B. STATISTICAL METHODS

There are two major types of data collection: (I) GRFP Follow-Up Survey administered to Fellows and their counterparts; and (II) institutional interviews that encompass in-person interviews with several respondents at 6 institutions (institutional site visit sample) and telephone interviews with respondents at 20 institutions (institutional phone interview sample). The sections below discuss subsections B.1-B.4 for each of the two data collections separately. A final section, B.5, provides key contact information for the study.

Providing information about the study helps legitimize it and recruit potential respondents. It also serves to keep the public informed of the study's purpose and scope. As such, NORC will develop a project-specific web page that briefly describes the goals of the project, the different data collection efforts, and the expected reports on findings. This page will be hosted on NORC's corporate web site (http://www.norc.org) and will be accessible to the public and interested respondents and potential respondents, who may navigate to it through NORC's website or by using a search engine. The project page will contain links to additional information about the GRFP (on NSF's website) as well as contact information for NORC's team.

I. GRFP FOLLOW-UP SURVEY

B.1. GRFP Follow-Up Survey: Sampling Methodology

The study will survey several cohorts of GRFP Fellows and their peers. To define the sample population, GRFP application records will be used to identify Fellows and peers ("Honorable Mentions" discussed below). In order to be considered eligible for the GRFP Fellowship award, all applicants in the sampling file meet the following three eligibility criteria:[1]

1. Applicants must be U.S. citizens, U.S. nationals, or permanent residents of the U.S.
2. Applicants must be in the early stages of their graduate education career, having completed no more than twelve months of full-time graduate study at the point of program application. This limit applies to all graduate education, not just current program of enrollment. If the applicant has previously completed a Master's degree, he/she would be ineligible unless it is documented that the applicant completed a one-year Master's degree program.
3. Applicants must be seeking a research-focused Master's or doctoral degree in an NSF-supported field.[2]

All applicants meeting the above eligibility requirements are reviewed by panelists with disciplinary expertise and assigned one of four quality group (QG1 to QG4) rankings based on NSF merit review criteria. In addition, reviewers take into consideration applicants' background characteristics, including their personal, professional, and educational experiences, as well as letters of reference.[3] Applicants who receive a rating of QG1 receive fellowships. Applicants receiving a QG2 rating are split into two additional groups; one group receives the fellowship award, while the other group is awarded the title "Honorable Mention" without the fellowship. **Recommendations for awards within QG2 help the program meet Congressional mandate for geography, and the program goal of broadening participation.**

---

1 http://www.nsfgrfp.org/how_to_apply/eligibility_guide; http://www.nsf.gov/pubs/2010/nsf10604/nsf10604.pdf

2 For NSF-supported fields of study, see http://www.nsf.gov/pubs/2010/nsf10604/nsf10604.pdf

3 http://www.nsf.gov/pubs/2010/nsf10604/nsf10604.pdf

In the vast majority of cases, applicants who receive a rating of QG3 receive the title "Honorable Mention." Finally, applicants who receive a QG4 ranking do not receive a fellowship award or the title of "Honorable Mention." This evaluation will focus exclusively on fellowship and honorable mention award recipients in QG1 and QG2. Fellowship award recipients are further restricted to those who have accepted their award and are GRFP Fellows.

The sampling data file will contain unit-record identifiers, application information and QG rankings for all eligible GRFP applicants who received the fellowship or honorable mention award from four cohorts based on program application year: 1994-1998 (Cohort 1), 1999-2004 (Cohort 2), 2005-2008 (Cohort 3), and 2009-2011 (Cohort 4).

The main sampling frame is the list of GRFP Fellows and Honorable Mentions for 1994 – 2011. As shown in Table B.1.1, we propose to randomly select a total sample of 13,188 cases from Cohorts 1 through 4 (1,099 QG1 Fellows, 1,099 QG2 Fellows, and 1,099 QG2 Honorable Mentions per cohort). An assumed 65% overall response rate for GRFP Fellows and Honorable Mentions will result in 8,568 completed questionnaires (714 QG1 Fellows, 714 QG2 Fellows, and 714 QG2 Honorable Mentions per cohort). Table B.1.2 additionally shows expected completes for specific sub-group populations without oversampling of minorities and disabled, pooling the four cohorts. The fewest expected completes are from disabled individuals, with expected 574 disabled and 7,994 others across the entire sample.

The sampling plan was designed to select a sample large enough to make statistically valid estimates of program outcomes in answering 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?*), and RQ4 (*Is the program design effective in meeting program goals?*). A variety of analytic techniques will be used to address the research questions. While the size of the analytic sample, minimum detectable effects, and statistical power vary with the specifications of a given comparison, it is important to broadly assess statistical power and minimum detectable effects for a given sample to determine if each is sufficient for answering the research questions.

