[OBJ]

Supporting Statement for Science Undergraduate Laboratory Internship (SULI) Long-term Follow-up Study

Part B: Collections of Information Employing Statistical Methods

OMB No. 1910-XXXX

Form DOE-F 413.42 / 413.43, Survey of Former SULI Program Participants and Applicants

February 2025

Table of Contents

B.1. Respondent Universe

Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used.

The respondent universe consists of (1) former participants of the SULI program and (2) a group of individuals that are similar to (1), but who did not participate in the SULI program, although they applied to the program, were eligible to participate, and otherwise complied with all application requirements.

Table B1 provides the number of SULI participants as well as non-participant applicants by semester and year.

Table B1: Former SULI Participants and Non-Participants Considered for Inclusion in Study Group and Comparison Group by Semester, Summary of Participants and Non-Participants

	Fall 2013	Spring 2014	Summer 2014	Fall 2014	Spring 2015	Summer 2015	Fall 2015	Spring 2016	Summer 2016	Total
Participants	70	73	531	80	107	590	81	101	614	**2,247**
Non-Participants out of Eligible and Compliant Applicants	105	135	1,218	39	83	1,091	39	56	989	**3,755**
Total	**175**	**208**	**1749**	**119**	**190**	**1681**	**120**	**157**	**1603**	**6,002**

Study Population (Former SULI participants)

The study will focus on the SULI program delivered by DOE in Fiscal Years (FY) 2014 through 2016. Given the size of the SULI population, the interest in disaggregating by multiple subgroups, and the availability of contact information for most participants, evaluators will aim to conduct a census study, with targeted follow-up for non-respondents.

In the data analysis phase, the evaluation team will conduct a set of descriptive analyses using data from SULI participants only. In these analyses, we will focus on three different categories of descriptive analyses: the composition of those who participated in the SULI program during the study period, the distribution of outcomes achieved by participants as indicated in survey responses, and how these outcomes vary by different background characteristics. These analyses will include multiple variables of interest, including career preparation, educational attainment, employment, research competencies/skills, professional collaborations, intellectual products and accomplishments, program endorsement, and alumni involvement. The results will be presented as descriptive statistics and cross-tabulations, and in cases where the question calls for an assessment of differences between cross-tabulated groups, we will present regression models and findings of inferential statistical tests.

Comparison Group (SULI applicants who did not participant in the program)

In order to estimate the impacts of the SULI program, the evaluation team will use other eligible applicants who did not participate in the SULI program as a point of comparison – denoted as "non-participants (out of eligible applicants)." The WDTS Application Review System (WARS) uses several

unique statuses to differentiate applications and, eventually, participation. The application status indicates whether the application was started, submitted, complete, or withdrawn; eligibility status indicates whether the application met eligibility requirements; and the compliance status indicates whether the application and materials were compliant with submission requirements.

Only applicants with complete, eligible, and compliant status are able to receive an offer for the SULI program. Eligible but non-compliant applications may be missing one or more required elements (such as transcripts and essays), or may have other non-compliant factors (such as personally identifiable information left on documentation), and these applicants are not eligible to receive an internship offer. Once an application meets the requirements for complete, eligible, and compliant status, an offer status is assigned. All of the SULI alumni participants will have an offer status of "accepted," as they accepted an offer and participated in SULI. For the purposes of constructing a comparison group, ORISE will consider only individuals with complete, eligible, and compliant applications, in order to select the most equivalent counterparts to SULI alumni participants.

Preliminary analyses will address and make a determination concerning how to treat "duplicate" participants who held more than one SULI appointment, or non-participant applicants during the SULI LFS study period who did hold a different SULI appointment outside of the study period. Additional techniques to ensure the equivalence of the comparison group are under consideration, but the comparison group will be drawn only from eligible applicants who did not receive an offer or participate in the SULI program.

