Research Experiences for Undergraduate Students Program Supporting Statement

B. Collections of information employing statistical methods

1. Respondent universe and sampling methods

- REU Data System

  Universe. The respondent universe for the REU Data System are:
  - College and university students interested in applying to the REU program (registration pilot) or to the Sites participating in the pilot (common application). These individuals will submit background information described in Part A as part of their application to the REU program.
  - Principal investigators (PIs) of the Sites *participating in the pilot or their designees (such as co-PIs or Site administrators).* These individuals will be asked to provide background information on their Sites (information useful to prospective applicants or NSF), admission decisions (who was admitted to their Site), and participation information (who actually participated in the REU program at their Site).

  Sampling methods. We will not use sampling as Sites need information from every applicant. Consequently, the system will collect basic information from the universe of about 8,436 students who wish to submit an application to an REU program Site in biology and earth sciences in summer 2019, and additional information from the subset of 6,865 students who wish to submit an application to an engineering or mathematics Site that is participating in the pilot (restricted to applications for summer 2019). A justification for the estimated numbers of students is provided in section 12 of Part A of this request. Note that collecting this information from the universe of applicants does not add cost or burden to the government or respondents. Indeed, it should reduce cost and burden on Sites and burden on respondents by providing an online, centralized alternative to the applications currently offered by individual Sites (see section 3 of Part A of this request for more details).

- **National student clearinghouse (NSC)**

  Universe. 1,578 rising seniors participating in the registration and common application pilots for the 2019 program cycle. These students are expected to graduate in 2020. In 2020, we will obtain information on educational outcomes (enrollment and graduation) from the NSC.

  Sampling methods. We will not use sampling given that the information we seek is available in a centralized location (the NSC) at zero burden for the participants and a low marginal cost for additional records ($.60 per student) after the initial 1,000 records are purchased at $1 per student. We also want to ensure adequate representation of groups that are traditionally underrepresented in STEM (minorities, students with disabilities, and others).

- **Employment Survey**

Universe. 1,113 rising seniors participating in the registration and common application pilots for the 2019 program cycle who, based on the NSC data, are not enrolled in undergraduate or graduate programs (or this outcome is unknown).

Sampling methods. We will not use sampling because (1) this is a relatively small population and (2) we seek to obtain an accurate measure of the expect response rate for the future. Note that the logic of this pilot is to test a system that can be easily implemented by NSF in the future. The information collected—including a non-response bias analysis comparing survey data to NSC and registration/common application data—will enable the study team to make a recommendation to NSF regarding whether and how to sample in future data collection rounds, if this approach to data collection is adopted.

2. Procedures for the collection of information
   a. Statistical methodology for stratification and sample selection

Sites participating in the pilot were selected in two stages. In the first stage, we identified disciplines that may be good candidates for the pilot (note that the REU program is organized into 11 mostly disciplinary groups, with Sites supported in each discipline ranging from 6 to 121). To do so, we analyzed NSF awards data and consulted with NSF POs who are knowledgeable about the Sites in their disciplines. We assessed disciplines to ensure (1) the discipline had a large enough number of Sites to offer informative findings that would be useful for other disciplines, (2) the disciplines were mostly self-contained (this was important to reduce potential confusion among applicants to disciplines included and excluded from the pilot, such as astronomy that has content overlaps with physics and Sites in both disciplines may receive applications from the same individuals), and (3) the discipline did not already have a common application that would be a confounding factor for the pilot. These conditions resulted in the exclusion of the following disciplines: computer sciences and advanced cyberinfrastructure; education and human resources; and social, behavioral, and economics sciences.

In the second stage, we offered NSF POs the possibility of contacting PIs across the remaining 8 disciplines to gauge interest in participating in the pilot (participation is voluntary). Five program officers invited PIs in their disciplines to take a brief survey and indicate their interest in participating in the pilot selected by the program officer (as indicated in Table 1; OMB control number 3145-0215). This yielded 4 disciplines selected for inclusion in the pilot based on two criteria: (1) high interest among PIs (37 to 72 percent of Site PIs responded to the survey and, on average, 92 percent of respondents expressed interest in participating) and (2) large enough numbers of Sites to run a robust pilot and receive feedback from representatives of key program constituencies. The disciplines selected for the pilot are: Biology and Geosciences (Earth Sciences[1]) for the registration, and engineering and mathematics for the common application.

