National Center for Science and Engineering Statistics

Title: Creating Longitudinal Panels for the NSCG

Date: December 2021
Final Report

Contractor Awardee: NORC
Contract Number: 49100420C0024

FINAL REPORT

DECEMBER 2021

NATIONAL SCIENCE FOUNDATION BROAD AGENCY AGREEMENT

Creating Longitudinal Panels for the NSCG

Presented by:
NORC at the University of Chicago

Presented to:
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230
(703) 292-5111

Authors
Donsig Jang
Josiane Bechara
Lance A. Selfa
Lisa Lee
Lee Florio
Quentin Brummet
Tom Hoffer

Table of Contents

List of Tables

List of Exhibits

Executive Summary

The March 12, 2020 Broad Agency Announcement (BAA) from the National Center for Science and Engineering Statistics (NCSES) identified "moving towards a longitudinal and condensed survey design for ongoing surveys" as a strategic research priority area. The ongoing surveys of particular concern include the three NCSES workforce surveys: the National Survey of College Graduates (NSCG), the Survey of Doctorate Recipients (SDR), and the Early Career Doctorate Survey (ECDS). Responding to this priority area, the NORC project addressed four shortcomings of the NCSES survey program identified in the BAA, most of which are also discussed in the landmark 2018 Committee on National Statistics report, "Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs"[1]:

- Individuals residing in the United States who earned research doctorates in science, engineering, and related health (SEH) fields of study from universities outside the U.S. are not included in the SDR sample frame. The numbers of these individuals are substantial, and a better understanding of their career pathways and outcomes is needed (NAS, 2018: pp. 44-45, Recommendation 3-2; pp. 74-78, Recommendations 4-2 and 4-3). These individuals are eligible for the NSCG sample but are not a separate sample stratum and are sampled at relatively low rates. They are also eligible for the ECDS, but sample eligibility is restricted to doctorate recipients working in U.S. academic institutions or federally funded research and development centers (FFRDCs), and who earned the doctorate within the last 10 years. Neither the foreign doctorate holders sampled by NSCG or ECDS are integrated with the SDR.
- The first 10 years of doctorate recipients' careers are particularly important to their overall career paths and outcomes, but the early career doctorate population is not fully represented in either the ECDS or the SDR (NAS, 2018: pp. 84-85). The SDR frame does not include the foreign-earned doctorate recipients, and the ECDS frame does not include individuals employed in nonacademic/non-FFRDC settings.
- Doctoral students are a vital segment of the SEH enterprise but are not targeted by any NCSES labor force surveys. Gaining a better understanding of who enrolls and who completes the doctorate, and what factors affect completion and career plans are of particular interest. Doctoral students are college graduates and therefore are eligible for the NSCG but are not explicitly sampled as a separate stratum and are not represented in sufficiently large numbers under the current sample design to support separate analyses. Sampling sufficient numbers of doctoral students may require expanding the NSCG frame with new data sources; the NAS report notes that the National Student Clearinghouse (NSC) is a promising resource in this regard (NAS, 2018: pp. 79-82).

[1] National Academies of Sciences, Engineering, and Medicine. *Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs*. Washington, DC: The National Academies Press; 2018.

- Longitudinal data collection, analysis, and reporting have not been part of the NCSES labor force surveys but are needed to address important questions about career paths and outcomes (NAS, 2018: pp. 41-44, Recommendation 3-1). SDR now has a longitudinal design component, but analysis and reporting have lagged, and the longitudinal design needs to be more fully realized in the survey instrumentation. NSCG has a rotating panel sample design but does not have a longitudinal design component. ECDS completed an initial round of data collection in 2017 and plans to follow up with additional cycles, but with a repeated cross-sectional design.

To help identify ways to address these shortcomings, NORC undertook a multifaceted exploratory project focused on the NSCG and investigating potential sources for improving the sample. As part of a comprehensive NCSES data collection program on the scientific and technical workforce, changes to the NSCG potentially have implications for other NCSES surveys, including the SDR and the ECDS. This executive summary reviews the main objectives and analytic findings and our recommendations based on those findings.

ES1 Summary of Project Objectives

The NORC project had four main objectives:

1. Assess the adequacy of NSCG sample coverage of individuals residing in the United States who earned a science, engineering, or health (SEH) doctorate outside the U.S. and, if not adequate, identify ways to improve the sample, particularly for supporting longitudinal analyses of the doctoral workforce consistent with the SDR design.
2. Extending the first objective, assess the adequacy of NSCG sample coverage of foreign-earned doctorate recipients in the early career stage—i.e., those who earned their doctorate within the last 10 years. Adequate coverage of this subpopulation would support a redesign of the ECDS project to draw from the SED and NSCG frames and support longitudinal analyses across the early career stage.
3. Assess the adequacy of NSCG sample coverage of individuals currently enrolled in research doctorate degree programs in the United States and, if not adequate, identify ways to improve the sample, particularly for longitudinal analyses.
4. Identify changes to the NSCG survey instrument that would be needed to collect the longitudinal data on foreign doctorate holders and doctoral students.

To meet the objectives related to sample design, the project conducted investigations of the American Community Survey (ACS) data used to define the NSCG sample frame and the National Student Clearinghouse (NSC) data as a possible additional sample frame for current doctoral students.

ES2 Summary of Main Findings

Task 1: Evaluation of NSCG Sample

The sample sizes of individuals who earned SEH doctorates outside the U.S. and doctoral candidates enrolled in U.S. institutions in the NSCG are too small to obtain sufficiently accurate estimates of these groups' characteristics, experiences, and outcomes. ES Table 1 shows that the cross-sectional numbers from the 2010-2019 NSCG cycles are relatively low for doctoral candidates and for those earning their doctorates outside the United States.

ES Table 1. Cross-sectional population estimates for selected SEH groups (sample sizes in parentheses), NSCG 2010-2019

SEH Populations with sample sizes (sample sizes in parentheses)	2010	2013	2015	2017	2019
Enrolled in PhD program	290,000 (1,445)	306,000 (2,545)	359,000 (1,764)	364,000 (1,319)	367,000 (1,405)
Earned PhD in U.S.	1,049,000 (2,712)	1,092,000 (5,445)	1,174,000 (5,113)	1,345,000 (4,929)	1,548,000 (5,843)
Earned PhD in U.S. between 2000 and 2010	331,000 (967)	313,000 (1,470)	290,000 (1,179)	316,000 (1,094)	333,000 (1,269)
Earned PhD outside U.S.	199,000 (1,211)	180,000 (1,014)	181,000 (801)	186,000 (801)	253,000 (1,608)
Earned PhD outside U.S. between 2000 and 2010	61,000 (225)	55,000 (311)	48,000 (254)	53,000 (243)	55,000 (446)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National College Graduate Study, 2010-2019 (public use files).

For our longitudinal analysis, we focused on the NSCG 2010 new cohort panel that was preselected to be retained through the 2017 survey round. The sample sizes in the longitudinal panel are much smaller and not sufficient to support analyses. For doctoral students, there were just 414 sample members in 2010 in the new cohort panel. The number of non-U.S.-earned SEH doctorate holders in the 2010 new cohort panel was only 302 eligible cases, and the number of cases determined to be eligible respondents declined to 195 cases in the 2017 sample, reflecting difficulties in locating and recontacting this subpopulation across the survey cycles.

Task 2: Sample Design with NSCG Sampling Frame for Foreign-Earned Doctorate Holders

Expanding the NSCG frame to include two years instead of one year of ACS data would probably yield enough foreign-earned doctorates to support a longitudinal panel. ACS does not directly ask respondents whether they earned a doctorate outside the United States but does collect data that allow identification with high probability. With the expanded two-year frame, the

ACS sample members in high-probability strata could be oversampled to yield sufficient numbers of foreign-earned doctorate holders in the population parallel to the SDR: individuals earning an SEH research doctorate and 75 years of age or younger.

The ACS also collects information on the year in which the doctorate was earned and, with the expanded frame and probabilistic oversampling, would support sample sizes sufficient to represent early-career foreign-earned doctorate holders.

However, the ACS frame is not likely able to yield enough current doctoral students for a longitudinal panel. There is not sufficient information collected in the ACS to identify students enrolled in research doctoral programs, making oversampling impractical for cross-sectional samples as well as for a longitudinal panel extending over four NSCG cycles (for example, 2010-2017). The highest-probability subpopulations identifiable in the ACS include all U.S. graduate and professional-school students regardless of degree sought (e.g., MA/MS, MD, JD, PsyD, PhD) and field of study. Linking the 2009 ACS with the 2010 NSCG response data, NORC estimates that current doctoral students represent <20 percent of graduate and professional school enrollments, meaning that oversampling to reach the desired sample size would require extensive screening and associated costs.

Task 3: Sample Design with NSC Data on Doctoral Students

Establishing the size and demographic composition of the U.S. doctoral student population that would be needed to expand the NSCG sampling frame presents challenges. Although the numbers and characteristics of current research doctoral students can be identified from the NCSES Survey of Graduate Students and Postdocs (GSS), the GSS is an institutional survey and does not collect individual-level records for the students. Ideally, a comprehensive list of current doctoral students could be obtained and a nationally representative sample drawn from the list. Alternatively, doctorate-granting institutions could be sampled and recruited from the GSS to provide lists of their doctoral students and the students sampled from those lists.

The project investigated the National Student Clearinghouse (NSC) list of current doctoral students as a possible resource for supplementing the NSCG sampling frame. The NSC provides student-level records of doctoral enrollments that, with the enrolling institution's permission, could be used to draw samples of doctoral students to augment the ACS frame. NSC claims to cover 97.1 percent of the 2019 postsecondary enrollment reported in the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS) comprehensive records for U.S. institutions.

To assess the adequacy of the NCS coverage of the doctoral student population, NORC compared the NCS totals to aggregate data from the NSF GSS data. GSS is an institution survey of universities with science and engineering (S&E) graduate programs. It provides counts of graduate students by program of study; institution and Carnegie class of institution; and student sex, race/ethnicity, and citizenship status. Comparing counts of doctoral candidates from GSS and NSC indicates agreement with respect to counts by most fields of study but significant divergence in some fields. The discrepancies we noted in the comparison are significant, but we believe that statistical adjustments, such as post-stratifying survey data to GSS population counts, are a viable solution to address them.

Task 4: Measures of Longitudinal Surveys

Assuming the sampling challenges can be met, the NSCG instrumentation would need to be revised in order to measure the constructs of primary interest for the proposed new longitudinal panels of doctoral students and non-U.S.-earned doctorate holders.

Longitudinal study of doctoral students. NSCG does not currently include questions about doctoral students' experiences in graduate school, progress toward the degree, and reasons for completing, or not completing, the doctoral program. Promising sources for these measures include the Association of American Universities Data Exchange (AAUDE) Doctoral Exit Survey and the gradSERU (Graduate Student Experience in the Research University) survey conducted by the SERU Consortium. Items from these studies could be integrated into the NSCG via a doctoral student module that would be repeated at each data collection cycle to follow the students longitudinally through completion of their graduate studies.

Longitudinal study of early-career doctorate recipients from U.S. and non-U.S. doctorate-granting institutions. The 2017 ECDS provides instrumentation that could be administered to both early-career doctorates from non-U.S. institutions in an expanded NSCG sample design and to ECDs sampled into the SDR. These instruments would be administered as ECD-specific modules. The 2017 ECDS sample was restricted to individuals employed at U.S. higher education institutions or at federally funded research and development centers (FFRDCs), thereby excluding the substantial contingent that was employed in private business and industry, nonprofit organizations, and government, as well as those living and working abroad. With those exclusions lifted, the instrument would need to include questions about work experiences relevant to nonacademic and non-FFRDC employment. These could be drawn from the SDR instrument, particularly as adapted to the recent longitudinal redesign.

Longitudinal study of doctorate recipients from U.S. and non-U.S. doctorate-granting institutions. A primary goal of expanding the NSCG to include a larger sample of individuals who earned doctorates from non-U.S. institutions is to make the NCSES doctorate labor-force survey samples, taken as a whole, fully representative of U.S.-resident doctorate population. The current SDR instrument generally fits for non-U.S.-earned doctorate holders. However, the key questions about employment do not capture the full longitudinal record, focusing instead on detailed snapshots tied to the survey reference date. This shortcoming applies to the SDR as well, even though it has transitioned to a longitudinal design.

The NORC project conducted a comprehensive review of NCSES and other labor-force surveys to assess options for enhancing the longitudinal data records regarding doctorate recipients. Promising longitudinal employment event history formats are used by the National Longitudinal Surveys of Youth (NLSY) and other labor force surveys; these models could be adapted to the NSCG and SDR.

ES3 Recommendations

The NORC study developed recommendations tied to each research question and corresponding task area for NCSES to consider.

Task 1: Evaluation of the NSCG Sample:

- We recommend augmenting the current NSCG sample to support longitudinal panels of foreign-earned doctorate holders and doctoral students. Increased sample will allow a greater ability to estimate fine field of degree, gender, and race/ethnicity for these groups (which generally show sample sizes of <1,000 in a sample of 130,000).
  - A 10 percent sampling rate relative to population size, as currently implemented in the SDR, would be an ideal goal for both groups. But the feasibility of such a goal must be evaluated in relation to the available frame sources as discussed in Tasks 2 and 3 below.

Task 2: Sample Design with NSCG Sampling Frame:

- The ACS frame should be used to oversample foreign-earned doctorates for a longitudinal panel.
- The ACS frame should not be used to oversample current doctoral students for a longitudinal panel.

Task 3: Sample Design with NSC Data:

- We recommend that the NSC be used to obtain the individual records to augment the ACS frame for oversampling doctoral students.
- We recommend using the GSS population counts to establish the target population of current doctoral students by institution and field of study. Drawing on the GSS population counts for guidance, post-stratification of the survey results from samples drawn from the NSC frame could be used to make statistical adjustments to improve population estimates.

Task 4: Measures of Longitudinal Surveys:

- We recommend developing event-history formats for employment data in the NSCG and SDR; the non-U.S.-earned doctorate panel could then be integrated with the SDR.
- For the proposed new doctoral student panel, we recommend developing a new module drawing mainly from the AAUDE, GradSERU, and SED-GSE projects to measure graduate student experiences and outcomes.
- Finally, NORC recommends a multistep process for developing and testing the new measures.

Chapter 1. Introduction

This introductory chapter situates the work that the NORC team performed at NCSES's direction and provides an overview of the report and its contents.

1.1 Background

This project aims to advance the National Center for Science and Engineering Statistics (NCSES) research goal of moving toward longitudinal and condensed designs for ongoing surveys. We will investigate adding three new longitudinal panels to the National Survey of College Graduates (NSCG): 1) doctorate holders from non-U.S. institutions who reside in the United States, 2) doctoral students at U.S. research universities, and 3) early career doctorate (ECD) recipients. We also propose revisions to the NSCG questionnaire to address research questions directed to these distinct longitudinal target populations.

These additions could benefit the NCSES research agenda in multiple ways:

- The revised NSCG panels could be integrated with the Survey of Doctorate Recipients (SDR) to provide complete coverage of the U.S. science, engineering, and health (SEH) workforce holding research doctoral degrees conferred either in the United States or abroad.
- Tracking samples of U.S.-earned (SDR) and non-U.S. earned (NSCG) doctorate recipients through the first decade of their post-graduation careers would allow NCSES to replace the Early Career Doctorate Survey (ECDS), thereby creating significant cost savings while giving NCSES a wider view of ECD careers than that provided by the ECDS alone.
- Analyzing a longitudinal panel of enrollees in U.S. doctoral programs would allow NCSES and other stakeholders to gain insight into graduate education experiences and outcomes.

The project also responds to the NCSES recommendation to seek efficiencies by collapsing surveys and integrating alternative data sources. First, collapsing the ECDS into the NSCG and SDR would bring a substantial efficiency to NCSES and would greatly reduce aggregate respondent burden. Second, augmenting the NSCG sample of students currently enrolled in research doctoral programs with a supplemental sample from the National Student Clearinghouse (NSC) database on individuals enrolled in postsecondary education degree programs also would allow more efficient direct sampling of doctoral program enrollees.

This research project responds to the Committee on National Statistics (CNSTAT) 2018 recommendations for NCSES to enhance the sample designs for its surveys to try to improve the ability of NCSES's data products to answer important research questions. Each of the design changes we investigate corresponds to a set of research questions of broad interest to SEH postsecondary education and to the SEH workforce. First, policymakers and researchers have expressed a need to understand U.S. workforce patterns of doctorate recipients who obtained their degrees abroad (CNSTAT, 2018). The numbers of non-U.S. doctorate recipients residing in the United States has steadily grown in recent years, and many of these individuals come to the U.S. as postdocs (Gupta, Nerad, and Cerny, 2003; Lin et al., 2009). Many leave this country after completing a postdoc or after a period of regular employment, although some remain in the United States. The circulation of PhDs has important implications for the U.S. human capital supply and related policies. The SDR follows international patterns for persons earning doctorates in the United States, but a complete picture of the careers of the U.S.-resident doctoral labor force also requires a more complete sampling and longitudinal data collection from those earning doctorates from another country.

The project also evaluated the creation of a longitudinal sample of U.S.-based doctorate candidates to examine the educational pipeline that is producing candidates for employment in the U.S. SEH workforce. Currently, researchers and policy makers know little about candidates who are in doctoral programs. More than one-quarter of doctoral students drop out before completing their degrees (Bates, 2011) and therefore are not included in the Survey of Earned Doctorates (SED) or eligible for the SDR. Researchers have recognized this phenomenon as a cause for concern, especially noting that women and underrepresented minority doctoral candidates are more likely to drop out than their male and white counterparts (Bates, 2011; Nettles and Millett, 2006). Even though the NSCG identified an estimated 558,000 students enrolled in doctoral programs on the 2017 reference date, the NSCG sample counts are relatively small and, therefore, do not adequately capture this target population. We propose a method to increase the sample size by drawing from the National Student Clearinghouse (NSC) database, following them longitudinally, and thereby setting the stage for survey modules to gather needed detail on doctoral students' graduate school experiences.

A longitudinal perspective is also crucial to understanding the career trajectory of early career research doctorates. The early career stage is increasingly recognized as a critical period for making decisions about pursuing: academic teaching and research paths; nonacademic research paths in not-for-profit organizations, business, industry, and government; and nonresearch paths. With an eye toward future employment, prospective students demand information to better inform their decisions regarding program enrollment and field concentration, and the paths they choose and their employment outcomes are of vital importance to graduate institutions (CNSTAT, 2018). To address this data need, a growing number of doctorate-granting institutions are collecting systematic information on early career pathways and making it available to students, university faculty, and administrators (Allum, Kent, and McCarthy, 2014). In addition, NCSES has developed the ECDS to provide in-depth coverage of postdocs (including those held by individuals earning doctorates outside the United States) and other PhD holders in the first 10 years after earning their doctorates. Our project seeks to expand and strengthen the NCSES survey program for early career research doctorate recipients by building a nationally representative longitudinal sample across all employment sectors, including doctorate recipients who are unemployed or out of the labor force.

1.2 Research Questions

This report answers four main research questions (RQ):

RQ1: How adequate is the current NSCG sample design for addressing the data gaps and supporting longitudinal analyses?

RQ2: Could the current American Community Survey (ACS)–based NSCG sampling frame be used to increase coverage of non-US doctorates and doctoral candidates?

RQ3: Is the NSC a viable option for augmenting the NSCG frame for doctoral candidates?

RQ4: How could the NSCG survey instrument be redesigned to cover non-U.S. doctorates and doctoral candidates and support longitudinal analyses?

1.2 Approaches Taken and Organization of the Report

The NORC team linked each of the above research questions to a project task and to a chapter of this report.

Task 1, following RQ1, assessed the current NSCG sample. Using the NSCG public use file (PUF) and restricted-use survey-specific files, we evaluated the extent to which the current NSCG sample would provide a sufficient number of cases to build a longitudinal panel for foreign-earned doctorate recipients or for doctoral candidates. Chapter 2 of this report will describe our evaluation of the current NSCG sample (Task 1).

Task 2, addressing RQ2, investigated and simulated the best sampling strategy to target intended longitudinal sample cases, using the existing ACS-based NSCG sampling frame. We consulted with the NSCG team to directly access NSCG data via a remote link to the Chicago-based Federal Statistical Research Data Center. We performed sample allocation analyses for two longitudinal panels to be included in the main NSCG sample: the panel of non-U.S.-earned doctorates and the panel of U.S.-based doctoral candidates. Chapter 3 will describe this analysis and report the findings on our consideration of the utility of the current NSCG sample frame—the Census Bureau's ACS—to fulfill requirements to provide an adequate sample for the two main longitudinal panels of focus here.

Task 3, covering RQ3, investigated an alternative approach using the data from the NSC database as a sampling frame for a longitudinal panel with doctoral candidates as a part of the NSCG longitudinal redesign. In an exploratory analysis, we assessed issues of coverage and data quality, comparing the NSC data to other benchmarks, such as the U.S. Department of Education's Integrated Postsecondary Data System (IPEDS) and the National Science Foundation's (NSF) Survey of Graduate Students and Post doctorates (GSS). Chapter 4 discusses the NSC's strengths and weaknesses as an alternative frame source for a sample of doctoral candidates that could be a treated as either a standalone longitudinal sample or as a part of NSCG.

We explored RQ4 under this project's Task 4. The NORC team investigated how to develop questionnaire modules aimed at supporting longitudinal panel research for both the NSCG and the SDR. NORC reviewed the SDR, NSCG, and ECDS questionnaires to identify survey items relevant to three longitudinal panels. We devised a plan for next steps on how best to develop and test the new questionnaire modules as a future phase for this effort. Chapter 5 provides further details to address RQ4.

The last chapter (Chapter 6) summarizes key findings from this research project and issues recommendations for NCSES to consider as it looks to improve its human resource survey programs.

Chapter 2. Task 1: Evaluation of NSCG Sample

2.1 Introduction

For Task 1, our aim was to assess the current state of the NSCG sample and the extent of its coverage of the population of all SEH doctorates living or working in the United States, along with those pursuing doctoral degrees at U.S. institutions. Although the SDR provides full coverage of U.S.-earned SEH doctorates, the NSCG is the only current NCSES survey that captures two other populations: doctoral candidates and non-U.S.-earned doctorate scientists and engineers working in the United States. Our work under this task explored the extent to which the current NSCG sample design would support the creation of longitudinal panels of non-U.S.-earned PhDs (overall and early career) and PhD candidates.

Since the 2010 survey round, the NSCG has adopted the U.S. Census Bureau's American Community Survey (ACS) as its main sampling frame. The sampling strategy relied on a rotating panel design in which a new cohort panel is rotated out of the survey after four survey cycles. For each survey cycle, this strategy produced approximate cross-sectional sample sizes as follows: 100,000 (2010); 143,000 (2013); 135,000 (2015); 124,000 (2017) and 130,000 (2019). Figure 2.1, reproduced from the 2017 NSCG sample design overview,[2] illustrates how cohort samples rotated out of the 2010-2019 survey rounds and shows the approximate sample sizes and frame sources for each NSCG survey. The focus of the analysis in this chapter is the 2010 new cohort sample (designated as "New Sample" in the 2010 NSCG), selected from the 2009 ACS. Our analysis tracks this specific sample of interest from the participants' selection into the NSCG in 2010 through their last participation in the survey panel in 2017.

---

[2] Memorandum, Jennifer G. Tancreto, U.S. Census Bureau, to Lynn Milan, National Science Foundation, "2017 National Survey of College Graduates Sample Design Overview (Document # NSCG17-SAMP-3)," Washington D.C., May 18, 2018.

Figure 2.1. NSCG rotating panel design

NSCG Year	Cohort/Panel: Returning Sample	Cohort/Panel: New Sample
2010	2008 NSCG and NSRCG (n ≈ 35,000)	2009 ACS (n ≈ 65,000)
2013	2010 NSRCG (n ≈ 13,000) 2009 ACS (n ≈ 47,000)	2011 ACS (n ≈ 83,000)*
2015	2010 NSRCG (n ≈ 13,000) 2009 ACS (n ≈ 23,000) 2011 ACS (n ≈ 57,000)	2013 ACS (n ≈ 42,000)*
2017	2009 ACS (n ≈ 23,000) 2011 ACS (n ≈ 28,000) 2013 ACS (n ≈ 25,000)	2015 ACS (n ≈ 48,000)*
2019	2011 ACS (n ≈ 28,000) 2013 ACS (n ≈ 25,000) 2015 ACS (n ≈ 29,000)	2017 ACS (n ≈ 48,000)*

Note: * indicates the inclusion of a young graduate oversample

The data sources required for this analysis—downloaded from the NCSES website as well as acquired through the NCSES and the NSCG team at the Census Bureau—include the 2010-2017 NSCG PUFs and the 2010-2017 survey-specific files (2010 new cohort and subsequent old cohort files). Because the survey-specific files were available only through the NORC-maintained Data Enclave, the NORC team worked inside that environment for that part of the analysis.

From the 2010-2017 NSCG PUF, we obtained population estimates (weighted counts) and the number of respondents (unweighted counts) for those who received a doctorate in an SEH field from an institution outside the United States, and for those who were enrolled in a doctoral degree program. We also produced (weighted and unweighted) counts of non-U.S.-earned doctorates by year of degree receipt, focusing specifically on persons who received their degrees between 2000 and 2010. These estimates provided population sizes and trends of foreign-earned PhDs for the 10 years prior to 2010, the baseline survey year for this longitudinal design research.

Next we performed an analysis of the 2010 NSCG survey-specific file (new cohort) and the response disposition codes in the successive NSCG old cohort files through the 2017 cycle. The results of this analysis provided a rough estimate of the rate of attrition (i.e., nonresponse and ineligibility) from one survey round to the next for the following cohorts of interest:

- Non-U.S.-earned SEH doctorates, compared to U.S-earned SEH doctorates

- Early career SEH doctorates (in the first 10 years following doctorate receipt): U.S.-earned compared to non-U.S.-earned
- Enrollees in U.S-based SEH doctoral programs

This assessment provided a cumulative attrition rate for the groups of interest that allowed us to understand if and by how much attrition differed between U.S.-trained and non-U.S.-trained SEH doctorates.

A second, but equally important, aim of this analysis was to estimate the number (weighted and unweighted) of individuals holding SEH doctorates earned outside the United States who responded to the survey in 2010 but who subsequently left this country. We intended to identify those who were discovered to be residing outside the United States in the 2013, 2015, and 2017 survey rounds and calculated the duration of their stays in the United States (based on their year of doctorate receipt or year of arriving in the U.S.) to capture the 2010 sample attrition that resulted from emigration.

2.2 Data Sources Used

As described earlier, this analysis depended on the following data sources:

- 2010-2017 PUFs. To complete the analysis using these datasets, we required access to the NCSES research Data Enclave to utilize the 2017 PUF that was keyed to the reference identification number, REFID.
- 2010 NSCG survey-specific file (new cohort) and the 2013-2017 survey-specific files (old cohort)

We created weighted estimates using the cross-sectional survey weights (WTSURVY) as well as the 2010 new cohort panel weight (FNLWGH10). WTSURVY is the final analysis weight created for the entire NSCG sample; to represent the NSCG cross-sectional population in each survey cycle, it included adjustments for eligibility and nonresponse. FNLWGH10 is the weight that accounts for the 2010 NSCG population with the >65,000 sample members selected for the 2010 NSCG panel from the 2009 ACS one-year data.

For our longitudinal analyses, we isolated the 2010 new cohort panel that was preselected to be retained through the 2017 survey round. To accomplish this, we identified all active sample cases in the 2015 survey-specific file, as these cases were retained in 2015 and 2017 after the NSCG team reduced the 2010 new cohort sample by half. As expected, this set of cases represented approximately one-half of the populations of interest from the 2010 new cohort panel.[3] To achieve relative consistency in population estimates from 2010 to 2017, we calculated longitudinal weights for the half of the 2010 new cohort sample retained for this analysis. For 2010 and 2013 (LONGWGH10 and LONGWGH13), we calculated longitudinal weights by multiplying each round's final panel weight (FNLWGH10 and FNLWGH13) by two. For 2015 and 2017, cohort-specific final weights (FNLWGH15 and FNLWGH17) were taken as longitudinal weights for 2015 (LONGWGH15) and 2017 (LONGWGH17). In the longitudinal analysis discussed below (i.e., RQs 2 and 3), we used the calculated longitudinal weights (LONGWGH10, LONGWGH13, LONGWGH15, and LONGWGH17) to produce estimates for 2010, 2013, 2015, and 2017, respectively.

