

AN EVALUATION OF THE NATIONAL SCIENCE FOUNDATION'S GRADUATE RESEARCH FELLOWSHIP PROGRAM: OVERVIEW OF PROPOSED STUDY

Program Overview

As part of the National Science Foundation's (NSF) continued commitment to graduate student education in the U.S., the Graduate Research Fellowship Program (GRFP), which began in 1952, seeks to promote and maintain advanced training in Science, Technology, Engineering and Mathematics (STEM) fields by annually awarding roughly 2,000 fellowships¹ to U.S. citizens, nationals, and permanent residents for graduate study in research-based programs. The goals of the program are:

- To select, recognize, and financially support individuals early in their careers with the demonstrated potential to be high achieving scientists and engineers, and
- To broaden the participation of underrepresented groups, including women, minorities and persons with disabilities, in science and engineering fields

Underpinning the program goals are NSF's broader strategic organization goals, including that of performing as a model organization. To achieve this goal and to become a model Federal steward, representing excellence in management and fiscal responsibility, NSF seeks to "*learn through assessment and evaluation of NSF programs, processes, and outcomes; continually improve them; and employ outcomes to inform NSF planning, policies, and procedures*" (http://www.nsf.gov/news/strategicplan/nsfstrategicplan_2011_2016.pdf, pp. 16-17, italics in original). Thus, excellence in management is an underlying goal of each NSF program, including the GRFP.

Purpose and Need for Study

NSF is seeking to conduct a study that has three purposes:

- Provide descriptive information related to the GRFP program goals on the demographics, educational decisions, career preparation, aspirations and progress, as well as professional productivity, of GRFP Fellows and comparable non-recipient applicants and national populations of graduate students and doctorate recipients.
- Provide rigorous evidence of the impact of the GRFP on individuals' educational decisions, career preparations, aspirations and progress, as well as professional productivity
- Provide an understanding of how the program is implemented by universities and whether and how specific program policies could be adjusted to make the program more effective in meeting its goals.

Previous studies of the GRFP were largely completed in the mid-1970s to the mid-1990s. The most recent study, published in 2002, examined GRFP Fellow cohorts through 1993, and is now dated. The NSF GRF program collects data on an ongoing basis through multiple sources that is used for program management and accountability purposes. These sources include reports from the GRFP Committee of Visitors, annual surveys of the review panelists, comments from Fellows and university GRFP coordinating officials, and data compiled from the applications. In addition, GRFP Fellows submit annual activity reports, the format for which was revised in 2010 to include activities that contribute to career preparation such as acquisition of research skills and other professional skills, data on career plans,

¹ The annual number of fellowships awarded increased from approximately 1,000 to 2,000 in 2010.

internships, and other sources of financial support. The data are tracked over multiple years to examine trends and identify gaps that need to be addressed in subsequent competitions. However, the data, while useful, offered limited information for prior years. In addition, they did not address program impact or implementation.

Thus, NSF needs current information on several fronts to inform future decisions about program structure and design that cannot be addressed either with NSF data or existing national databases. These include: (a) how GRFP Fellows differ from their peers in terms of demographics, educational trajectories, and career outcomes; (b) the impact of the GRF program on Fellows in terms of educational trajectories, career outcomes, and professional productivity; (c) the effect of the GRF program on institutions in terms of student diversity and quality in STEM graduate programs; and (d) program implementation. The current study, being conducted for NSF by NORC at the University of Chicago, is intended to address each of these areas.

Research Questions and Study Approach

Research Questions

The study focuses on the following research questions:

RQ1. *What is the impact of the GRFP fellowship on the graduate school experience?*

RQ2. *What is the impact of the GRFP fellowship on career outcomes?*

RQ3. *What are the effects of the GRFP on institutions?*

RQ4. *Is the program design effective in meeting program goals?*

While RQ1 and RQ2 are framed in terms of impact, a necessary component of the research is examining how the Fellows compare with peers in terms of demographics, aspirations, educational trajectories, career outcomes, and professional productivity, to help address the program goals. RQ3 and RQ4 are designed to address both the GRF program goals as well as the underlying NSF strategic goal of excellence in management.

Data Sources

To address the research questions, the study will use both primary and secondary data sources. In terms of primary data collection, the study will:

- Collect data from Fellows and carefully-matched counterparts (QG2 Honorable Mentions)² through a survey (GRFP Follow-Up Survey) that asks about graduate school experiences, educational attainment, career outcomes, employment characteristics, and professional productivity (RQ1 and RQ2). The survey also asks Fellows about the influence of program elements (choice, flexibility, and monetary value) on their decision to enroll in and successfully complete STEM graduate programs (RQ4).
- Collect data from two samples of institutions:
 - a. In-depth data from six institutions (“institutional site visit sample”) gathered from site visits which will encompass in-person interviews with administrators, faculty, and staff to understand: (1) the current climate, (2) perceived impact of the program on Fellows,

² The quality groupings (QG) refer to the categories assigned to each GRFP participant upon applying to the fellowship program. QG1 is the highest ranking an applicant can receive. The study sample includes the highest two categories: QG1 and QG2 applicants.

institutions, and programs, (3) program implementation, and (4) GRFP policies. The in-depth data will be used to address RQ3 and RQ4.

- b. Targeted data from a larger sample of 20 institutions (“institutional phone sample”) gathered from shorter phone interviews more narrowly focused on implementation and specific GRFP policies (RQ4 and, to a more limited extent, RQ3).

These two institutional samples serve two different purposes, each suited to the type of data to be collected. By limiting the site visit sample to six institutions and broadening the institutional phone sample to include 20 institutions, we balance the need for in-depth data collection with the goals of minimizing respondent burden and collecting data from a broader pool of institutions.

- Review and analyze similar federal fellowship programs using data collected from websites, program materials, and interviews with program officers managing these programs. This part of the study will help inform GRFP policies and best practices (RQ4). The findings will be valuable in understanding how best to support Fellows and help develop a more diverse STEM workforce.

Secondary data sources include the Doctorate Records File to provide a national context and national comparison group.