In some cases, comparisons will be based on the full sample pooled across all four cohorts, such as when examining the overall impact of the GRFP fellowship on the graduate school experience (RQ1). Here comparisons will be made among all current and former graduate students, spanning all four cohorts. In other cases, comparisons will be between sub-samples defined according to specific cohorts or population characteristics. For example, when examining the impacts of the GRFP fellowship on career outcomes (RQ2), comparisons will be made within a sub-sample comprised of former graduate students (Cohort 1 and Cohort 2 applicants). When examining GRFP program goals (RQ4), comparisons will be made within a given cohort, or between sub-populations defined according to population characteristics (e.g., minority vs. other, female vs. male, disabled vs. other).[4]

Table B.1.3 provides information on the effect sizes detectable based on the expected number of respondents for pooled samples and sub-samples, following conventional standards of an 80 percent level of statistical power and a 95 percent confidence level (alpha=0.05) for different comparisons. For comparisons based on the pooled sample of 2,856 completed questionnaires in each comparison group (Cohorts 1 – 4 QG1 Fellows, QG2 Fellows, and QG2 Honorable Mentions) and an expected estimate of 50 percent for a particular outcome across the full sample of cases, we would be able to detect a 3.7 percentage point difference between two groups. If the expected estimate for a particular variable of interest is 90 percent, we could detect a 2.1 percentage point difference.

[4] See Table C.1 for additional details on data sources per specific analysis.

Table B.1.1. Population Counts and Sample Estimates

Cohort	Fellowship Status	Population	Sample	Expected Number of Respondents
C1: 1994-1998 cohort (GRF+MGF)[a]	QG1 Fellows	2,580	1,099	714
C1: 1994-1998 cohort (GRF+MGF)[a]	QG2 Fellows[b]	2,038	1,099	714
C1: 1994-1998 cohort (GRF+MGF)[a]	QG2 Honorable Mentions	1,490	1,099	714
C2: 1999-2004 cohort (GRF)	QG1 Fellows	2,897	1,099	714
C2: 1999-2004 cohort (GRF)	QG2 Fellows[b]	2,561	1,099	714
C2: 1999-2004 cohort (GRF)	QG2 Honorable Mentions	2,220	1,099	714
C3: 2005-2008 cohort (GRF)	QG1 Fellows	2,035	1,099	714
C3: 2005-2008 cohort (GRF)	QG2 Fellows[b]	1,725	1,099	714
C3: 2005-2008 cohort (GRF)	QG2 Honorable Mentions	2,233	1,099	714
C4: 2009-2011 cohort (GRF)	QG1 Fellows	~2,708	1,099	714
C4: 2009-2011 cohort (GRF)	QG2 Fellows[b]	~2,536	1,099	714
C4: 2009-2011 cohort (GRF)	QG2 Honorable Mentions	~5,228	1,099	714
	Total	TBD	13,188	8,568

Source: 1989-2009 merged data from NSF applicant record data files, 1989 - 1998 Ci data non-pii resort.xls; 1989 – 1998; CI 99-04 data -Non- PII.xlsx; 05-09 data -Non-PII.xlsx. Population counts for the 2010 and 2011 application years were based on personal correspondents with NSF-GRFP program directors.

[a] Through the 1998 GRFP application year, students could apply for the Graduate Research Fellowship and the Minority Graduate Research Fellowship (MGF) programs. For sampling purposes, no distinction will be made between the two groups. The MGF was discontinued in 1998.

[b] A small share of GRFP QG3 applicants were awarded the fellowship due to program exceptions. For sampling purposes, these cases will be treated as QG2 Fellows.

Table B.1.2. Expected Sample Composition by Sub-group Populations

Variable	Level	Rate	Expected Number
Sex	Men	54.73%	4689
	Women	45.27%	3879
Race/Ethnicity	Minority	8.82%	756
	Other	91.18%	7812
Ph.D. completion by 2012	Completer	70.00%	5998
	Non-completer	30.00%	2570
Disability	Disabled	6.70%	574

	Other	93.30%	7994

Table B.1.3. Minimum Detectable Effect by Comparison Group and Estimated Outcome Value