2.3 Analysis

The basic mode of analysis the team used to assess each research question was that of comparing weighted (i.e., population estimates) and unweighted (i.e., sample size) frequencies, either cross-sectionally or longitudinally linked by REFID, as required. The analysis focused on doctorate or professional degree holders (both U.S.-earned and non-U.S.-earned), as well as the 2010 cohort of doctoral candidates, as denoted by persons who reported being enrolled in a U.S. doctoral program in the 2010 survey round. The constructs used, along with NSCG variable logic to classify cases, are reported in Table 2.3 at the end of this chapter.

Research question 1: What are the cross-sectional estimates of the target and comparison populations in science, engineering, and health fields from 2010 to 2017?

As illustrated in Table 2.1, the population estimates of each U.S.-based cohort of interest were relatively stable over the 2010 to 2017 period, whereas sample sizes underlying those estimates varied more widely after 2010. Persons reporting enrollment in SEH doctoral programs increased from approximately 290,000 in 2010 to 364,000 in 2017, a 25 percent increase. In contrast, those reporting having earned a doctorate from a non-U.S. institution hovered at approximately 180,000 from 2013 to 2017, after the NSCG estimated approximately 199,000 non-U.S. trained doctorates residing in the United States in 2010. The number of professional degree holders from non-U.S. institutions was approximately 50 percent higher than those reporting non-U.S. doctorate degrees. Almost all SEH professional degree holders were in health fields.

[3] To identify the 2010-2017 longitudinal cohort from among the 7,556 cases in our populations of interest in the 2010 NSCG new sample, we flagged 3,615 cases from the 2010 new cohort that were retained in the 2015 NSCG. This number represented approximately 48 percent of the total of 7,556 cases. If 124 cases that were dropped from the sample in 2013-2017 are further excluded from the total, the 3,615 cases in the 2010-2017 longitudinal cohort represent just under 49 percent of the population of interest—a result that is roughly in line with our expectations.

Sample sizes varied in opposite directions, with the numbers of those seeking or having earned degrees in U.S. institutions generally showing an upward trend after 2010, while sample sizes of those reporting non-U.S. degrees declined.

Table 2.1. Cross-sectional population estimates and sample sizes of selected SEH groups

SEH populations and sample sizes (in parentheses)	2010	2013	2015	2017	2019
Enrolled in PhD program	290,000 (1,445)	306,000 (2,545)	359,000 (1,764)	364,000 (1,319)	367,000 (1,405)
Earned PhD in U.S.	1,049,000 (2,712)	1,092,000 (5,445)	1,174,000 (5,113)	1,345,000 (4,929)	1,548,000 (5,843)
Earned PhD in U.S. between 2000 and 2010	331,000 (967)	313,000 (1,470)	290,000 (1,179)	316,000 (1,094)	333,000 (1,269)
Earned PhD outside U.S.	199,000 (1,211)	180,000 (1,014)	181,000 (801)	186,000 (801)	253,000 (1,608)
Earned PhD outside U.S. between 2000 and 2010	61,000 (225)	55,000 (311)	48,000 (254)	53,000 (243)	55,000 (446)
Enrolled in a professional degree program	180,000 (344)	202,000 (849)	190,000 (473)	212,000 (362)	232,000 (399)
Earned professional degree in U.S.	1,168,000 (1,683)	1,275,000 (1,918)	1,368,000 (1,650)	1,444,000 (1,433)	1,462,000 (1,404)
Earned professional degree in U.S. between 2000 and 2010	337,000 (439)	386,000 (491)	413,000 (358)	360,000 (294)	316,000 (249)
Earned professional degree outside U.S.	243,000 (773)	280,000 (440)	245,000 (330)	252,000 (315)	229,000 (325)
Earned professional degree outside U.S. between 2000 and 2010	54,000 (60)	64,000 (95)	57,000 (67)	77,000 (70)	51,000 (68)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National College Graduate Study, 2010-2019 (public use files)

Table 2.1 focuses on the topline estimates and sample sizes for the domains of interest. The tables in Appendix A reproduce all estimates and sample sizes for these domains, by field of degree (SEH or non-SEH), gender, and race/ethnicity from the full public-use datasets for each round. Enrollment in U.S. doctoral programs shows a rough parity between men and women in the 2010-2017 period. Over the same period, those who earned doctorates outside the U.S. tended to be more heavily weighted toward men. With respect to race and ethnicity, non-Hispanic whites and non-Hispanic Asians were the two largest racial/ethnic groups in both the U.S.-educated and non-U.S.-educated doctorate populations. But all other racial/ethnic groups were heavily underrepresented in the non-U.S. groups, with their estimates based on only a few dozen respondent cases. Finally, only the U.S.-based doctorate and professional populations have sufficient sample sizes for field of degree estimates, at least at the major field (seven categories) if not at the minor (32 categories) levels. These points are more evident from a review of data in Appendix B that reproduces the Appendix A tables, except for limiting the case set to only cases that were identified as the 2010-2017 panel, sampled from the 2009 ACS, as denoted by the NSCG's COHORT variable.

Research question 2: What is the response behavior for the 2010 target populations in science, engineering, and health fields from 2013 to 2017?

As noted in Figure 2.1 above, the overall 2010 new cohort sample included approximately 65,000 cases selected from the ACS. When sampling the 2010 new cohort, the NSCG sampling plan assigned approximately one-half of these cases to be retained in the panel through 2017. More than 47,000 of the initial sample of 65,000 were respondents in the 2010 survey and were included in the 2013 old cohort sample. A smaller group of 2010 temporarily ineligible cases (n = 670) was assigned to the 2013 new cohort sample so that its members could receive the NSCG new cohort questionnaire. In the two subsequent survey rounds, the NSCG team implemented the planned reduction of the original 2010 new cohort sample (selected from the 2009 ACS) from approximately 47,000 cases in 2013 to approximately 23,000 in 2015 and 2017. This sample maintenance procedure accommodated the addition of other post-2010 cohorts, while limiting the full sample to a target of 130,000 cases. We expect to see similar trends in sample sizes for our doctorate-focused populations of interest.

Our analysis focuses on discrete subsets of the 2010 new cohort sample, selected from those enrolled in a doctoral or professional degree program in 2010 (n = 1,366); doctorate degree holders in 2010 (n = 3,875); and professional degree holders in 2010 (n = 2,345). Because we are most interested in the cohorts of non-U.S.-earned SEH doctorates and candidates for SEH doctorates, most of the analysis in this chapter focuses on these two groups. However, Table 2.2 and the tables in Appendices A and B show data for all of these groups.

This 2010-2017 longitudinal cohort,[4] a subset of the approximately 23,000 sample cases retained in 2015 and 2017, represents a total of 3,615 cases, weighted to a population of approximately 5.2 million with the 2010 calculated longitudinal weight, LONGWGH10. This group includes doctorate holders, professional degree holders, and doctorate (and professional degree) candidates in 2010. Although the focus of our analysis is on specific populations (e.g., non-U.S. trained doctorates and doctoral candidates in 2010), the total of 3,615 cases includes comparison groups such as U.S.-trained doctorates and professional degree holders. Table 2.2 presents the response status of these cases—all of whom were respondents in 2010—over the following three survey rounds until that set of cases rotated out of the NSCG sample. The table shows the longitudinal response profile for groups comprising the 3,615 cases in the 2010-2017

[4] This includes all 2010 new cohort cases who fall into one or more of the following categories: enrolled in a doctoral or professional degree program in the U.S. in 2010, doctoral or professional degree holder earned either in the U.S. or outside the U.S. Both SEH and non-SEH degree holders are included in this cohort, but the analysis concentrates on SEH doctoral candidates and degree holders.

longitudinal cohort, exhibiting response combinations from 2013 through 2017 for all groups of ≥10 cases. As noted in the table, the modal statuses are respondent and “eligibility unknown,” the default categorization for sample members who were noncontact cases[5] in a survey year. Given the inclusion of non-U.S.-earned doctorates in this population, it is noteworthy that sample attrition due to temporary ineligibility resulting from emigration does not appear to affect a large number of cases. Nevertheless, approximately 20 percent of these sample members were not successfully contacted (i.e., were classified as “eligibility unknown”) for at least two rounds of the survey. The NSCG dropped approximately two percent of these cases from the sample in 2017 (those with a 2017 status of “out of sample”), most likely because of their being found to be permanently ineligible (e.g., older than age 75) for the survey even after responding to previous surveys.

Table 2.2. Response status of 2010-2017 longitudinal cohort, populations of interest

2013 status	2015 status	2017 status	Sample count	Weighted frequency	Percent, sample	Percent, population
Respondent	Respondent	Respondent	2,131	2,966,300	58.9	56.6
Eligibility unknown	Eligibility unknown	Eligibility unknown	299	497,500	8.3	9.5
Respondent	Eligibility unknown	Eligibility unknown	279	459,200	7.7	8.8
Respondent	Respondent	Eligibility unknown	244	371,000	6.7	7.1
Respondent	Eligibility unknown	Respondent	225	287,900	6.2	5.5
Eligibility unknown	Respondent	Respondent	69	110,200	1.9	2.1
Eligibility unknown	Eligibility unknown	Respondent	68	128,400	1.9	2.5
Respondent	Respondent	Out of sample	51	77,100	1.4	1.5
Eligibility unknown	Respondent	Eligibility unknown	29	60,600	0.8	1.2
Respondent	Ineligible, other	Out of sample	26	42,300	0.7	0.8
Respondent	Nonrespondent	Respondent	13	47,100	0.4	0.9
Respondent	Eligibility unknown	Ineligible, emigrant	13	7,500	0.4	0.1
Respondent	Respondent	Ineligible, emigrant	12	9,400	0.3	0.2
Respondent	Eligibility unknown	Ineligible, other	10	13,900	0.3	0.3
Ineligible, emigrant	Eligibility unknown	Eligibility unknown	10	7,600	0.3	0.1
	All other sample cases		136	155,700	3.8	3.0
Total			**3,615**	**5,241,600**	**100.0**	**100.0**

SOURCE: National Science Foundation, National Center for Education Statistics, National Survey of College Graduates, 2010-2017 (survey-specific files).

---

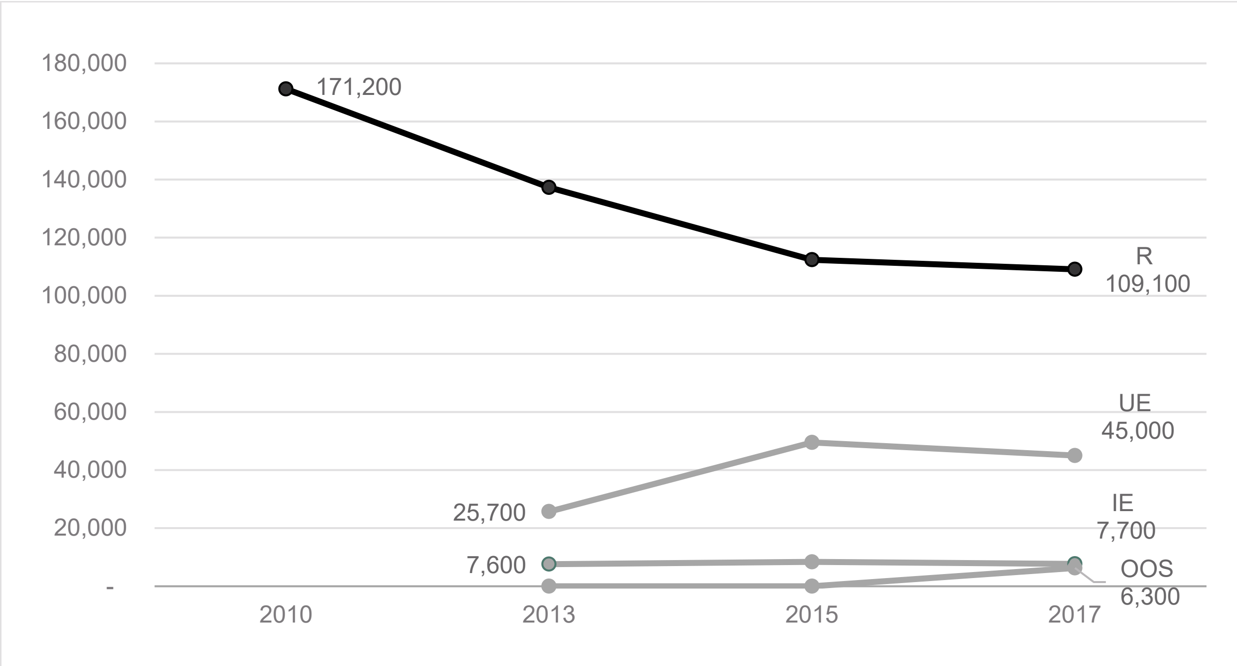
[5] “Noncontact” means that: 1) sample members failed to respond to numerous attempts to contact them, or 2) survey managers were unable to locate or to contact them.

NOTE: “Populations of interest” are all 2010 new cohort cases who fall into one or more of the following categories: enrolled in a doctoral or professional degree program in the U.S. in 2010, doctoral or professional degree holder earned either in the U.S. or outside the U.S. Both SEH and non-SEH degree holders are included in this cohort. All cases were respondents in 2010; cases weighted by calculated 2010 longitudinal weight (LONGWGH10).

In the four-round panel rotation design, the presence in the sample of each new cohort from a previous round declines over time. This reflects both “natural” survey behavior of nonresponse and ineligibility and also the design decision of shrinking each cohort to accommodate newer cohorts in subsequent survey rounds. Figures 2.2, 2.3, and 2.4 illustrate these trends for two of our main cohorts of interest: those who received SEH doctorates from institutions located outside the United States and those who reported being doctoral candidates in 2010.

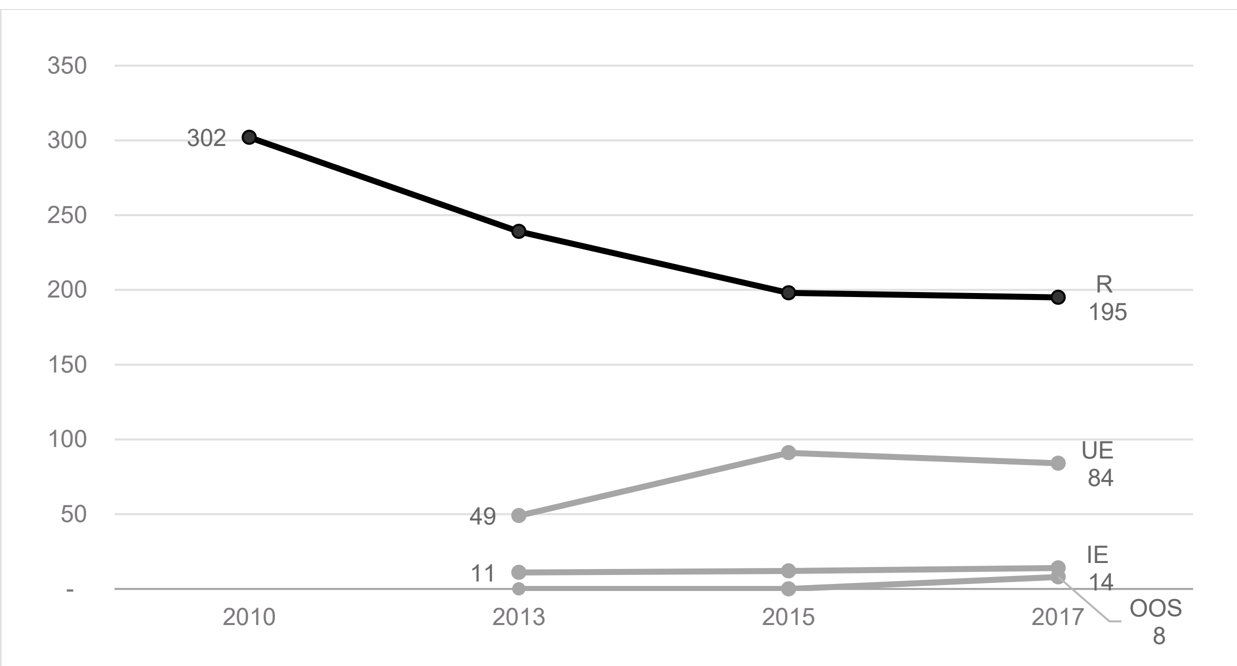
Figures 2.2 and 2.3 follow 2010 respondents who reported earning SEH doctorates outside the U.S. (n = 302) through the 2017 survey round. The figures reported for 2017 represent the numbers counted in each of the main response rate classifications: R (respondents), IE (ineligible), UE (nonrespondents with unknown eligibility), NR (eligible nonresponse), OOS (out of sample, or deselected). The NSCG tracks three forms of ineligibility: ineligibility due to degree (i.e., the sample member does not hold a bachelor’s or higher degree), other forms of permanent ineligibility (e.g., deceased), and temporary ineligibility due to location outside the United States on the reference date or emigration. In both figures, the IE estimates in 2013 and 2017 represent the emigrant-ineligibles only, as other ineligible groups were not large enough for their numbers to be meaningful. In 2015, the ineligibles were approximately evenly split between those who were ineligible due to emigration and those ineligible due to other reasons. Eligible nonrespondents and other classifications (e.g., OOS) have been omitted from both figures because their numbers are negligible.

Figure 2.2. Response behavior, 2010-2017, non-U.S.-earned doctorate holders in 2010-2017 longitudinal cohort (population estimates)



SOURCE: National Science Foundation, National Center for Education Statistics, National Survey of College Graduates, 2010-2017 (survey-specific files). Weight: LONGWGH10.

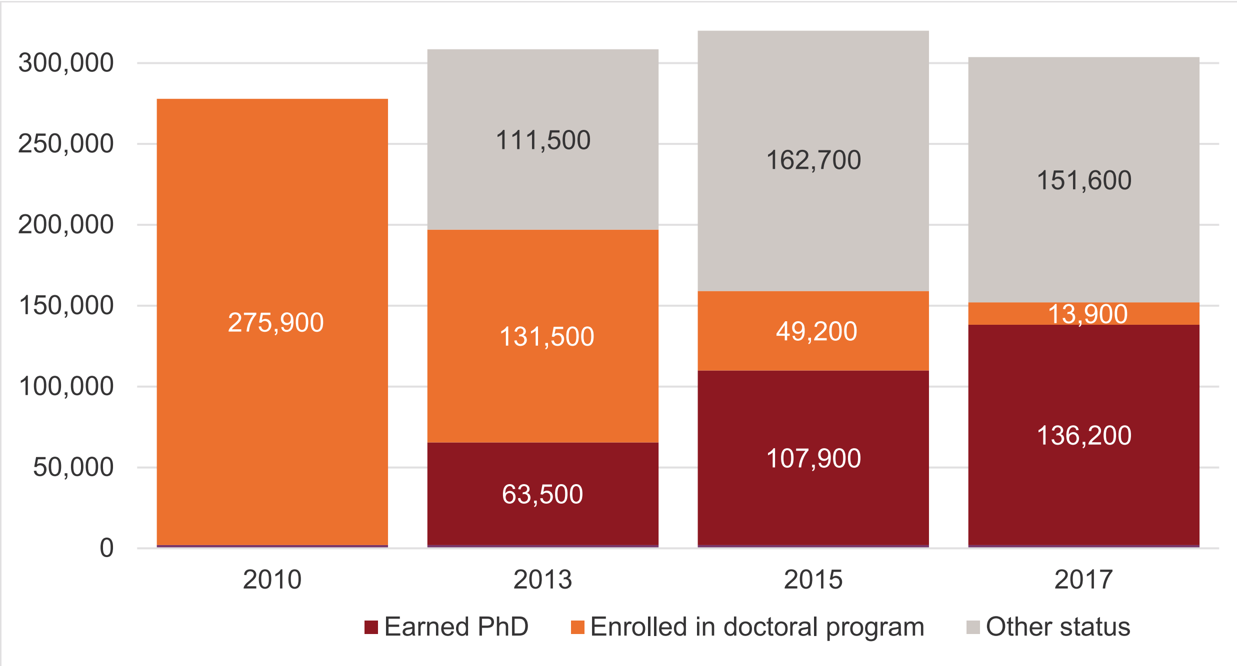
Figure 2.3. Response behavior, 2010-2017, non-U.S.-earned doctorate holders in 2010-2017 longitudinal cohort (sample sizes)



SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National College Graduate Study, 2010-2017 (survey-specific files)

Figure 2.4 shows: 1) completion of PhD degrees, 2) still in PhD programs, 3) out of PhD programs without completion over the 2013-2017 period among the 2010 doctoral candidates who were members of the 2010-2017 longitudinal panel. In the figure, the height of the entire bar represents the weighted estimate of respondents in each survey round who had reported being SEH doctoral candidates in 2010. The height of the bar varies slightly in each round after 2010 as weighted counts were calculated, with the final panel weight for each survey round accounting for respondents and ineligibles of those who were candidates in 2010 in each survey round, i.e., LONGWGH10 for 2010 NSCG, LONGWGH13 for NSCG13, FNLWGH15 for NSCG15, and FNLWGH17 for NSCG17. This longitudinal segment is a small domain of the entire NSCG cross-sectional population for which a weight was created for representation, which might also result in variation of small domain estimation such as PhD candidates in 2010. Nevertheless, the 2010-2017 longitudinal panel's weighted estimate of respondents who were doctoral candidates in 2010 was tightly varied between 301,700 (2017) and 319,800 (2015). The red segment of each bar represents the number of those 2010 doctoral candidates who reported holding an SEH doctorate in that survey round. As is evident, a little fewer than one-quarter (i.e., 63,500) of the 2010 doctoral candidates who responded in 2013 reported earning a doctorate. That number climbed to 107,900 in 2015 and to 136,200 in 2017. The 2017 estimate indicates that approximately one-half of those who were enrolled in an SEH doctorate program in 2010 earned their doctorates by 2017. These estimates rest on a sample size of 414 (the total number of NSCG 2010-2017 longitudinal panel respondents reporting enrollment in an SEH doctoral program in 2010). The number of sample members in this group who reported earning SEH doctorates increased steadily from 94 in 2013 to 145 (2015) to 177 (2017). Nevertheless, about one-half of the doctoral candidates from 2010 held a nondoctorate status (neither enrolled in a doctoral program nor having earned a doctorate) by 2015.

Figure 2.4. SEH doctorates and continuing doctoral candidates in 2013-2017 who were 2010 SEH doctoral candidates in 2010-2017 NSCG longitudinal cohort



SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National College Graduate Study, 2010-2017 (survey-specific files)

Research question 3: How does the response behavior of the non-U.S.-earned doctorate respondents in the 2010 new cohort compare to that of the U.S.-earned doctorate respondents in follow-up surveys?

Because approximately one-half of the cases included in the 2010 new cohort were deselected in 2015, we performed sample reduction by design, with the 2013 to 2017 new cohort samples added to retain balance. With the exception of the 2013 round, when ineligibility due to emigration accounted for fewer than five percent of the non-interviews, known eligible nonresponse had negligible impacts on sample attrition. Meanwhile, approximately one in six 2010 new cohort cases are never located and contacted and therefore are classified as "eligibility unknown."

We compared the response classifications of SEH doctorate holders with degrees from non-U.S. institutions to those of SEH doctorate holders with degrees from U.S. institutions. Because early career doctorates (ECD)—defined here as those who received their degree between 2000 and 2010—constitute a subset of the total number of doctorates in the full 2010 new cohort, we wanted to discern whether there was an appreciable difference in response behavior and sample attrition for more recent graduates than for the sample of doctorate holders as a whole.

Figure 2.5 illustrates these comparisons, showing graphically the percentage distribution of three major response classifications from 2013 to 2017: respondents, those classified as temporarily ineligible emigrants, and those with unknown eligibility. The figure omits other response classifications because of their relatively small numbers. The percentages reported in the figure are those of the 2010-2017 longitudinal panel, where the 2010 and 2013 longitudinal weights are used to estimate the 2013 and 2015 response patterns, respectively, and where the 2015 final weight is used to estimate the 2017 response patterns.[6] The total population sizes of the four groups displayed in Figure 2.5 vary by year, according to weight used and sample attrition. Below are the minimum and maximum population sizes (with sample sizes noted in parentheses) estimated over the survey rounds. The totals illustrate relative stability in the underlying populations:

- U.S.-earned SEH doctorates (U.S. PhD, *n* = 1,164): 984,800 – 1,032,100
- U.S.-earned SEH early career doctorates (U.S. ECD, *n* = 401): 306,600 – 339,300
- Non-U.S.-earned SEH doctorates (non-U.S. PhD, *n* = 302): 153,800 – 171,200
- Non-U.S.-earned SEH early career doctorates (non-U.S. ECD, *n* = 107): 48,300 – 56,700

Figure 2.5 illustrates two observations about the location of doctorate institution and the recency of doctorate receipt. First, U.S.-earned doctorate holders in the 2010-2017 longitudinal panel were more likely to be located and to respond to follow-up surveys when compared to non-U.S. earned doctorate holders. Second, ECD cohorts appear to be less likely to complete the follow-up surveys and more likely to be classified with unknown eligibility than were the cohorts of doctorate holders as a whole.[7] Nevertheless, the relatively high percentage of “eligibility unknown” cases suggests difficulty in locating these recently graduated populations, who may be more mobile in their early careers than they are later in their working lives.[8]

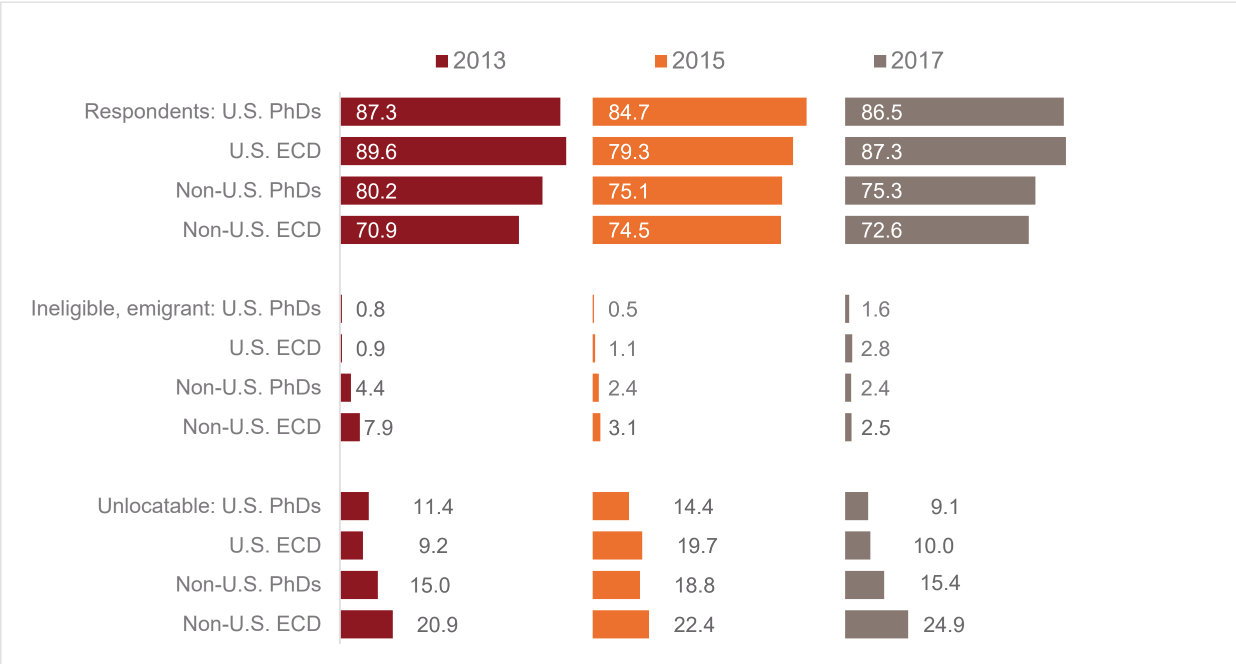
---

[6] Because each round’s longitudinal weight is used as the base weight for the following round’s survey, LONGWGH10 is used to weight the 2013 estimates, LONGWGH13 is used to weight the 2015 estimates, and FNLWGH15 is used to weight the 2017 estimates.

[7] These comparative patterns between U.S.-earned and non-U.S.-earned doctorate holders are suggestive only, as small sample sizes and incomplete data for calculation of design-based variances precluded tests of statistical significance.

[8] This observation points to a difference with the response behavior of the earliest cohorts in the Survey of Doctorate Recipients (SDR). In the SDR, the most recent graduates are located and contacted more successfully than are midcareer doctorates. See Table 7.14.f in Grigorian, K., et al., *2017 Survey of Doctorate Recipients Methodology Report* (Alexandria, Va., 2019): 72.