Analysis

Descriptive analysis will be used to examine the composition, experiences, and outcomes of Fellows, non-recipients, and national peers and will provide evidence of GRFP participation of underrepresented groups and trends in program selection, recognition, and financial support of early career scientists and engineers. To measure impact, the study will model outcomes using quasi-experimental methods to compare outcomes of the treatment group (Fellows) with outcomes of plausibly similar control groups (QG2 Honorable Mentions). These methods are widely accepted as the best methods on which to base causal inferences in the absence of a randomized experiment (i.e., when it is not feasible to randomly assign participants to treatment and control groups). To examine implementation, the study will use qualitative methods to code and analyze the institutional interviews and to draw out lessons learned from the interviews with program officers.

Findings and Dissemination

The data collected in this study and the analytic reports will provide a comprehensive look at the GRFP, its impact on Fellows, institutions, and the science and engineering workforce, and the extent to which the program is meeting its goals. In particular, the findings will provide information on:

- The influence the GRFP has had on the decisions, experiences, academic attainment, and career outcomes of Fellows compared with carefully-matched peers ;
- The extent to which the program has broadened the participation of underrepresented groups in STEM fields at the graduate level;
- The perceived effects on institutions in terms of student financing, enrollment, diversity and quality (among others);
- Whether specific design elements—choice, flexibility, and monetary value—are working as intended and the extent to which they are valued; and
- The need (if any) for changes in the way the program is structured to make it more effective.

Overall, the study findings will provide valuable insights to NSF on the impact of its investments in the GRFP and inform its program management. In conjunction with findings from the review of similar federal fellowship programs, the findings may prove valuable to the larger community of program

officers administering these programs as well as to the graduate education policy community in understanding how best to support graduate education and help develop a more diverse STEM workforce.

Study results will be reported to the Division of Graduate Education, Education and Human Resources (EHR)/NSF, distributed within the community of universities who participate in GRFP, and published on the NSF website. Limited print copies of the full report will be made available to NSF as well as 500 copies of a printed executive summary that can be disseminated more widely and that will be useful to a variety of public audiences. A policy brief reporting on the review and analysis of federal fellowship programs will be made available to all federal fellowship program managers and may be of interest to the larger foundation community as well. Findings of more general policy or methodological interest will be distributed more broadly, through conference presentations and submissions for publication in peer-reviewed journals.

SUPPORTING STATEMENT FOR PAPERWORK REDUCTION SUBMISSION

An Evaluation of the National Science Foundation's Graduate Research Fellowship Program

SECTION A. JUSTIFICATION

A.1. Circumstances Requiring the Collection of Data

As part of the National Science Foundation's (NSF) commitment to graduate student education in the U.S., the Graduate Research Fellowship Program (GRFP), which began in 1952, seeks to promote and maintain advanced training in Science, Technology, Engineering and Mathematics (STEM) fields³ by annually awarding approximately 2,000 fellowships⁴ to U.S. citizens, nationals, and permanent residents for graduate study in research-based programs. Accompanying this award is the expectation that NSF Graduate Fellows (referred to as "Fellows" in the remainder of the document) complete their degree and become scientists and engineers with the skills and knowledge necessary to contribute to research, teaching, and/or innovation in STEM fields.

NSF is seeking to conduct a study that has three purposes:

- Provide descriptive information related to the GRFP program goals on the demographics, educational decisions, career preparation, aspirations and progress, as well as professional productivity, of GRFP Fellows and comparable non-recipient applicants and national populations of graduate students and doctorate recipients.
- Provide rigorous evidence of the impact of the GRFP on individuals' educational decisions, career preparations, aspirations and progress, as well as professional productivity
- Provide an understanding of how the program is being implemented by universities and whether and how specific program policies could be adjusted to make the program more effective in meeting its goals.

There have been several previous studies of the GRFP. The National Research Council conducted four major studies from the mid-1970s to the mid-1990s, focusing on traditional measures of academic career success such as completion rates, time to degree, faculty appointment, success in obtaining research grants, and publications and citations.⁵ These studies used secondary data sources (such as NSF's annual Survey of Earned Doctorates (SED), Survey of Doctorate Recipients (SDR), and the NSF/NIH (National Institutes of Health) postdoctoral and research grant files) to examine completion rates and career plans of several cohorts of Fellows (1952-1972; 1967-1976; 1972-1981; and 1979-1981 cohorts respectively). The authors of these studies acknowledged limitations in the data used in the reports including limited measures used for career outcomes, lack of a credible comparison group, and need for primary data collection from students and faculty.

³ For a full list of NSF-supported fields of study, see <http://www.nsf.gov/pubs/2010/nsf10604/nsf10604.htm#appendix>

⁴ The number of annual fellowships awarded increased from 1,000 to 2,000 in 2010.

⁵ Harmon, L. R. (1977). *Career achievements of NSF graduate fellows: The awardees of 1952-1972*. Washington, D.C.: Commission on Human Resources, NRC; Baker, J. (1994). *Career paths of the National Science Foundation graduate fellows of 1972-1981*. Washington, D.C.: Office of Scientific and Engineering Personnel (OSEP), National Research Council (NRC); Baker, J. (1995). *Minority science paths: National Science Foundation Minority Graduate Fellows of 1979-1981*. Washington, D.C.: OSEP, NRC; Snyder, J. (1988). *Early career achievements of National Science Foundation graduate fellows, 1967-1976*. Washington, D.C.: OSEP, NRC.

The most recent comprehensive evaluation of the GRFP (Goldsmith, Presley, and Cooley, 2002)⁶ was based on a mixed-method analysis of data from several sources. This included: (a) secondary data analysis of the 1979-1993 Fellows using data from the SED and NSF's Cumulative Index (CI); (b) analysis of a graduate student follow-up survey of the 1989-1993 cohort of Fellows and graduate student peers in four disciplines at the Fellows' institutions; and (c) analysis of interviews with administrators, faculty, staff, and students conducted during site visits to six research universities. The 2002 study compared the highest quality award recipients (QG1) to lower-quality recipients (QG2) and non-recipients and used the survey data to compare Fellows to their disciplinary peers. While the study produced important and useful information, it was primarily descriptive and is now dated.