Comparison Groups[a]	Expected Number of Respondents within Groups	Estimated Value for Outcome	Minimum Detectable Effect (percentage points)
Pooled sample			
QG1 F vs. QG2 F vs. QG2 HM[b]	2856 per group	50%	3.7
		90%	2.1
Male vs. Female	4689 vs. 3879	50%	3.0
		90%	1.8
Minority vs. Other	756 vs. 7812	50%	5.3
		90%	3.0
Disabled vs. Other	574 vs. 7994	50%	6.0
		90%	3.2
Sub-sample: Former graduate students			
QG1 F vs. QG2 F vs. QG2 HM	1428 per group	50%	5.5
		90%	3.1
Sub-sample: Within cohort			
QG1 F vs. QG2 F vs. QG2 HM	714 per group	50%	7.4
		90%	4.0

Source: http://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx

[a] The pooled sample includes all 8568 Cohort 1 – 4 applicants. The sub-sample of former graduate students includes Cohort 1 and 2 applicants. The within-cohort sub-sample represents applicants within a single cohort.

[b] QG1=Quality Group 1, QG2=Quality Group 2, F=Fellows, HM=Honorable Mentions

The fewest expected completes from a key comparison group is among disabled students, (n=574) compared with nondisabled respondents (n=7,812). Based on these estimates, we could detect a 6.0 percentage point difference if the expected outcome estimate is 50 percent and a 3.2 percentage point difference if the expected outcome estimate is 90 percent. Table B.1.3 provides equivalent information on other comparison groups.

To provide context for the minimum detectable effects in our study, we examined past studies of GRFP applicants to see what differences have been reported in the literature and so determine if our sample size will be sufficient to support the planned analyses. While the following review focuses on Ph.D. completion rates as a useful point of reference to past studies, it is important to note that this

evaluation will examine a sizable number of different outcomes related to graduate education, careers, and professional productivity.

The most recent comprehensive evaluation of the GRFP (Goldsmith, et al., 2002) provided evidence of mean differences between QG1 Fellows, QG2 Fellows, and QG2 Honorable Mention recipients. Results indicate that, for example, Ph.D. completion rates by 1999 among 1979-1988 GRFP applicants were 75.3 percent for QG1 Fellows, 69.4 percent among for QG2 Fellows, and 66.0 percent for QG2 Honorable Mentions, suggesting a 3.4 percentage point program effect among comparable QG2 applicants.[5] This difference would not be significant at the 95% confidence level, given the power of the proposed sample to detect differences between quality groups within a single cohort. However, if differences of similar size were found in two or more of the 5-year cohorts the pooled cohort comparisons would have sufficient power for the difference to be significant at the 95% confidence level. The GRFP evaluation will examine Ph.D. completion rates among all but the most recent cohort of applicants.

The Goldsmith, et al. (2002) evaluation reported gender differences in Ph.D. completion rates by 1999 among the 1979-1988 GRFP applicants. The results showed 70.5 percent of male QG1 Fellows completed their Ph.D. by 1999, while 70.2 percent of QG2 Fellows and 67.5 percent of QG2 Honorable Mentions did so, indicating a 2.7 percentage point program effect. A difference of this magnitude among QG2 males would not attain 95% confidence even if all cohorts were pooled. The differences in Ph.D. completion rates were larger among females, with a 9.8 percentage point difference between QG2 Fellows and Honorable Mentions (68.3 vs. 58.5 percent); female completion rates were 73.3 percent for QG1 Fellows, 68.3 percent for QG2 Fellows, and 58.5 percent for QG Honorable Mentions.[6] In contrast to males, differences of this size would be detected with 95% confidence even within cohorts.

Baker (1998) also examined gender and race differences in Ph.D. completion by 1988 among 1972 – 1981 GRFP applicants. Baker's findings indicate that rates of Ph.D. completion favored males GRFP applicants over their female counterparts by 4.5 percentage points among QG1 Fellows (77.4 vs. 72.9 percent), 9.1 percentage points among QG2 Fellows (73.6 vs. 64.5 percent), and 9.8 percentage points among QG2 Honorable Mentions (67.0 vs. 57.2 percent). Among QG2 applicants, male Fellows differed from male Honorable Mentions by 6.6 percentage points (73.6 vs. 67.0 percent), while female Fellows and Honorable Mentions differed by 7.3 percentage points (64.5 vs. 57.2 percent), again indicating a larger program effect among female applicants. Differences of this magnitude would be detected with the pooled sample of cohort 1 and cohort 2 in the current design.

Because the evaluation will include several different outcome measures and the analyses will compare a variety of different groups, Table B.1.4 presents the sample sizes required to detect a range of anticipated effect sizes based on mean differences between groups. Cohen's *d* metric for effect sizes (calculated as the mean difference between groups divided by the overall sample standard deviation) is frequently used to estimate sample sizes, where a smaller value necessitates a larger sample size. Following Cohen's conventions, a mean difference between two groups of 0.20 is considered small and would be detectable for samples of 393 or more cases within each comparison group.