_____

[1] Earth Sciences, a subdivision of Geosciences, ultimately preferred to be considered for the registration pilot.

Table 1. Principal Investigator Survey Results (selected disciplines in bold)

Discipline	Pilot	Total Active Sites[a]	Surveys Completed	Volunteer to Participate
Biology	**Registration**	**121**	**76**	**72**
Engineering	**Common Application**	**98**	**71**	**25**
Mathematics	**Common Application**	**58**	**29**	**25**
Geosciences	**Common Application**	**71**	**26**	**18**
Chemistry	Registration	72	20	18
Astronomy	Registration	21	0	
Physics	Registration	50	0	
Materials	Registration	64	0	

[a] Total active awards were obtained from FastLane and represent sites active as of May 31, 2016.

b. Data collection

- **REU Data System.** This is a web-based system to which users will have access for different purposes and at different points in time:

  Student applicants will be instructed to access the electronic system to complete their registrations or applications. Information will be disseminated widely through the NSF website, the Site websites, emails sent to all REU PIs (who will, in turn, inform prospective applicants), and announcements made at PIs meetings (PIs in some disciplines, such as biology, hold discipline-wide meetings virtually, in person near NSF, or at national/disciplinary conferences widely attended by academics).

  The system will be launched in the Fall of 2018 to coincide with the earliest opening of applications by Sites that start accepting applications in September and October. Applications to individual Sites will be accepted within the dates specified by PIs (to respect the current practice of allowing Site PIs to determine the application period).

- **NSC.** Mathematica will prepare and submit to the NSC a request for data on REU participant enrollment and graduations a year after participation in the pilot.

- **Employment Survey.** In alignment with current best practices in data collection, we will use a multi-mode contact protocol to guide respondents to a web-based employment survey. We will attempt to reach participants via email, text (if they consent) and, if needed, postal mail to take them to a survey instrument that will be dynamically optimized for a range of screen sizes, including computers and mobile devices (Dillman 2017; Finamore and Dillman 2013).

3. Methods to maximize response rates and deal with nonresponse

The pilot is designed to determine what response rate is possible using only low-cost outreach strategies while maximizing data quality.

- **REU Data System.** Non-response is not a concern for the common application, as students will be asked by PIs to submit their applications through the system. Non-response is a concern, however, among registration users as it requires enforcement by PIs, who will be asking students to complete their Sites' application processes in addition to the registration. Participating PIs will be asked to (1) adapt their applications to avoid redundancies, as they will have access to student information submitted online by students who later apply to their Sites, and (2) require that all applicants complete the registration and provide them with their user ID. PIs will need the students' IDs to obtain information through the system and record admissions and participation decisions. To encourage adherence with these guidelines, we will host webinars with participating PIs to introduce them to the system, share detailed guidelines regarding participation in the pilot, and answer questions. We will also share guidelines in writing among all participating PIs. The pilot will test the success of this approach and elicit feedback to consider improving the system in the future if NSF choses to adopt it.

- **NSC.** Non-response is not a concern, but non-matches is. The purpose of the pilot is, in part, to test the feasibility of obtaining educational outcomes information on REU participants through the NSC.

- **Employment Survey.** We will ask NSF to send (or ask PIs to send) an email announcing the survey to former REU participants and encouraging participation. Mathematica will then send an initial welcome email to all sample students that includes information about the study and a link to access the survey. Students will be given 8 weeks to complete the survey and nonresponders will receive several reminders. If participants consent, we would also like to send text message invitations. In addition, in weeks 4 and 8 of data collection, we will forward to the REU program a list of nonrespondents, requesting its support in encouraging the students to complete the survey. We will also consider sending a mailed paper invitation with a link and PIN in a later contact. Research suggest that, although this strategy does not increase responses rates significantly, it may encourage responses from different types of respondents who are not well represented among the initial Internet responses (Dillman 2017). A postal service mail contact attempt may also have a better chance of being opened and read by respondents who receive a large volume of email.