B.2. Statistical Methods

Describe the procedures for the collection of information including:

The sampling design for the respondent universe will be a census approach, with the objective of maximizing participation among all eligible respondents. Across the 6,000 potential respondents, ORISE has considered the ease of contacting them, based on whether program records include multiple email addresses and based on the characteristics of those addresses. The most difficult contact group is typically individuals who at the time of their participation in SULI only had a single *.gov* email address, since access to those accounts is cut immediately after a role change; in the case of SULI, this group is negligible (6 individuals). The second most difficult group may be the 40% of potential respondents who only provided one email address, and it is a *.edu* email address, which might be out of date or inaccessible. About 37% of potential respondents have multiple email addresses noted, increasing ORISE's ability to contact them for the LFS. About 44% have at least one Gmail address, the most likely to be persistent and accessible years after participation. Thus, the first challenge of administering this long-term follow-up survey will be contacting past participants and comparison group members.

ORISE will use a broad approach to securing contact information. Prior to data collection, ORISE will work through WDTS to request that laboratory points of contact provide any updates to alumni contact information. Web scraping from publicly-available information, such as employer websites, university websites, LinkedIn profiles, or CVs posted online will also be used to update contact information. We anticipate that the comparison group may be the most difficult to reach, and we will therefore plan to use paid contact tracing services to find the most up-to-date email and telephone information available; additional tracing may be necessary after initial pre-invitation notifications are sent out, once ORISE is

able to determine which addresses will cause bounce backs. This broad approach will form the backbone of our contact database.

ORISE will follow a broad approach to securing responses, including phone calls. During data collection, the evaluators will note "bounce backs" or undelivered emails, as well as unanswered phone calls or other contact attempts, so that SULI can differentiate between individuals who chose not to respond to the survey and individuals who were not reached with an invitation. The contact protocol will be as follows:

1. Pre-invitation notice
   a. Alternative contact information will be tried for bounce backs
   b. Additional contact information will be sought after all contacts fail
   c. Phone calls will be used to obtain email information after all contacts fail
2. Invitation and two reminders delivered via email, first administration wave
3. Invitation and two reminders delivered via email to a subset, second administration wave
4. Invitation and two reminders delivered via email to a subset, third administration wave
5. Telephone follow-up.

Quasi-Experimental Design with Propensity Score Estimation

To estimate the impacts of the SULI program, the evaluation team will utilize a quasi-experimental design that incorporates a comparison group similar to the treatment group on baseline characteristics. First, to achieve reasonable similarity with the SULI treatment group, the comparison group will be limited to other eligible applicants who did not participate in the SULI program. Limiting to only those who also applied to the program is the first layer of protection to ensure similar levels of familiarity with and interest in SULI prior to exposure to the program.

In order to further adjust for observed differences between the treatment and comparison groups, the evaluation team will employ propensity score estimation (Austin, 2011; Rosenbaum & Rubin, 1983). This technique, with four decades of application and rigorous testing in studies similar to this long-term follow-up evaluation, will support a study without randomized control assignment, while still accounting for differences between the groups. This technique is intended for quasi-experimental studies in which researchers were not able to assign individuals to treatment (SULI participation) and non-treatment (no SULI experience) groups. Propensity score estimation techniques will allow the study team to weight the data associated with the non-participant group, and carry these weights through all analyses, to ultimately draw conclusions about the outcomes of SULI participants as compared to those who did not benefit from the program (Curtis et al., 2017; Lee, Lessler & Stuart, 2011; Li, Morgan & Zaslavsky, 2018).

Propensity scoring techniques use available pre-treatment baseline variables (such as GPA at time of application) and stable variables (such as sex, race or date of birth) to predict the likelihood that a given member of the analytic sample participated in the program. The predicted likelihood is then represented as a propensity score, which can be used to weight data from the comparison such that non-participant applicants who are most similar to the treatment group are given greater weight in the analyses than those who look less like the treatment group. Essentially, propensity score techniques focus on achieving balance between the intrinsically non-equivalent SULI participant group and the non-participant comparison group by weighting them differently. The data weighted by propensity score will then be carried through all analyses. Using this technique, we will then estimate differences between

those who participated in the SULI program versus those who did not on outcomes of interest including career preparation, educational attainment, employment, research competencies/skills, professional collaborations, and intellectual products/ accomplishments.