NSF currently collects data from different sources to inform program management and for accountability purposes. First, it collects data on the composition of program applicants by outcome of the application process (awardees (QG1 and QG2; Fellows), those ranked as high quality but not offered the award (QG2; Honorable Mentions), and declinations) and reports trends over time in terms of distribution by field of study, level of academic preparation, demographics, geographic representation, and baccalaureate institution (disaggregated by minority-serving institution status). Second, NSF collects data on the composition of the GRFP panelist pool by demographics, institution type, professional rank, geography, and by new versus return panelist status. The panelists complete a survey that is used to inform the review process and future outreach efforts to underrepresented populations. Third, NSF collects annual activity reports from Fellows while enrolled in graduate school on their research, professional productivity (including papers, patents, and inventions), research and teaching appointments, activities that integrate research and education, international achievements, and their activities to help broaden participation. The report format was revised in 2010 to gather additional information on career plans, internships, sources of financial support during the Fellowship, and acquisition of research skills and professional skills. Fellows are also asked to summarize, for public dissemination, their accomplishments over the past year and the intellectual merit and broader impact of their work. The new format implemented in 2010 will make these activity reports a rich source of data going forward. There is limited data for prior years.

There are three major drivers for the new data collection. First, the overall climate for graduate education has changed over the past two decades along with the characteristics of students enrolled in college⁷ and we need to better understand the current environment and its effect on program outcomes and on institutions hosting Fellows. Second, addressing issues of impact requires a more rigorous and sophisticated modeling approach than has been used in previous studies. Better estimates of the program's impact can help inform NSF's policies and program review. This cannot be addressed with the data that NSF collects from Fellows because issues of impact require a counterfactual (a comparison group) against which experiences and outcomes of Fellows can be compared. Third, there have been changes to the program since 2010 (including increased number of annual awards and policies related to permitted service, no concurrent federal fellowships, and affiliation with U.S. institutions, among others), largely based on earlier studies and program review. NSF now needs information on the implementation of the revised policies and their effect on institutions and students to inform future decisions about program structure and design. Again, these cannot be addressed by the data currently collected by NSF because that is not the focus of the activity reports submitted by Fellows.

⁶ Goldsmith, S.S., J.B. Presley, & E.A. Cooley (2002). *National Science Foundation Graduate Research Fellowship Program final evaluation report*. Arlington, VA: National Science Foundation.

⁷ Wendler, C., Bridgeman, B., Cline, F., Millett, C., Rock, J., Bell, N., and McAllister, P. (2010). *The path forward: The future of graduate education in the United States*. Princeton, NJ: Educational Testing Service.

To address these needs, the study will use both primary and secondary data sources and a mix of rigorous quantitative and qualitative analytic techniques. In terms of primary data collection, the study will:

- Collect data from Fellows and carefully-matched counterparts (QG2 Honorable Mentions)⁸ through a survey (GRFP Follow-Up Survey) that asks about graduate school experiences, educational attainment, career outcomes, employment characteristics, and professional productivity. The survey also asks Fellows about the influence of program elements (choice, flexibility, and monetary value) on their decision to enroll in and successfully complete STEM graduate programs.
- Collect data from two samples of institutions:
 - a. In-depth data from six institutions (“institutional site visit sample”) gathered from site visits which will encompass in-person interviews with administrators, faculty, and staff to understand: (1) the current climate, (2) perceived impact of the program on Fellows, institutions, and programs, (3) program implementation, and (4) GRFP policies.
 - b. Targeted data from a larger sample of 20 institutions (“institutional phone sample”) gathered from shorter phone interviews more narrowly focused on implementation and specific GRFP policies.

These two institutional samples serve two different purposes, each suited to the type of data to be collected. By limiting the site visit sample to six institutions and broadening the institutional phone sample to include 20 institutions, we balance the need for in-depth data collection with the goals of minimizing respondent burden and collecting data from a broader pool of institutions.

- Review and analyze similar federal fellowship programs using data collected from websites, program materials, and interviews with program officers managing these programs. This part of the study will help inform GRFP policies and best practices. The findings will be valuable in understanding how best to support Fellows and help develop a more diverse STEM workforce. Note that, because the respondents will be interviewed in their official capacity as program officers, this data-collection effort is exempt from OMB review and is not discussed further in this OMB submission. It is mentioned as it is an important piece of the study that will help inform the program.

Secondary data sources include the Survey of Earned Doctorates (SED) and the Survey of Doctorate Recipients (SDR). These will be used to provide a national context and national comparison groups.

Descriptive analysis will be used to examine the composition, experiences, and outcomes of Fellows, non-recipients, and national peers and will provide evidence of GRFP participation of underrepresented groups and trends in program selection, recognition, and financial support of early career scientists and engineers. To measure impact, the study will model outcomes using quasi-experimental methods to compare outcomes of the treatment group (Fellows) with outcomes of plausibly similar control groups (QG2 Honorable Mentions). These methods are widely accepted as the best methods on which to base causal inferences in the absence of a randomized experiment (i.e., when it is not feasible to randomly assign participants to treatment and control groups). To examine implementation, the study will use

⁸ The quality groupings (QG) refer to the categories assigned to each GRFP participant upon applying to the fellowship program. QG1 is the highest ranking an applicant can receive. The study sample includes the highest two categories: QG1 and QG2 applicants.

qualitative methods to code and analyze the institutional interviews and to draw out lessons learned from the interviews with program officers.

The GRFP evaluation will be the first comprehensive evaluation of this program since 2002 and the first to look at more recent cohorts (1994-2011) of Fellows and peers. While other NSF data collection such as the SED and SDR examine the graduate experience and career trajectories of doctoral recipients in STEM research fields, the GRFP evaluation is the only study to specifically assess the impact of this program on Fellows (both those in doctoral programs and master's degree programs) and institutions. In addition, the GRFP evaluation is the only current study that will examine the program impact from multiple angles in regards to Fellows (e.g., graduate experience, career trajectories) while also gaining an external perspective from academic institutions. Although the several previous studies noted above inform this study, the present study's approach will contribute to and significantly advance the current state of knowledge regarding the program, its implementation, outcomes, and impact.