Figure 2.5. Response behavior of U.S.-earned doctorate holders and non-U.S.-earned doctorate holders (percent by population), including early career doctorates (ECD), 2010 NSCG longitudinal cohort, 2013-2017



SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National College Graduate Study, 2013-2017 (survey-specific files).

NOTE: For brevity of presentation, the table omits other statuses of lower frequency, such as non-respondent, ineligible other, and out of sample.

2.4 Discussion and Next Steps

This brief descriptive review of the current NSCG sample lays the groundwork for further research and statistical work toward creating longitudinal panels of two main groups: 1) SEH doctorate recipients from non-U.S. institutions (a subset of whom would be early career SEH doctorate recipients from non-U.S. institutions), and 2) SEH doctoral candidates at U.S. institutions. We began by first assessing whether the NSCG public use file (PUF) provides sufficient insight into these populations and then focused on the 2010 NSCG new cohort as it rotated through the subsequent three survey rounds (2013-2017) as part of the current rotating panel design.

The analysis encountered some limitations stemming from the available data. First, the PUFs did not allow for easy linkage across all the survey rounds, as the case identification number (REFID) was not available in the 2017 NSCG PUF. This required work in the NCSES research folder in the NSF Data Enclave, which had unforeseen impacts on the timeliness of the analysis. Second, sampling and weighting information needed to identify and properly assess the 2010-2017 longitudinal panel was not available in the 2010-2017 NSCG survey-specific files analyzed in the Data Enclave. As a result, we had to identify this cohort deductively, assuming that all 2010 new cohort cases retained in the 2015 NSCG survey-specific file constituted this cohort. Moreover, without an explicitly identified weight for this cohort, we had to derive longitudinal weights for 2010 and 2013 NSCGs (LONGWGH10 and LONGWGH13) from existing panel weights (FNLWGH10 and FNLWGH13). Although these two work-arounds provided reasonable tools for this analysis, having the actual case set and weights that NSCG survey contractor used would have improved the precision of our findings.

The main findings of the research can be summarized as follows:

- Public use NSCG data (exhibited in Appendix A tables) show a rough parity in enrollment in U.S. doctoral programs between men and women in the 2010-2019 period. Over the same period, a higher proportion of those who earned doctorates outside the United States were men rather than women. With respect to race and ethnicity, non-Hispanic whites and non-Hispanic Asians were the two largest racial/ethnic groups in both the U.S.-educated and non-U.S.-educated doctorate populations. Among the non-U.S. groups, all other racial-ethnic population estimates were based on only a few dozen respondent cases.
- The set of non-U.S.-earned SEH doctorate holders in the 2010 new cohort panel declined from an initial sample of 302 cases (population estimate = 171,200) to a 2017 sample of 195 (pop. est. = 115,200) respondents. We observed a similar pattern among non-U.S.-earned ECDs.
- Almost one of four of those who reported being enrolled in an SEH doctorate program in 2010 completed their degree by 2013. More than 400 sample members in the 2010-2017 longitudinal panel (n = 414) in the 2010 new cohort reported enrollment in an SEH doctoral program. The number of these sample members reporting earning a doctorate in subsequent survey rounds increased from 94 (in 2013) to 147 (2015) and 180 (2017).
- A comparison of the 2010 new cohort populations of doctorates by location of doctoral institution (U.S./non-U.S.) and recency (i.e., ECD/non-ECD), suggests that non-U.S.-earned doctorate holders and ECDs are more likely to be classified as “unknown eligible” than U.S.-earned doctorate holders, but this pattern may not represent a statistically significant difference.

With these observations in mind, we recommend further work to establish a sampling plan that augments the current samples of the doctorate-related populations discussed here. Increased sample size will allow a greater ability to estimate fine field of degree, gender, and race/ethnicity for these groups (which generally show sample sizes of <1,000 in a sample of 130,000). Fewer than 300 cases initially sampled as SEH doctoral candidates in 2010 earned a PhD in the subsequent three survey rounds. Further research will be required to determine sample sizes to afford analysis of these populations within acceptable levels of precision. For the population of doctoral candidates, we will also explore an alternate data source to select the panel sample.

Creating a sustainable longitudinal panel of these populations would need to draw a larger sample, perhaps ideally based on the approximately 10 percent sampling rate that the Survey of

Doctorate Recipients uses. Evaluating the feasibility of a longitudinal sample with such a high sampling rate depends on greater understanding of the nature of the NSCG's ACS-based frame.

To facilitate this research, Task 2 of this project, described in the next chapter, investigated sample modifications and augmentations using NSCG frame data based on the ACS. We analyzed available population data and sampling rates for the populations of interest, using the U.S. Census Bureau's secure data access area through the Federal Statistical Research Data Center. As noted above, the NORC team also investigated the NSC (see Chapter 4) as an alternative data source to afford a richer understanding of the educational and career decisions of SEH doctoral candidates at U.S. universities.

Table 2.3. Analysis variable descriptions and definitions

Description	NSCG variable definition
Enrolled in Ph.D program	ACDRG = 3
Earned Ph.D. in US	DGRDG = 3 and (HDDGRUS=Y)
Earned Ph.D. in US between 2000 and 2010	DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)
Earned Ph.D. outside US	DGRDG = 3 and (HDDGRUS=N)
Earned Ph.D. outside US between 2000 and 2010	DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)
Enrolled in a professional degree program	ACDRG = 4
Earned professional degree in US	DGRDG = 4 and (HDDGRUS=Y)
Earned professional degree in US between 2000 and 2010	DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)
Earned professional degree outside US	DGRDG = 4 and (HDDGRUS=N)
Earned professional degree outside US between 2000 and 2010	DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)
Degree is in SEH field	SEH = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Degree is in non-SEH field	Non-SEH = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)
Gender	Men: GENDER = 'M'; Women: GENDER = 'F'
Ethnicity and Race	Hispanic/Latino: RACETHM = '4'; Non-Hispanic Asian: RACETHM = '1'; Non-Hispanic Black/African American: RACETHM = '3'; Non-Hispanic White: RACETHM = '5'; All others: RACETHM in ('2','6','7')

Description	NSCG variable definition
Indicator of origin of sample case in the 2009 ACS sample from 2010	COHORT = ‘D’
Cross-sectional survey weight used in calculations from public use files	WTSURVY
Final panel weight (survey specific files)	FNLWGH10, FNLWGH13, FNLWGH15, FNLWGH17
Calculated longitudinal weight for longitudinal estimates of 2010-2017 panel in 2010 and 2013	LONGWGH10 for NSCG10 (i.e., FNLWGH10 × 2); LONGWGH13 for NSCG13 (i.e., FNLWGH13 × 2)

Chapter 3. Task 2: Sample Design with NSCG Sampling Frame

3.1 Introduction

Goals of Task 2

Given the findings from Task 1, the goal of Task 2 is to determine whether and how we could alter the National Survey of College Graduates (NSCG) sample design to more precisely and efficiently select members of the three populations of interest: 1) foreign-earned doctorate recipients, 2) recent foreign-earned doctorate recipients, and 3) current doctoral students. We consider four potential strategies:

1. Maintaining current NSCG stratification but increasing sampling rates of strata that yield our target populations
2. Creating two additional strata: one for likely foreign doctorates and another for doctoral students, while maintaining the other NSCG strata
3. Utilizing the 2019 NSCG oversampling scheme
4. Identifying and incorporating American Community Survey (ACS) variables not currently considered on the NSCG frame that are highly predictive of target populations

Results

Results from our analysis show that using a propensity score to create additional strata for targeting foreign-earned doctorates may be a more efficient way to sample this population, but the overall yield remains small. The NCSES should consider using an additional year of ACS data as a frame to increase the yield of the foreign-earned doctorate population. For the enrolled doctorate population, although a propensity score analysis using 2010frame variables appears likely to be ineffective, the unused grade-attending variable from the ACS may help in targeting current PhD students.

3.2 Data and Methods

3.2.1 Data

Building on the work conducted for Task 1, analysis for Task 2 takes the 2010 NSCG as a starting point. However, instead of working with the NSCG public-use files and restricted-use NSCG survey-specific files as in Task 1, NORC conducted the analysis in this task directly on ACS and NSCG microdata via the Chicago Federal Statistical Research Data Center (FSRCD).

These data include the full 2010 NSCG sampling frame file, full 2010 NSCG sample file, full 2010 NSCG survey operations file, and the main 2010 NSCG questionnaire file. In addition to the 2010 survey files, NORC also obtained access to the 2009 unswapped ACS via the Chicago FSRDC, which comprised the basis for the 2010 NSCG frame.

3.2.2 Methods

We begin with a descriptive analysis of response rate by NSCG sampling strata. For the relevant stratum, we calculate the frame size, sample count, and response rate. Among completes in each stratum, we calculate counts and percentages of: 1) foreign-earned doctorates in a science/engineering/health (SEH) field, 2) foreign-earned doctorates in an SEH field with <10 years since receiving the doctoral degree (i.e., ECD), and 3) doctoral students in SEH. This descriptive analysis will show which ACS-based sampling variables are most predictive of foreign-earned doctorates (overall and early career) and doctoral students.

We continue with a series of analyses to investigate whether additional strata or ACS variables could help to predict target populations. We first attempt to target foreign doctorates or doctoral students by constructing propensity scores for each group. For each target population, we fit a logistic model to determine whether we can identify high-propensity cases *within* currently existing survey strata. Using model results, we estimate predicted probabilities associated with each frame case and generate an expected yield by strata.

We then evaluate the results from the propensity score analysis against the yield from the 2019 NSCG, which utilized an oversample to target individuals with a foreign-earned doctorate. We assess the precision of the propensity score approach and offer recommendations for how a propensity analysis may be used to better target individuals within currently existing strata.

Finally, we also explore the possibility of incorporating unused variables from the ACS to better flag possible individuals in one of the three target populations. Specifically, we investigate the utility of the 2009 ACS "grade-attending" variable which reports current enrollment in a graduate or professional degree program in identifying current doctoral students in the 2010 NSCG.

3.2.3 Working in the FSRDC

The analyses for Task 2 were conducted exclusively within the Chicago FSRDC, a secure data facility operated by the Census Bureau in partnership with the Federal Reserve Bank of Chicago, the University of Chicago, Northwestern University, and the University of Illinois Chicago. All work for Task 2 was conducted by a pair of researchers at NORC with Special Sworn Status, a level of Census Bureau-granted clearance for outside researchers.

Although the FSRDC network makes it possible for outside consultants, like those at NORC, to perform highly detailed analyses such as those entailed in Task 2, working in the FSRDC comes with several limitations, as described below.

Set-Up Costs

The first set of limitations comes in the form of set-up costs. NORC is currently an institutional member of the Chicago FSRDC, which requires an annual fee to use the facility but means that NORC does not pay for each specific project in the FSRDC environment. If researchers at a nonaffiliated institution wish to work with the Chicago FSRDC, the annual cost per project is $20,000. In addition, any institutional affiliation only applies to a single FSRDC and not to the entire network; for example, researchers with an institutional membership at the Chicago FSRDC who wish to work at an FSRDC in another city would need to pay seat fees to the second FSRDC as well.[9]

Moreover, Special Sworn Status is time-consuming to obtain. It involves rigorous background screening that can take months, depending on the current workload of Census Bureau staff. In the case of Task 2, both NORC employees Lee Fiorio and Quentin Brummet had already obtained Special Sworn Status before starting this work—Fiorio as part of his graduate work and Brummet for a prior NORC project. In future collaborations with outside consultants, the NSF will either need to target organizations that already employ individuals with up-to-date Special Sworn Status or include in the budget the time required for researchers to obtain it.

Accessing an FSRDC Facility

A second set of limitations is the requirement that restricted research be performed in a physical FSRDC facility (in the case of Chicago, located in the Chicago Federal Reserve Building). The Chicago FSRDC is one of 31 FSRDC locations throughout the United States, and to conduct analyses like those conducted for Task 2, an organization must be located within commuting distance of one of those locations. For NORC employees, this is not an issue, as the Chicago-FSRDC is only a 15-minute walk from NORC's central office. However, accessing the Chicago FSRDC requires entry to the Federal Reserve Building, which adds a layer of administrative costs. In addition to being issued a badge for accessing the FSRDC by the Census Bureau, Chicago FSRDC researchers must also obtain a badge for entering the Federal Reserve Building and may have their access limited by the schedule or guidelines of that specific building.

Due to the extenuating circumstances of the Covid-19 pandemic, the Census Bureau temporarily granted remote access to some FSRDC researchers, including Fiorio and Brummet. This has made conducting Task 2 analyses more convenient, but how long this remote-working arrangement will persist remains unclear. Current Census Bureau policy is that remote access will be temporary, and future analyses of this kind will likely need to be conducted within a physical FSRDC space.

---

[9] Note that the specifics of seat fees at FSRDCs vary by FSRDC location and change over time.

Disclosure

Because the analysis of Task 2 involves Title 13 data, all results must adhere to Census Bureau Title 13 disclosure guidelines and be evaluated for disclosure risk by the Census Bureau before they can be released from the FSRDC. For the specific analyses presented in this chapter, the first iteration of output went before the full Census Bureau disclosure review board (DRB) to ensure that our use of the 2009 unswapped ACS was justified and appropriate. The DRB convenes each Monday and requires that output be submitted by the previous Monday to provide enough time for review.

For each subsequent release, we were able to make use of the Census Bureau's bypass procedure. Therefore, our results were reviewed by the disclosure avoidance officer (DAO) designated for our project instead of by the full DRB. This typically allows for a quicker review but makes reviews reliant on the schedules of the DAO and other Census Bureau employees—e.g., one disclosure request for this project took almost a month due to a series of overlapping vacations around the July 4 holiday.

The disclosure requirements also add time to research when not everyone on a research team has Special Sworn Status. Results cannot be discussed with those who do not have FSRDC access until they have been disclosed, so additional iterations of output may be required. For this project, Task 2 required four rounds of disclosure requests.

Finally, the Census Bureau disclosure rules for Title 13 data are quite stringent. All counts must be rounded following census rounding rules and all counts <15 must be suppressed. All statistics such as model coefficients must be rounded to four significant digits. Researchers must also pay careful attention to achieve complementary suppression—e.g., a sum total cannot be released, even if rounded, if it would result in the disclosure of a suppressed cell that contributes to the sum.

As discussed in the next section, limitations regarding rounded counts constrained our descriptive analysis and, therefore, our modeling results. The populations of interest (foreign-earned doctorates, recent foreign-earned doctorates, and current doctoral students) are all small, whereas the stratification scheme employed by the NSCG is highly detailed. If the frame size in a particular stratum is small, the sample size, count of completes, and yields of foreign-earned doctorates or current doctoral students for that cell will be even smaller. Therefore, we needed to collapse survey strata in a way that would strike a balance between the disclosure risk and the meaningfulness of the results.

3.2.4 Collapsed Strata

As described in Chapter 2 of this report, the 2010 NSCG was stratified on three variables: a three-part highest degree earned variable, a 25-part detailed occupation variable, and a nine-part demographic variable. This means that the 2010 NSCG utilized 675 strata. Given that the unweighted yield of each of the three target populations in the 2010 NSCG was small—

approximately 600 individuals with foreign-earned doctorate, 200 with foreign-earned doctorates received in the last 10 years, and 800 current doctoral students—we decided to collapse the strata from 675 to 24 for the purposes of disclosure, using the following scheme:

- Three highest degree categories became two categories: doctorate; less than doctorate.
- *Twenty-five detailed occupation categories became three categories*: biological/medical scientists and physicist/physical scientist (combined); postsecondary teachers; all others.
- *Nine demographic group categories became four categories:* low likelihood of U.S.-earned degree, Hispanic; low likelihood of U.S.-earned degree, Asian; low likelihood of U.S.-earned degree, other; high likelihood of U.S.-earned degree.

We chose to use these collapsed strata because they were most relevant to identifying members of the three target populations, and we report our descriptive findings using these collapsed strata. Similarly, for the propensity score analysis we use these collapsed strata when reporting expected yield. Coefficients from our models in the appendix, however, do not use these collapsed cells since they do not summarize counts by strata. Throughout this analysis, we note instances in which the collapsed strata may be preventing a more robust interpretation of our results.

3.3 Analysis of Foreign-Earned Doctorate Panel

3.3.1 Overview of 2010 NSCG sample design as it relates to foreign-earned doctorates

The 2010 NSCG design includes an indicator variable, DEMGROUP, to stratify frame cases with a low likelihood of having a U.S.-earned degree. This indicator variable is a composite of seven ACS frame variables: 1) race, 2) ethnicity, 3) disability status, 4) U.S. citizenship at birth (USCAB), 5) date of birth, 6) year of entry into the United States, and 7) highest degree earned, the last four of which are used to assign dichotomous likelihood (low or high) of having a U.S.-earned doctorate. This is done by estimating age at entry among individuals who were not USCAB by subtracting year of birth from year of entry. If an individual's highest earned degree is a bachelor's degree and they entered *after age 20*, then they are estimated to have a low likelihood of a U.S.-earned degree. If an individual's highest earned degree is a master's degree and they entered *after age 24*, then they are estimated to have a low likelihood of a U.S.-earned degree. If an individual's highest earned degree is a doctorate and they entered *after age 26*, then they are estimated to have a low likelihood of a U.S.-earned degree.

All remaining individuals who are not USCAB (i.e., those who entered before age 20, 24, or 26 depending on highest earned degree) are grouped with those who are USCAB and are estimated to have a high likelihood of having a U.S.-earned degree. Combined with the race, ethnicity and disability status variables from the ACS frame, this gives nine demographic group categories:

1. Hispanic (USCAB or high likelihood of U.S.-earned degree)
2. Black (USCAB or high likelihood of U.S.-earned degree)
3. Asian (USCAB or high likelihood of U.S.-earned degree)
4. American Indian/Alaska native (AIAN) and Native Hawaiian/Pacific Islander (NHPI) (USCAB or high likelihood of U.S.-earned degree)
5. Disabled (USCAB or high likelihood of U.S.-earned degree)
6. White/other (USCAB or high likelihood of U.S.-earned degree)
7. Hispanic (low likelihood of U.S.-earned degree)
8. Asian (low likelihood of US-earned degree)
9. Remaining (low likelihood of U.S.-earned degree)

Of these categories, three correspond with a low likelihood of having a U.S.-earned degree—i.e., a high likelihood of having a foreign-earned degree. When considered together with the highest earned degree strata variable, most of the foreign-earned doctorate recipient yield would be most likely to occur in cells where the highest degree earned is a doctorate and demographic group corresponds to one of the low likelihood of U.S.-earned degree categories.

3.3.2 Descriptive Findings

Table 3.1 presents frame counts together with weighted and unweighted sample sizes, completes, and target population yields. This table establishes a few key features of the ACS frame and its ability to yield foreign-earned doctorates: 1) The current unweighted yield is small. In the 2010 NSCG, approximately 600 of 47,000 respondents (~1.3 percent) had earned their doctorates at non-U.S. institution. 2) As expected, a large majority of this yield came from low likelihood of U.S.-earned degree strata, specifically those in combination with PhD as highest degree strata. At least 440 (possibly more, given cell suppression) of the approximately 600 foreign-earned doctorate respondents came from these cells. 3) The low likelihood of U.S.-earned degree/PhD cells were sampled at a relatively high rate. To provide context, for the entire 2009 NSCG, approximately 65,000 individuals were sampled from frame of 822,000 (~7.9 percent). Among cells with a low probability of having a U.S.-earned degree and PhD as highest earned degree, approximately 1,420 individuals were sampled from a frame of 5,510 (25.8 percent).

So, at a very basic level, the ACS frame appears to be somewhat constrained in its ability to deliver a sizable panel of foreign-earned doctorate recipients. Even if the individuals in the low likelihood of U.S.-earned doctorate cells were sampled at 100 percent, approximately 1,700 total foreign-earned doctorate recipients would be yielded, holding eligibility and response rates constant.

Table 3.1. 2010 NSCG frame counts, sample size, completes and yield of foreign-earned doctorate, recent foreign-earned doctorate and doctoral enrollees by collapsed survey strata

			Frame	Sample		Completes		Foreign-Earned PhD		Recent Foreign-Earned PhD		PhD Enrollees	
HIDEG	DEMGROUP	OCC_DETAIL	N	N	Weighted	N	Weighted	N	Weighted	N	Weighted	N	Weighted
Less than PhD	High likelihood U.S.-earned degree	All others	711,000	46,000	39,070,000	33,000	35,720,000	(D)	(D)	(D)	(D)	200	97,500
Less than PhD	High likelihood U.S.-earned degree	Biolog/med scientists OR phys/physical scientist	3,900	2,200	163,000	1,700	155,000	(D)	(D)	(D)	(D)	100	11,000
Less than PhD	High likelihood U.S.-earned degree	Postsecondary teachers	6,000	2,200	245,000	1,700	226,000	(D)	(D)	(D)	(D)	300	40,500
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	All others	9,500	1,000	762,000	550	538,000	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog/med scientists OR phys/physical scientist	40	40	3,100	30	2,700	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	80	60	6,200	50	5,700	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	All others	33,500	4,400	2,085,000	3,000	1,762,000	(D)	(D)	(D)	(D)	40	15,000
Less than PhD	Low likelihood U.S.-earned degree; Asian	Biolog/med scientists OR phys/physical scientist	350	300	20,000	200	18,000	20	2,500	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	650	300	42,000	250	38,000	(D)	(D)	(D)	(D)	80	12,500
Less than PhD	Low likelihood U.S.-earned degree; other	All others	23,500	3,500	1,693,000	2,200	1,344,000	20	12,500	(D)	(D)	20	12,500
Less than PhD	Low likelihood U.S.-earned degree; other	Biolog/med scientists OR phys/physical scientist	200	200	13,000	100	10,500	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	350	200	21,500	150	19,000	(D)	(D)	(D)	(D)	30	4,900
PhD	High likelihood U.S.-earned degree	All others	21,000	2,300	1,149,000	1,800	1,121,000	20	6,300	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Biolog/med scientists OR phys/physical scientist	1,600	350	101,000	300	101,000	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Postsecondary teachers	4,300	700	230,000	600	229,000	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	All others	300	70	21,000	50	18,500	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog/med scientists OR phys/physical scientist	50	30	2,800	(D)	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	60	20	3,600	20	3,400	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	All others	1,600	350	92,500	250	76,000	70	21,000	20	8,500	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	Biolog/med scientists OR phys/physical scientist	550	200	31,000	150	30,000	80	15,000	40	7,700	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	450	150	27,500	100	27,000	40	8,300	20	3,900	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	All others	1,600	300	91,500	200	80,500	100	27,500	20	6,100	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	Biolog/med scientists OR phys/physical scientist	400	200	25,000	150	24,000	100	18,000	50	8,700	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	500	100	28,000	90	26,000	50	12,500	(D)	(D)	(D)	(D)
		Total	***822,000***	***65,000***	***45,930,000***	***47,000***	***41,570,000***	***600***	***145,000***	***200***	***59,000***	***800***	***204,000***

(D) – Suppressed for disclosure avoidance

3.3.3 Targeting Foreign-Earned Doctorates Using Propensity Scores

Although 1,700 can be understood as a rough estimate of the upper bound yield of foreign-earned doctorates from the ACS frame under the current stratification scheme, there may be methods to be more targeted in sampling. As an alternative to a crude oversampling approach, we explored the extent to which likely foreign-earned doctorates could be targeted through a propensity score analysis. The goal of this analysis is to identify individuals ***within*** existing strata who could be flagged as “high-propensity” foreign-earned doctorate recipients. Our ability to do so is contingent on variation in predicted probabilities within currently existing survey strata. If all frame cases in a particular stratum are equally likely to be foreign-earned doctorates given our model, oversampling the cell at a uniform rate may be the best approach. However, if some cases in a stratum are more likely than others to be foreign-earned doctorates, high-propensity cases identified by the model potentially could be treated as a new stratum to be oversampled independently.

To perform this analysis, we fit a logistic model onto the universe of approximately 47,000 survey completes from the 2010 NSCG ACS new cohort sample. We modeled whether a respondent reported having earned a doctorate abroad as a function of their highest degree earned, detailed occupation, age during survey year (2010), U.S. citizenship status at birth, age at entry into the United States (assigned to 0 for those born in the U.S.), disability status, sex, and race/ethnicity. Appendix Table C1 shows coefficients and standard errors from this model.

After fitting the model, we estimated predicted probabilities for all 822,000 frame cases. Table 3.2 shows frame count and expected yield by collapsed survey strata and predicted probability range. For example, there are approximately 500 frame cases that have a doctorate as the highest degree earned, are employed as postsecondary teachers, have a low likelihood of a U.S.-earned degree, and are neither Asian nor Hispanic (see last non-total row in Table 3.1 and Table 3.2). Of these 500 frame cases, the predicted probabilities vary: 20 have predicted probabilities <0.1, 50 have predicted probabilities between 0.1 and 0.3, 200 have predicted probabilities between 0.3 and 0.5, and 200 have predicted probabilities >0.5. For the last two range bins, summing the predicted probabilities of all frame cases produces expected yields of 70 and 150, respectively. In other words, if we were to oversample cases with a predicted probability >0.5, we would expect a yield of 150 foreign-earned doctorates from a sample of 200 completes.

Table 3.2 underscores the potential to use a model to identify high-propensity cases within currently existing NSCG strata. With the results from a simple logistic model, it may be possible to more precisely target foreign-earned doctorates within currently existing high-yield strata. Section 3.5, which compares the hypothetical yield of foreign-earned doctorates using a propensity score approach with the oversampling that was conducted in the 2019 NSCG, discusses some of the limitations of this analysis. These expected yields, however, do not factor in response rate.

Table 3.2. Model 1 of foreign-earned doctorates – Frame cases and expected yield by predicted probability ranges by collapsed survey strata

			Frame Count by Predicted Probability				Expected Yield by Predicted Probability			
HIDEG	DEMGROUP	OCC_DETAIL	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]
Less than PhD	High likelihood U.S.-earned degree	All others	711,000	(D)	(D)	(D)	80	(D)	(D)	(D)
Less than PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	3,900	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	High likelihood U.S.-earned degree	Postsecondary teachers	6,000	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	All others	9,500	50	(D)	(D)	40	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	40	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	80	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	All others	33,500	200	20	(D)	150	30	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	350	20	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	650	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	All others	23,000	400	70	(D)	200	70	30	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	150	20	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	350	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	All others	20,500	450	20	(D)	150	70	(D)	(D)
PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	1,400	100	90	30	50	20	40	20
PhD	High likelihood U.S.-earned degree	Postsecondary teachers	4,000	300	60	(D)	60	50	20	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	All others	100	100	40	30	(D)	20	20	20
PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	(D)	(D)	20	30	(D)	(D)	(D)	20
PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	(D)	40	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	All others	450	800	250	100	20	150	100	80
PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	(D)	20	200	300	(D)	(D)	80	200
PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	(D)	250	150	40	(D)	60	60	30
PhD	Low likelihood U.S.-earned degree; other	All others	300	450	400	400	(D)	90	150	350
PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	(D)	(D)	(D)	250	(D)	(D)	(D)	300
PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	20	50	200	200	(D)	(D)	70	150
		Total	***815,000***	***3,300***	***1,600***	***1,700***	***850***	***600***	***600***	***1,200***

(D) – Suppressed for disclosure avoidance

3.4 Analysis of Recent Foreign-Earned Doctorate Panel

3.4.1 Descriptive Findings

In addition to adding a panel of foreign-earned doctorates to the NSCG, our analysis also aimed to add a panel of *recent* foreign-earned doctorates, specifically those who earned their doctorates within the last 10 years. By this definition, this population will be a subset of the full foreign-earned doctorate population. Table 3.1 shows that the same strata that yield the majority of foreign-earned doctorates also yield recent foreign-earned doctorates. Nevertheless, a disproportionate number of the unweighted yield of recent foreign-earned doctorates came from the collapsed cell of biological/medical scientists and physicist/physical scientists. This suggests that individuals who earned their doctorate abroad in the last 10 years may be concentrated in these specific occupations reflecting the demand of the U.S. job market.

3.4.2 Targeting Recent Foreign-Earned Doctorates Using Propensity Scores

Mirroring our approach for the foreign-earned doctorate population, we conducted a propensity analysis for the recent foreign-earned doctorate group. As before, we fit a logistic model on the universe of approximately 47,000 survey completes from the 2010 NSCG ACS new cohort sample. We then modeled whether a respondent reported having earned a doctorate abroad *in the last 10 years* as a function of their highest degree earned, detailed occupation, age during survey year (2010), U.S. citizenship status at birth, age at entry to U.S. (assigned to 0 for those born in the U.S.), disability status, sex, and race/ethnicity. Appendix Table C2 contains coefficients and standard errors from this mode analysis.