While this quasi-experimental approach will provide us with estimates of SULI's impact, it is worth noting that propensity score estimation may not fully eliminate differences between the SULI participants and the comparison group, and may not fully achieve the desired balance. In particular, because the comparison group is primarily made up of applicants who were not offered admission to the program, it is likely that the alumni group and comparison group were not equivalent at time of application. These differences, including in background and preparation, were likely correlated with later outcomes. Further, these non-equivalent characteristics are not fully captured by the data available from the application system, since a wide range of applicant characteristics may have an influence on selection into SULI as well as future outcomes. The only technique to fully eliminate this non-equivalence would be a carefully designed randomized controlled experiment. Given that an RCT is not a viable option for WDTS programs, propensity score estimation is the strongest and most rigorous tool to address this risk.

Like all analysis approaches, propensity score estimation techniques allow for some fine-tuning, discretion, and specific selection between methodological options for how to treat the data and how to conduct the estimation of propensity scores and weighting of the data. We anticipate calculating propensity scores and using these values as weights on the non-participant datapoints during all analyses. Specific methodological determinations depend on the characteristics and distributions of data in each group, and therefore on access to the full dataset for participants and non-participants. These detailed decisions will be explored and discussed by the ORISE evaluation team, in coordination with WDTS, once the study described in this proposal is approved and commences. Further methodological decisions will only be possible once all data have been collected and are available for detailed examination.

B.3. Maximizing Response Rates

Describe methods to maximize response rates and to deal with issues of non-response.

Target Response Rate

Long-term follow-up research faces intrinsic challenges with achieving high response rates, particularly when potential respondents have not had ongoing contact with the program or organization seeking a response (Blaney et al. 2019). The survey response literature has examined multiple reasons for non-response that are relevant to the SULI LFS including lack of interest, being too busy, the time burden of completing the survey, privacy concerns, or the voluntary/opt out nature of the survey (National Research Council, 2013). A reasonable response rate for this study may be 30%, and ORISE plans to strive for a 40% response rate or above using best practices and additional methods to track and contact alumni. In Techniques to Secure and Increase Response Rate, we discuss our planned approach to maximize response rates and to minimize non-response bias.

Techniques to Secure and Increase Response Rate

Internet survey response rates can be increased by securing high-quality contact information, by providing a quality survey invitation and response experience, and by providing incentives. In this study, we will also use tracers and web scraping efforts to contact potential respondents. The willingness for a

respondent to follow through on a survey invitation, represented through survey response rates, is influenced by their interests, the structure of the survey, the communication methods used for invitations and reminders, rewards/financial incentives, and assurance of confidentiality (Saleh & Bista, 2017). Because no single factor will have a strong effect on the response rate, we will combine multiple approaches that can together provide a substantial result (Trespalacios & Perkins, 2016).

Fan & Yan (2010) provide an overview of best practices to increase response rates on web surveys, including:

- official sponsorship by an academic or government agency (e.g., WDTS/SULI);
- reminding potential respondents of their interest/the relevance of the topic;
- reducing the time burden as much as possible;
- providing pre-notification of an incoming survey from the recognizable survey sponsor (e.g., SULI) followed by an email invitation;
- sending reminders two days after the initial invitation;
- personalizing invitations;
- avoiding attachments and/or HTML in the invitation;
- clearly communicating the survey's purpose and the task in the invitation;
- an identification of how the researchers located the email address (e.g., program records);
- assurances that respondent details and their responses will be kept confidential and findings will only be reported in the aggregate; and
- an incentive program.

Achieving a response rate above 40% will reduce the chances of non-response bias, which results when invitees who choose to respond differ from those who do not choose to respond in ways that bias the results (e.g., demographics factors, outcomes like employment or educational achievement). In this study, we will conduct thorough response/non-response bias analyses.