A.2. Purpose and Uses of the Data

GRFP's goals are:

- To select, recognize, and financially support individuals early in their careers with the demonstrated potential to be high achieving scientists and engineers, and
- To broaden participation in STEM fields of underrepresented groups, including women, minorities, and persons with disabilities.

Underpinning the program goals are NSF's broader strategic organization goals, including that of performing as a model organization. To achieve this goal and to become a model Federal steward, representing excellence in management and fiscal responsibility, NSF seeks to "*learn through assessment and evaluation of NSF programs, processes, and outcomes; continually improve them; and employ outcomes to inform NSF planning, policies, and procedures*" (http://www.nsf.gov/news/strategicplan/nsfstrategicplan_2011_2016.pdf, pp. 16-17, italics in original). Thus, excellence in management is an underlying goal of each NSF program, including the GRFP.

To be conducted by NORC at the University of Chicago, the overall purpose of the study is to help NSF evaluate the impact of the program on Fellows, institutions, and the science and engineering workforce and the extent to which the program elements are effective in meeting the program goals. More specifically, the research questions (RQ) addressed by the study include:

RQ1. What is the impact of the GRFP fellowship on the graduate school experience?

RQ2. What is the impact of the GRFP fellowship on career outcomes?

RQ3. What are the effects of the GRFP on institutions?

RQ4. Is the program design effective in meeting program goals?

While RQ1 and RQ2 are framed in terms of impact, a necessary component of the research is examining how the Fellows compare with peers in terms of demographics, aspirations, educational trajectories, career outcomes, and professional productivity, to help address the program goals. RQ3 and RQ4 are designed to address both the GRF program goals as well as the underlying NSF strategic goal of excellence in management.

The study approach is summarized in Table A.2.1, which presents a crosswalk between the RQs and the data sources and analyses to be used to address them.

Table A.2.1. An Overview of the Study Approach: Crosswalk between Research Questions and Proposed Data Sources and Analyses

Data Source	Analysis
RQ1. What is the impact of the GRFP fellowship on the graduate school experience?	
<p><u>Primary data</u></p> <ul style="list-style-type: none"> GRFP Follow-Up Survey of Fellows and similar but non-awarded GRFP applicants in the 1994-1998, 1999-2004, 2005-2008, and 2009-2011 cohorts <p><u>Secondary data</u></p> <ul style="list-style-type: none"> Survey of Earned Doctorates (SED), selected years 	<ul style="list-style-type: none"> Compare demographic composition (gender and race/ethnicity) and graduate student experiences, participation in STEM graduate study, selection of institution, professional productivity, career aspirations, graduate degree attainment, and time-to-degree) of Fellows with those of a matched comparison group of similar but non-awarded GRFP applicants. For Fellows who graduated from doctoral programs, compare their demographic composition and experiences with those of a matched comparison group of doctoral graduates nationally from the annual Survey of Earned Doctorates (SED). To the extent possible, analyze outcomes by demographic subgroups.
RQ2. What is the impact of the GRFP fellowship on career outcomes?	
<p><u>Primary data</u></p> <ul style="list-style-type: none"> GRFP Follow-Up Survey of Fellows and similar but non-awarded GRFP applicants in the 1994-1998, 1999-2004, 2005-2008, and 2009-2011 cohorts <p><u>Secondary data</u></p> <ul style="list-style-type: none"> Survey of Doctorate Recipients (SDR), selected years 	<ul style="list-style-type: none"> Compare career outcomes (for example, in terms of academic and non-academic career choices, science and engineering careers versus careers in other fields, job characteristics, and professional productivity) of Fellows with those of a matched comparison group of similar but non-awarded GRFP applicants. Compare career outcomes of Fellows who graduated from doctoral programs with those of other national populations of doctoral graduates. To the extent possible, analyze outcomes by demographic subgroups.
RQ3. What are the effects of the GRFP on institutions?	
<p><u>Primary data</u></p> <ul style="list-style-type: none"> Institutional site visit sample: in-person interviews with university administrators, staff, and faculty involved with the program or Fellows <i>(To a more limited extent):</i> Institutional phone sample: phone interviews with university administrators, staff, and faculty involved with the program or Fellows 	<ul style="list-style-type: none"> Assess effects of the GRFP on graduate institutions with respect to student diversity and student quality, scholarly productivity and research, the extent to which Fellows participate in departmental teaching and research (“service to the department”), and financial aspects (for example, adequacy of the cost-of-education allowance, ability to free up resources to provide funding to other students, etc.). Use data from the institutional phone sample

	interviews where appropriate and relevant.
--	--

Table A.2.1. (Continued)

Data Source	Analysis
RQ4. <i>Is the program design effective in meeting program goals?</i>	
<p><u>Primary data</u></p> <ul style="list-style-type: none"> • Institutional phone sample: phone interviews with university administrators, staff, and faculty involved with the program or Fellows • Institutional site visit sample: in-person interviews with university administrators, staff, and faculty involved with the program or Fellows • GRFP Follow-Up Survey of Fellows in the 1994-1998, 1999-2004, 2005-2008, and 2009-2011 cohorts 	<ul style="list-style-type: none"> • Analyze responses from several different sets of respondents to address the RQ, including: <ul style="list-style-type: none"> – Institutional administrators and faculty about how the program is working and whether and how the program could be improved. – Program officers managing similar federal fellowship programs about what they have learned from their programs regarding implementation and practices. – Fellows about the impact of the GRFP on their decision to attend graduate school in a STEM field, the impact of program elements on choice, flexibility, and ability to fund and complete their graduate programs, and their career trajectories in terms of employment and professional productivity in STEM fields.¹

Note: ¹Our samples will be drawn from the GRFP applicant files which have information on each applicant's final award status. This allows students to be classified as Fellows or QG2 non-recipients. Since a small number of students refuse the fellowship, the GRFP Follow-Up Survey includes a screening question to ask awardees whether they refused the Fellowship. Applicants who refused the Fellowship will also be asked why, and then screened out of the survey. Because no other data will be collected from them, they will be excluded from any subsequent analyses. Data on why some students refused the Fellowship may be useful to NSF. See Section B.1 on p.16 for additional information on sampling.