In combination with the previous tables, it is evident that the proposed sample will be sufficient to detect effect sizes (mean differences between two groups) as small as 0.10 for pooled sample comparisons between QG1 Fellows, QG2 Fellows, and QG2 Honorable Mentions, and for comparisons between males and females. For pooled sample comparisons between disabled versus others, minorities versus other cases, or other sub-sample comparisons, the proposed sample will be sufficient to detect effect sizes between 0.10 and 0.20.

---

[5] See Goldsmith, et al. (2002), Table G14, p.141. http://www.nsf.gov/pubs/2002/nsf02080/nsf02080.pdf.

[6] See Goldsmith, et al. (2002), Table G9, p.136. http://www.nsf.gov/pubs/2002/nsf02080/nsf02080.pdf

Table B.1.4. Sample Size Requirements by Effect Sizes abased on mean differences between groups.

Detectable effect sizes (*d*, mean differences between groups) [a]	Required sample size per comparison group [b, c]
0.10	1571
0.20	393
0.30	175
0.40	99
0.50	64
0.60	45
0.70	33
0.80	26

[a] An effect size of 0.20 is considered small, 0.50 is considered medium, and 0.80 is considered large (Cohen, 1988).

[b] Assumptions: power=0.80, alpha=0.05.

[c] The sample sizes shown are for continuous variables. Comparable numbers exist for dichotomous variables (1560 for *d* =0.10, 392 for *d*=0.20, 174 for *d*=0.30, etc.).

An important component of this evaluation is to compare GRFP participants who ultimately completed a doctoral program to a nationally representative sample of "other Ph.D. recipients." For this purpose we will identify a comparison group from subsets of the Doctorate Records File (DRF) and the 2006 Survey of Doctorate Recipients (SDR) as national peer groups of "other Ph.D. recipients." The DRF, collected through the annual Survey of Earned Doctorates (SED), is a census of research doctorate recipients from U.S. universities. The SED data are collected at the doctorate recipients' point of doctorate receipt and explore a number of issues related to the doctorate recipients' graduate experience (i.e., interdisciplinary activities, time-to-degree, sources of financial support and indebtedness).

These data provide excellent benchmarks on which to compare similar data gathered from the GRFP Follow-Up Survey. The SDR is a national sample survey of doctorate scientists and engineers that focuses on career paths, further education and employment-related data. As such, it can provide a solid comparison of "other Ph.D. recipients" career-related data with those gathered from the career outcomes component of the GRFP Follow-Up Survey. To establish as close a comparison group as possible to the sample in the GRFP surveys, subsets of these datasets that approximate the cohorts and criteria for application to the NSF Fellows program will be used. The median elapsed time from graduate school entry (roughly the application point for NSF Fellows) to Ph.D. receipt for STEM graduates is 8.2 years (Hoffer, Welch, et al., 2006).[7] We assume that SED and SDR respondents who received their doctorates in academic years 1996 to 2006[8] will approximate the NSF Fellow cohorts of interest. Further

[7] Hoffer, T.B., V. Welch, Jr., K. Webber, K. Williams, B. Lisek, M. Hess, D. Loew, & I. Guzman-Barron (2006). *Doctorate Recipients from United States Universities: Summary Report 2005*. Chicago: National Opinion Research Center.

[8] The 2006 SDR includes sample members who received their doctorates through AY 2005. The SDR 2008, sampling doctorate cohorts up to AY 2007, was launched in the fall of 2008. If the SDR 2008 data are available in time for the GRF Survey comparison study, we propose using 2008 SDR instead of the 2006 SDR.

selection criteria to be applied are completion of a degree in a STEM field of study and U.S. citizenship or permanent resident alien status. In addition, we will drop from these datasets any records that we can match from the sample of GRFP participants. Using these filters, we believe we can create a national peer group of "other Ph.D. recipients" to provide a valid comparison to the surveyed Fellows who completed doctoral programs. Because these comparisons will require record matching based on the restricted-use SED and SDR microdata files, NORC work closely with the GRFP COTR, SED COTR, and SDR COTR in following the National Center for Science and Engineering Statistics guidelines for obtaining and executing a restricted-use license (see http://www.nsf.gov/statistics/license/start.cfm).

Note that this comparison excludes those who graduated with master's degrees. For this group, comparisons will focus on Fellows and Honorable Mentions, and extant data sources are being researched to determine the feasibility of obtaining a nationally representative sample of M.A. recipients in research-focused STEM fields.

