Anderson	NSF COR	National Science Foundation 2415 Eisenhower Avenue Alexandria, VA 22314	703-292-7216