<u>Pre-Invitation Notice, Invitations, Reminders</u>

ORISE will coordinate with SULI/WDTS and with the contact individuals at individual DOE labs to arrange for pre-invitation notices to be send to potential respondents. This will serve multiple purposes: first, it will come from a recognizable source and prompt potential respondents to expect the email with the survey. Secondly, this will give us an early warning system for the number of undeliverable email addresses, and will allow ORISE to begin seeking alternative contact information for those individuals. This pre-notice should be inviting, should be professional and clearly indicate the sponsor, should be personalized, should mention the incentive strategy, and should mention the response collection window. The pre-invitation notice should be sent from an "authority figure" or known personnel/organizations (Saleh & Bista, 2017).

Invitations will indicate that a response is requested within two weeks of receipt. However, responding within this timeframe is not a requirement of participation. Automated reminders will be distributed four days after the initial invitation. This initial wave will be followed up with an additional three-week period to make contact with individuals with failed email addresses and non-respondents with an

additional available email address. Supplemental invitations and reminders will be sent for subsequent waves. With each wave lasting a minimum of two weeks, plus ample extra administration time between waves, the data collection period will last a minimum of 15 weeks, for an estimated total data collection duration of approximately 15-18 weeks.

Administration Waves

ORAU's survey distribution strategy will pursue multiple waves. There are two different reasons for us to reach back out to participants after initial contact:

1.   To make first contact with individuals whose available contact email-addresses failed; and

2.   To solicit responses from individuals who received, but did not respond to, the initial invitation.

We anticipate that certain cases will require more effort to contact and to convince to respond (Calderwood, 2016): SULI non-participants; individuals without Gmail addresses or with only a school/employer address at time of participation, etc. Further, we are aware that the study's reliability depends on having a representative sample . We will focus our contact interventions by increasing effort to locate and incentive values (Calderwood, 2016).

First contact will be attempted through a rigorous contact protocol. The number of failed contacts will not be known until the pre-invitation notice goes out. Evaluators will gather "bounce back" email notifications, document which emails were unreachable, and then exhaust all available email addresses in an effort to reach the participant. If an individual cannot be reached via any of these addresses, we will provide their information to a contact tracing service in an attempt to retrieve a more recent email address. If this fails, we will contact them via phone and use a standard phone script to solicit their participation in the survey. Similarly, two weeks after the first wave, evaluators will attempt second and third email addresses for non-respondents. Exhausting all contact methods and giving each potential respondent a full three-week period to receive and respond to the invitation will comprise the first wave of survey administration. This approach will help us not only secure as many responses as possible, but eliminate a source of non-response bias, as we expect that there will be differences between those with useable email addresses and those without, such as their degree advancement at time of last contact (Fowler et al., 2019).

Between administration waves, the evaluation team will also conduct the random selection process for the incentives in the previous waves (selecting both respondents and non-respondents). Non-respondents who were selected will be contacted with information about the gift card, as well as a final reminder with the link to the survey. We will reserve some incentives for those individuals who are reached later in the cycle, so that every potential participant has equivalent access to the pool of incentives.

Once the first wave is complete, we will shift our attention to soliciting responses from non-respondents (both SULI alumni and the comparison group of non-participants) in two additional waves. We will use two research-backed strategies: responsive design, which modifies data collection plan based on monitoring incoming data (Czajka & Beyler, 2016), and a "prevention" approach in which we attempt to

convert non-respondents into respondents (Calderwood, 2016; this can result in up to a 25% conversion rate).

In these additional contact waves, we will attempt to address non-response bias through our incentive strategy. We will assess non-response bias in demographic categories as described above, then selectively reach out to a stratified random sampling of non-respondents to encourage them to re-consider. We anticipate focusing on approximately 1,000 SULI participants and 1,000 non-participants in the second wave, and 500 of each in the third wave. In the second wave, we will modify the invitation to emphasize SULI and DOE's needs and interests. In the third wave we will also add telephone contact. Specific counts will depend on the number of non-respondents and the degree of non-response bias measured following the previous wave.

Incentives

Incentives are a key tool for increasing response rates (Czajka & Beyler, 2016; Saleh & Bista, 2017), and will be especially important in the SULI contact given that these are long-term follow-up surveys for participants who have not otherwise had ongoing contact with the program and non-participants who are otherwise not intrinsically motivated to assist.