We will draw from other available data resources as needed. Among the other data we will incorporate are the Integrated Postsecondary Education Data System (IPEDS) for measures of institutional characteristics and Carnegie Classification, *Barron's Profiles of American Colleges* for measures of institutional selectivity or prestige, and possibly the *National Faculty Directory*[9] for measures of faculty employment status.

B.2. GRFP Follow-Up Survey: Procedures for the Collection of Information

The GRFP Follow-Up Survey will be conducted as a Web-based instrument, accessible to respondents using a combination of personalized PIN and password. NORC will administer the GRFP Follow-Up Survey to Fellows and Honorable Mentions sampled in Cohorts 1 through 4 (1994-2011). A copy of the GRFP Follow-up Survey is provided in Appendix B.

In preparation for data collection, NORC will prepare and mail letters inviting sample members to participate in the study. Letters will be sent through the U.S. Postal Service. The mailings will describe the study and its purpose, and the measures taken to assure confidentiality. Included in the mailings will be a unique PIN and password to use for accessing the survey online. Upon receiving this advance letter, sample members will be able to go to the survey website and complete the questionnaire. The advance letter will also include a study toll-free number and email address through which respondents can directly contact project staff to verify study authenticity, ask questions about their participation, or receive technical assistance.

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 so a mouse is not necessary to respond to the survey. For those with disabilities, we will offer the option of a telephone or paper survey.

B.3. GRFP Follow-Up Survey: Methods to Maximize Response Rates

To target a 65% response rate overall, and no less than a 60% response rate within each cohort, NORC will follow up the advance letter mailings with a series of letter and postcard prompts as reminders to complete the survey, with special emphasis on monitoring and prompting the Honorable Mention sample members to ensure adequate response rates. NORC will consider offering incentives to Honorable Mention sample members if necessary to obtain a sufficient response rate. When NORC receives information that a sample member no longer resides at a particular location, additional steps will be taken to locate the individual (see *Locating* section below).

---

[9] http://www.gale.cengage.com/pdf/facts/NFDonGDL.pdf

In addition to the standard letter prompts, NORC will employ both phone and email prompting. Approximately one month after data collection begins, NORC will use phone prompting to encourage sample members to participate in the survey. NORC will also use two methods of email prompting to gain participation. Weekly batch emails will be sent out during data collection to all non-respondents. These emails will come from a GRFP study email address and will contain the respondent's personalized PIN and password. The second method will involve a more targeted and personalized email strategy to boost response rates for those groups having lower than expected response rates. These emails will come from a personal email address and will contain the respondent's PIN, password, and a direct link into the survey.

To ensure the confidentiality of sample members the survey Web page and all advance and prompting materials will contain generic branding referring to a "graduate student follow-up study" rather than referring to any particular group (e.g., Fellows and Honorable Mentions). This will not affect the types of questions included in the survey, and specific paths through the survey will be based on participants' fellowship status, determined upfront by the survey login PIN and password each is assigned.

To further enhance user-friendly access for sample members, NORC will maintain a survey-specific web page throughout data collection. In addition to a link to the secure web instrument, the study's web site will serve as a source of information for potential respondents. This web site will be cited in all advance and prompting materials sent to sample members and will also be accessible through the main NORC website and general search engines, such as Google. The Web page will contain links allowing a sample member to:

- Access a detailed description of the survey
- Review the GRFP Follow-Up Survey's Frequently Asked Questions
- Obtain contact information for NORC survey staff
- Review the Privacy Policy for the GRFP Follow-Up Survey
- Review citations and/or publications of previous surveys' data
- Link to the GRFP Follow-Up Survey Web Questionnaire
- Send an e-mail message to the GRFP Survey in-box
- Call a 1-800 number for information on the survey.

Locating

Accurate address and telephone contact information are essential for notifying sample members of their selection in the study and further prompting for survey completion. Because we will be using the GRFP applicant records as the data source for sample member contact information captured at the time of applying to the program, the locating strategy has been designed to handle varying degrees of outdated information. Past NORC studies have found that 80% of located cases ultimately go on to complete the survey when following such a prompting strategy. Therefore to target a 65% response NORC will need to locate 75-85% of sample members within each cohort. To accomplish this location rate, NORC will use a multi-stage strategy for locating sample members that will be responsive to varying amounts of locating information within any given cohort.

As noted, NSF will provide contact information from GRFP applicant records, including names and birthdates for all cases. For many cases, available information will also include social security numbers, address information, phone numbers, email addresses, and educational institutions (i.e., intended

graduate school, current or previous college or university). Cases with incomplete contact information and cases with outdated contact information will be submitted for locating.