The data will be useful in two major ways. First, the study will provide information regarding how the GRFP influences: educational decisions, experiences, and graduate degree attainment of U.S. students enrolled in STEM graduate programs; workforce participation and career outcomes; professional productivity; and graduate school institutions in terms of recruitment, funding, reputation, diversity and quality of students participating in STEM fields, and professional development opportunities offered to students. The analyses and findings will help NSF evaluate how the program is meeting its mission and help inform future program policies and initiatives. Graduate institutions may also find the information useful in understanding how best to support their Fellows and to help develop a robust and diverse U.S. science and engineering workforce.

Second, the study will shed light on how institutions are implementing the program and the extent to which specific design elements are valued or working as intended by NSF. In addition, the review and analysis of other similar federal fellowship programs may help point to best practices in terms of program design and structure. Both of these should inform NSF and the managers of similar federal fellowship programs in terms of program evaluation, review, and improvement.

The analysis and sampling plans have been designed to optimize reliability and validity throughout the study. Reliability (or, the consistency of the measures used) will be enhanced by utilizing data reduction techniques such as factor analysis to group questionnaire items into scaled measures of like concepts, and by adapting items from instruments previously used in education studies. Validity is often defined as the best available approximation to the truth or falsity of a given inference or conclusion⁹ and researchers need to be concerned with both internal and external validity. Internal validity refers to the rigor with which the study was conducted and the extent to which the study has taken into account alternative explanations for causal relationships that are the focus of the study. External validity refers to the extent to which the results of a study are generalizable. In our study, internal validity will be addressed in two important ways. First, the sampling plan focuses on awarded Fellows from the highest two quality groupings (QG1 and QG2) and non-awarded Honorable Mention applicants of similar quality (QG2) so that resulting program or treatment effects are not confounded by variability in the backgrounds or academic preparation of the sample. Second, quasi-experimental analytic methods such as regression discontinuity (RD) and propensity score matching (PSM) will reduce bias from non-random assignment of individuals to treatment and control groups. In terms of external validity, it is important to note that the results of the study will be generalizable to similar populations of academically accomplished graduate students in the STEM disciplines. Conclusions based on the study results should therefore not be applied to broader graduate student populations.

Study results will be reported to the Division of Graduate Education, EHR/NSF, distributed within the community of universities who participate in GRFP, and published on the NSF website. Limited print copies of the full report will be made available to NSF as well as 500 copies of a printed executive summary that can be disseminated more widely and that will be useful to a variety of public audiences. A policy brief reporting on the review and analysis of federal fellowship programs will be made available to all federal fellowship program managers and may be of interest to the larger foundation community as well. Findings of more general policy or methodological interest will be distributed more broadly, through conference presentations and submissions for publication in peer-reviewed journals.

A.3. Use of Information Technology to Reduce Burden

In order to reduce respondent burden, internet-based surveys will be used to collect information from participants. As the populations being surveyed in this study are graduate students in STEM fields, or professionals trained as scientists and engineers, they are expected to have easy access to and be fluent in the use of web-based technologies. The use of web-based systems facilitates accuracy, completeness, and speed of data entry, and helps reduce respondent burden. Web-based surveys employ user-friendly features, such as automated tabulation, data entry with custom controls such as checkboxes, data verification with error messages for online correction, standard menus, and predefined charts and graphics. Survey skip patterns reduce time burden on respondents by automatically moving them to the next appropriate section, simplifying the survey-taking experience. Web-based surveys also allow for easy identification of non-respondents and facilitate follow-up.

In addition, data entered by participants can be automatically uploaded into standard analysis software, eliminating the additional data entry step, thus increasing the efficiency of the analysts conducting the

⁹ Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Rand McNally, Chicago, Illinois.

study. Email will be used to invite participants to complete the survey and to follow up with the non-respondents to encourage their participation.

The survey will offer the same accommodation for those with disabilities as the Survey of Doctorate Recipients. There will be added navigation functionality on the Web survey so a mouse is not necessary for responding to the survey. Those with disabilities will be offered the option of a telephone or paper survey.

A.4. Efforts to Identify Duplication

This evaluation does not duplicate other NSF efforts [See response to A.1].

A.5. Small Business Involvement

No small businesses will be involved in this study nor is data being collected from any small businesses.

A.6. Consequences of Not Collecting the Information or Less-Frequent Data Collection

As this is the first program evaluation of GRFP since the last study published in 2002 and the only evaluation to date to examine more recent cohorts (i.e., from 1994 on), failure to conduct the study will leave a knowledge gap at a time when the GRFP is experiencing growth and increased attention as a Foundation-wide program to support STEM workforce development. If this information is not collected, NSF will not have needed evidence to meet its accountability requirement for independent evaluations to document the effectiveness and broader impacts of STEM education programming. It would also prevent NSF from learning what policies and practices are effective in meeting GRFP program goals, identifying effective strategies adopted by other federal fellowship programs, disseminating lessons learned to the broader STEM community, and obtaining valuable information about implementation and specific design elements to help inform future policies and programs.

The scope of the current proposed study is a one-time data collection effort, thus the issue of less-frequent data collection is moot.

A.7. Special Circumstances Justifying Inconsistencies with Guidelines in 5 CFR 1320.6

The project will fully comply with the guidelines of 5 CFR 1320.5. No special circumstances apply to this data collection.

A.8. Federal Register Notice and Consultation outside the Agency

Two notices to the public soliciting comments on this information collection prior to OMB submission were published in the Federal Register (75 FR 36697, Monday, June 28, 2010; 75 FR 56596, Thursday, September 16, 2010). A copy of the text of both notices is attached as Appendix A. No public comments were received in response to the notice during the 60 days that each appeared in the Federal Register.