Like the corresponding table derived from the model of foreign-earned doctorates, Table 3.3 tabulates frame counts and expected yield by collapsed survey strata and predicted probability range for the model of recent foreign-earned doctorates. Similar to the previous analysis, there appears to be variation in predicted probability within survey strata that could be used to better target recent foreign-earned doctorates. Even within the biological/medical scientists plus physicist/physical science strata, in which recent foreign-earned doctorates are represented disproportionately, some individuals seem to be more likely than others to belong to this specific target population.

Table 3.3. Model 2 of recent foreign-earned doctorates – Frame cases and expected yield by predicted probability ranges by collapsed survey strata

			Frame Count by Predicted Probability				Expected Yield by Predicted Probability			
HIDEG	DEMGROUP	OCC_DETAIL	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]
Less than PhD	High likelihood U.S.-earned degree	All others	711,000	(D)	(D)	(D)	90	(D)	(D)	(D)
Less than PhD	High likelihood U.S.-earned degree	Biolog./med. sscientist OR phys./physical scientist	3,900	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	High likelihood U.S.-earned degree	Postsecondary teachers	6,000	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	All others	9,500	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	40	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	80	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	All others	33,500	(D)	(D)	(D)	50	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	350	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	650	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	All others	23,500	(D)	(D)	(D)	50	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical Scientist	200	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	350	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	All others	21,000	90	(D)	(D)	50	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	1,500	80	30	(D)	20	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Postsecondary teachers	4,200	70	(D)	(D)	20	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	All others	250	40	20	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	20	(D)	(D)	20	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	40	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	All others	1,300	200	80	20	20	40	30	(D)
PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	200	150	70	90	(D)	30	30	60
PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	300	100	40	40	(D)	20	20	20
PhD	Low likelihood U.S.-earned degree; other	All others	1,300	200	150	50	20	40	50	40
PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	150	90	50	100	(D)	20	20	90
PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	300	90	40	70	(D)	20	20	40
	Total		***820,000***	***1,200***	***500***	***450***	***350***	***200***	***200***	***300***

(D) – Suppressed for disclosure avoidance

3.5 Comparison with 2019 NSCG Oversample of Low-Likelihood U.S.-Earned Degree Cells

3.5.1 Results from 2019 NSCG Oversample

To recruit a larger sample of foreign-earned doctorates, the 2019 NSCG oversampled low-likelihood U.S.-earned degree strata. Table 3.4 shows the frame counts, sample size, completes, and count of yields for relevant strata in the 2019 NSCG. In many instances, strata were sampled at 100 percent. This oversample resulted in a yield of 1,200 foreign-earned doctorates, nearly twice what was yielded in the 2010 NSCG. However, this increased yield came at the cost of increasing the sample size and number of completes.

Table 3.4. 2019 NSCG yield of foreign-earned doctorates after oversampling

HIDEG	DEMGROUP	OCC_DETAIL	Frame (N)	Sample (N)	Completes (N)	Yield (N)
Doctorate	Hispanic low	Biological/medical scientist + physicist and other physical scientists	60	60	30	20
Doctorate	Hispanic low	Postsecondary teacher, science and engineering (S&E) FOD	100	80	50	20
Doctorate	Hispanic low	Mathematician + computer information scientists	(D)	(D)	(D)	(D)
Doctorate	Hispanic low	Agricultural and other life scientists + chemist, except biochemist	(D)	(D)	(D)	(D)
Doctorate	Hispanic low	Social scientists	20	20	(D)	(D)
Doctorate	Hispanic low	Engineers	(D)	(D)	(D)	(D)
Doctorate	Hispanic low	S&E-related health and nonhealth occupation	100	20	(D)	(D)
Doctorate	Hispanic low	Secondary or postsecondary teacher, non-S&E FOD	40	20	(D)	(D)
Doctorate	Hispanic low	Non-S&E high/low-interest occupation, S&E FOD	90	40	20	(D)
Doctorate	Hispanic low	Non-S&E occupation, non-S&E FOD + not working	150	(D)	(D)	(D)
Doctorate	Asian low	Biological/medical scientist + physicist and other physical scientists	600	450	300	150
Doctorate	Asian low	Postsecondary teacher, S&E FOD	700	550	350	100
Doctorate	Asian low	Mathematician + computer information scientists	300	150	100	30
Doctorate	Asian low	Agricultural and other life scientists + chemist, except biochemist	60	60	40	(D)
Doctorate	Asian low	Social scientists	20	20	(D)	(D)
Doctorate	Asian low	Engineers	300	250	150	50
Doctorate	Asian low	S&E-related health and nonhealth occupation	400	40	30	(D)
Doctorate	Asian low	Secondary or postsecondary teacher, non-S&E FOD	150	90	60	(D)
Doctorate	Asian low	Non-S&E high/low-interest occupation S&E FOD	500	250	150	50
Doctorate	Asian low	Non-S&E occupation, non-S&E FOD + not working	550	40	30	(D)
Doctorate	Remaining low	Biological/medical scientist +physicist and other physical scientists	450	400	250	200
Doctorate	Remaining low	Postsecondary teacher, S&E FOD	650	500	350	200
Doctorate	Remaining low	Mathematician + computer information scientists	150	90	70	50
Doctorate	Remaining low	Agricultural and other life scientists + chemist, except biochemist	40	40	30	20
Doctorate	Remaining low	Social scientists	50	50	30	(D)
Doctorate	Remaining low	Engineers	100	100	80	40
Doctorate	Remaining low	S&E-related health and nonhealth occupation	350	40	20	(D)
Doctorate	Remaining low	Secondary or postsecondary teacher, non-S&E FOD	250	100	70	30
Doctorate	Remaining low	Non-S&E high/low-interest occupation, S&E FOD	600	250	150	100
Doctorate	Remaining low	Non-S&E occupation, non-S&E FOD + not working	600	50	30	20
		Total	***7500***	***3900***	***2500***	***1200***

3.5.2 Comparing 2019 and 2010 Foreign-Earned Doctorate Yields

Table 3.1 shows that in low-probability U.S.-earned degree strata, the 2010 NSCG frame count was 5,500 and the sample size was 1,420. From this sample, approximately 1,000 completes yielded approximately 440 foreign-earned doctorates. Table 3.4 shows that in low-probability U.S.-earned degree strata, the 2019 NSCG frame count was 7,500 and the sample size was 3,900. From this sample, approximately 2,500 completes yielded approximately 1,200 foreign-earned doctorates. The response rates across these strata in the two surveys were comparable: approximately 70 percent in 2010 vs. 64 percent in 2019. The hit rates were also comparable: 44 percent of completes were foreign-earned doctorates in 2010 vs. 48 percent in 2019.

3.5.3 A Counterfactual: Using a Propensity-Score Model to Target Foreign-Earned Doctorates in the 2010 NSCG

In section 3.3.3, we evaluated whether foreign-earned doctorates can be further targeted using a propensity score approach. The propensity model showed that the expected yield and observed yield was similar for the 2010 set of completes. Table 3.4 shows 1,700 frame cases with a predicted probability >0.5, which, by summing their predicted probabilities, would yield approximately 1,200 foreign-earned doctorates[10], based on the model.

Assuming an equivalent response rate as observed across low-probability U.S.-earned degree strata in the 2010 and 2019 NSCG, sampling these 1,700 frame cases at or close to 100 percent would generate approximately 1,100 completes, giving an expected yield of approximately 800 foreign-earned doctorates. This number falls below the 1,200 yielded in the 2019 NSCG oversample, but it would be considerably more efficient (a hit rate of approximately 70 percent vs. 48 percent). By lowering the target threshold from 0.5 to 0.45, for example, more sample could increase the total yield, but likely at a cost to overall efficiency.

3.5.4 Limitations and Future Research

Without access to the 2019 NSCG frame, our ability to evaluate the propensity score approach is limited. Further evaluations could be conducted by applying the propensity model from 2010 to the 2019 frame to assess the likely yield from an approach that targets likely foreign-earned doctorates, using a propensity score model vs. the oversample that was conducted.

Moreover, our discussion of the propensity score approach is complicated by use of collapsed survey strata, which was necessary for disclosure reasons. For example, the collapsed survey strata with the greatest variation in predicted probability in Table 3.2 and Table 3.3 are those that correspond to the All Other detailed occupation category. This collapsed cell contains 22 of the 25 detailed occupations, and therefore it is possible that much of the variation in predicted

---

[10] Standard errors of expected number of yields can be calculated using ACS's successive difference replication weights. However, ACS replicate weighs were not available in the FSRDC for this project.

probabilities within this collapsed stratum is already captured by the detailed occupation strata in the actual NSCG survey. Therefore, although these results indicate that propensity score approaches have potential, before using them in production, it would be advisable to investigate potential improvements to this propensity score model using these more detailed cell

3.6 Recommendations for Foreign-Earned Doctorates and Recent Foreign-Earned Doctorates

Based on our analysis, it appears that a propensity score approach could help more efficiently target foreign-earned doctorates and recent foreign-earned doctorates. However, the potential yields remain small due to the relative rarity of these populations. Therefore, we recommend considering expanding the NSCG frame for the purpose of targeting foreign-earned doctorates and recent foreign-earned doctorates. The simplest way to do this would be to use two years of ACS data. Because the NSCG is collected every two years and the ACS is sampled in such a way as to avoid year-to-year overlap to allow for the pooling to generate five-year estimates, it may be possible to incorporate a second year of ACS data into the frame without having to vastly alter the current NSCG design. Doubling the frame size in combination with a propensity score approach could allow for the effective targeting of foreign-earned doctorates and recent foreign-earned doctorate panels.

3.7 Analysis of Current Doctoral Student Panel

3.7.1 Descriptive Findings

Unlike the foreign-earned doctorate population, the NSCG does not stratify with respect to whether a frame case is likely enrolled in a doctoral program. Given the strata variables used in the 2010 NSCG—highest degree earned, demographic group, detailed occupation—most of the current doctoral student population will likely come from the bachelor's or master's highest-degree-earned strata. Current doctoral students may also be concentrated in certain occupation like postsecondary teacher, since many doctoral students are employed as teaching assistants.

Table 3.1 shows that the majority of the unweighted current population yield came from the collapsed strata corresponding to persons with less than a PhD who have a high likelihood of a U.S.-earned degree. Notably, 300 of the 800 current doctoral students in the 2010 NSCG were postsecondary teachers. Unlike the foreign-earned doctorate population which is efficiently stratified with the low likelihood of U.S.-earned degree indicator, the yield rates among the collapsed strata generating the most current doctoral students are low. For example, only 300 out of 1,700 completed interviews from the postsecondary teacher strata yielded current doctoral students. Therefore, although increasing the sampling rate of this stratum would produce more current doctorate completes, it would come at considerable cost in terms of nondoctoral student interviews.

3.7.2 Targeting Current Doctoral Students Using Propensity Scores

As before, we conducted a propensity score analysis to determine whether a model using frame variables would enable us to better target our sampling within currently existing survey strata. To perform this analysis, we fit a logistic model to the universe of approximately 47,000 survey completes from the 2010 NSCG ACS new cohort sample. We modeled whether respondents reported being enrolled in an SEH doctoral program as a function of their highest degree earned, detailed occupation, age during survey year (2010), U.S. citizenship status at birth, age at entry into the United States (assigned to 0 for those born in the U.S.), disability status, sex, and race/ethnicity. Appendix Table C3 contains coefficients and standard errors from this model.

After fitting the model, we used the model results to estimate predicted probabilities for all 822,000 frame cases. Table 3.5 shows frame count and expected yield by collapsed survey strata and predicted probability range. The results from this analysis are less promising than those from the foreign-earned doctorate models. There are appears to be little variation within survey strata. Current doctoral students are a rare population that are poorly identified under the current NSCG sampling stratification scheme. The one exception is that our model may be somewhat effective at targeting likely current doctoral students among the postsecondary teacher strata.

Table 3.5. Model 3A of current doctoral students – Frame cases and expected yield by predicted probability ranges by collapsed survey strata

			Frame Count by Predicted Probability				Expected Yield by Predicted Probability			
HIDEG	DEMGROUP	OCC_DETAIL	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]
Less than PhD	High likelihood U.S.-earned degree	All others	710,000	650	(D)	(D)	2,700	90	(D)	(D)
Less than PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	3,000	800	50	(D)	80	150	20	(D)
Less than PhD	High likelihood U.S.-earned degree	Postsecondary teachers	3,100	1,600	1,100	200	80	350	400	150
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	All others	9,500	(D)	(D)	(D)	50	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	20	20	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	30	30	20	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	All others	33,500	100	(D)	(D)	250	20	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	250	90	20	(D)	(D)	20	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	80	100	250	150	(D)	20	100	100
Less than PhD	Low likelihood U.S.-earned degree; other	All others	23,500	50	(D)	(D)	150	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	150	50	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	100	80	100	50	(D)	20	50	30
PhD	High likelihood U.S.-earned degree	All others	21,000	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	1,600	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High likelihood U.S.-earned degree	Postsecondary teachers	4,300	(D)	(D)	(D)	30	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	All others	300	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	50	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	60	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	All others	1,600	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	550	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	450	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	All others	1,600	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	400	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	500	(D)	(D)	(D)	(D)	(D)	(D)	(D)
	Total		***816,000***	***3,600***	***1,600***	***450***	***3,400***	***650***	***600***	***250***

(D) – Suppressed for disclosure avoidance

3.8 Incorporating Other ACS Variables

3.8.1 ACS Grade Attending

The 2010 NSCG took the 2009 ACS as its sampling frame, but only a few ACS variables were used in stratification. NORC investigated the possibility of incorporating additional variables to help better identify members of our target populations. After review, we determined that the ACS variables for stratifying people with foreign-earned doctorates—such as citizenship status, year of entry, age, and educational attainment—were already being used in the low-likelihood U.S.-earned degree indicator variable. There was, however, one variable that seemed likely to have obvious potential for targeting current doctorate variables, i.e., ACS grade attending.

The ACS grade-attending variable reports the grade or level of recent schooling for people who attended “regular school or college” any time in the past three months. The highest level that a respondent can choose is “Graduate or professional school beyond a bachelor’s degree *(for example: MA or PhD program, or medical or law school).*” Although it does not precisely identify doctoral students, this variable should help to narrow our selection of individuals likely to be enrolled in a doctoral program.

3.8.2 Descriptive Findings

Table 3.6 shows frame counts, sample sizes, competes, and yields of current doctoral students by collapsed strata and 2009 ACS grade-attending variable. The grade-attending variable clearly increases our specificity in identifying current doctoral students. Of the approximately 800 doctoral students who were among the 47,000 survey completes, approximately 650 (81.3 percent) appeared on the 2009 ACS as having attended a graduate or professional degree program in the last three months.

As expected, the ACS grade-attending variable still lacks precision, specifically among the occupation strata who are not postsecondary teachers, biological/medical scientists, or physicists/physical scientists. Table 3.6 shows that only 150 (6.5 percent) of the 2,300 completes from the All Other collapsed occupational stratum who reported attending a graduate or professional degree program on the 2009 ACS were current doctoral students on the 2010 NSCG. This is a marked improvement over the yield rate from the All Other collapsed occupation stratum in Table 3.1, in which 200 of 33,000 completes were current doctoral students (0.6 percent), but the yield rate is still quite small.

Table 3.7 shows the 2010 NSCG response among the approximate 4,000 survey completes who were attending a graduate or professional degree program according to the 2009 ACS. Just over 17 percent were working toward a science and engineering doctorate, while the remaining 82.7 percent were either not enrolled in any program or were enrolled in a non-science and engineering doctorate or in a bachelor’s, master’s, or professional degree program. Overall, the ACS grade-attending variable has specificity (>80 percent of the current doctoral student yield from the 2010 NSCG also reported attending a graduate or professional degree program on the 2009 ACS), but it lacks precision (<20 percent of the 2010 NSCG survey completes who reported attending a graduate or professional degree program in the 2009 ACS reported attending a science and engineering doctorate program in the 2010 NSCG).

Table 3.6. 2010 NSCG frame counts, sample size, completes and yield of doctoral enrollees by collapsed survey strata and ACS grade attending

				Frame	Sample		Completes		Current PhD Students	
GRADE_ATTENDING	HIDEG	DEMGROUP	OCC_DETAIL	N	N	Weighted	N	Weighted	N	Weighted
Not enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	All others	670,000	43,000	36,320,000	31,000	33,300,000	60	37,500
Not enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	Biolog./med. scientists OR phys./physical scientist	3,300	1,900	133,000	1,500	127,000	20	1,500
Not enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	Postsecondary teachers	3,800	1,400	144,000	1,000	134,000	30	5,000
Enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	All others	41,000	3,300	2,746,000	2,300	2,419,000	150	60,000
Enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	Biolog./med. scientists OR phys./physical scientist	550	300	30,000	250	27,500	90	9,600
Enrolled in PhD or professional degree	Less than PhD	High prob. U.S.-earned degree	Postsecondary teachers	2,200	850	101,000	650	92,000	250	35,500
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Hispanic	All others	450	60	27,000	40	26,000	(D)	(D)
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Hispanic	Biolog./med. scientists OR phys./physical scientist	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Hispanic	Postsecondary teachers	30	20	2,900	20	2,500	(D)	(D)
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Asian	All others	1,900	300	138,000	200	115,000	40	14,000
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Asian	Biolog./med. scientists OR phys./physical scientist	50	50	4,500	30	3,600	(D)	(D)
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; Asian	Postsecondary teachers	450	200	32,500	150	29,500	70	10,500
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; other	All others	1,400	200	114,000	150	94,000	20	12,000
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; other	Biolog./med. scientists OR phys./physical scientist	30	30	2,500	20	2,300	(D)	(D)
Enrolled in PhD or professional degree	Less than PhD	Low prob. U.S.-earned degree; other	Postsecondary teachers	200	100	12,500	60	11,500	30	4,600

(D) – Suppressed for disclosure avoidance

Table 3.7. 2010 NSCG completes – 2009 ACS grade attending (Professional or Graduate Degree Program) by 2010 NSCG response

2009 ACS Grade Attending	2010 NSCG Response	N	%
Graduate or professional degree program	Attending - working toward bachelor's or master's	800	19.8
Graduate or professional degree program	Attending - working toward non-S&E PhD	200	4.9
Graduate or professional degree program	Attending - working toward professional degree	200	4.9
Graduate or professional degree program	Attending - working toward S&E PhD	700	17.3
Graduate or professional degree program	Not attending - has bachelor's or master's	1,700	42.0
Graduate or professional degree program	Not attending - has PhD	300	7.4
Graduate or professional degree program	Not attending - has professional degree	150	3.7
	Total	***4,050***	***100.0***

3.8.3 Targeting Current Doctoral Students Using Propensity Scores and ACS Grade-Attending Variable

One option would be to incorporate the ACS grade-attending variable as a new stratification variable. Another option would be to leave the current NSCG stratification in place but use the ACS grade-attending variable to target likely doctoral students within current NSCG strata, using a propensity score approach.

To test this approach, we fit a logistic model onto the universe of approximately 47,000 survey completes from the 2010 NSCG ACS new cohort sample. We modeled whether respondents reported being enrolled in an SEH doctoral program as a function of whether they were enrolled in a graduate or professional degree program as well as their highest degree earned, detailed occupation, age during survey year (2010), U.S. citizenship status at birth, age at entry into the U.S. (assigned to 0 for those born in the U.S.), disability status, sex, and race/ethnicity. Appendix Table 3C contains coefficients and standard errors from this model.

We used results from this model to estimate predicted probabilities for all frame cases. Table 3.8 shows the counts of frame cases and expected current doctoral student yields by collapsed survey strata and propensity score ranges. Compared with Table 3.5—which shows the same output but from the model that does not include the ACS grade-attending variable—it is apparent that the grade-attending variable adds little to our ability to target current doctoral students. The overall expected yield for high predicted probability cases is larger for the model that uses the ACS grade-attending variable but not by much, so we concluded that the ACS grade-attending variable lacks the precision to be used in a propensity score approach.

Table 3.9 shows the distribution of ACS grade-attending response (enrolled in a graduate or professional degree program vs. not) across the entire 2010 NSCG frame by collapsed strata. From this table, there are approximately 50,000 cases on the 2010 NSCG frame who responded as attending a graduate or professional degree program on the 2009 ACS. If the proportion of these cases who are enrolled in a science and engineering doctoral program (17.3 percent) is the same as among the 2010 NSCG completes (Table 3.7), then we might estimate that there are >8,500 current science and engineering doctoral students in the frame, which would provide a large enough number for a longitudinal panel. However, introducing a new stratification variable based on the ACS grade-attending variable alone would likely disrupt the NSCG sample design due to the lack of precision in the grade-attending variable. For example, given a yield rate of 17.3 percent, an additional 5,780 survey completes would be needed to achieve 1,000 science and engineering doctoral survey completes.

Table 3.8. Model 3B of current doctoral students with ACS grade attending – Frame cases and expected yield by predicted probability ranges by collapsed survey strata

			Frame Count by Predicted Probability				Expected Yield by Predicted Probability			
HIDEG	DEMGROUP	OCC_DETAIL	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]	(0, 0.1)	[0.1, 0.3)	[0.3, 0.5)	[0.5, 1]
Less than PhD	High prob. U.S.-e.arned degree	All others	706,000	4,900	100	20	2,000	700	40	(D)
Less than PhD	High prob. U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	3,400	200	250	70	40	50	90	40
Less than PhD	High prob. U.S.-earned degree	Postsecondary teachers	3,900	400	1,200	550	60	80	500	350
Less than PhD	Low prob. U.S.-earned degree; Hispanic	All others	9,500	70	(D)	(D)	30	(D)	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	30	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Hispanic	Postsecondary teachers	50	(D)	(D)	20	(D)	(D)	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Asian	All others	33,000	750	20	(D)	100	100	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	300	(D)	20	20	(D)	(D)	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Asian	Postsecondary teachers	200	(D)	150	300	(D)	(D)	60	150
Less than PhD	Low prob. U.S.-earned degree; Other	All others	23,500	300	(D)	(D)	80	50	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Other	Biolog./med. scientist OR phys./physical scientist	150	(D)	(D)	(D)	(D)	(D)	(D)	(D)
Less than PhD	Low prob. U.S.-earned degree; Other	Postsecondary teachers	150	20	70	100	(D)	(D)	30	60
PhD	High prob. U.S.-earned degree	All others	21,000	(D)	(D)	(D)	20	(D)	(D)	(D)
PhD	High prob. U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	1,600	20	(D)	(D)	(D)	(D)	(D)	(D)
PhD	High prob. U.S.-earned degree	Postsecondary teachers	4,200	90	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Hispanic	All others	300	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	50	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Hispanic	Postsecondary teachers	60	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Asian	All others	1,600	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	500	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; Asian	Postsecondary teachers	450	30	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; other	All others	1,600	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	400	(D)	(D)	(D)	(D)	(D)	(D)	(D)
PhD	Low prob. U.S.-earned degree; other	Postsecondary teachers	500	(D)	(D)	(D)	(D)	(D)	(D)	(D)
	Total		***812,000***	***6,800***	***1,800***	***1,100***	***2,400***	***1,000***	***750***	***650***

(D) – Suppressed for disclosure avoidance

Table 3.9. 2010 frame cases by 2009 ACS grade-attending variable (Attending Graduate or Professional Degree Program vs. Not)

			Frame Cases		
HIDEG	DEMGROUP	OCC_DETAIL	Not Attending	Attending	Total
Less than PhD	High likelihood U.S.-earned degree	All others	670,000	41,000	711,000
Less than PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	3,300	550	3,850
Less than PhD	High likelihood U.S.-earned degree	Postsecondary teachers	3,800	2,200	6,000
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	All others	9,100	450	9,550
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	30	(D)	30
Less than PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	50	30	80
Less than PhD	Low likelihood U.S.-earned degree; Asian	All others	31,500	1,900	33,400
Less than PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	300	50	350
Less than PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	200	450	650
Less than PhD	Low likelihood U.S.-earned degree; Other	All others	22,000	1,400	23,400
Less than PhD	Low likelihood U.S.-earned degree; Other	Biolog./med. scientist OR phys./physical scientist	150	30	180
Less than PhD	Low likelihood U.S.-earned degree; Other	Postsecondary teachers	150	200	350
PhD	High likelihood U.S.-earned degree	All others	20,000	900	20,900
PhD	High likelihood U.S.-earned degree	Biolog./med. scientist OR phys./physical scientist	1,500	80	1,580
PhD	High likelihood U.S.-earned degree	Postsecondary teachers	4,100	150	4,250
PhD	Low likelihood U.S.-earned degree; Hispanic	All others	300	20	320
PhD	Low likelihood U.S.-earned degree; Hispanic	Biolog./med. scientist OR phys./physical scientist	50	(D)	50
PhD	Low likelihood U.S.-earned degree; Hispanic	Postsecondary teachers	60	(D)	60
PhD	Low likelihood U.S.-earned degree; Asian	All others	1,500	100	1,600
PhD	Low likelihood U.S.-earned degree; Asian	Biolog./med. scientist OR phys./physical scientist	500	20	520
PhD	Low likelihood U.S.-earned degree; Asian	Postsecondary teachers	400	40	440
PhD	Low likelihood U.S.-earned degree; other	All others	1,600	80	1,680
PhD	Low likelihood U.S.-earned degree; other	Biolog./med. scientist OR phys./physical scientist	400	(D)	400
PhD	Low likelihood U.S.-earned degree; other	Postsecondary teachers	500	20	520
	Total		***771,000***	***49,700***	***820,700***

3.9 Recommendations for Current Doctoral Student Panel

The current sample design yields only a small number of currently enrolled doctoral students. Based on our analysis, a propensity score approach does not appear to help better target this population. Incorporating the unused ACS grade-attending variable helps target doctoral students, but its effectiveness is limited. The highest category from the grade-attending variable combines doctoral programs with all other graduate and professional degree programs. Therefore, the grade-attending variable offers sufficient specificity (most 2010 NSCG completes who responded as being enrolled in a doctoral program also indicated that they were in a doctoral or professional degree program in the 2009 ACS) but relatively poor precision (approximately two-thirds of NSCG respondents in the target ACS grade-attending category were not enrolled in a doctoral program in the 2010 NSCG). Therefore, we recommend using an alternative frame for targeting this population.

Chapter 4. Task 3: Sample design with NSC Data

4.1 Introduction

As noted in the Introduction, one of the chief concerns for policy analysts over the next decade is the “pipeline” of doctorate-level scientists and engineers emerging from graduate schools and taking their place in the U.S. science and engineering (S&E) workforce. The National Science Foundation (NSF) has devoted considerable resources to increasing and diversifying this pool of highly educated members of the workforce. Yet, as our preliminary look (Chapter 2) at the NSCG’s 2010-2017 longitudinal panel showed, approximately half of the persons reporting enrollment in a science/engineering/health (SEH) doctoral program in 2010 reported a nondoctoral status (neither completing the doctorate nor remaining enrolled in a doctoral program) by 2015. This observation raises multiple questions that a longitudinal sample of doctoral candidates could begin to answer: What are the factors that lead some doctoral candidates to complete their degrees while others leave their programs? What are the career paths of those SEH doctorates, compared to those who obtain non-SEH doctorates and those who leave their programs before obtaining a degree? What is the role of NSF support programs in influencing degree completion and career choice?

A longitudinal panel of doctoral candidates in U.S. institutions would provide the clearest insight into these and similar questions, as they would be followed through their graduate school careers into their post college employment. This design would have the merit of allowing researchers and policy analysts to compare outcomes between individuals who obtained doctorates and those who did not and between those who obtained SEH degrees and those who obtained non-SEH degrees.

Assuring an adequate sample size to meet analysis objectives and to compensate for longitudinal panel attrition is the first challenge to sample design. The Chapter 2 analysis of doctoral candidates depended on a sample of 414 NSCG members reporting enrollment in a U.S. SEH doctoral program in 2010. A sample of that size is inadequate to produce robust estimates of the number of contextual (e.g., graduate institution and workplace climate) and outcome measures that researchers and policy analysts will want to explore. However, this small sample is based on National Survey of College Graduates (NSCG) sample members’ 2010 self-reporting of enrollment in an SEH doctoral program. This means that identifying this group of doctoral candidates was achieved through data analysis rather than through intentional selection. In chapter 3, we considered how we might obtain a larger sample of doctoral candidates from the existing American Community Survey (ACS) frame. But we concluded that the utility of the existing frame to provide a longitudinal sample of U.S. doctoral candidates was limited.