Our primary locating tools will include Accurint and LinkedIn searches. Accurint is a locating service that maintains a database of national information. When there is a match, the Accurint search yields address, phone, and/or email information. Previous NORC studies with populations similar to GRFP applicants have been able to locate 60% of their sample using Accurint searches. Because Accurint searches rely on SSN and birthdate information, LinkedIn will additionally be used for cases that lack these critical information fields. LinkedIn searches have been found to be highly successful in locating profiles of professionals when using academic institution information. NORC estimates that an additional 15-20% of the sample can be located through LinkedIn searches. LinkedIn is a professional networking web site where individuals create profiles listing their academic and professional credentials. Often these profiles list current employers, and/or contact emails and phone numbers. NORC has developed methods of searching LinkedIn profile pages using educational institution information listed on profile pages. These searches will be used to identify cases where sample members profiles list a current employer, current educational institution, or where there is contact information listed. Locators will then enter this information into our cases management system. This information will also be used to guide academic directory and employer directory searches.

These locating strategies will be employed during two phases of the study: prior to data collection (i.e., Pre-field), and during data collection.

Pre-field Locating

The locating strategy is based on the assumption that current address information may be incomplete for a portion of the sample. To account for this, pre-field locating will be conducted using Accurint. In addition, with NSF's support, NORC may contact coordinating officials at GRFP-sponsoring institutions to request updated contact information for more recent cohorts who may still be enrolled graduate students.

Locating during Data Collection

Once the survey is in the field, we will mail letters to addresses obtained through Accurint searches, or from applicant data, and send out an email prompt. Cases having mail returned as undeliverable or invalid email addresses or phone numbers will be designated for more intensive locating treatments. These cases will be forwarded to NORC's locating department, where staff will conduct additional Accurint individual searches.

In additionally, NORC will use LinkedIn searches using educational institution data contained in the applicant files (including BA institution, current institution, and intended graduate institution at the time they submitted the GRFP application) to locate sample members on LinkedIn using automated search techniques. Locators will manually review LinkedIn profiles for new contact information, including: email, phone, current employer, and current location. For older cohorts, LinkedIn searches will be necessary to determine updated location information. For example, where LinkedIn searches produce an affiliation with an educational institution, the information will be used to conduct academic directory searches.

NORC expects that more recent cohorts will have more up-to-date information and will require less intensive locating efforts in comparison to older cohorts. For older cohorts, employing LinkedIn searches will likely be necessary to locate 75% to 85% of sample members. Location rates within cohorts, as well within awardee status group will be monitored throughout the data collection period, and the locating strategies applied as necessary.

Case Management

All sample member information, updated locating data, case history and status will be maintained in NORC's Case Management System (CMS). This comprehensive database maintains records for all incoming and outgoing contacts with sample members along with complete address history information. Every update in the CMS records the status and date of the update, as well as the staff member who made the update. The CMS acts as a central 'brain' of the GRFP system design, holding much of the case data and directing the processing flow of the other component systems.

B.4. GRFP Follow-Up Survey: Tests of Procedures

Pilot and Cognitive Testing

The survey was time tested with five individuals. Complete pilot testing with up to nine respondents will occur in December and will gather respondent comments on directions, clarity of items and overall logic of the programmed Web survey. Results from this pilot test will be used to refine the survey.

Cognitive testing also will be used as a tool that explores the respondent's understanding of the survey questions and the cognitive processing to formulate an answer. The scripted and unscripted cognitive probing during the interview will be directed towards understanding these issues. NORC will conduct five cognitive interviews. After each interview, respondents will be asked to provide feedback on the interview including respondent's overall interview experience, suggestions for improving the survey, and an open question and answer period for the respondent and interviewer.

The factors that will be examined during the cognitive interviews include: respondent understanding of the task/questions, respondent burden, interview timings, incorporating feedback from interviewers/respondents on problems with the instruments. Some specific questions that will guide the cognitive testing include:

- Do respondents have any difficulty comprehending the survey questions?
- Are there any survey questions that can be improved, clarified?
- Are there any additional survey questions that should be included?
- How burdensome is the survey and has burden been reduced as much as possible?
- Can respondents provide accurate responses to survey questions that ask about events that may be more than a few years in the past?
- Has all relevant feedback from respondents and cognitive interviewers been incorporated?
- Is the timing of the instruments within the appropriate parameters?

The cognitive interviews will bring to light problems that exist with the GRFP survey. Some of these problems will be ones that we anticipated based on our review of the instruments, while other issues may be revealed by the cognitive interviews. Following the set of five interviews, data will be examined and materials will be revised in order to address the issues that emerge from testing.