NSF contracted NORC at the University of Chicago to design and conduct the study of the GRFP. NSF and an External Advisory Panel provided consultation on the study design. The GRFP evaluation team convened one meeting of its advisory group in March 2009. Advisory group members come from research organizations and universities and include the following individuals:

- Dr. Ronald G. Ehrenberg - Irving M. Ives Professor of Industrial and Labor Relations and Economics; Director - Cornell Higher Education Research Institute, Cornell University
- Dr. Lisa Frehill - Executive Director, Commission on Professionals in Science and Technology (CPST)
- Dr. Lewis Siegel - Dean of the Graduate School and Vice Provost for Graduate Education, Duke University

- Dr. William Trent - Professor, Department of Educational Policy Studies, University of Illinois at Urbana-Champaign.

The advisory group was responsible for reviewing the evaluation plan and the framework for the questionnaire.

The GRFP Follow-Up Survey was piloted with five individuals in June-July 2010; respondents took an average of 35 minutes to complete the survey. Since then the survey has been revised and shortened. Further pilot testing of the GRFP Follow-Up Survey with additional nine or fewer respondents will be conducted in December 2011 and respondents will be asked to comment on the clarity of directions and survey items, and the overall logic of the programmed Web survey. Results from this pilot test will be used to refine the survey.

The interview protocols for administrators, faculty, and staff (for both the institutional site visit and the phone interview samples) were based on similar interview protocols used in previous studies and have yet to be piloted. Estimated times to complete the interviews used to calculate respondent burden in Section A.12 are based on results in earlier studies and the contractor's experience with similar data collection efforts. A pilot test of these interview protocols will be conducted with nine or fewer respondents in December 2011 and respondents will be asked to comment on the clarity of questions and the flow of the interview. Results from this pilot test will be used to refine the protocols.

A.9. Payments or Gifts to Respondents

No payments or gifts will be provided to respondents.

A.10. Assurance of Confidentiality

All respondents will be advised that any information on specific individuals will be maintained in accordance with the Privacy Act of 1974. Data collected will be made available to the study contractors, contractors hired to manage data and data collection software, and, at the aggregate level, to NSF staff. Data will be processed in accordance with federal and state privacy statutes. Detailed procedures for making information available to various categories of users are specified in the Education and Training System of Records (63 Fed. Reg. 264, 272 January 5, 1998). The system limits access to personally identifiable information to authorized users. Data submitted will be used in accordance with criteria established by NSF for monitoring research and education grants, and in response to Public Law 99-383 and 42 USC 1885c. The information requested may be disclosed to qualified contractors in order to coordinate programs and to a federal agency, court or party in court, or federal administrative proceeding, if the government is a party.

Individuals responding to the GRFP Follow-Up Survey—Fellows and similar but non-awarded GRFP applicants—will be assured that the information they provide will not be released in any form that identifies them, and that their responses will be kept confidential to the extent provided by law. The contractor will be expected to maintain the confidentiality, security, and integrity of the survey data. The web-based survey data will be maintained on a secure server with appropriate levels of password and other types of protection. Proposed procedures for protecting the data and privacy of respondents will be reviewed by the contractor's Institutional Review Board (IRB) prior to data collection.

Individuals interviewed as part of the institutional data collection (i.e. institutional administrators, faculty, and staff who are selected to participate in the institutional site visit and the phone interview samples) will be asked for informed oral consent. They will be assured the information they provide will not be attributed to them and that all data will be reported in aggregated form. Direct quotations will not be attributed to any individuals or their institutions. These data will be identified only by site and interviewee codes and will be kept in locked cabinets or password-protected data files. In addition, any

crosswalk between the interviews and identifying information will be maintained separately from the actual interview notes and files.

A.11. Questions of a Sensitive Nature

The GRFP Follow-Up Survey contains very few sensitive questions, with perhaps the exception of salary. However, a respondent can choose to not answer this or other questions they deem sensitive. All survey questions will be reviewed by the contractor's IRB prior to fielding. Any public reporting of sensitive data will be in aggregate form.

The interview protocols (for both the institutional site visit and phone interview samples) also contain few sensitive questions because they focus on program implementation and the effect of hosting GRFP Fellows on the institutions. Respondents will be informed of their right to not answer specific questions if they so wish.

Copies of the survey and protocols can be found in Appendix B. These include:

- GRFP Follow-Up Survey
- Institutional site visit sample: Interview protocols for (a) university administrators; (b) faculty; and (c) university staff
- Institutional phone interview sample: Interview protocol (*one common protocol*).

A.12 Estimates of Response Burden

The total estimated number of respondents is shown below in Table A.12.1. Things to note about the table include: (a) Some students are likely to be still enrolled in the programs at the time the survey is administered, while others may have graduated or dropped out of their program. We refer to these groups as "current" and "former" graduate students, respectively. The table assumes a 65% response rate among the two survey groups; (b) Among respondents who will be interviewed during site visits or over the telephone, the table assumes a 100% response rate; (c) There is likely to be a mix of professors of different ranks (full, associate, perhaps assistant) involved with the GRFP Fellows.

Since average salaries differ by rank, we wanted to calculate an upper bound for the cost of the response burden by assuming that the institutional interviews would include faculty only at the full and associate professor ranks and would be more heavily weighted towards the more senior faculty. Thus, the site visit respondent sample assumes a mix of 4 full and 3 associate professors at each institution, and the phone interview respondent sample includes 2 full professors and 1 associate professor at each institution.

To calculate the respondent burden, the following assumptions regarding completion times were used:

- GRFP Follow-Up Survey:
 - 0.50 hours for current graduate students
 - 0.67 hours for former graduate students
- Institutional Site Visit Sample:
 - 0.33 hours for university administrators
 - 1 hour for full professors
 - 1 hour for associate professors
 - 0.75 hours for university staff
- Institutional phone interview sample:
 - 0.33 hours for university administrators
 - 0.50 hours for full professors

- o 0.50 hours for associate professors
- o 0.50 hours for university staff

A.12.1. Number of Respondents, Frequency of Response, and Annual Hour Burden

Table A.12.1 below indicates the total sample size and expected number of responses for each category of respondent type and the time demand these instruments will place on each individual respondent and on all respondents in aggregate. The total number of respondents is estimated to be 8,728, resulting in an estimated response burden for the study of approximately 5,098 person hours.