The incentive strategy will follow three different tracks for the first administration wave and for subsequent waves. Based on a review of incentive plans for federally funded surveys with similar data collection methodologies, WDTS and ORISE selected a plan that includes three different tracks for the first administration wave, with increasing incentive amounts for two of these tracks on the second and third waves. In the first administration wave, potential respondents will be randomly assigned into three tracks: 25% will be offered $30, 25% will be offered $20, and 50% of potential respondents will not be offered an incentive. On the second wave, potential respondents who were not previously offered an incentive will be offered 20$, and those who were previously offered 20$ will be offered $25. On the final administration wave, those who were offered $20 or $25 on the previous wave will be offered $30. The maximum total incentive a respondent can receive is $30.

Table B2. Incentive Amounts by Administration Wave

Wave	Incentive Strategy for Alumni	Incentive Strategy for Non-Participants
First Wave – Initial Email Invitation + 2 Reminders	Initial Email addresses $30 (25%) $20 (25%) $0 (50%)	Initial Email addresses $30 (25%) $20 (25%) $0 (50%)
Second Wave – Targeted Email Invitation + 2 Reminders	$30 (25%*) $25 (25%*) $20 (50%*)	$30 (25%*) $25 (25%*) $20 (50%*)
Final Wave – Targeted Phone Call + Email Invitation + 2 Reminders	$30	$30

*Approximate percentages

In the second and third administration waves, we will attempt to address non-response bias through our incentive strategy. We will assess non-response bias in demographic categories determined by previous analysis of the demographic composition of the study population, then selectively reach out to a stratified random sampling of non-respondents to encourage them to re-consider. Evaluators will select a balanced stratified random sample of remaining non-respondents in both the SULI alumni and the comparison group, and contact them for the following wave of administration with increased incentive values and a modified invitation. The second wave will target approximately 2,000 non-respondents (1,000 SULI participants and 1,000 non-participants) in this way, and the third wave will target an additional 1,000 (500 SULI participants and 500 non-participants). Specific counts will depend on the number of non-respondents and the degree of non-response bias measured following the previous wave. Non-respondents who are in a track with increasing incentives and who are selected into the stratified random sample for the second wave will be contacted separately.

ORISE will follow standard procurement procedures to identify potential third-party vendors and secure bids to manage the process of distributing incentives to respondents. At this time ORISE is aware of 3 different vendors that provide this type of service. ORISE will seek guidance from legal counsel, from Paperwork Reduction Act guidance, and from the ORSIRB to determine the specific details of the incentive structure, including allowable costs and the feasibility of increasing incentives in later administration waves. ORISE understands that WDTS will have its own clearance process, and that the proposed structure may need substantial revisions and overhauls in order to comply with WDTS, ORISE, and OMB requirements.

ORISE A&E will also calculate response rates post-stratified by incentive plan. For this analysis, the sampling frame will be stratified into groups using two different post-stratifications. The first will correspond to the 3 different incentive levels. The second will correspond to the 3 administration waves. Crossing these two stratifications will yield a 3 X 3 matrix, corresponding to 9 (potentially different) response rates. In addition to using this matrix to characterize response rates, ORISE A&E will also investigate what impact, if any, the incentive plan has on nonresponse bias. To estimate any impact on nonresponse bias, ORISE A&E will examine whether respondents receiving different amounts, or different timing of incentives, look different across a range of demographics and baseline factors. These analyses will be carried out after all data have been collected, and they are separate from the planned ongoing nonresponse analysis that will be conducted to identify any high non-response groups that might be more responsive with targeted incentives. These findings will be shared with the study sponsor and can be provided to the IRB either as an interim finding or as part of study closure.

Non-Response Adjustment

While a non-response analysis and an outcome validation analysis will provide the evaluators with insight about the effects of non-response once data collection is completed, we will also plan to address non-response bias during the administration waves.