4. Tests of Procedures or Methods to be Undertaken

- **REU Data System.** We will conduct two types of pre-tests—alpha testing to ensure the system is working properly, and beta- or usability-testing to ensure the content works as intended, that is, items are eliciting the responses we seek. Specifically:

  Alpha-testing. The study team will conduct testing of the web-based system to ensure that 1) the site functions properly, 2) the user interface is friendly and clear, and 3) the business logic aligns with specifications. Site functionality testing will involve testing the flow of

the site between pages, verifying the ability to enter and retrieve appropriate information based on user role, and ensuring that the data collection capabilities of the site work properly and securely with website's attached database. Site user interface testing ensures that the site's buttons, tabs, scroll bars, text boxes and other inputs are laid out in an intuitive way and function properly. Business logic verification ensures that all text content and status reports are aligned with the business needs for the instrument and the specifications which document those business needs. The goal of these tests is to ensure that the products—the registration and the common application—work properly before we move to the stage of beta-testing.

Beta- or usability-testing. We will pre-test the registration and common application with intended users. Both were designed based on existing REU applications and modeled after the NSF GRFP common application (OMB control number 3145-0023). We will request fast-track clearance using NSF's existing active generic approval for information collection (3145-0215). We plan to reach out to 10 college students (5 will test the registration and 5 will test the common application) and 5 PIs (at least one in each of the disciplines included in the pilot).We will also share a link to the system with the 29 NSF POs associated with this work or the REU program (this includes our contracting officer representative and NSF REU POs) for them to review the system if they wish to do so (all are government employees). Before launching the collection, we will make revisions to the system based on feedback received from students, PIs, and NSF POs.

- **NSC.** Not applicable as, for many years, the NSC has been collecting and providing the data we will be requesting.

- **Employment survey**. Pretesting the web-based employment survey is vital to the integrity of data collection. We will conduct alpha and beta testing similar to that described earlier for the REU Data System. The goal of the alpha testing will be to ensure that the survey behaves as intended. The goal of the beta-testing will be to assess how respondents understand the terms and questions presented in the survey, assess the accuracy and relevance of the questions, determine if any important questions are missing, and determine the length of time the survey takes to complete. We will test the survey with 8 former REU participants, 2 in each of the 4 participating disciplines.

5. **Individuals consulted on statistical aspects of the design**

Dr. Michael Sinclair, senior fellow at Mathematica, is a statistician who was formerly with the U.S. Department of Justice and Census Bureau and has extensive expertise in areas such as sampling methodology, survey weighting and imputation, and data linkage.

Andrew Weiss, vice president at Mathematica, is an expert in data collection systems, research operations, and survey methodology.

REFERENCES

Baum, S., and P. Steele. "Who Goes to Graduate School and Who Succeeds?" 2017. Available at

https://www.urban.org/sites/default/files/publication/86981/who_goes_to_graduate_school_and_who_succeeds_1.pdf. Accessed February 1, 2018.

Dillman, D. "The Promise and Challenge of Pushing Respondents to the Web in Mixed-Mode Surveys." *Survey Methodology*, vol. 43, no. 1, 2017, pp. 3–30.

Finamore, J., and D.A. Dillman. "How Mode Sequence Affects Responses by Internet, Mail, and Telephone in the National Survey of College Graduates." Presentation to the European Survey Research Association, Ljubljana, Slovenia, July 18, 2013.

Galesic, M., and M. Bosnjak. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly*, vol. 73, no. 2, 2009, pp. 349–360.

National Science Foundation. "REU Program Solicitation NSF 13-542." 2013. Available at https://www.nsf.gov/pubs/2013/nsf13542/nsf13542.pdf. Accessed February 1, 2018.

U.S. Department of Education. "Employment and Enrollment Status of Baccalaureate Degree Recipients 1 Year After Graduation: 1994, 2001, and 2009." 2016. Available at https://nces.ed.gov/pubs2017/2017407.pdf. Accessed February 1, 2018.

Zuckerman, B., J. Doyle, A. Mudd, T. Jones, and G. Davis. "Assessment of the Feasibility of Tracking Participants from the National Science Foundation's Research Experiences for Undergraduates (REU) Sites Program." Final report. Washington, DC: STPI, 2016.