A.12.2. Hour Burden Estimates by Each Form and Aggregate Hour Burdens

As each respondent will complete the survey or interview once, the annual burden and the aggregate burden will be the same as shown in Table A.12.1.

Table A.12.1. Number of Respondents, Burden Hours per Respondent, and Total Person Hours, by Respondent Type

Type of Data-Collection and Respondent Type	Total Sample Size	Total Number of Respondents*	Burden Hours Per Respondent	Total Person Hours**
<i>1. GRFP Follow-Up Survey: web-based survey of Fellows and peers</i>				
Current Graduate Students	6594	4284	0.50	2142
Former Graduate Students	6594	4284	0.67	2870
<i>2. Institutional site visit sample: in-person interviews</i>				
University administrators	6	6	0.33	2
Full professors	24	24	1.00	24
Associate professors	18	18	1.00	18
University staff	12	12	0.75	9
<i>3. Institutional phone sample: phone interviews</i>				
University administrators	20	20	0.33	7
Full professors	40	40	0.50	20
Associate professors	20	20	0.50	10
University staff	20	20	0.50	10
Total	13348	8728	--	5112

Notes: *The table assumes a 65% response rate for the survey groups based on previous studies and a 100% response rate among the interviewees. The table also assumes that the total sample and the number of respondents will be evenly split between current and former graduate students.

**Rounded to nearest whole number.

A.12.3. Estimates of Annualized Cost to Respondents for the Hour Burdens

The overall annualized cost to respondents is **\$96,510**. Table A.12.2 shows the estimated total annual costs to each group of respondents over one year for the surveys and for the interviews. The assumptions underlying the table are discussed below.

Table A.12.2. Annualized Cost to Respondents, by Respondent Type

Respondent Type	Total Number of Respondents	Burden Hours Per Respondent	Total Person Hours	Hourly Salary Estimate	Estimated Cost per Respondent	Estimated Overall Cost*
<i>1. GRFP Follow-Up Survey: web-based survey of Fellows and peers</i>						
Graduate Students	4284	0.50	2142	\$16.00	\$8.00	\$17,136
Graduates	4284	0.67	2870	\$39.49	\$26.33	\$75,567
<i>2. Institutional site visit sample: in-person interviews</i>						
University administrators	6	0.33	2	\$89.72	\$29.57	\$59
Full professors	24	1.00	24	\$66.30	\$66.30	\$1,591
Associate professors	18	1.00	18	\$44.11	\$44.11	\$794
University staff	12	0.75	9	\$24.07	\$18.05	\$162
<i>3. Institutional phone sample: phone interviews</i>						
University administrators	20	0.33	7	\$89.72	\$29.57	\$197
Full professors	40	0.50	20	\$66.30	\$33.15	\$663
Associate professors	20	0.50	10	\$44.11	\$22.06	\$221
University staff	20	0.50	10	\$24.07	\$12.04	\$120
Total	8728	--	5112	--	--	\$96,510

Notes: *Rounded to nearest whole dollar.

The assumptions used in the table are the following:

- A work year consists of 240 days.
- For graduate students, the table uses the hourly salary paid to graduate assistants. These range from about \$12 to \$17—the table uses \$16 as a reasonable approximation (see, for example, http://finweb.rit.edu/controller/graduate/job_classifications.html)
- For graduates, the table uses the **average** of the 2006 median salaries for all full-time doctoral scientists and engineers employed 0-5 years and 6-10 years, adjusted to 2011 dollars, using the Bureau of Labor Statistics CPI calculator. (Available at <http://www.nsf.gov/statistics/nsf09317/> and http://www.bls.gov/data/inflation_calculator.htm).
- For university administrators, the table uses the average salary for deans of graduate programs in doctoral institutions (available at <http://chronicle.com.proxy.uchicago.edu/article/Median-Salaries-of-Senior/126455/>).
- For university senior faculty, the table uses the average salary of full professors; for junior faculty, the table uses the average salary of associate professors (available at <http://chronicle.com.proxy.uchicago.edu/article/Faculty-Salaries-Vary-by/127073>).

- For university staff, the table uses the average salary for academic-support-center coordinating officials (available at <http://chronicle.com.proxy.uchicago.edu/article/Median-Salaries-of-Midlevel/126834/>).

The above sources were all accessed on June 20, 2011.

A.13. Estimate of Total Capital and Startup Costs/Operation and Maintenance Costs to Respondents or Record Keepers

There is no overall annual cost burden regarding capital, operation, or maintenance costs to respondents that results from this study, other than the time spent responding to the survey.

A.14. Estimates of Costs to the Federal Government

The total estimated cost of the GRFP evaluation is \$2,639,512.17. This cost includes development of data-collection instruments (GRFP Follow-Up Survey and interview protocols for both the institutional site visit and phone interview samples), management of data-collection efforts, data collection through surveys, site visits, and phone interviews, cleaning and preparation of data files for evaluation, and data analysis and report writing to summarize the findings, implications, and lessons learned from the evaluation.

A.15. Changes in Burden

The GRFP evaluation is a new, one-time data collection from respondents.

A.16. Plans for Publication, Analysis, and Schedule

The four research questions the study is designed to address are:

RQ1. What is the impact of the GRFP fellowship on the graduate school experience?

RQ2. What is the impact of the GRFP fellowship on career outcomes?

RQ3. What are the effects of the GRFP on institutions?

RQ4. Is the program design effective in meeting program goals?

Appendix C (Tables C.1-C.4) provides a crosswalk between the research questions underlying the study and the data being collected through the surveys, the site visits, and the telephone interviews. The tables also briefly outline the kinds of analyses that will be used to address each question. More details can be found in the discussion below and the next section on Sampling and Estimation.