As a result of the analysis in Chapter 3, the task that this chapter discusses started from the presumption that another source for a sample of doctoral candidates would be desirable, and there are two main ways to create such a sample: 1) Researchers could use a traditional two-stage sample design in which a statistically selected sample of graduate institutions (first stage) supplies a frame or list of graduate enrollees (second stage), from which the sample can be selected. The NSF-sponsored National Survey of Recent College Graduates, discontinued after the 2010 survey round,[11] followed this method to create a biennial cross-sectional sample of bachelor's and master's degree holders. Although this method of sample selection is perfectly defensible, it is also time-consuming and costly. 2) To alleviate these drawbacks, researchers can use a second method: that is, sampling from an existing list of graduate enrollees. That is the method that this chapter discusses.

We analyze the utility of student-level data warehoused at the National Student Clearinghouse (NSC), a nonprofit educational organization that obtains and stores data on U.S. students from kindergarten through graduation into the workforce. The NSC is the nation's leading source that employers and government agencies use to verify educational enrollment and degree completion. Its coverage of the postsecondary landscape is nearly complete. The NSC reports that the data it collects from U.S. postsecondary institutions cover 97 percent of currently enrolled U.S. postsecondary students, covering 99 percent of U.S. public and private nonprofit institutions. It records data on 94 percent of degrees awarded in U.S. institutions.[12]

The NSC provides information on college persistence and degree completion that could augment existing NSCG and Survey of Doctorate Recipients (SDR) survey data. We investigated the possibility of using the NSC for a longitudinal panel of doctoral candidates as a part of (or as a supplement to) the NSCG longitudinal redesign. To create this part of the longitudinal panel for the main NSCG sample described above, we investigated the use of the NSC as a sampling frame to select a sample for this longitudinal panel. NSC tracks individual students through their educational careers from enrollment to graduation, including transfers and withdrawals. Therefore, it can provide a frame from which an individual person-level sample can be drawn. An additional benefit of using NSC as the sampling frame will be to identify doctoral candidates who did not complete their programs. This group represents a segment of the population that is currently not captured in either the NSCG or SDR.

As we will detail later in this chapter, we obtained from NSC aggregate-level counts of persons enrolled in doctoral programs in the 2019-2020 school year. NSC provided summary data of this universe, with subtotals on several important factors: gender, race/ethnicity, field of degree (using the U.S. Department of Education's Classification of Instructional Programs [CIP] 2010 classification), and year of original enrollment back to 2015-2016. Our analysis of these data

---

[11] For more information on the National Survey of Recent College Graduates, see https://www.nsf.gov/statistics/srvyrecentgrads/.

[12] These data on the NSC's coverage come from the organization's fact sheet, that can be accessed at: https://studentclearinghouse.info/onestop/wp-content/uploads/NSCFactSheet.pdf.

intended to determine if NSC could provide a promising frame for a longitudinal panel of doctoral candidates at U.S. institutions. NSC's nearly complete coverage of U.S. postsecondary educational institutions and their students can supplement the ACS-based sample to help NCSES fulfill the Center for National Statistics' (CNSTAT) recommendation to "collect nationally representative data on the education, skills, and workforce characteristics" of the skilled technical workforce.

4.2 Feasibility of NSC-Based Frame for Longitudinal Sample of Doctoral Candidates

To assess the feasibility of NSC-provided data for use as a frame for a future study of doctoral candidates, we performed descriptive analyses on the NSC data to verify their coverage of the institution universe, as well as to provide summary statistics on proposed sampling stratification variables.

The NORC team worked collaboratively with NSC, whose data analysts provided several files of summary statistics. According to NSC, "The final cohort in this data file includes one person per year over the five-year timeframe of this study, with the earliest year of PhD enrollment being the one considered (i.e., the counts for each year are the count of earliest enrollment so the students who started in 2019-2020 are students in their 1st year; while students who started in 2015-16 are students in their 5th year as they were first enrolled in Year 1 of this study and still enrolled in 2019-2020 (Year 5 of this study)."[13] At the broadest level, NSC data tabulated >554,577 enrollees in doctoral programs in 2019-2020, a figure comparable to that reported in the 2019 NSCG, when an estimated 559,000 individuals reported being enrolled in a doctoral program on the reference date of February 1, 2019.[14] NSC data provided NORC aggregate counts, in which NSC classified field of study information according to the U.S. Department of Education's 2010 CIP.[15] NSC provided these six breakdowns:

- Doctoral candidate total counts by year of enrollment
- Aggregate counts by student demographics (i.e., counts based on the rollup of enrollment year, field of study/CIP code, race/ethnicity, and gender)
- Aggregate counts by enrollment year and field of study/CIP code

---

[13] This note is taken directly from notations on the data files (Excel files) that NSC provided to NORC. This means that the sample frame is based on all who are enrolled in a doctoral program in academic year (AY) 2019-2020 and who were first enrolled, at the earliest, in AY 2015-2016. Postsecondary institutions report doctoral program enrollment to NSC by field of degree sought, as the U.S. Department of Education's Classification of Instructional Programs defines them. As noted later in this chapter, the NSC enrollment data might include candidates for both research doctorates and for professional doctorates.

[14] This estimate is the weighted frequency of those reporting "doctorate" as the degree they were seeking when enrolled on the reference date in 2019, documented in the NSCG public use file codebook for variable ACDRG.

[15] For more information on the Classification of Instructional Programs, see the documentation at the National Center for Education Statistics at: https://nces.ed.gov/ipeds/cipcode/.

- Aggregate counts by field of study/CIP code, race/ethnicity, and gender (regardless of enrollment year)
- Aggregate race/ethnicity counts by field of study/CIP code
- Aggregate counts by field of study/CIP code

NSC demographic data generally followed U.S. Department of Education convention, with race/ethnicity and citizenship (i.e., nonresident alien) data paralleling those reported in the National Center for Education Statistics Integrated Postsecondary Education Data Systems (IPEDS). NSC also noted that it used a first-name-matching algorithm to impute missing gender data where needed.

The data NSC provided to NORC had been subjected to disclosure review so that data were suppressed in particular tabulations. For example, NSC reported that "CIP codes with student counts <10 were excluded from the 'Aggregate by Student Demos' demographic data (i.e., the second bullet listed above) for de-identification purposes to prevent identification of students broken into smaller groups." Also, in the "Aggregate by CIP code" tab (i.e., the last bullet in the list above), NSC suppressed counts of CIP codes with student counts fewer than or equal to five.

The NORC team used the NSC-provided data untransformed except for the CIP field of study data, in which case we implemented a procedure to recode the CIP-based field of study data, first, to the field of study code frame the Survey of Graduate Students and Postdocs in Science and Engineering (GSS) uses (i.e., GSS_code), and then, to the NCSES Taxonomy of Disciplines (ToD) code frame. We created field of study variables from the aggregate tallies of doctoral program enrollees the NSC provided. We linked the six-digit 2010 CIP field of study code NSC provided to a vetted CIP-to-GSS_code crosswalk (used in the 2019 GSS) that is found in methodological documentation on the GSS website. Since the GSS data code only science, engineering, and health fields, we assigned the CIP codes that did match to the GSS crosswalk into a residual non-SEH category. To assure that we did not misclassify any of these fields, we reviewed the nonmatched CIP codes to verify that they could be legitimately classified as non-SEH. Once we had assigned all CIP codes to a GSS code or to the non-SEH residual, we aggregated these data further, following GSS documentation, to mimic the eight-category "major" field of degree classification based on the NSCES ToD code frames used to summarize field of degree data in the NSCG and SDR.[16]

After creating an analysis database with the NSC data, we conducted two major validation checks to evaluate its appropriateness as a frame for a sample of U.S. doctoral candidates. The first of these checks considered the institution coverage that NSC provides. The second entailed comparing NSC's doctorate population coverage to that of NSF's GSS. We chose the GSS because it provides counts of enrolled graduate students in science, engineering, and health

---

[16] The CIP to GSS crosswalk is available in Table A-16, "Crosswalk between 2010 Classification of Instructional Program (CIP) codes and 2019 GSS Codes." The mapping of GSS codes to the eight major science, engineering and health fields is available in Table A-17, "Mapping of 2019 GSS Codes and Fields." See https://ncses.nsf.gov/pubs/nsf21318#technical-tables.

fields that allow comparisons on key factors: field of study and demographic characteristics. Although the GSS does not collect data on non-SEH graduate students, its data on degree sought and SEH field of degree are far more extensive than those available from IPEDS enrollment datasets.[17]

4.3 Validation check 1: institution coverage

NSC also provided NORC with the list—separate from the student count data—of the institutions from which it compiled the doctoral enrollee statistics. We used these data, keyed to IPEDS UNITID, to compare NSC's institution coverage to that of three other sources of postsecondary enrollment data: IPEDS and the institution universes of the Survey of Earned Doctorates (SED) and the GSS. We obtained the list of SED research-doctorate-granting institutions from the NCSES data tool,[18] limiting them to institutions that reported at least one doctorate graduate between 2010 and 2019 (n = 453). The GSS contractor at RTI International provided NORC with a current list of the institutions from which it collects GSS data, highlighting those with at least one unit (i.e., department, school, or other) supporting doctoral students in the institution (n = 448). Using the 2019 IPEDS Completions dataset and focusing on the 74,524 research/scholarly doctorate awards that IPEDS recorded from U.S. institutions, we classified all IPEDS institution identification numbers (UNITID) reporting a nonzero count for research/scholarly doctorates as to whether they appeared in the NSC, GSS, and/or SED universes of institutions. Figure 4.1 shows the whole-number percentage distribution of 2019 IPEDS completions for research/scholarly doctorates classified according to the institution lists on which their institutions appear. As is evident, ≥83 percent of institutions providing data to NSC are also represented in both the GSS and SED institution universes (legend = NSC+,GSS+,SED+). An additional 5 percent of the NSC institutions are also included in either the GSS (legend = NSC+,GSS+,SED-) or the SED (legend = NSC+,GSS-,SED+). Note that institutions appearing only in the NSC universe accounted for approximately 10 percent of the research/scholarly doctorates that IPEDS reported in 2019.

From this validation exercise, we concluded that the data that NSC collect cover the vast majority of institutions from which the GSS and SED also collect data. Slightly different definitions of the doctorate universe that each data source uses most likely account for the discrepancies noted here. NSC and SED collect data on both SEH and non-SEH doctorates; GSS focuses only on SEH doctorates. SED includes only "research doctorates" in its universe;

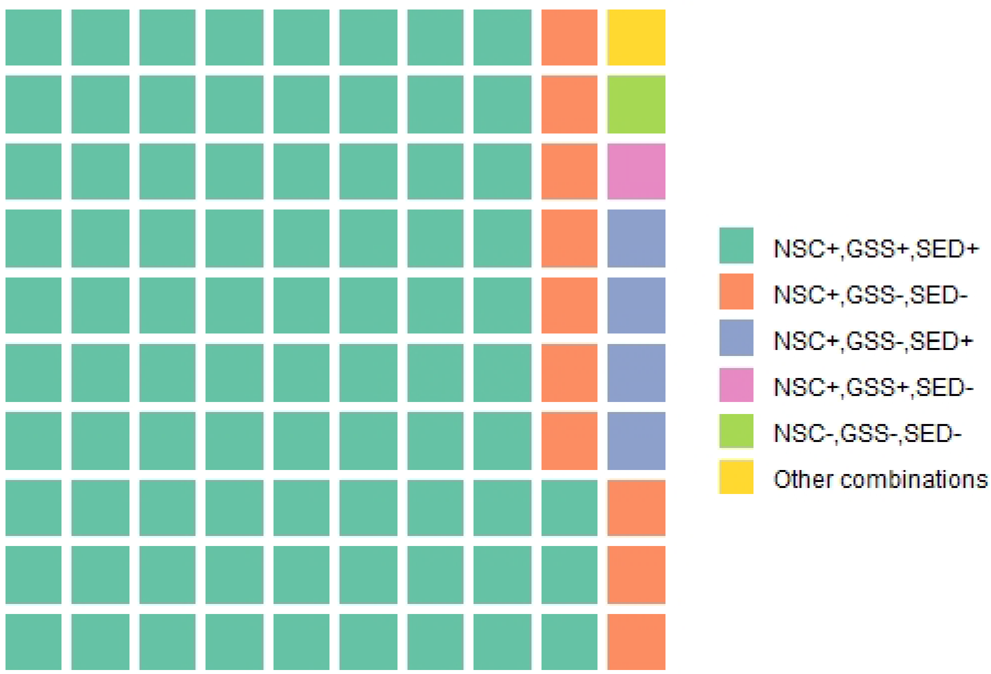
---

[17] IPEDS regularly reports data from two enrollment surveys: the 12-month and the fall enrollment surveys. The 12-month survey does not report data by program or degree sought. Only in even-numbered years, the fall enrollment survey provides counts of enrollees in nine fields of study monitored for civil rights enforcement. Neither survey distinguishes doctoral candidates from the overall population of graduate students. See further documentation of the IPEDS enrollment surveys at: https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?goToReportId=7.

[18] We downloaded from the NCSES Interactive data tool (https://ncsesdata.nsf.gov/home) an Excel data file listing counts by year for all institutions reporting in the SED. We restricted the institution universe to those institutions (n = 453) that reported at least one doctorate graduate between 2010 and 2019.

the NSC does not impose that definition. At most, institutions not represented on the NSC list accounted for only approximately 2 percent of all the research/scholarly doctorates that the 2019 IPEDS Completions reported. When subset only to institutions that, according to IPEDS, granted at least one research/scholarly doctorate in 2019, the NSC institution universe (N= 394) covers about 97 percent of the GSS institution universe (N = 408). In general, the NSC provides strong institution-level coverage of the U.S. doctorate-earning population.

Figure 4.1. Institution coverage of research/scholarly doctorate completions, by source of institution



SOURCE: Data used to identify doctorate-granting institutions was derived from National Center for Education Statistics, Integrated Postsecondary Data System, 2019 Completions database available for download at https://nces.ed.gov/ipeds/datacenter/DataFiles.

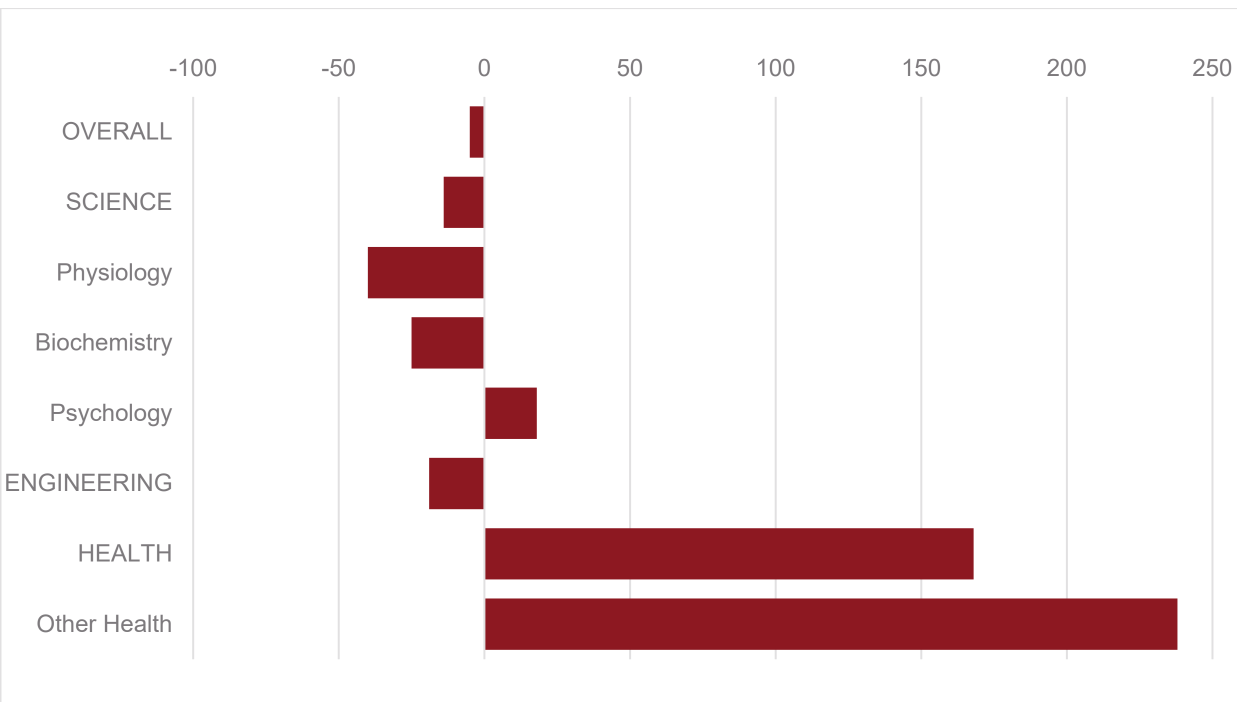
NOTE: Since IPEDS completions represents a near-census of postsecondary education graduates at all levels, we used those data to identify institutions reporting at least one research/scholarly doctorate graduate in 2019. Using the IPEDS institution identification number (UNITID) as a key, we matched doctorate-granting institutions identified in IPEDS to the set of UNITIDs represented in the NSC, SED and GSS data. The IPEDS universe of institutions granting research/scholarly doctorates in 2019 is fully represented in this diagram.

To more closely conform to the doctorate-granting institution universe covered by NCSES surveys, at NCSES's suggestion, we asked NCS to limit the data tabulations to the institutions from which GSS collects data with a nonzero count for doctoral program students across the five enrollment years. Limiting the institution universe reduced the universe of doctoral candidates from >550,000 to 411,035 The data analysis that follows, as well as a proposed sample design for an NSC-based longitudinal sample of doctoral students, is based on the fall 2019 GSS universe of U.S. postsecondary institutions.

4.4 Validation check 2: doctoral candidate population coverage

Once we resolved to align the institution universes of the GSS and the NSC data, it became apparent that we should also assess the NSC's coverage of the doctoral candidate population. At NCSES's request, the GSS team prepared a comparison of the totals of doctoral students in SEH fields of degree derived from the NSC data, to the totals obtained from the 2019 GSS public use file, and published in GSS's detailed Table 4-1 (which can be found here). The GSS team provided two comparisons: 1) "original," that is NSC doctoral student counts to GSS student counts for SEH fields of degree that aligned between the two data sources, based on the U.S. Department of Education's CIP; 2) the same comparison as the "original" comparison, but limited to institutions that report to the GSS ("GSS schools only"). The full comparison of all 107 aggregated and fine fields can be found in Appendix Table 4.1. Figure 4.2 summarizes these differences, presenting selected fields of degree, focusing on the GSS institutions-only comparison. Although the NSC student count total is approximately 95 percent of the GSS total (shown in the graph as 5 percentage points below the zero point, which would represent a perfect aligning of NSC and GSS), the two data sources vary widely across the SEH fields otherwise. In the bar graph, science, engineering, and health are the aggregated major fields of degree. These are compared to selected fine fields (e.g., physiology, other health) that are part of these major fields.

Figure 4.2. NSC population as percentage of GSS population, by field of study (restricted to institutions included in GSS only)



SOURCE: Comparison of counts of enrollees in doctoral programs, 2019 GSS and NSC, 2015 -2020 (special tabulation), GSS team at RTI International

GSS analysts provided five main observations from this comparison[19]:

- Although the aggregate sums between the two data sources are close, these hide significant differences in distribution of counts across fields. With the exception of the psychology field, the NSC counts are lower than the GSS counts for most sciences.
- The NSC counts are substantially lower [than GSS counts] for many of the detailed fields in the biosciences.
- NSC data likely include doctors of psychology (PsyD), who are excluded from the GSS.
- Most engineering fields seem underrepresented in the NSC data. GSS analysts hypothesized that this may be due to an underinclusion of temporary visa holders.
- NSC estimates are higher than the GSS estimates in all health fields, but especially the “other health” fields, which GSS analysts hypothesized result from the inclusion of students in professional doctoral degree programs.

These observations largely held up through both the “original” and the “GSS schools only” comparisons. NORC requested further documentation from NSC on the degrees and fields that its data cover. NSC reported that it can provide data indicating the type of doctorate that students are pursuing so that survey managers could devise a frame of research doctoral candidates. This would further limit the numbers of doctoral candidates that the NSC-based frame could enumerate, making the total target population accessible through NSC smaller than the “GSS schools only” total that the GSS team identified.

However, when assessing these comparisons between NSC and GSS, some caveats are in order. One relates to data disclosure limitations in the NSC aggregate data, where data are suppressed for specific fields of degree and excluded from totals. Moreover, a comparison would need to ascertain whether there are any differences in the definition of enrollees reported to NSC and GSS. Finally, it is important to establish a comparison that is as time-specific as possible. Reporting to NSC lags throughout a reporting year, so it is important to compare aggregate statistics at a defined reference date.[20] The NSC-GSS comparisons reported here compare data from GSS’s reference period of fall 2019 to NSC’s five-year cumulative totals compiled through the 2019-2020 school year (considered August 15, 2019, to August 15, 2020). Therefore, it is likely that a number of doctoral students that GSS counted in fall 2019 were not

[19] These takeaways are slightly edited recapitulations of the GSS team’s annotations that accompanied Appendix Table 4.1.

[20] NSC benchmarks its enrollment figures in the fall term to IPEDS fall enrollment data. In 2019, NSC reported that its enrollment figures accounted for 97.1 percent of all postsecondary enrollees IPEDS reported in its fall enrollment survey for that year. For details on these figures, see the “enrollment workbook” at https://nscresearchcenter.org/workingwithourdata/.

enrolled in graduate studies by August 2020. Of course, NSC might have accounted for students that GSS did not. At the very least, comparisons of coverage from these two data sources need definitions of enrollees and account for reference periods when enumerating a moving target such as the population of doctoral candidates at U.S. institutions.

A second caveat in evaluating NSC coverage of the doctoral candidate population is the NSC data's ability to provide reliable and complete data on key analysis domains for use in stratification for sample selection. The chief stratification data we considered, besides field of degree, were demographic data on doctoral candidates' gender, race/ethnicity, and citizenship. Table 4.1 summarizes NSC-provided classifications of race/ethnicity, citizenship, and gender.

Table 4.1. NSC categories of race/ethnicity, citizenship, and gender

Race/ethnicity/citizenship	■ A: nonresident alien ■ AN: Asian ■ B: Black, non-Hispanic ■ IA: American Indian/Alaska Native ■ H: Hispanic ■ HP: Native Hawaiian or Other Pacific Islander ■ W: White, non-Hispanic ■ TM: two or more race/ethnicity categories ■ U: race/ethnicity/unknown
Gender	■ F: female ■ M: male ■ Blank: gender unknown/missing

As noted above, NSC suppressed data that did not meet their disclosure requirements. In addition, NSC assigned other data cells to “unknown” or blank when data were not available. To gain a sense of the level of missingness, we calculated the proportions of missing data for those cells that were not subject to data disclosure suppression (total n = 337,040). This dataset provided demographic data by enrollment year and field of degree. We estimated that approximately 33 percent of the total number of doctoral enrollees had unknown or missing race/ethnicity, approximately 7 percent were missing gender data, and approximately 6.6 percent were missing both race/ethnicity and gender. By way of comparison, approximately 7.1 percent of graduate student headcounts reported in IPEDS enrollment data are also missing race/ethnicity data.[21]

Another NSC source dataset provided student counts by gender, race/ethnicity, and field of degree, but omitted enrollment year. Even though NSC suppressed small student counts in this dataset, the total number of students classified by race/ethnicity and gender totaled >398,000. Table 4.2 uses these data to stratify the five-year total NSC student counts by field of study (using the eight-category NCSES ToD for SEH fields, with an additional non-SEH residual category) and demographic group (gender and race/ethnicity). We classified gender into two categories: male/unknown and female. We coded race/ethnicity into three categories: White, non-Hispanic; underrepresented minority (Hispanic/Latino, and non-Hispanic Black/African American and non-Hispanic American Indian/Alaska Native) and Asian and other groups (including Native Hawaiians and other Pacific Islanders, those reporting two or more races, missing or unknown race, and nonresident aliens). Note that NSC, like IPEDS, includes a citizenship category: nonresident aliens. Given this, we assumed that all students with a

---

[21] The figure reported represents the percentage of graduate students whose race/ethnicity data are recorded as unknown. See “Table 6. Unduplicated headcount enrollment and percentage distribution at Title IV institutions, by control of institution, student level, level of institution, gender, and race/ethnicity: United States, 2019–20” at https://nces.ed.gov/ipeds/Search?query=&query2=&resultType=all&page=1&sortBy=date_desc&overlayTableId=28462.

definitive racial/ethnic category are U.S. citizens and permanent residents. But this assumption most likely undercounts non-U.S. citizens in the doctoral candidate population. It is also likely that much of the missing data on race/ethnicity may be associated with non-U.S. citizenship.

Table 4.2. NSC counts for graduate students in doctoral programs, 2015-2020

	White, Non-Hispanic		Underrepresented Minority[a]		Asian and Other Groups[b]		All
NCSES ToD Major Field Code/Name	**Male/unk**	**Female**	**Male/unk**	**Female**	**Male/unk**	**Female**	**Total**
1: Biological, agricultural, and environmental life science	8,917	11,203	1,906	2,695	11,922	10,639	**47,282**
2: Computer and information sciences	2,321	669	436	170	8,587	2,309	**14,492**
3: Mathematics and statistics	2,697	957	428	134	5,468	1,631	**11,315**
4: Physical sciences, geosciences, atmospheric sciences	8,803	4,895	1,372	919	13,186	5,333	**34,508**
5: Psychology	2,579	7,213	911	2,984	2,364	5,141	**21,192**
6: Social sciences	4,508	4,677	1,138	1,539	8,525	6,568	**26,955**
7: Engineering	10,814	3,984	2,115	962	32,259	8,429	**58,563**
8: Health	8,369	16,694	2,574	5,343	7,113	10,919	**51,012**
9: Nonscience/engineering/health	21,411	31,965	6,956	13,711	26,141	32,896	**133,080**
Total	**70,419**	**82,257**	**17,836**	**28,457**	**115,565**	**83,865**	**398,399**

[a] Under-represented minority includes Hispanic/Latino, and non-Hispanic Black/African American and non-Hispanic American Indian/Alaska Native.

[b] Includes other races (Native Hawaiians and other Pacific Islanders, those reporting two or more races), missing or unknown race, and nonresident aliens.

This brief accounting of the nature and coverage of the NSC data and its comparison to the GSS provide a foundation for a usable frame for a longitudinal sample of doctoral candidates. This will allow NCSES the flexibility to consider either of two options: 1) create a standalone survey of doctoral candidates, or 2) create a doctoral candidate supplement to the NSCG. Either option presents operational and statistical challenges that we will discuss later in this chapter.

Based on the proposed frame counts in Table 4.2, we provide insight on design effects, precision requirements, and proposed sample sizes for a longitudinal panel of doctoral candidates in Tables 4.3 to 4.5. Key assumptions for these simulations are:

- Stratification is based on 54 cells, defined by:
  - Fields of doctoral degree (nine): eight SEH cells and one non-SEH cell
  - Race/ethnicity (three): White, URM (Hispanic, Black, Native American), other (Asian, missing, nonresident alien, other race)
  - Sex (two): male/unknown sex and female
- Precision is based on a coefficient of variation for p = 0.5 estimate.
- A baseline survey response rate of 70 percent is assumed to inflate the sample size to account for nonresponse in the initial baseline survey.

Assuming that the panel will be followed for four more cycles, the cumulative response rate after the baseline survey is assumed to be 60 percent.