Nonresponse Bias Analysis

High nonresponse overall or differential response rates between the treatment and control groups, and/or between older and newer cohorts can jeopardize the integrity of a study. The contractor plans to conduct a series of comparisons to assess the extent to which nonresponse has resulted in the respondent sample being different from the original baseline sample.

NORC will employ different approaches to examining non-response bias and accounting for it in the final analysis. Two potential options for our non-response bias tests are the Confidence Interval Bias Test and

Cochran's Bias Test.[10] For the Confidence Interval Bias Test, variables such as demographic characteristics will be compared for the original baseline sample and the respondent sample. A confidence interval (CI) around the mean value among the responders will be calculated and compared to the mean value of the baseline sample—if the mean of the baseline sample falls within the CI, this suggests that the responders are sufficiently similar to the baseline sample and that it is not necessary to correct for nonresponse bias.

A more rigorous test to detect bias is the Cochran Bias Test, which is calculated by taking the difference of the mean of the responders and the mean of the baseline sample and dividing it by the standard error of the responders. A resulting bias of greater than 0.10 is considered problematic. It is important to note, however, that the Cochran test is extremely sensitive and leads to the conclusion of bias on most factors being tested. Thus, it is important to consider more than one type of bias test to determine if bias exists.

If non-response bias appears to be an issue, NORC plans to re-weight the sample data according to each respondent's likelihood or propensity of being a respondent. A logistic regression using baseline characteristics is used to predict the probability of being a respondent. Respondents in the sample with characteristics that most often are associated with nonresponse would effectively receive a higher weight to make up for their low incidence in the sample. These steps are essential to ensuring that our final estimates are not biased by the under-representation of any important subgroups, particularly the older cohorts. Where such methods are used, they will, of course, be carefully noted in the final report.

The next section discusses sections B.1-B.4 for the institutional data collection.

II. INSTITUTIONAL DATA COLLECTION

As noted earlier, two different samples of institutions will be selected to address RQ3 and RQ4 that focus on the effects of hosting GRFP Fellows on institutions and program implementation. The first, *the Institutional Site Visit Sample*, consists of six institutions that will participate in a site visit during which the NORC team will conduct in-depth, in-person interviews with up to 10 administrators, faculty and staff to ask about the effect of the GRFP on the institution and students as well as implementation of the GRFP and recommended changes. The second, *the Institutional Phone Interview Sample*, consists of 20 institutions where up to five administrators, faculty, and staff will be asked—via short telephone interviews—about implementation of specific design and policy elements of the GRFP and recommended changes to the GRFP.

B.1. Institutional Data Collection: Sampling Methodology

The contractor will work with NSF to put together a sampling frame for institutions where Fellows in Cohorts 1-4 enrolled. This sampling frame will contain the name of the institution, location, and total number of Fellows enrolled at that institution from 1994 through 2011. The contractor will add institutional characteristics to this sampling frame—size of graduate student population, Census region, type of institution in terms of public/private and Carnegie classification, among others.

Institutional site visit sample: We hypothesize that effects on faculty, students, and the institution are likely to require some threshold number of Fellows. Thus, institutions will be ranked according to the total number of Fellows they have hosted in each of the cohorts. NSF and NORC will then jointly select a purposive sample of six institutions from among these institutions with perhaps greater weight given to those that rank highest for the most recent cohorts.

Institutional phone interview sample: We wish to obtain a more representative and diverse sample so that we can understand more broadly how the specific policies are affecting institutions with different characteristics. Thus institutions will be ranked on several dimensions (for example, size of graduate

[10] Cochran, William G. (1977). *Sampling techniques* (Third ed.). NY: Wiley.

enrollment in STEM fields, number of Fellows enrolled, type, reputation, and geographical location, among others) and selected into the sample based on characteristics of interest. Although not a random sample, the sample will be balanced to some degree with respect to the variables where we might expect some variation in responses. Greater weight may be given to institutions that are hosting Cohort 4 Fellows because we need to capture information on recently-changed policies and their effect on institutions and students. NSF and NORC staff will jointly select the sample of 20 institutions.

B.2. Institutional Data Collection: Procedures for the Collection of Information

For both samples: NSF will send a letter to the graduate dean/GRFP coordinating official informing them of the study and the purpose of the data collection and encouraging them to participate in the study. NORC will then follow-up with the dean/GRFP coordinating official to solicit participation in the study through an email that reiterates the purpose of the study and ask for assistance in identifying relevant faculty and staff and a contact person in the graduate office with whom they could work to set up the interviews. If needed, the dean will receive a follow-up phone call. As noted, we propose to interview up to five faculty and staff (institutional phone interview sample) and up to 10 faculty and staff (institutional site visit sample) at each institution. These will likely include deans, program chairs, program administrators, and faculty.