Each of the three administration waves will last at least two weeks, with up to two weeks in between for planning and implementing the next wave of the survey. During and following the first two

administration waves, the evaluation team will monitor both the SULI alumni responses and the comparison group for disproportionate responses in the following demographic categories:

- Year of participation in/application to SULI
- Sex
- Race
- Ethnicity
- Disability or impairment
- GPA at time of application
- Community college attendance and/or Carnegie classification of institution at time of application

Non-response bias will be assessed using t-test and multivariate models to compare respondents to non-respondents on the baseline characteristics described above. To account for any bias that is detected, the evaluators will select a balanced stratified random sample of remaining non-respondents in both the SULI alumni and the comparison group, and contact them for the following wave of administration with increased incentive values and a modified invitation. The second wave will target an additional 1,000 potential respondents in this way, and the third wave will target an additional 500 respondents. Between administration waves, the evaluation team will also conduct the random selection process for the gift cards in the previous waves (selecting both respondents and non-respondents). Non-respondents who were both selected for a gift card in the previous wave and who were selected into the stratified random sample for the second wave will be contacted separately. At the end of data collection, the evaluation team will also compare all data from early responders to later responders. Because it is assumed that those who responded later after multiple reminders and incentives are likely more similar to non-respondents than those who responded at the beginning of the survey administration, this will provide insight into any potential remaining differences due to non-response.

B.4. Test Procedures and Form Consultations
Describe any tests of procedures or methods to be undertaken.

Beta tests were conducted to assess survey functionality and visual appeal/readability.

- Survey functionality – We have tested all survey buttons (e.g., next, back, submit), save and resume functionality, response options (e.g., checkboxes, radio buttons, dropdown lists, text fields), skip logic, to ensure they all function as intended and provide the correct information in the data file when downloaded.

- Visual appeal/readability - We tested the survey display to ensure the text is sized and colored appropriately for readability. We reviewed survey content to optimal formatting (e.g., dividing up a question into two separate questions to not lose column headers, breaking up text into different paragraphs, using italicization, etc).

B.5. Statistical Consultations
Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s) or other person(s) who will actually collect and/or analyze the information for the agency.

The SULI LFS survey will be conducted by ORISE, a private non-profit contractor, selected in a competitive bid process. Understanding of survey analytics and project management are important components in the selection of a contractor. Individuals consulted in the design of the study include:

Ann Martin, PhD, ORISE Senior Evaluation Specialist
Ann.Martin@orau.org
(607) 296-5698

Tony Garcia, PhD, ORISE Senior Evaluation Specialist
Tony.Garcia@orau.org
(865) 722-1459

Erin Burr, PhD, ORISE Senior Evaluation Specialist
Erin.Burr@orau.org
(865) 440-6852

References

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46(3), 399-424.

Blaney, J. M., Sax, L. J., & Chang, C. Y. (2019). Incentivizing longitudinal survey research: The impact of mixing guaranteed and non-guaranteed incentives on survey response. The Review of Higher Education, 43(2), 581-601

Curtis, L. H., Hammill, B. G., Eisenstein, E. L., Kramer, J.M., & Anstrom, K. J. (2007). Using inverse probability weighted estimators in comparative effectiveness analyses with observational databases. Medical Care, 45(10), 103-107.

Czajka, J. L., & Beyler, A. (2016). Declining response rates in federal surveys: Trends and implications (Mathematica Policy Research Reports) [Background Paper]. Mathematica Policy Research. https://EconPapers.repec.org/RePEc:mpr:mprres:a714f76e878f4a74a6ad9f15d83738a5

Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. PloS One, 6(3), e18174. http://dx.doi.org/10.1371/journal.pone.0018174

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. Journal of the American Statistical Association, 113(521), 390-400.

National Science Foundation, National Center for Science and Engineering Statistics, Scientists and Engineers Statistical Data System (SESTAT). (n.d.). Available at https://www.nsf.gov/statistics/sestat/

Paperwork Reduction Act of 1980. Pub. L. No. 96-511, 94 Stat. 2812, codified at 44 U.S.C. §§ 3501–3521 (1980)

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Saleh, A., & Bista, K. (2017). Examining factors impacting online survey response rates in educational research: Perceptions of graduate students. Journal of Multidisciplinary Evaluation, 13(2), 63-74.