Data from the GRFP Follow-Up Survey: As shown in Table A.2.1, these data will be used to address RQ1, RQ2, and, to some extent, RQ4. The analyses of the survey data will include both descriptive and multivariate analyses. Part of this evaluation is to provide information describing the demographic composition, graduate school experiences, educational outcomes, and career progression of GRFP participants in the national context. Descriptive approaches will be used for examining the characteristics of the overall sample population, as well as racial/ethnic, sex, and other sub-populations (e.g., STEM field; graduate degree programs, graduate institution type). In addition to calculating the relevant means, standard deviations, and frequencies of the variables under investigation, we will use cross-tabulations for categorical outcomes and ANOVA for continuous outcomes to examine whether significant differences occur across groups and across variations of educational settings. Descriptive analysis will also present information on the sample and sub-sample populations in comparison to national benchmarks, such as those obtained from the SED, SDR and other data sources. This phase of

the analysis will provide a general understanding of group differences and will inform the interpretation of results from the multivariate stage of analysis.

Multivariate techniques will enable us to isolate the effects of the GRFP Fellowship award on program outcomes by statistically controlling for differences in a variety of confounding factors. Multivariate methods will include logistic and linear regression, as well as regression discontinuity (RD) and propensity score matching (PSM). Logistic and linear regression will facilitate the testing of differences among groups, while PSM and RD will account for the non-random, non-experimental program design in estimating the effects of the GRFP Fellowship award.

Data from Interviews with Participants in the Institutional Site Visit and Phone Interview Samples: These data will be used to address RQ3 and RQ4. The interview data will be analyzed using qualitative methods. Each interview will be recorded, transcribed, and coded for content relevant to the research questions underpinning the study. Mock interviews and staff trainings will be utilized along with monitoring inter-rater reliability to maintain consistency when coding interview data. Once coded, NORC staff will review the responses associated with each research question to identify the major type or types of answers, as well as any interesting individual responses. The analysis will indicate not only major trends, but also the strength of each trend (the proportion of interviewees with similar responses), the presence of any responses counter to that trend, and the source of the trend (if it comes from one particular type of respondent, for example, faculty advisors, or if it comes from multiple sources). Discussions of trends will include particularly illuminating or otherwise interesting quotes, when available. In addition to analysis based solely on site visit interviews, NORC's analysis will be informed by responses to the GRFP Follow-Up Survey described above to understand the perspectives of students.

The contractor, NORC at the University of Chicago, will prepare a major technical report on the results of the study that provides details regarding the sampling, methodology, and analysis. In addition, the contractor will produce a short 3-5 page research brief that provides highlights of the study targeted towards the research questions and that can be disseminated to policymakers. The contractor will also prepare a separate policy brief that will report on the findings of the review and analysis of federal fellowship programs, supplemented by data from the institutional interviews on implementation of the GRFP program.

As stated earlier, GRFP's goals are (a) to select, recognize, and financially support individuals early in their careers with the demonstrated potential to be high achieving scientists and engineers, and (b) to broaden the participation of underrepresented individuals, including women, underrepresented minorities, and persons with disabilities, in science and engineering fields. In addition, each NSF program seeks excellence in management and continuous improvement. The data collected in this study and the analytic reports will provide a comprehensive look at the GRFP, its impact on Fellows, institutions, and the science and engineering workforce, and the extent to which it is meeting its goals and speak directly to the program goals as well as the NSF's strategic goal of performing as a model organization. In particular, the findings will provide information on:

- The influence the GRFP has had on the decisions, experiences, academic attainment, and career outcomes of Fellows compared with carefully-matched peers;
- The extent to which the program has broadened the participation of underrepresented groups in STEM at the graduate level;
- The perceived effects on institutions in terms of student financing, enrollment, diversity and quality (among others);

- Whether specific design elements—choice, flexibility, and monetary value—are working as intended and the extent to which they are valued; and
- The need (if any) for changes in the way the program is structured to make it more effective.

Thus, overall, the study findings will provide valuable insights to NSF on the impact of its investments in the GRFP and will inform program management. In conjunction with findings from the review of similar federal fellowship programs, the findings may prove valuable to the larger community of program officers administering these programs as well as to the graduate community in understanding how best to support Fellows and help develop a more diverse STEM workforce.

Table A.16 shows the timeline for the study.

A.17. Approval to Not Display Expiration Date

The data collection instruments will display the OMB clearance number and expiration date.

A.18 Exceptions to Item 19 of OMB Form 83-I

No exceptions are sought.

Table A.16. Project Schedule

Task	Start Date	End Date
Sample draw	7/20/11	7/27/11
OMB submission	9/29/11	12/29/11
Review of Federal fellowship programs		
Phone calls with program officers; review of materials	10/12/11	11/14/11
Analysis of data and drafting policy brief	11/17/11	1/17/12
<i>Submit policy brief</i>	1/18/12	1/18/12
GRFP Follow-Up Survey		
Pilot test with 9 or fewer respondents	11/12/11	12/14/11
Pre-field locating for GRFP Follow-Up Survey	11/12/11	1/11/12
GRFP Follow-Up Survey programming	11/12/11	1/4/12
GRFP Follow-Up Survey: Advance letter	1/13/12	1/13/12
GRFP Follow-Up Survey: Advance email	1/17/12	1/17/12
GRFP Follow-Up Survey: Postcard reminder	1/27/12	1/27/12
GRFP Follow-Up Survey: 1st email prompt	2/10/12	2/10/12
GRFP Follow-Up Survey: 1st prompt letter	3/2/12	3/2/12
GRFP Follow-Up Survey: Data collection	1/16/12	6/16/12
Institutional Phone Interview Sample		
Selection of sample of 20 institutions	12/2/11	12/14/11
Initial contact (via phone and email)	12/30/11	1/11/12
Phone interviews and data collection	1/17/12	3/30/12
Analysis of data and report-writing	4/2/12	5/14/12
<i>Submit Preliminary Implementation Findings (based on phone interviews, review of federal fellowship programs, completed site visits, and early survey data)</i>	5/15/12	5/15/12
Institutional Site Visit Sample		
Selection of sample of 6 institutions	12/5/11	12/19/11
Initial contact (via phone and email)	1/3/12	1/13/12
Site visits and data collection	1/17/12	6/15/12
Analysis of survey and interview data and report-writing	6/15/12	11/26/12
<i>Submit Draft Report</i>	11/27/12	11/27/12
<i>Submit Final Report and Research Brief</i>	2/28/13	2/28/13

Note: Deliverables in italics.