Rather than rely entirely on the dean, NORC will focus on representing different departments and faculty and staff positions, with the goal of recruiting those potential participants most likely to have insights on how the presence of Fellows affects the institution (deans, department chairs, and GRFP Fellows' faculty advisors). We will look to Principal Investigators of grants and GRFP coordinating officials on campus, and will refer to information from GRFP's administrative records (such as those gathered from the FastLane web page) for current representatives at each institution with existing knowledge of the GRFP from a management perspective. We will target these individuals to gain cooperation on each campus. In addition, when selecting respondents we will focus on individuals who have a history of interacting with Fellows.

Institutional phone interview sample: Each of the potential respondents will be contacted and informed of the purpose of the study. Sample respondents will be told that this is an information-gathering exercise aimed at understanding how GRFP policies affect institutions and students, how institutions implement those policies, and whether they have recommendations for changes in policy. Once a mutually-convenient time is decided, the interviews will be conducted via phone by a team of two NORC staff members. NORC will use a semi-structured protocol with potential follow-up questions and probes targeted to the type of respondent (Appendix B). Respondents will be asked for verbal informed consent and be assured of the confidentiality of their responses. They will be informed that neither institutions nor respondents will be identified in the final report and briefs, and data will be presented only in aggregate form. Illustrative quotes will be presented with the speaker described in a non-identifiable fashion. Interviewers will pre-code many of the responses while taking detailed notes, by identifying whether they fit into one of the expected types of responses for that question.

Institutional site visit sample: Sample respondents will be contacted and informed of the purpose of the study and that NSF is interested in understanding the value of the GRFP to both students and institutions and the larger effects of GRFP on institutions, students, and career outcomes. Once a mutually-convenient time for the site visit is decided, an agenda and interview schedule will be developed. For the site visits, the team will consist of two to three NORC staff members. The interviews will begin with an introduction to the study and ask for verbal informed consent. NORC will use semi-structured protocols with potential follow-up questions and probes (Appendix B). In-person interviews

will be recorded and transcribed for analysis, with the participants' permission (if participants refuse, a team member will take detailed notes about the interview in lieu of a recording and transcription).

B.3. Institutional Data Collection: Methods to Maximize Response Rates

We believe the contractor can obtain 100% response rates among the sample, especially with the cooperation of the dean or the GRFP coordinating official and given that both samples are fairly small. Most institutions benefitting from the GRFP are generally willing to participate in studies that may help inform the program. If a particular institution refuses to participate, they will be replaced, with NSF input.

B.4. Institutional Data Collection: Tests of Procedures

Interview protocols will be tested via cognitive interviews with faculty and administrators at graduate institutions similar to those selected for the site visit. This iterative cognitive interviewing process will allow NORC's qualitative research experts to quickly identify which questions yield answers relevant to the identified research questions, and which need to be revised or replaced to improve clarity and flow. NORC will pilot test the three interview protocols with at least two participants per protocol (i.e., at least six total participants) prior to the site visits. NORC will consult with NSF before selecting an appropriate local institution for pilot testing to make sure that we do not select an institution that would more appropriately be included in the full study.

The section below provides key contact information for the study.

B.5. Key Contact information

Key personnel who have been involved in the statistical aspects and who will be involved in collecting and analyzing data are presented in the table below (Table B.5). The contractor for collection and analysis of data in this study is NORC at the University of Chicago, Chicago, IL. Staff with experience in evaluation of research programs, expertise in scientific research, and knowledge of statistical methods, was involved in the design. NSF program staff members familiar with the programs have been included in the design of the evaluation.

Table B.5. Individuals Consulted

Name	Role	Phone
NORC at the University of Chicago		
Marie Halverson	Project Director	(312) 759-4041
Hee-Choon Shin	Sampling Statistician	(773) 256-6150
Gregory Wolniak	Task Leader	(312) 759-2356
Jake Bartolone	Task Leader	(312) 759-4002
Lisa Setlak	Task Leader	(312) 357-3774
Tom Hoffer	Senior Scholar	(773) 256-6097
Sheila Nataraj Kirby	Senior Scholar	(301) 634-9397
National Science Foundation		
Carol Stoel	Program Officer, Division of Graduate Education	(703) 292-8630
Gisele Muller-Parker	Program Director, Division of Graduate Education	(703) 292-7468
Gilbert John	Program Director, Division of Graduate Education	(703) 292-2343
Roosevelt Johnson	Program Director & COTR, Division of Research and Learning	(703) 292-5152