Table 4.3. Design effect for a sample of graduate students in doctoral programs, 2015-2020

	White, Non-Hispanic			Underrepresented Minority[a]			Asian and Other Groups[b]			All		
NCSES ToD Major Field Code	Male/ unk	Female	Total	Male/ unk	Female	Total	Male/ unk	Female	Total	Male/ unk	Female	Total
1: Biological, agricultural, and environmental life sciences	1.00	1.00	1.01	1.00	1.00	1.03	1.00	1.00	1.00	1.31	1.23	**1.26**
2: Computer and information sciences	1.00	1.00	1.31	1.00	1.00	1.19	1.00	1.00	1.33	1.85	1.76	**2.43**
3: Mathematics and statistics	1.00	1.00	1.23	1.00	1.00	1.27	1.00	1.00	1.29	1.52	1.46	**1.92**
4: Physical sciences, geosciences, atmospheric sciences	1.00	1.00	1.08	1.00	1.00	1.04	1.00	1.00	1.18	1.39	1.29	**1.54**
5: Psychology	1.00	1.00	1.22	1.00	1.00	1.28	1.00	1.00	1.14	1.14	1.11	**1.34**
6: Social sciences	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.02	1.41	1.24	**1.34**
7: Engineering	1.00	1.00	1.21	1.00	1.00	1.14	1.00	1.00	1.34	1.71	1.47	**2.19**
8: Health	1.00	1.00	1.11	1.00	1.00	1.12	1.00	1.00	1.04	1.17	1.18	**1.28**
9: Nonscience/engineering/health	1.00	1.00	1.04	1.00	1.00	1.11	1.00	1.00	1.01	1.20	1.11	**1.18**
Total	***1.53***	***2.05***	***1.84***	***1.91***	***2.63***	***2.56***	***1.53***	***1.91***	***1.70***	***2.07***	***2.33***	***2.19***

[a] Under-represented minority includes Hispanic/Latino, non-Hispanic Black/African American, and non-Hispanic American Indian/Alaska Native.
[b] Includes other races (Native Hawaiians and other Pacific Islanders, those reporting two or more races), missing or unknown race, and nonresident aliens.

Table 4.4. Final sample size[a] for graduate students in doctoral programs, 2015-2020

	White, Non-Hispanic			Underrepresented Minority[b]			Asian and Other Groups[c]			All		
NCSES ToD Major Field Code	Male/ unk	Female	Total	Male/ unk	Female	Total	Male/ unk	Female	Total	Male/ unk	Female	Total
1: Biological, agricultural, and environmental life sciences	165	165	335	165	165	340	165	165	332	648	608	**1,255**
2: Computer and information sciences	165	165	432	165	165	394	165	165	440	917	872	**2,413**
3: Mathematics and statistics	165	165	406	165	165	421	165	165	427	753	722	**1,904**
4: Physical sciences, geosciences, atmospheric sciences	165	165	358	165	165	344	165	165	390	691	638	**1,532**
5: Psychology	165	165	405	165	165	424	165	165	376	568	553	**1,331**
6: Social sciences	165	165	331	165	165	338	165	165	336	699	614	**1,325**
7: Engineering	165	165	401	165	165	377	165	165	444	847	731	**2,169**
8: Health	165	165	367	165	165	371	165	165	345	581	584	**1,267**
9: Nonscience/engineering/health	165	165	344	165	165	366	165	165	335	596	552	**1,170**
Total	***2,274***	***3,050***	***5,477***	***1,819***	***2,507***	***7,609***	***1,458***	***1,818***	***5,070***	***5,471***	***6,168***	***19,598***

[a] Sample sizes were determined to meet precision requirements (in CV) set for longitudinal measures as follows: $n = 1/(CV)^2 \times deff \times (1/R_B) \times (1/R_C)$, where CV is the coefficient of variation for $P(=0.5)$, deff is the design effect due to weight variation, R_B is the baseline survey response rate, and R_C is the cumulative response rate of multiple survey rounds after the baseline survey. Target CVs are presented in Table 4.5.

[b] Under-represented minority includes Hispanic/Latino, non-Hispanic Black/African American, and non-Hispanic American Indian/Alaska Native.

[c] Includes other races (Native Hawaiians and other Pacific Islanders, those reporting two or more races), missing or unknown race, and nonresident aliens.

Table 4.5. Requirements of baseline and longitudinal precision requirements in coefficients of variation (CV[d]), by domain

Domain	Coefficient of variation (percent)	
	Baseline	Longitudinal
Race/ethnicity × gender × field of degree	9.3	12.0
Field of degree × race/ethnicity	6.6	9.0
Field of degree × gender	5.4	9.0
Field of degree	3.8	8.0
Race/ethnicity × gender	3.1	5.0
Race/ethnicity	2.2	4.0
Gender	2.3	3.0
Total	***1.3***	***2.0***

[d] CV is calculated for a proportion estimate, p. Then, given the CV value, the required minimum sample size can be calculated as $n = CV^2 \times p/(1-p)$. For sample size determinations here, we assume $p = 0.5$, and thus $n = CV^2$.

Our exercise in sample allocation and calculation of required precision for the baseline survey and longitudinal surveys leads us to estimate an overall sample in the baseline year of approximately 20,000. Table 4.4 shows stratum-specific sample sizes. In our estimation, design effects based on combinations of sample strata vary from 1.00 to 2.63. These reflect differential sampling rates across sampling strata to meet domain-specific precision requirements calculated based on the NSC counts provided.

We apply common baseline and cumulative response rates to all strata in the tables above. But because we anticipate differential response and attrition rates across key domains, such as race/ethnicity and some fields of degree, we will adjust sampling rates to reflect differential response propensities. However, our estimated sample size should provide at least the order of magnitude of sample needed to meet precision goals set out in Table 4.5. Like other survey designs, the final sample size can be determined based on the budget available and the precision goals aligned to study goals.

4.5 Issues to Address

This brief consideration of the National Student Clearinghouse data repository as the frame for a longitudinal sample of U.S. doctoral candidates has shown that it is indeed a resource that NCSES should investigate. Although the NSC is a promising data source, it has some limitations that NCSES will need to consider: 1) First is the issue of target population coverage, as we noted in the discussion of the comparison between NSC data and those of the GSS report. The discrepancies we noted in the comparison are significant, but we believe that statistical adjustments, such as post-stratifying survey data to GSS population counts, are a viable solution to address them. 2) Second is an operational consideration—i.e., the use of NSC data for sampling purposes will require approval from institutions that provide the data to the

NSC, which is currently working with institutions to streamline this process so that data will become more readily available for research use. Survey researchers can also develop methods to gain survey cooperation that can minimize the requirements for disclosure of student contact information. For example, a student sample can be drawn based on an anonymized database with key stratification variables. Only after the sample is drawn will survey researchers need to reach out to NSC for individual student contact data and to institutions for permission to use them. Moreover, NSC maintains student contact information, such as email address, that can facilitate matches to commercial databases (e.g., Accurint) that maintain address information.

In whatever way NCSES decides to address these coverage and operational matters, it will face other statistical and methodological decisions about how to incorporate estimates of the doctorate-seeking population into its broader survey program agenda. Some of the key issues to be considered include:

- If we assume that a longitudinal panel of doctoral candidates becomes part of the NSCG, should it be part of the NSCG cross-sectional sample as well? If so, this will require multiplicity adjustments in the weighting of the NSCG.
- However, because the frame for these data differs from that of the main NSCG, the longitudinal panel of doctoral candidates should not be part of the main NSCG rotation panel, which, in any event, is designed primarily for cross-sectional estimates. NCSES will have to determine eligibility and panel rotation for the longitudinal panel separately from the main NSCG panel.
- Members of the longitudinal panel who obtain doctorates will overlap with the SDR's target population, including its longitudinal panel. To maintain accurate national estimates of the SEH doctorate population, we will need to implement statistical and operational adjustments to address this overlap. The NSF will have to address the issue of redundancy if it develops two longitudinal panels of the U.S. doctorate-seeking and -earning populations.

The use of the NSC data as a frame for a longitudinal panel of doctoral candidates will need to be carefully considered. However, the benefits of creating a sustainable panel of doctoral candidates are many. In the conception proposed here, the longitudinal panel will have several unique features:

- It will succeed in providing longitudinal data, including educational and occupational outcomes, that are nationally representative of both SEH and non-SEH doctoral candidates. Insight from these comparisons should provide solid evidence to policymakers on the utility of graduate education and occupational training to the U.S.'s 21st-century scientific enterprise.
- It will provide policymakers, employers, and researchers with an excellent resource to compare the career paths of those who completed their doctoral programs and those who did not complete them for one reason or another.
- A longitudinal panel of doctoral candidates will help NSF to fulfill the CNSTAT report's goal for NCSES to provide "longitudinal data that follow individuals through the pathways of their careers in higher education and the workforce [that] can help researchers understand their

reasons for changing those pathways and the precursors and antecedents of such changes."[22]

Chapter 5. Task 4: New Topics and Longitudinal Measures for a Redesigned NSCG

5.1 Introduction

In the previous chapters, we presented a proposal for longitudinal sample designs for the National Survey of College Graduates (NSCG) that include panels for foreign-earned (and early career foreign-earned) doctorate holders as well as U.S.-trained doctoral students. The new sample designs would not only allow for enhanced data collection on populations that are not well represented in current National Center for Science and Engineering Statistics (NCSES) surveys but also would allow for collection of longitudinal data. The longitudinal component is particularly valuable for collecting employment data that will allow researchers to better understand career trajectories.

In this chapter, we describe our recommendations for candidate longitudinal measures and topic modules for the proposed longitudinal panel surveys. We focus mainly on key longitudinal employment measures and topic modules that can stand alone or be incorporated into the current survey. Since the NSCG covers the entire science and engineering workforce in the United States, we assume that the proposed longitudinal panels will be built within the NSCG, and our recommendation for questionnaire design reflects this panel structure. However, if NCSES chooses to build a standalone longitudinal panel, our recommendations can be readily adapted to this design.

The recommendations we present here are limited to the new proposed sample designs for panels of foreign-earned doctorates and U.S.-trained doctoral students. However, some of the measures and topical modules we recommend for this redesign of the NSCG questionnaire are relevant to other populations as well. For example, some of the employment measures that we suggest be collected for foreign-earned doctorates, such as jobs beyond the principal job and spells of unemployment, are not currently collected for U.S.-earned doctorates. Incorporating these measures and topical modules into other NCSES surveys, such as the Survey of Doctorate Recipients (SDR), would enable the collection of equivalent measures from the entire population of U.S. doctorates.

[22] National Academies of Sciences, Engineering, and Medicine. *Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs*. Washington, DC: The National Academies Press; 2018.

In this chapter we describe candidate longitudinal measures and topic modules that we feel would be of greatest interest to researchers to add to the NSCG. We present proposals for how the new measures could be constructed, based on a combination of existing NSCG questions, questions drawn from other NCSES surveys, and new items that would need to be developed. We discuss the implications of the proposed new longitudinal measures and topic modules for the design of the NSCG in terms of periodicity of data collection, preloaded data that would be required to implement dependent interviewing, and question branching. We then present a potential work plan for developing and testing the new longitudinal measures and topic modules and provide recommendations for next steps.

5.2 NSCG Core Measures for New and Returning Sample

5.2.1 Employment Measures and Populations Currently Covered in NCSES Surveys

The Committee on National Statistics (CNSTAT) report indicated the need to expand data collection on the U.S. science, engineering, and health (SHE) workforce to include both: 1) longitudinal data collection to support research on SEH career pathways (Recommendations 3-1, 4-1, 6-4, and 6-6), and 2) collection of data on persons entering the U.S. workforce from abroad (Recommendation 3-2), a population that is not adequately covered in current NCSES data collections.[23]

NCSES surveys collect a wealth of information on individuals who received a doctorate degree at U.S. institutions. Three surveys focus on the doctorate-level science and engineering workforce: the Survey of Earned Doctorates (SED), the SDR, and the Early Career Doctorates Survey (ECDS). SED is a census conducted annually to collect information from all doctorate recipients at U.S. institutions. The SED collects data on the educational history, demographic characteristics, and postgraduation plans of doctorate recipients. SDR collects information on educational history, demographics, and employment status and occupation for a sample of U.S. SEH doctorate recipients. The ECDS, launched in 2017, includes individuals who received their first doctorate within the last 10 years and work at institutions that are in the sample for the 2016 Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS) or work at Federally Funded Research and Development Centers (FFRDCs). The ECDS collects retrospective data detailing employment history; educational history; work-life balance; mentoring, training, and research opportunities; and career paths and plans.

It is noteworthy that graduate students are not included in the samples for SED and SDR. Furthermore, the SED and SDR focus on U.S.-trained doctorates, thereby omitting an important part of the U.S. SEH workforce who possess foreign-earned doctorates and future doctorates in

[23] National Academies of Sciences, Engineering, and Medicine. Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs. Washington, DC: The National Academies Press; 2018.

the pipeline at U.S. doctorate-granting institutions. In addition, the employment information collected by these surveys provides a snapshot of current status. SDR (as well as NSCG, discussed below) uses a biennial repeated panel survey design that provides employment status at each survey time point. This allows for analysis of change in employment from wave to wave but leaves gaps in the data regarding changes in employment between waves. In contrast, the ECDS, launched in 2019, does collect employment history detail up through the time point of the survey, but this retrospective data collection is burdensome and may be subject to recall error. A panel design that includes both graduate students at U.S. institutions and individuals from abroad with foreign-earned doctorates and fuller data collection of employment history at each wave would enable capture of more robust data on members of the U.S. SEH workforce and their career trajectories.

5.2.2 Enhancing the NSCG Questionnaire to Collect Employment History

The NSCG collects detailed information about several employment topics. Respondents who are employed are asked about their current principal employer, the type of work that they do, whether the job requires a bachelor's degree or above, and whether the job is related to their highest earned degree. In addition, data are collected on work activities, satisfaction, and hours worked and earnings. The NSCG includes a rotating panel design with a four-panel rotation. Sample members complete a baseline survey and three biennial follow-up surveys while they are members of the panel. The set of employment questions that is asked of new and returning sample members is very similar, providing a snapshot of employment at each point in time. Appendix E lists the NSCG employment topics.

Although the NSCG collects detailed information on the respondent's current principal job in each survey round, the design of the survey yields employment data only at the survey time points, resulting in some gaps in the employment data and limitations in the data available to examine employment outcomes of doctorates in the U.S. workforce. To identify areas in which the NSCG questionnaire could provide fuller information, we compared items in the NSCG to the set of longitudinal measures for SDR provided by NCSES.[24] This comparison showed data gaps both in the topics that are covered in the NSCG (such as detail on academic employment or work environment) and in the breadth of the data that are collected (such as a complete employment history). The survey does not collect data on all jobs held and does not collect information about any jobs or gaps in employment between survey rounds. Rather, the survey collects only cross-sectional employment data on the respondent's current employment status and details about one principal employer.

Furthermore, although the NSCG obtains fuller coverage of the U.S. science and engineering workforce—including U.S.- and foreign-earned doctorates as well as graduate students enrolled in U.S. doctoral programs—the sample sizes of these populations provided by the NSCG are

---

[24] Chang WY. *Candidate Longitudinal Estimates_SDR* [Excel spreadsheet]. National Center for Science and Engineering Statistics 2020.

not sufficient to obtain reliable estimates. The longitudinal panel designs that we propose are intended to address this coverage issue.

The ability to collect longitudinal employment data and data on specific topics of interest would enhance the utility of NCSES's data on the SEH workforce in several ways. Capturing all jobs will allow for research on the factors that influence career trajectories. In addition, employment histories that could be collected would be especially rich for those without regular full-time employment (such as doctorate holders who work part-time in adjunct positions or who have temporary positions). The multiple jobs that those without permanent full-time employment hold and employment gaps may not be apparent from cross-sectional data alone. The impact of issues related to career trajectory could also be studied more comprehensively with data that could be collected in topic modules. Factors such as work-life balance, workplace climate, and graduate school experiences may affect the U.S. SEH pipeline and the career trajectory of members of the SEH workforce and need to be more fully understood.

5.2.3 Structuring of Questions to Collect Employment History

Although the NSCG does not collect all the data that are needed to understand employment outcomes for doctorate recipients, other NCSES surveys include relevant measures that could be drawn into the NSCG. For example, SDR and ECDS include questions on academic employment (such as academic position and tenure status) that NSCG does not include. However, to provide a complete employment history, the NSCG will need to collect information on all jobs rather than just the principal job and data on employment events (jobs or periods of unemployment) between survey rounds.

An approach to collecting a more complete employment history would be to mirror the methodology of the National Longitudinal Surveys of Youth 1979 and 1997 (NLSY79 and NLSY97, respectively) in collecting employment data. The NLSY79 and NLSY97 are longitudinal studies that follow two samples of youth in the United States who were born between 1957 and 1964 (NLSY79) and between 1980 and 1984 (NLSY97). The surveys cover a broad range of topics, including schooling and training, marriage and family, health, and other topics. The survey includes a detailed section on employment that collects information for all jobs the respondent has held as well as on all time gaps in which the respondent has not reported a job.

As part of the interview, the NLSY employs dependent interviewing to confirm and update information about all jobs that the respondent held at the prior interview. Information is then collected on any new jobs that the respondent began since the last interview. For each job, respondents report information on the job and dates that they started and ended the job (if no longer working there). If there is a period during which the respondent has not worked at any job, the respondent is asked reasons for not working during this gap.

The flowcharts presented in Appendix F illustrate how longitudinal employment data can be collected for the proposed longitudinal survey. Under this new design, the baseline survey will identify respondents who are either current graduate students or who possess a foreign-earned doctorate. These individuals will be asked to report information about their current employment. They will be asked about all jobs they currently hold, and, if not working, they will be asked about why they are not working and the last time they worked for pay (Appendix F, Figure F1: Employment History Baseline Flowchart). In each follow-up survey, respondents will be asked to update the status of each job they held at the time of the previous survey and to report any additional jobs they held since the previous survey (Appendix F, Figure F2: Employment History Follow-up Flowchart).

In creating the flowcharts, we assume that new respondents will join the panel as graduate students or new doctorate recipients, making it possible to collect full employment data on these groups. Although it would be possible to collect retrospective data to fill in gaps in employment history for individuals who join the panel later in their careers, the current flowcharts do not reflect the question series that would be needed to collect these data. However, the ECDS would be a model for how a prebaseline employment history could be collected. To limit respondents' burden, identifying what data on prior employment should be collected in retrospective questioning as part of the baseline survey will be necessary. Both the time and the cognitive difficulty of recalling employment history could add to their burden.

Including dependent interviewing in the NSCG methodology would help to provide the data needed to fill in the gaps between survey rounds. However, the impact of this method on survey programming, data processing, and respondent burden should be considered. Dependent interviewing requires that a substantial number of variables from the prior round of data collection be preloaded before the next survey round. This adds the complexity of additional survey programming and data processing. Furthermore, although the NLSY is interviewer-administered, NCSES surveys are primarily self-administered surveys, in which respondents may complete a paper or web-based questionnaire. Collecting retrospective data via dependent interviewing in a self-administered web survey could increase respondent burden. In addition, for surveys that are also collected via paper, the impact of survey mode on the ability to collected detailed employment history needs to be explored to develop the optimal design for collecting the data of interest. Limiting respondent burden and the data processing required entails prioritizing which measures of employment history are most important to collect retrospectively.

If longitudinal data collection for the new proposed panels is included in the NSCG, which is self-administered, design considerations will be important to ensure that the employment data collected are as complete and accurate as possible. In planning for longitudinal collection of employment data in the NSCG, there will be several important impacts on the survey: 1) It would substantially lengthen the interview to collect the above information about additional jobs that the respondent held between rounds. 2) Given the implications on burden, it will be necessary to select which survey items to include. The basic questions about each job are likely the most relevant to include for all jobs in the roster. NORC plans to consult with NCSES and other stakeholders to identify which other survey items are a priority to keep and which can be discontinued. Furthermore, NORC proposes to examine research produced from NCSES surveys to identify the variables that are most and least used in analysis by the research community to provide further information on how to prioritize variables.

5.2.4 Topic Modules for the NSCG

The proposed revised NSCG questionnaire (or standalone survey) provides an opportunity to expand the data collected about graduate students at U.S. universities and the foreign-trained doctorate workforce. To identify the topical areas in which more information is needed, the NORC team reviewed materials provided by NCSES as well as the CNSTAT report on the science and engineering workforce.[25] NORC also met with two of its experts, Norman Bradburn[26] and Dan Black[27], to discuss data needs for a revised NSCG survey. Additional information about these experts is in the Appendix.

From these sources of information, a consensus has emerged on the topics that are most useful to include in a revised NSCG in order to understand the specific experiences of U.S.-trained graduate students and the foreign-earned doctorate workforce and how they relate to future outcomes and earnings. For each of the following topics, we discuss the types of measures that would be useful to include in the NSCG. Some of the measures of interest apply to the entire group (that is, all graduate students or all foreign-earned doctorates), whereas others may be more appropriate for a subgroup (such as female graduate students or early-career foreign-earned doctorates):

- Work activities and training
- Work environment (such as harassment, discrimination, experiences of women in the workplace)
- Work-life balance
- Graduate school experiences

These modules could be fielded either during each round or occasionally, and to the entire sample or to a subset, as relevant. Implementation of these modules will need to balance utility of the information collected with the burden associated with adding modules to the NSCG.

---

[25] National Academies of Sciences, Engineering, and Medicine. Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs. Washington, DC: The National Academies Press; 2018.

[26] Professor Norman Bradburn is Tiffany and Margaret Blake Distinguished Service Professor Emeritus at University of Chicago and former University Provost. He is also a Senior Fellow in the Academic Research Centers at NORC. Associated with NORC since 1961, he has been its Director and President of its Board of Trustees.

[27] Professor Dan A. Black is deputy dean and professor at the University of Chicago Harris School of Public Policy. He is also a Senior Fellow in NORC's Economic, Labor, and Population Studies department.

5.2.4 Work Activities and Training

Recommendation 3-3 in the CNSTAT report highlights the need to obtain in-depth information on training that doctorate recipients engage in and the skills they need for their jobs. The CNSTAT recommendation is that these topics may be best suited to topic modules administered to subgroups in the sample, with the focus of the module varying across rounds. NCSES surveys currently collect some information on respondent work activities. The current NSCG, the ECDS, and the SDR all include general questions on work activities at a respondent's principal job. However, NCSES surveys do not include questions on the skills respondents feel they need to perform their jobs and to acquire new knowledge and skills as the demands of the workplace change. In addition, our experts noted that detailed information on what doctorate workforce participants actually do at their jobs is lacking.

A variety of workforce trainings is available to workers who want to enhance their skills. For example, SDR and NSCG ask whether respondents have taken trainings and their reasons for doing so. This set of questions could be expanded by adding topics and measures that are included in other federal data collections—e.g., the National Center for Education Statistics (NCES) Adult Training and Education (ATES) survey collects detailed information on training related to certifications or licenses, certificates, and work experience.[28]

As the CNSTAT report notes, the types of training that are available to workers have increased. Outside of the traditional educational pathways offered at colleges and universities, there are other trainings that provide workers with shorter-term and often more targeted trainings, such as boot camps offered by private training companies, Coursera, and other entities.[29] The types of nontraditional trainings that doctorate recipients pursue and their reasons for doing so are areas in which data are lacking.

We propose that NCSES consider developing measures related to the following topics in order to fill information gaps on the training and skills needed by doctorate holders in the SEH workforce:

- Types of skills SEH workforce members need and obtain
  - Non-science and engineering (S&E) skills needed or acquired (e.g., management, subject matter expertise outside doctoral area)
  - Retooling, upgrading skills
- Funding for the training
- Reasons for taking the training
  - Promotion, raise, upskilling, and others

---

[28] More information on the ATES survey can be found at: https://nces.ed.gov/nhes/ates.asp.

[29] A recent NORC study provides more information on the variety of workforce trainings that workers participate in: https://www.norc.org/Research/Projects/Pages/research-on-long-term-efficacy-of-career-programs-survey-of-education-attainment.aspx.

  - For personal reasons (e.g., personal enjoyment)
- Usefulness of training (impact on ability to perform current job)
- Impact on career (promotion, new job, raise, and so forth)
- Credentials earned (certifications, licenses, certificate, additional degrees, and others)
- Other training experiences (employer-provided training, continuing education, short courses, boot camps, internships, and others)

5.2.5 Work Environment

Recommendation 3-6 of the CNSTAT report indicates that core questions as well as an in-depth module on harassment and discrimination should be created. These new questions would provide much-needed information on the climate in the workplace and its impact on work performance and career trajectories for the doctoral-level SEH workforce. These issues are particularly important for gaining an understanding of potential barriers faced by members of minority groups and women in the workforce. However, for graduate students also, a better understanding of the climate on campuses and in a student's particular department is needed to measure the impact of the graduate school environment on academic progress, degree attainment, and early-career outcomes.

We propose the following topics for a module on the work environment:

- Climate and culture of workplace
  - Intellectual climate
  - Social climate
- Diversity in the workplace
- Sense of community
- Experiences with harassment and discrimination
- Impact of graduate school environment on career opportunities and choices
- Impact of work environment on career opportunities and choices

Work-Life Balance

Balancing the demands of work and personal life can have an impact on career choices. Recommendation 3-5 of the CNSTAT report notes that the NCSES should assess the new questions in the ECDS on working conditions and work-life balance and consider adding these items to the NSCG.

Questions in the ECDS ask about times that the respondent took time off from work for six or more months and the reasons for doing so. The ECDS also asks about work stress and balancing work and personal life. However, these questions do not cover other issues with work-life balance that may directly impact work performance and decisions about work. For example, our experts noted that it is important to know about day-to-day issues related to childcare, not just leaves from work, and whether parents, particularly mothers, have resources available to assist with childcare. In addition, for many respondents, elder-care responsibilities can impact career opportunities and decisions. The types of professional activities being impacted—such as the ability to produce academic or professional products (e.g., publish, write grants and proposals), collaborate with colleagues, attend networking events, and so on—need to be better understood. Potential topics for a module on work-life balance include:

- Balancing personal and professional responsibilities
- Stress of position
- Taking time off
  - Gap or time off (incidence and duration)
  - Reasons for gap or time off
- Spouse or partner educational background and occupation
- Childcare options
- Impact of work-life balance (e.g., childcare, eldercare, spouse's job) on career choices

5.2.6 Graduate Student Experiences

Because graduate students are the pipeline through which workers enter the SEH workforce, understanding their experiences at U.S. institutions is important to obtain a fuller picture of the S&E workforce. Current NCSES data collections either do not include the graduate student population or do not fully cover the topics of interest for this population. Some NCSES data collections, such as the SDR and ECDS, do not include graduate students in the study population. The SED does include graduate students, but only those who are nearing completion of their degrees. Some graduate students will be sampled into the current NSCG, but this survey does not include in-depth questions on the graduate student experience.

A GSE module will provide information on the pipeline to the SEH workforce:

- The GSE module will provide information on why some graduate students do not complete their degrees and whether or not they continue to pursue careers in S&E. Understanding graduate students' experiences and the reasons why some do not finish their degrees will provide information that could be useful to improving retention of S&E graduate students.
- The GSE module would also provide data on how graduate school experiences may be associated with career outcomes. Data on experiences with mentoring, workforce experiences, skills acquired, and more can be linked to future career outcomes.

Existing GSE Surveys. Some institutions and consortiums of institutions already field surveys on the graduate student experience. To inform the content of a GSE module, we examined publicly available information on two of these surveys, the Doctoral Exit Survey being conducted by the Association of American Universities Data Exchange (AAUDE; https://www.aaude.org/) and the gradSERU (Graduate Student Experience in the Research University) survey conducted by the SERU Consortium (https://sites.google.com/umn.edu/gradseru/home).

NORC conducted an online search to identify the most recently available versions of the AAUDE and gradSERU surveys, which cover topics in which many graduate institutions are likely interested. Table 5.1 lists topics that are covered in the AAUDE Exit Survey and the gradSeru. The content of these surveys was determined based on material identified in online searches.

Table 5.1. Graduate student exit survey content

AAUDE Exit Survey	GradSERU Exit Survey

▪ Overall satisfaction
▪ Training program/program quality
  ► Orientation, expectations, academic progress
▪ Various forms of support
▪ Faculty mentoring and advising
  ► Helpfulness and timeliness of advice on dissertation, academic and nonacademic careers
  ► Advisor and another faculty member mentor
▪ Professional development
  ► Teaching assistant position
  ► Research assistant position
  ► Scholarly presentations on campus and away from campus, and travel funding
  ► Publications, both accepted for publication and under review
▪ Satisfaction, training program/program quality
▪ Climate
▪ Postgraduation plans
▪ Demographics
	▪ Student background
	▪ Reasons for choosing program
	▪ Financial support
	▪ Experiences with primary advisor
	▪ Educational experience (e.g., access to needed resources, ability to learn and participate)
	▪ Research and training experiences
	▪ Teaching experience
	▪ Program climate (including for diverse groups)
	▪ Career plans (e.g., higher education, government, industry, and so forth)
	▪ Obstacles (requirements, advising and feedback, advisor, and so forth)
	▪ Health and well-being (physical and mental health, access to food and housing)
	▪ Overall satisfaction
	▪ Demographics
	
	The 2021 survey also included questions on the impact of the COVID-19 pandemic on students.

Note: This content was determined through examination of publicly available materials from the Massachusetts Institute of Technology and the University of Texas at Austin: https://ir.mit.edu/doctoral-exit-results web.mit.edu/surveys/grad/aau/p1.html web.mit.edu/surveys/grad/aau/p2.html web.mit.edu/surveys/grad/aau/p3.html web.mit.edu/surveys/grad/aau/p4.html web.mit.edu/surveys/grad/aau/p5.html web.mit.edu/surveys/grad/aau/p6.html https://reports.utexas.edu/content/doctoral-exit-survey	Note: This content was determined through examination of publicly available materials at: https://drive.google.com/file/d/1iUyT_o3G7m_Lw2xtHqpfg7LdVQP9SCTj/view https://drive.google.com/file/d/1DG78c6f1mRRtraIBREfBVJTi7CFyPrBX/view https://seru.tamu.edu/assessment/media/Survey-Instruments/gradSERU-2021-Instrument-Survey-Core.pdf

NORC's Research on GSE. On behalf of NCSES, NORC conducted research on a potential GSE module for the SED.[30] We conducted this research to determine whether there was interest in a GSE module and the topics that would be useful to include. NORC conducted focus groups, in-depth interviews, and a web survey to obtain information from deans, graduate department heads, and data users on this topic. The research confirmed the interest that SED institutions have in GSE data. As part of the research, NORC asked deans at SED institutions to rate potential GSE topics on how useful the information would be. Deans examined broad topics and also selected specific subtopics within each topic that would be of greatest interest to them. Table 5.2 lists broad topics and subtopics that deans considered useful to include.

Table 5.2. Topics that deans at SED institutions considered useful to include in GSE survey

Professional Developmen t Experiences	Experiences with Faculty and Staff	Academic Progress	Program and Campus Climate	Scholarly Products and Presentations	Curriculum and Instruction	Academic Support Services	Overall Satisfaction
Attending/ presenting at professional conferences	Support from faculty during dissertation process	Quality of mentoring and advising	Satisfaction with climate and culture of program	Opportunities to present papers and posters at conferences, department seminars, etc.	Overall quality of instruction	Career guidance and placement services (other than advice from faculty members)	Academic/educa tional experience
Research assistantship s	Satisfaction with support from advisors/other faculty	Financial problems/fina ncial support	The intellectual climate in the program	Number and types of publications	Relevance of coursework to professional developmen t goals	Support intended to develop writing and public speaking skills	Overall experience in program
Grant writing	Extent to which faculty/program kept pace with developments and trends in field	Physical or mental health	Sense of community within the program	Financial support for attending conferences	Inclusion of diverse perspectives	Services for international students (e.g., visa and documentation)	Whether they would recommend the university/progra m to a friend
Teaching assistantship s	Advice, guidance, and encouragement to publish from department faculty	Feelings of academic or social isolation	Diversity within the program	Number and types of presentations		Support intended to develop teaching skills	Whether they would select the same university, department or field of specialization, and advisor again
Internships	Support from faculty in finding employment outside academia	Access to affordable housing	The social climate in the program			Services for students with disabilities	Overall experience at university

[30] NORC. Assessing the potential value of adding GSE items to the SED: Focus group research. NORC; 2017.

NORC. Assessing the potential value of adding GSE items to the SED: One-on-one interviews. NORC; 2017.

Neishi K, Gleicher D, Lee L, Zimowski M, Hoffer T. Assessing the potential value of adding GSE items to the SED: Graduate dean survey. NORC; 2018.

Lee L, Neishi K, Zimowski M, Gleicher D, Hoffer T. Assessing the potential value of adding GSE items to the SED; final report. NORC; 2018.

Professional Developmen t Experiences	Experiences with Faculty and Staff	Academic Progress	Program and Campus Climate	Scholarly Products and Presentations	Curriculum and Instruction	Academic Support Services	Overall Satisfaction
Collaboration with students and/or faculty	Support from faculty in finding employment in academia	Family obligations or issues/help from family	Diversity on campus			Support intended to develop networking and job search skills	Professional development opportunities
		Degree requirements	Satisfaction with climate and culture on campus			Services for English language learners/English as a second language students	Student life experience
		Social environment/peer group support	Sense of community on campus				
			Intellectual climate on campus				

Proposed Content for GSE Module in the NSCG. Below are candidate measures for the proposed doctoral student panel. These measures are drawn from a variety of sources, including NORC's GSE research and an examination of several existing GSE surveys. We have categorized these measures into those that would be collected at baseline (first survey round) and those to be collected in subsequent rounds. For longitudinal measures, the intent is to collect a full history of employment between survey rounds.

Information for Baseline/Cross-sectional Measures

- Baccalaureate information: degree, field, academic measures
- Doctoral institution, field of study, start date (and any other basic academic information)
  - Confirm doctorate institution, field, and continued enrollment
- Academic progress toward degree (e.g., meeting requirements on schedule)
- Financial support (coverage for tuition/fees/housing/health)
- Professional development experiences
  - Opportunities and financial support for publications, conferences, grant-writing experiences, collaboration
- Experiences with faculty and staff
  - Support from faculty during dissertation
  - Support in finding academic employment
  - Support in finding nonacademic employment
  - Overall satisfaction with support from faculty and staff
- Climate
- Barriers (financial, academic advising, health, other)
  - Nonacademic skills graduate students seek to develop (e.g., management of people, intellectual property rights, and others)

Information for Longitudinal Measures

- Professional development experiences
    - RA/TA positions
    - Other work experiences
    - Time off from graduate work due to workforce experience
- If left program without degree
    - Date left program
    - Academic status (progress toward degree)
    - Reasons for leaving (financial, academic difficulty, advising, climate, change in interests, and so forth)
    - Employment (employed or not, field—SEH or not)

Other

- If completed, doctoral program
    - See career trajectory measures

5.3 Additional Considerations

Expert Feedback

NORC internal experts noted two specific barriers that stand in the way of using longitudinal data if they were available for the NSCG. First, in our experts' experience, data users are not familiar with how to conduct longitudinal analysis of the data. They noted that there is a need for training in order to help data users learn how to use the longitudinal data. They also noted that an additional barrier to the use of longitudinal NCSG data is that the structure of the data files is not conducive to longitudinal analysis. Furthermore, in order to appropriately conduct longitudinal analysis, it will be necessary to calculate longitudinal weights, which are not currently available for the NSCG. As part of future work on the development of longitudinal panel designs and questionnaires for the NSCG, file structure and weighting issues should also be addressed and data user guides developed to assist users in making full use of the longitudinal NSCG data.

A consideration that NORC's internal experts noted was that a full understanding of the experiences and career outcomes of SEH doctoral students and doctorate recipients would be enriched through a comparison with non-SEH sample members. This comparison non-SEH sample could be substantially smaller than the SEH sample. Many analysts who are interested in labor force data would want to have data from the workforce as a whole, rather than just one sector/segment of it. Understanding of the characteristics and unique features of the SEH workforce requires a comparison group to the non-SEH workforce.

The experts further noted that to reach the goal of increasing the numbers employed in the SEH workforce, it is important to consider the large numbers of doctoral students in non-SEH fields. Education and business are two major fields with a large number of graduate students. To increase the SEH workforce, graduate students would need to be drawn in from those who are pursuing study in other fields or from those who are not currently pursuing a doctoral degree.

Also, the experts indicated that a full understanding of the SEH workforce depends on understanding the training and background of these members of the SEH workforce. There are non-SEH doctorates who work in SEH jobs. Questions on how they entered SEH fields, any additional training they received to prepare, their work activities, and their experiences and trajectories are issues that could be addressed if both SEH and non-SEH doctoral students and doctorate recipients are included in the NSCG sample.

Expanding measurement to the broader population of U.S.-trained doctorates

Since the NSCG is fielded to a national sample and the new panels constitute only a small part of the SEH workforce, it is important to consider the relevance of the proposed new content to the full NSCG sample. Some of the proposed new topic module content for the NSCG pertains specifically to the new panels on foreign-earned doctorates and doctoral (or at least graduate) students at U.S. institutions. However, much of the data that would be available on employment history and topics such as work-life balance and experiences with harassment and discrimination in the workplace would be relevant to the broader sample. Those data could provide insight into the career trajectories and labor force experiences of nondoctorate holders across fields of study and for specific subgroups such as by gender and race/ethnicity.

5.4 Recommendations for Next Steps: Development Process for Longitudinal Questionnaire

In this section we present a plan for designing and testing a longitudinal questionnaire (either stand-alone or a revised form of NSCG) to accompany the new sample design. The proposed development process includes background research in the form of a literature review as well as qualitative research to gather additional information on the data needs that the NSCG should address.

Task 1: Determine Survey Design

In collaboration with NCSES, we will identify data needs and prioritize the survey measures needed in the NSCG. Based on this discussion, we will recommend content for the core NSCG survey and topic modules and fielding schedule for topic modules. We will also determine the mode(s) in which the survey will be fielded and a target length for the interview.

Task 2: Background Research

We will conduct a review of the literature and existing surveys to identify survey measures that would be important to include in the NSCG and the questions that are needed to calculate each measure. We propose to conduct qualitative research (such as in-depth interviews or focus groups) with stakeholders and experts, including NCSES staff, university leadership, researchers, early- and mid-career doctorate recipients, and graduate students (both U.S.- and foreign-earned degrees) to learn more about the data that are needed to understand issues related to degree completion and career trajectories.

Task 3: Develop Draft Questionnaires

The background research will provide the foundation for developing draft questionnaires that will meet NCSES data needs. We will draft questionnaires for the core NSCG and each topic module under consideration. We will also examine the feasibility of incorporating dependent interviewing in the NSCG. The draft baseline questionnaire and topic modules will be reviewed by the expert panel to be convened.

Task 4: Convene Expert Panel

To ensure that the redesigned NSCG questionnaire fully accounts for the data needed to understand the characteristics of and the factors affecting career trajectories of the US SEH workforce, we propose to convene an expert panel. The expert panel will be charged with detailed review of the topics and measures being considered for the revised NSCG and the baseline and follow-up questionnaires.

Task 5: Cognitive testing

The NSCG questionnaires will undergo testing in multiple rounds of cognitive interviews. In conducting recruitment for this cognitive testing, we will work in conjunction with NCSES to identify the sample characteristics that are most relevant for testing the functioning of the questionnaires. We anticipate that recruited respondents will represent a broad range of demographic characteristics graduate student experiences, and career fields and trajectories.

Task 6: Final Recommendations

With input from NCSES and the expert panel, we will finalize the baseline and follow-up questionnaires, including topic modules. We will summarize the questionnaire design work in a final report that includes draft questionnaires ready for pilot testing. This report will provide comprehensive background and documentation on the developmental tasks for the NSCG questionnaire.

Chapter 6. Conclusions and Recommendations

To conclude the NORC project for the 2020 National Science Foundation's (NSF) Broad Agency Announcement (BAA), this chapter summarizes the project goals, findings, and recommendations. The project addressed two important challenges related to the BAA strategic priority area of "moving towards a longitudinal and condensed survey design for ongoing surveys":

- How might the National Center for Science and Engineering Statistics (NCSES) workforce surveys—the National Survey of College Graduates, Early Career Doctorates Survey, Survey of Doctorate Recipients (NSCG, ECDS, and SDR, respectively)—improve coverage of individuals residing in the United States but having earned a science, engineering, or health (SEH) doctorate at a non-U.S. university? NCSES is interested in collecting longitudinal data from samples comparable to the SDR and to the early-career doctorate recipients in the ECDS. Is it possible to condense the survey designs of NSCG, ECDS, and SDR to include these individuals in an efficient way and to follow them longitudinally?
- How might the NSCG be modified to include representative samples of current doctoral students enrolled in U.S. universities and to follow them longitudinally through graduate school and into the workforce?

6.1 Summary of Study Objectives and Main Findings

The NORC project had four main objectives:

1. Assess the adequacy of NSCG sample coverage of individuals residing in the United States who earned an SEH doctorate outside the U.S. and, if not adequate, identify ways to improve the sample, particularly for supporting longitudinal analyses of the doctoral workforce consistent with the SDR design.
2. Extending the first objective, assess the adequacy of NSCG sample coverage of foreign-earned doctorate recipients in the early career stage—i.e., those who earned their doctorate within the last 10 years. Adequate coverage of this subpopulation would support a redesign of the ECDS project to draw from the Survey of Earned Doctorates (SED) and NSCG frames and support longitudinal analyses across the early career stage.
3. Assess the adequacy of NSCG sample coverage of individuals currently enrolled in research doctoral degree programs in the United States and, if not adequate, identify ways to improve the sample, particularly for longitudinal analyses.

4. Identify changes to the NSCG survey instrument that would be needed to collect the longitudinal data on foreign doctorate holders and doctoral students.

To meet the objectives related to sample design, the project conducted investigations of the American Community Survey (ACS) data used to define the NSCG sample frame and the National Student Clearinghouse (NSC) data as a possible additional sample frame for current doctoral students.

6.2 Summary of Main Findings

Task 1: Evaluation of NSCG Sample

Starting with an analysis of the NSCG public use files, we found that the sample sizes of individuals who earned SEH doctorates outside the U.S. and doctoral candidates enrolled in U.S. institutions in the NSCG are too small to obtain sufficiently accurate estimates of these groups' characteristics, experiences, and outcomes. This is particularly the case for the 2010-2017 rotating panel, which we focused on in order to investigate possible support for longitudinal analyses over four survey cycles.

Task 2: Sample Design with NSCG Sampling Frame for Foreign-Earned Doctorate Holders

Expanding the NSCG frame to include two years instead of one year of ACS data would probably yield enough foreign-earned doctorates to support a longitudinal panel. ACS does not directly ask respondents whether they earned a doctorate outside the U.S. but does collect data that, judging from propensity analyses of the NSCG survey data, allow identification with high probability. ACS sample members in high-probability strata could be oversampled to yield enough foreign-earned doctorate holders in the population parallel to the SDR—i.e, individuals earning an SEH research doctorate and 75 years of age or younger.

The ACS also collects information on the year in which the highest degree (in this case, the doctorate) was earned and, with probabilistic oversampling, may help support larger sample sizes of early-career foreign earned doctorate holders. This is significant in that it suggests the ECDS target population could be more effectively sampled through a modified design of the NSCG to cover non-U.S. doctorate recipients coupled with the SED to cover U.S.-earned doctorate recipients.

However, the ACS frame is not likely able to yield enough current doctoral students for a longitudinal panel. There is not sufficient information collected in the ACS to identify students enrolled in research doctoral programs, making oversampling impractical for cross-sectional samples as well as for support of a longitudinal panel. The highest probability subpopulations identifiable in the ACS include all U.S. graduate and professional-school students regardless of degree sought (e.g., MA/MS, MD, JD, PsyD, PhD) and field of study. Current doctoral students represent <20 percent of graduate and professional school enrollments, meaning that oversampling to reach the desired sample size would require extensive screening and associated costs.

Task 3: Sample Design with NSC Data on Doctoral Students

Establishing the size and demographic composition of the U.S. doctoral student population needed to expand the NSCG sampling frame presents challenges. Although the numbers and characteristics of current research doctoral students can be identified from the NCSES Survey of Graduate Students and Postdocs (GSS), the GSS is an institutional survey and does not collect individual-level records for the students. Ideally, a comprehensive list of current doctoral students could be obtained and a nationally representative sample drawn from the list. Alternatively, doctorate-granting institutions could be sampled and recruited from the GSS to provide lists of their doctoral students and the students sampled from those lists.

The project investigated the National Student Clearinghouse (NSC) list of current doctoral students as a possible resource for supplementing the NSCG sampling frame. The NSC provides student-level records of doctoral enrollments that, with the enrolling institution's permission, could be used to draw samples of doctoral students to augment the ACS frame. NSC claims to cover 97.1 percent of the 2019 postsecondary enrollment reported in the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS) comprehensive records for U.S. institutions.

To assess the adequacy of the NCS coverage of the doctoral student population, NORC compared the NCS totals to aggregate data from the NSF GSS data. GSS—an institution survey of universities with science and engineering (S&E) graduate programs—provides counts of graduate students by program of study; institution and Carnegie class of institution; and students' sex, race/ethnicity, and citizenship status. Comparing counts of doctoral candidates from GSS and NSC indicates agreement with respect to counts by most fields of study but significant divergence in some fields. The discrepancies we noted in the comparison are significant, but we believe that statistical adjustments, such as post-stratifying survey data to GSS population counts, are a viable solution to address them.

Task 4: Measures of Longitudinal Surveys

Assuming that the sampling challenges can be met, the NSCG instrumentation would need to be revised in order to measure the constructs of primary interest for the proposed new longitudinal panels of doctoral students and non-U.S.-earned doctorate holders.

Longitudinal study of doctoral students. NSCG does not currently include questions about doctoral students' experiences in graduate school, progress toward the degree, and reasons for completing, or not completing, the doctoral program. Promising sources for these measures include the Association of American Universities Data Exchange (AAUDE) Doctoral Exit Survey and the gradSERU (Graduate Student Experience in the Research University) survey conducted by the SERU Consortium. Prior research that NORC conducted for NSF on a potential graduate student experiences (GSE) module for the SED could also be useful. Items from these studies could be integrated into NSCG via a doctoral student module that would be repeated at each data collection cycle to follow the students longitudinally through completion of their graduate studies.

Longitudinal study of early-career doctorate recipients from U.S. and non-U.S. doctorate-granting institutions. The 2017 ECDS provides instrumentation that could be administered both to early-career doctorates from non-U.S. institutions in an expanded NSCG sample design and to ECDs sampled into the SDR. These instruments would be administered as ECD-specific modules. The 2017 ECDS sample was restricted to individuals employed at U.S. higher education institutions or at federally funded research and development centers (FFRDCs), thereby excluding the substantial contingent that was employed in private business and industry, nonprofit organizations, and government, as well as persons living and working abroad. With those exclusions lifted, the instrument would need to include questions about work experiences relevant to nonacademic and non-FFRDC employment. These could be drawn from the SDR instrument, particularly as adapted to the recent longitudinal redesign.

Longitudinal study of doctorate recipients from U.S. and non-U.S. doctorate-granting institutions. A primary goal of expanding the NSCG to include a larger sample of individuals who earned doctorates from non-U.S. institutions is to make the NCSES doctoral labor-force survey samples, taken as a whole, fully representative of U.S.-resident doctoral population. The current SDR instrument generally fits for non-U.S.-earned doctorate holders. However, the key questions about employment do not capture the full longitudinal record, focusing instead on detailed snapshots tied to the survey reference date. This shortcoming applies to the SDR as well, even though it has transitioned to a longitudinal design.

The NORC project conducted a comprehensive review of NCSES and other labor-force surveys to assess options for enhancing the longitudinal data records regarding doctorate recipients. Promising longitudinal employment event history formats are used by National Longitudinal Surveys of Youth (NLSY) and other labor force surveys; these models could be imported into the NSCG and SDR.

6.3 Recommendations

The NORC study developed recommendations tied to each research question and corresponding task area for NCSES to consider:

Task 1: Evaluation of the NSCG Sample:

- We recommend augmenting the current NSCG sample to support longitudinal panels of foreign-earned doctorate holders and doctoral students. Increased sample will allow a greater ability to estimate fine field of degree, gender, and race/ethnicity for these groups (which generally show sample sizes of <1,000 in a sample of 130,000).
- We recommend approximately 10 percent sampling rates (currently implemented in the SDR) for both groups.

Task 2: Sample Design with NSCG Sampling Frame:

- The ACS frame should be used to oversample foreign-earned doctorates for a longitudinal panel.
- The ACS frame should not be used to oversample current doctoral students for a longitudinal panel.

Task 3: Sample Design with NSC Data:

- We recommend that the NSC be used to obtain the individual records to augment the ACS frame for oversampling doctoral students.
- We recommend using the GSS population counts to establish the target population of current doctoral students by institution and field of study. Drawing on the GSS population counts for guidance, we could use post-stratification of the survey results from samples drawn from the NSC frame to make statistical adjustments to improve population estimates.

Task 4: Measures of Longitudinal Surveys:

- We recommend developing event-history formats for employment data in the NSCG and SDR; the non-U.S.-earned doctorate panel could then be integrated with the SDR.
- For the proposed new doctoral student panel, we recommend developing a new module drawing mainly from the AAUDE, GradSERU, and SED-GSE projects to measure graduate student experiences and outcomes.
- Finally, NORC recommends a multistep process for developing and testing the new measures.

6.4 Additional Considerations for Longitudinal Analyses

Expanding measurement to the broader population of U.S.-trained doctorate recipients. Since the NSCG is fielded to a national sample and the proposed new panels (i.e., non-U.S.-earned doctorate recipients and U.S. doctoral students) comprise only a small part of the SEH workforce, it is important to consider the relevance of the proposed new content to the full NSCG sample. Some of the proposed new topic module content for the NSCG discussed in Chapter 5 pertains specifically to the new panels on foreign-earned doctorates and doctoral students at U.S. institutions. However, much of the data that would be available on employment history and topics such as work-life balance and experiences with harassment and discrimination in the workplace would be relevant to the broader sample. The data could provide insight into the career trajectories and labor force experiences of non-doctorate holders across fields of study and for specific subgroups, such as by gender and race/ethnicity. This points to developing a true longitudinal design for the NSCG as a whole, with the proposed new panels as parts of that comprehensive whole.

Supporting longitudinal study of the SEH workforce. As a final part of the project, NORC sought out experts in longitudinal analysis of the workforce from NORC senior research staff and convened a meeting to discuss how longitudinal analyses of the NCSES workforce surveys might be advanced. The NORC group identified two specific barriers that stand in the way of use of longitudinal data if they were available for the NSCG:

- In our experience, many data users are not familiar with how to conduct longitudinal analyses of the data. They noted that the data users community would likely benefit from additional training opportunities to learn how to use the longitudinal data.
- Use of longitudinal NCSG data may be reduced because the structure of the data files is not conducive to longitudinal analysis—e.g, if individual records are not linked across survey cycles or longitudinal weights are not directly available. As part of future work on the development of longitudinal panel designs and questionnaires for the NSCG, file structure and weighting issues should also be addressed and data user guides developed to assist users in making full use of the longitudinal NSCG data.

The NORC team also noted that an understanding of the experiences and career outcomes of SEH doctoral students and doctorate recipients would be enriched through a comparison with non-SEH counterparts. This non-SEH comparison sample could be substantially smaller than the SEH sample.

In a related vein, the NORC team notes that a more comprehensive understanding of the SEH doctoral workforce depends on understanding the training and background not only of individuals with SEH educational backgrounds but also of those with non-SEH degrees who find their way into the SEH workforce. Questions on how they entered SEH fields, any additional training they received to prepare, their work activities, and their experiences and trajectories are issues that could be addressed for both SEH and non-SEH doctoral students and doctorate recipients who are included in the NSCG sample.

Appendix A: Cross-sectional Estimates of Selected NSCG Cohorts

Table A1 Cross-sectional estimates of selected NSCG cohorts (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	**Description**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**
ACDRG = 3	Enrolled in Ph.D program	290,000 (1,445)	242,000 (405)	306,000 (2,545)	217,000 (574)	359,000 (1,764)	239,000 (482)	364,000 (1,319)	194,000 (343)	367,000 (1,405)	193,000 (373)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	1,049,000 (2,712)	541,000 (1,332)	1,092,000 (5,445)	567,000 (1,517)	1,174,000 (5,113)	600,000 (1,315)	1,345,000 (4,929)	598,000 (1,290)	1,548,000 (5,843)	678,000 (1,314)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	331,000 (967)	193,000 (457)	313,000 (1,470)	173,000 (472)	290,000 (1,179)	160,000 (372)	316,000 (1,094)	156,000 (360)	333,000 (1,269)	191,000 (348)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	199,000 (1,211)	20,000 (57)	180,000 (1,014)	23,000 (81)	181,000 (801)	48,000 (89)	186,000 (801)	42,000 (83)	253,000 (1,608)	43,000 (130)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	61,000 (225)	8,000 (20)	55,000 (311)	5,000 (27)	48,000 (254)	17,000 (30)	53,000 (243)	22,000 (33)	55,000 (446)	14,000 (41)
ACDRG = 4	Enrolled in a professional degree program	180,000 (344)	94,000 (147)	202,000 (849)	101,000 (280)	190,000 (473)	111,000 (141)	212,000 (362)	106,000 (87)	232,000 (399)	147,000 (140)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	1,168,000 (1,683)	1,570,000 (2,407)	1,275,000 (1,918)	1,591,000 (2,313)	1,368,000 (1,650)	1,616,000 (1,883)	1,444,000 (1,433)	1,689,000 (1,779)	1,462,000 (1,404)	1,602,000 (1,698)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	337,000 (439)	509,000 (712)	386,000 (491)	476,000 (630)	413,000 (358)	457,000 (458)	360,000 (294)	429,000 (425)	316,000 (249)	362,000 (389)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	243,000 (773)	26,000 (38)	280,000 (440)	27,000 (49)	245,000 (330)	37,000 (45)	252,000 (315)	41,000 (59)	229,000 (325)	45,000 (64)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	54,000 (60)	3,000 (7)	64,000 (95)	2,000 (13)	57,000 (67)	9,000 (10)	77,000 (70)	13,000 (19)	51,000 (68)	12,000 (20)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A2 Cross-sectional estimates of selected NSCG cohorts: Men (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D program	147,000 (755)	112,000 (180)	157,000 (1,314)	102,000 (254)	190,000 (883)	98,000 (219)	176,000 (712)	85,000 (150)	186,000 (817)	84,000 (149)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	681,000 (1,714)	330,000 (749)	704,000 (3,430)	309,000 (799)	744,000 (3,205)	333,000 (666)	824,000 (3,111)	304,000 (647)	900,000 (3,649)	346,000 (665)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	187,000 (531)	106,000 (213)	181,000 (814)	81,000 (217)	167,000 (668)	72,000 (157)	164,000 (608)	63,000 (154)	165,000 (742)	96,000 (164)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	146,000 (890)	14,000 (39)	127,000 (692)	15,000 (57)	125,000 (557)	29,000 (56)	129,000 (552)	29,000 (52)	171,000 (1,131)	24,000 (80)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	41,000 (138)	5,000 (10)	36,000 (183)	4,000 (19)	34,000 (159)	12,000 (17)	37,000 (159)	15,000 (18)	32,000 (291)	9,000 (24)
ACDRG = 4	Enrolled in a professional degree program	87,000 (151)	40,000 (78)	94,000 (357)	49,000 (155)	77,000 (203)	60,000 (64)	89,000 (160)	22,000 (38)	116,000 (170)	79,000 (69)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	753,000 (1,013)	914,000 (1,368)	766,000 (1,085)	954,000 (1,324)	802,000 (904)	968,000 (1,082)	837,000 (773)	972,000 (998)	823,000 (745)	936,000 (956)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	161,000 (184)	234,000 (322)	191,000 (202)	228,000 (291)	182,000 (154)	211,000 (209)	164,000 (119)	206,000 (188)	150,000 (109)	202,000 (177)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	143,000 (467)	5,000 (19)	169,000 (239)	10,000 (18)	138,000 (180)	18,000 (20)	153,000 (181)	20,000 (29)	129,000 (176)	11,000 (28)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	34,000 (30)	D	31,000 (45)	D	28,000 (32)	8,000 (5)	41,000 (35)	11,000 (12)	28,000 (36)	5,000 (9)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A3 Cross-sectional estimates of selected NSCG cohorts: Women (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D program	143,000 (690)	130,000 (225)	149,000 (1,231)	114,000 (320)	169,000 (881)	140,000 (263)	188,000 (607)	109,000 (193)	181,000 (588)	108,000 (224)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	368,000 (998)	211,000 (583)	388,000 (2,015)	258,000 (718)	430,000 (1,908)	266,000 (649)	521,000 (1,818)	294,000 (643)	649,000 (2,194)	331,000 (649)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	145,000 (436)	87,000 (244)	133,000 (656)	92,000 (255)	122,000 (511)	88,000 (215)	152,000 (486)	93,000 (206)	168,000 (527)	94,000 (184)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	53,000 (321)	5,000 (18)	53,000 (322)	9,000 (24)	55,000 (244)	19,000 (33)	57,000 (249)	14,000 (31)	82,000 (477)	18,000 (50)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	20,000 (87)	3,000 (10)	20,000 (128)	1,000 (8)	14,000 (95)	5,000 (13)	17,000 (84)	7,000 (15)	23,000 (155)	4,000 (17)
ACDRG = 4	Enrolled in a professional degree program	94,000 (193)	54,000 (69)	108,000 (492)	52,000 (125)	113,000 (270)	51,000 (77)	122,000 (202)	83,000 (49)	116,000 (229)	68,000 (71)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	415,000 (670)	656,000 (1,039)	509,000 (833)	638,000 (989)	566,000 (746)	648,000 (801)	608,000 (660)	717,000 (781)	639,000 (659)	666,000 (742)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	176,000 (255)	275,000 (390)	195,000 (289)	248,000 (339)	231,000 (204)	246,000 (249)	196,000 (175)	223,000 (237)	166,000 (140)	160,000 (212)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	100,000 (306)	21,000 (19)	111,000 (201)	17,000 (31)	107,000 (150)	19,000 (25)	98,000 (134)	22,000 (30)	100,000 (149)	33,000 (36)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	20,000 (30)	3,000 (5)	33,000 (50)	2,000 (10)	29,000 (35)	1,000 (5)	36,000 (35)	2,000 (7)	24,000 (32)	7,000 (11)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A4 Cross-sectional estimates of selected NSCG cohorts: Hispanic or Latino (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	**Description**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**
ACDRG = 3	Enrolled in Ph.D program	22,000 (184)	12,000 (52)	30,000 (330)	23,000 (90)	28,000 (200)	16,000 (61)	28,000 (119)	15,000 (47)	32,000 (146)	20,000 (60)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	45,000 (226)	23,000 (107)	52,000 (486)	21,000 (116)	53,000 (444)	17,000 (92)	72,000 (392)	24,000 (94)	99,000 (453)	37,000 (110)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	18,000 (96)	11,000 (45)	19,000 (140)	10,000 (48)	15,000 (107)	5,000 (36)	18,000 (99)	5,000 (32)	20,000 (97)	7,000 (31)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	18,000 (72)	4,000 (5)	12,000 (58)	3,000 (7)	11,000 (50)	3,000 (10)	16,000 (50)	1,000 (5)	16,000 (100)	7,000 (11)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	5,000 (18)	D	4,000 (24)	D	2,000 (13)	D	3,000 (16)	D	2,000 (30)	D
ACDRG = 4	Enrolled in a professional degree program	7,000 (30)	6,000 (24)	13,000 (88)	11,000 (62)	11,000 (37)	9,000 (26)	24,000 (34)	9,000 (13)	40,000 (65)	32,000 (38)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	63,000 (141)	84,000 (272)	73,000 (142)	91,000 (263)	89,000 (145)	89,000 (223)	83,000 (104)	106,000 (190)	94,000 (104)	101,000 (196)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	21,000 (47)	26,000 (108)	25,000 (35)	35,000 (86)	25,000 (25)	18,000 (56)	21,000 (18)	37,000 (46)	18,000 (16)	38,000 (45)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	37,000 (132)	9,000 (5)	54,000 (113)	11,000 (9)	63,000 (77)	15,000 (12)	67,000 (72)	13,000 (12)	73,000 (67)	23,000 (15)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	5,000 (14)	D	10,000 (24)	D	8,000 (14)	D	4,000 (11)	2,000 (5)	7,000 (11)	4,000 (7)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A5 Cross-sectional estimates of selected NSCG cohorts: Non-Hispanic Asian (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	**Description**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**
ACDRG = 3	Enrolled in Ph.D program	63,000 (342)	12,000 (54)	71,000 (653)	16,000 (64)	101,000 (439)	18,000 (52)	98,000 (395)	20,000 (44)	102,000 (458)	33,000 (52)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	164,000 (697)	31,000 (128)	193,000 (1,335)	34,000 (132)	221,000 (1,235)	38,000 (104)	263,000 (1,248)	36,000 (107)	317,000 (1,841)	49,000 (173)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	75,000 (309)	11,000 (57)	73,000 (465)	12,000 (50)	66,000 (359)	10,000 (34)	71,000 (321)	13,000 (33)	78,000 (468)	28,000 (56)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	70,000 (486)	3,000 (13)	67,000 (372)	3,000 (15)	70,000 (300)	7,000 (18)	73,000 (310)	5,000 (18)	103,000 (621)	8,000 (27)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	23,000 (86)	2,000 (6)	21,000 (124)	1,000 (7)	25,000 (111)	2,000 (7)	29,000 (105)	3,000 (8)	33,000 (200)	2,000 (11)
ACDRG = 4	Enrolled in a professional degree program	46,000 (112)	7,000 (28)	39,000 (257)	20,000 (42)	51,000 (143)	16,000 (28)	47,000 (125)	3,000 (12)	43,000 (115)	6,000 (18)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	135,000 (269)	55,000 (229)	171,000 (282)	61,000 (254)	159,000 (247)	62,000 (207)	189,000 (220)	78,000 (213)	228,000 (245)	65,000 (220)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	79,000 (112)	34,000 (108)	80,000 (96)	29,000 (107)	66,000 (64)	27,000 (82)	58,000 (48)	25,000 (80)	50,000 (47)	20,000 (80)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	98,000 (366)	11,000 (14)	115,000 (159)	9,000 (13)	90,000 (119)	10,000 (10)	96,000 (126)	8,000 (12)	80,000 (140)	5,000 (11)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	16,000 (26)	D	27,000 (37)	D	28,000 (31)	D	40,000 (33)	D	23,000 (29)	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A6 Cross-sectional estimates of selected NSCG cohorts: Non-Hispanic Black (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D program	22,000 (148)	46,000 (95)	25,000 (241)	44,000 (107)	25,000 (173)	58,000 (82)	25,000 (104)	44,000 (62)	35,000 (93)	30,000 (66)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	38,000 (157)	52,000 (148)	37,000 (284)	52,000 (157)	43,000 (258)	61,000 (123)	54,000 (230)	70,000 (127)	81,000 (272)	89,000 (119)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	18,000 (68)	30,000 (63)	17,000 (90)	23,000 (60)	16,000 (67)	17,000 (41)	17,000 (72)	23,000 (39)	26,000 (84)	15,000 (29)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	4,000 (17)	D	3,000 (20)	D	5,000 (19)	D	4,000 (25)	D	20,000 (37)	1,000 (5)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	D	D	1,000 (6)	D	1,000 (7)	D	1,000 (7)	D	1,000 (11)	D
ACDRG = 4	Enrolled in a professional degree program	9,000 (42)	12,000 (16)	15,000 (99)	7,000 (27)	18,000 (56)	11,000 (14)	22,000 (32)	59,000 (10)	30,000 (55)	5,000 (15)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	46,000 (113)	64,000 (272)	51,000 (89)	73,000 (224)	63,000 (78)	85,000 (162)	82,000 (72)	82,000 (138)	59,000 (60)	101,000 (114)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	16,000 (41)	25,000 (79)	21,000 (38)	24,000 (53)	24,000 (19)	25,000 (29)	24,000 (16)	20,000 (32)	12,000 (6)	17,000 (33)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	13,000 (22)	D	16,000 (19)	D	10,000 (15)	D	8,000 (8)	D	12,000 (13)	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	D	D	4,000 (5)	D	D	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A7 Cross-sectional estimates of selected NSCG cohorts: Non-Hispanic White (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	**Description**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**
ACDRG = 3	Enrolled in Ph.D	162,000 (707)	161,000 (183)	166,000 (1,229)	126,000 (288)	197,000 (893)	139,000 (265)	201,000 (653)	105,000 (174)	188,000 (656)	105,000 (173)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	787,000 (1,559)	428,000 (922)	794,000 (3,209)	445,000 (1,072)	837,000 (3,054)	477,000 (967)	935,000 (2,930)	450,000 (931)	1,030,000 (3,137)	489,000 (884)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	215,000 (466)	136,000 (279)	200,000 (735)	122,000 (297)	188,000 (616)	124,000 (251)	205,000 (569)	110,000 (249)	204,000 (586)	137,000 (228)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	107,000 (633)	11,000 (36)	98,000 (560)	18,000 (57)	94,000 (428)	37,000 (60)	92,000 (409)	36,000 (57)	112,000 (842)	28,000 (87)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	33,000 (118)	5,000 (11)	30,000 (154)	4,000 (17)	20,000 (120)	14,000 (19)	20,000 (110)	19,000 (23)	19,000 (201)	11,000 (28)
ACDRG = 4	Enrolled in a professional degree program	113,000 (142)	62,000 (73)	129,000 (370)	60,000 (131)	102,000 (204)	71,000 (61)	111,000 (141)	35,000 (45)	110,000 (144)	98,000 (61)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	902,000 (1,120)	1,330,000 (1,540)	951,000 (1,351)	1,326,000 (1,475)	1,039,000 (1,133)	1,334,000 (1,213)	1,054,000 (998)	1,390,000 (1,155)	1,036,000 (945)	1,295,000 (1,087)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	214,000 (225)	408,000 (379)	249,000 (302)	374,000 (351)	291,000 (232)	363,000 (264)	242,000 (200)	332,000 (234)	214,000 (167)	277,000 (204)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	89,000 (241)	6,000 (18)	93,000 (145)	7,000 (26)	81,000 (119)	11,000 (21)	81,000 (109)	19,000 (32)	65,000 (105)	16,000 (35)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	23,000 (14)	1,000 (5)	22,000 (29)	2,000 (11)	21,000 (21)	8,000 (7)	33,000 (25)	10,000 (12)	16,000 (25)	7,000 (10)

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table A8 Cross-sectional estimates of selected NSCG cohorts: Other Race, non-Hispanic (sample size in parentheses)

		2010		2013		2015		2017		2019	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D	21,000 (64)	12,000 (21)	15,000 (92)	8,000 (25)	9,000 (59)	7,000 (22)	12,000 (48)	11,000 (16)	10,000 (52)	4,000 (22)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	15,000 (73)	6,000 (27)	15,000 (131)	14,000 (40)	21,000 (122)	7,000 (29)	21,000 (129)	19,000 (31)	21,000 (140)	14,000 (28)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	6,000 (28)	4,000 (13)	5,000 (40)	7,000 (17)	5,000 (30)	4,000 (10)	5,000 (33)	6,000 (7)	6,000 (34)	D
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	D	D	D	D	D	D	1,000 (7)	D	1,000 (8)	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	D	D	D	D	D	D	1,000 (5)	D	D	D
ACDRG = 4	Enrolled in a professional degree program	5,000 (18)	7,000 (6)	6,000 (35)	3,000 (18)	8,000 (33)	4,000 (12)	8,000 (30)	1,000 (7)	8,000 (20)	6,000 (8)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	21,000 (40)	38,000 (94)	29,000 (54)	40,000 (97)	17,000 (47)	46,000 (78)	36,000 (39)	33,000 (83)	45,000 (50)	40,000 (81)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	7,000 (14)	17,000 (38)	11,000 (20)	13,000 (33)	8,000 (18)	25,000 (27)	15,000 (12)	16,000 (33)	22,000 (13)	10,000 (27)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	6,000 (12)	D	D	D	D	D	D	D	D	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	D	D	D	D	D	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Appendix B: Cross-sectional Estimates of Selected NSCG Cohorts, 2009 ACS Sample Only

Table B1 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	204,000 (816)	209,000 (268)	110,000 (457)	115,000 (175)	59,000 (109)	90,000 (54)	39,000 (65)	55,000 (38)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	1,013,000 (2,439)	458,000 (787)	552,000 (2,342)	287,000 (760)	371,000 (1,096)	221,000 (331)	316,000 (1,122)	134,000 (338)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	311,000 (804)	162,000 (267)	168,000 (682)	80,000 (228)	100,000 (306)	63,000 (97)	81,000 (306)	37,000 (96)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	145,000 (610)	18,000 (39)	89,000 (487)	5,000 (26)	46,000 (198)	4,000 (14)	37,000 (197)	3,000 (11)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	59,000 (202)	8,000 (17)	29,000 (140)	2,000 (10)	15,000 (63)	2,000 (6)	13,000 (62)	D
ACDRG = 4	Enrolled in a professional degree program	131,000 (198)	79,000 (84)	62,000 (129)	19,000 (38)	19,000 (21)	4,000 (8)	8,000 (17)	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	620,000 (805)	1,154,000 (1,274)	590,000 (716)	813,000 (1,113)	439,000 (295)	579,000 (458)	301,000 (309)	475,000 (460)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	215,000 (237)	396,000 (382)	184,000 (184)	279,000 (314)	146,000 (65)	182,000 (124)	76,000 (66)	141,000 (117)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	157,000 (243)	24,000 (32)	129,000 (195)	16,000 (24)	58,000 (69)	16,000 (9)	39,000 (68)	15,000 (13)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	52,000 (54)	3,000 (7)	31,000 (41)	D	9,000 (10)	D	7,000 (11)	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B2 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Men (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	102,000 (437)	97,000 (115)	54,000 (229)	42,000 (71)	21,000 (54)	42,000 (25)	15,000 (28)	33,000 (17)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	666,000 (1,571)	283,000 (446)	354,000 (1,490)	172,000 (419)	245,000 (700)	135,000 (172)	194,000 (705)	73,000 (172)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	178,000 (442)	92,000 (132)	93,000 (370)	38,000 (106)	59,000 (173)	32,000 (37)	46,000 (169)	16,000 (39)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	104,000 (422)	13,000 (27)	63,000 (334)	4,000 (18)	35,000 (142)	3,000 (8)	27,000 (138)	2,000 (7)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	40,000 (121)	5,000 (9)	19,000 (80)	1,000 (6)	11,000 (40)	D	10,000 (38)	D
ACDRG = 4	Enrolled in a professional degree program	64,000 (83)	33,000 (49)	27,000 (45)	10,000 (19)	6,000 (7)	3,000 (6)	2,000 (5)	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	390,000 (480)	664,000 (722)	337,000 (400)	454,000 (627)	224,000 (154)	326,000 (253)	167,000 (156)	251,000 (262)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	108,000 (96)	183,000 (174)	94,000 (75)	124,000 (140)	57,000 (28)	63,000 (54)	33,000 (25)	50,000 (55)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	90,000 (131)	4,000 (16)	87,000 (107)	3,000 (11)	40,000 (39)	2,000 (5)	24,000 (38)	4,000 (6)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	33,000 (29)	D	21,000 (21)	D	3,000 (5)	D	3,000 (6)	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B3 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Women (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D program	103,000 (379)	112,000 (153)	56,000 (228)	73,000 (104)	38,000 (55)	49,000 (29)	23,000 (37)	21,000 (21)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	347,000 (868)	175,000 (341)	198,000 (852)	116,000 (341)	126,000 (396)	86,000 (159)	122,000 (417)	61,000 (166)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	133,000 (362)	70,000 (135)	74,000 (312)	42,000 (122)	40,000 (133)	31,000 (60)	34,000 (137)	20,000 (57)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	41,000 (188)	5,000 (12)	25,000 (153)	1,000 (8)	11,000 (56)	1,000 (6)	9,000 (59)	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D.outside US between 2000 and 2010	19,000 (81)	3,000 (8)	10,000 (60)	D	4,000 (23)	D	4,000 (24)	D
ACDRG = 4	Enrolled in a professional degree program	67,000 (115)	47,000 (35)	35,000 (84)	9,000 (19)	12,000 (14)	D	6,000 (12)	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	230,000 (325)	490,000 (552)	253,000 (316)	359,000 (486)	215,000 (141)	253,000 (205)	134,000 (153)	224,000 (198)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	108,000 (141)	213,000 (208)	90,000 (109)	155,000 (174)	89,000 (37)	118,000 (70)	43,000 (41)	91,000 (62)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	67,000 (112)	20,000 (16)	42,000 (88)	13,000 (13)	18,000 (30)	D	15,000 (30)	11,000 (7)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	19,000 (25)	3,000 (5)	10,000 (20)	D	5,000 (5)	D	4,000 (5)	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B4 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Hispanic or Latino (sample size in parentheses)

		2010		2013		2015		2017	
Definition	**Description**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**	**S E H**	**Non-S E H**
ACDRG = 3	Enrolled in Ph.D. program	16,000 (92)	10,000 (31)	14,000 (62)	15,000 (32)	1,000 (8)	3,000 (8)	4,000 (6)	D
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	44,000 (199)	19,000 (64)	25,000 (190)	13,000 (66)	18,000 (106)	7,000 (31)	16,000 (104)	7,000 (32)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	17,000 (81)	9,000 (24)	11,000 (73)	7,000 (23)	7,000 (38)	1,000 (9)	5,000 (34)	1,000 (8)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	15,000 (40)	D	7,000 (32)	D	2,000 (10)	D	2,000 (9)	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	4,000 (15)	D	2,000 (12)	D	D	D	D	D
ACDRG = 4	Enrolled in a professional degree program	3,000 (13)	5,000 (11)	4,000 (11)	4,000 (11)	D	D	D	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	40,000 (79)	62,000 (143)	40,000 (64)	43,000 (118)	26,000 (30)	32,000 (51)	20,000 (24)	30,000 (54)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	15,000 (29)	18,000 (52)	12,000 (17)	14,000 (39)	D	4,000 (11)	D	5,000 (11)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	23,000 (59)	9,000 (5)	20,000 (51)	D	13,000 (21)	D	6,000 (22)	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	5,000 (13)	D	4,000 (12)	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B5 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Non-Hispanic Asian (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	50,000 (222)	9,000 (34)	21,000 (109)	7,000 (13)	7,000 (21)	D	3,000 (10)	D
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	158,000 (626)	25,000 (69)	96,000 (589)	18,000 (67)	64,000 (273)	15,000 (27)	52,000 (271)	10,000 (27)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	72,000 (266)	8,000 (34)	35,000 (210)	5,000 (29)	22,000 (97)	3,000 (12)	18,000 (99)	2,000 (10)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	53,000 (238)	3,000 (8)	31,000 (184)	1,000 (7)	18,000 (77)	D	14,000 (73)	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	23,000 (83)	2,000 (5)	8,000 (51)	D	7,000 (23)	D	5,000 (19)	D
ACDRG = 4	Enrolled in a professional degree program	38,000 (67)	5,000 (16)	11,000 (33)	D	4,000 (5)	D	D	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	78,000 (106)	35,000 (120)	85,000 (106)	33,000 (113)	40,000 (44)	22,000 (48)	44,000 (54)	15,000 (45)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	53,000 (49)	23,000 (49)	43,000 (37)	19,000 (41)	24,000 (13)	12,000 (20)	16,000 (13)	10,000 (19)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	59,000 (91)	10,000 (10)	38,000 (65)	5,000 (6)	20,000 (24)	D	17,000 (24)	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	14,000 (22)	D	3,000 (13)	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B6 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Non-Hispanic Black (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	15,000 (71)	39,000 (57)	11,000 (43)	14,000 (34)	5,000 (13)	17,000 (9)	2,000 (7)	15,000 (8)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	37,000 (133)	45,000 (89)	19,000 (118)	22,000 (86)	16,000 (49)	20,000 (35)	11,000 (52)	10,000 (39)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	17,000 (53)	27,000 (40)	9,000 (45)	8,000 (31)	5,000 (15)	4,000 (12)	3,000 (16)	5,000 (15)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	2,000 (6)	D	D	D	D	D	D	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	D	D	D	D	D	D	D	D
ACDRG = 4	Enrolled in a professional degree program	6,000 (21)	11,000 (9)	6,000 (16)	2,000 (7)	D	D	18,000 (18)	21,000 (45)
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	23,000 (44)	43,000 (158)	28,000 (45)	35,000 (137)	34,000 (12)	25,000 (52)	6,000 (6)	12,000 (7)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	9,000 (21)	18,000 (42)	10,000 (19)	15,000 (38)	D	13,000 (11)	D	D
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	11,000 (9)	D	9,000 (8)	D	D	D	D	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	D	D	D	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B7 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Non-Hispanic White (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	106,000 (394)	141,000 (132)	54,000 (225)	72,000 (87)	45,000 (63)	62,000 (33)	28,000 (40)	31,000 (21)
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	761,000 (1,420)	364,000 (548)	404,000 (1,385)	231,000 (522)	265,000 (642)	175,000 (233)	230,000 (664)	104,000 (232)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	199,000 (382)	113,000 (161)	109,000 (335)	58,000 (137)	64,000 (149)	51,000 (62)	53,000 (149)	27,000 (61)
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	75,000 (324)	10,000 (27)	49,000 (265)	4,000 (18)	25,000 (108)	4,000 (9)	20,000 (111)	2,000 (6)
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	31,000 (102)	5,000 (11)	18,000 (76)	1,000 (7)	8,000 (36)	2,000 (5)	8,000 (39)	D
ACDRG = 4	Enrolled in a professional degree program	84,000 (88)	52,000 (42)	38,000 (62)	12,000 (16)	10,000 (11)	2,000 (5)	6,000 (9)	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	469,000 (559)	985,000 (796)	426,000 (482)	673,000 (694)	336,000 (205)	474,000 (285)	214,000 (206)	397,000 (292)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	135,000 (131)	324,000 (216)	115,000 (105)	220,000 (176)	96,000 (41)	135,000 (72)	47,000 (40)	105,000 (68)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	59,000 (78)	6,000 (17)	59,000 (67)	4,000 (14)	25,000 (23)	D	16,000 (22)	6,000 (8)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	23,000 (13)	1,000 (5)	19,000 (13)	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Table B8 Cross-sectional estimates of selected NSCG cohorts, 2009 ACS sample only: Other Race, non-Hispanic (sample size in parentheses)

		2010		2013		2015		2017	
Definition	Description	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H	S E H	Non-S E H
ACDRG = 3	Enrolled in Ph.D. program	17,000 (37)	11,000 (14)	11,000 (18)	7,000 (9)	D	D	D	D
DGRDG = 3 and (HDDGRUS=Y)	Earned Ph.D. in US	13,000 (61)	6,000 (17)	8,000 (60)	4,000 (19)	8,000 (26)	4,000 (5)	6,000 (31)	3,000 (8)
DGRDG = 3 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned Ph.D. in US between 2000 and 2010	6,000 (22)	4,000 (8)	3,000 (19)	2,000 (8)	2,000 (7)	D	2,000 (8)	D
DGRDG = 3 (HDDGRUS=N)	Earned Ph.D. outside US	D	D	D	D	D	D	D	D
DGRDG = 3 and HDAY5 in (2000 or 2005) and (HDDGRUS=N)	Earned Ph.D. outside US between 2000 and 2010	D	D	D	D	D	D	D	D
ACDRG = 4	Enrolled in a professional degree program	1,000 (9)	7,000 (6)	2,000 (7)	D	D	D	D	D
DGRDG = 4 and (HDDGRUS=Y)	Earned professional degree in US	11,000 (17)	29,000 (57)	10,000 (19)	28,000 (51)	D	24,000 (22)	5,000 (7)	12,000 (24)
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=Y)	Earned professional degree in US between 2000 and 2010	3,000 (7)	13,000 (23)	4,000 (6)	11,000 (20)	D	18,000 (10)	D	10,000 (12)
DGRDG = 4 and (HDDGRUS=N)	Earned professional degree outside US	5,000 (6)	D	D	D	D	D	D	D
DGRDG = 4 and (HDAY5: 2000 or 2005) and (HDDGRUS=N)	Earned professional degree outside US between 2000 and 2010	D	D	D	D	D	D	D	D

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Survey of College Graduates, 2010-2017
D = suppressed to avoid disclosure of confidential information. (* = Value < 500)
S E H = Science, Engineering, Health (For enrolled degree, NACEDNG = 11 to 61; for earned degree, NDGMENG = 11 to 61)
Non-S E H = Not Science, Engineering, Health (For enrolled degree, NACEDNG = 62 to 76; for earned degree, NDGMENG = 62 to 76)

Appendix C: Model Coefficients and Standard Errors

Table C1 Coefficients and Standard Errors for Model 1: Model of Foreign Earned PhDs

Description: This table shows the estimated coefficients and standard errors from a logistic model predicting whether 2010 NSCG respondent holds a foreign earned PhD as a function of ACS frame variables.

		Model 1- Foreign Earned PhDs		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig.**
(Intercept)	-	-9.217	0.7368	***
Age	-	-0.0257	0.0059	***
Masters' Degree	Bachelor's or Professional Degree	-0.9026	0.2667	***
Doctorate Degree	Bachelor's or Professional Degree	3.669	0.1662	***
Computer and information scientists	Mathematical scientists	0.01073	0.3538	
Biological/medical students	Mathematical scientists	1.444	0.3221	***
Agriculture and other life scientists	Mathematical scientists	-0.1891	0.6157	
Chemists, except biochemists	Mathematical scientists	0.7583	0.4759	
Physicists and other physical scientists	Mathematical scientists	1.032	0.3291	**
Psychologists	Mathematical scientists	-1.178	0.8166	
Economics	Mathematical scientists	0.08458	0.5375	
Other social sciences	Mathematical scientists	-0.566	0.6867	
Chemical engineers	Mathematical scientists	-15.07	585	
Civil and architectural engineers	Mathematical scientists	-1.389	0.8417	.
Electrical and computer engineers	Mathematical scientists	-0.3843	0.4977	
Mechanical engineers	Mathematical scientists	-0.2168	0.6168	
Other engineers	Mathematical scientists	-0.7413	0.4418	.
S&E-related health occupations	Mathematical scientists	0.4203	0.456	
S&E-related non-health occupations	Mathematical scientists	-0.7267	0.6265	
Postsecondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	0.2978	0.3171	
Postsecondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	-0.5812	0.4145	
Secondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	0.2535	0.5046	
Secondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	-15.11	563.5	
Non-S&E high interest occupations, S&E bachelor's FOD	Mathematical scientists	-0.3205	0.334	
Non-S&E low interest occupations, S&E bachelor's FOD	Mathematical scientists	-0.9088	0.4272	*
Non-S&E occupations (high and low) non-S&E bachelor's FOD	Mathematical scientists	-2.696	1.224	*
Not working, S&E bachelor's FOD	Mathematical scientists	-0.1352	0.439	
Not working, non-S&E bachelor's FOD	Mathematical scientists	-15.41	581.5	
Black	Hispanic	-1.041	0.5011	*
Asian	Hispanic	0.06513	0.2278	
AIAN	Hispanic	1.864	0.8021	*
NHPI	Hispanic	-14.26	1506	
White or other	Hispanic	1.008	0.2263	***
Not Disabled	Disabled	1.329	0.5031	**
Female	Male	-0.2382	0.1228	.

		Model 1- Foreign Earned PhDs		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig.**
Born in PR or Outlying Areas; or Born Abroad of American Parents	US Citizen At Birth	-2.016	1.096	.
Not Citizen or Naturalized Citizen	US Citizen At Birth	0.3561	0.3247	
Age at Entry into US		0.1254	0.00777	***

N = 47000, significance: *** < 0.0001, ** <0.001, * <0.05, . <0.1

Table C2 Coefficients and Standard Errors for Model 2: Model of Recent Foreign Earned PhDs

Description: This table shows the estimated coefficients and standard errors from a logistic model predicting whether 2010 NSCG respondent holds a foreign earned PhD as a function of ACS frame variables.

		Model 2 - Recent Foreign Earned PhDs		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig**
(Intercept)	-	0.8343	1.407	
Age	-	-0.2868	0.02211	***
Masters' Degree	Bachelor's or Professional Degree	-1.378	0.6279	*
Doctorate Degree	Bachelor's or Professional Degree	4.078	0.2964	***
.Computer and information scientists	Mathematical scientists	-0.2566	0.6597	
Biological/medical students	Mathematical scientists	1.18	0.522	*
Agriculture and other life scientists	Mathematical scientists	1.263	0.7789	
Chemists, except biochemists	Mathematical scientists	1.23	0.7165	.
Physicists and other physical scientists	Mathematical scientists	0.8503	0.5339	
Psychologists	Mathematical scientists	-15.36	1448	
Economics	Mathematical scientists	1.206	0.7631	
Other social sciences	Mathematical scientists	0.4215	0.9254	
Chemical engineers	Mathematical scientists	-15.83	1447	
Civil and architectural engineers	Mathematical scientists	-14.67	1074	
Electrical and computer engineers	Mathematical scientists	0.00456	0.8115	
Mechanical engineers	Mathematical scientists	-0.0692	1.153	
Other engineers	Mathematical scientists	-1.005	0.8934	
S&E-related health occupations	Mathematical scientists	-0.6707	0.9511	
S&E-related non-health occupations	Mathematical scientists	0.07577	1.135	
Postsecondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	0.3301	0.5262	
Postsecondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	-0.8576	0.8075	
Secondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	-0.4719	1.224	
Secondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	-14.64	1402	
Non-S&E high interest occupations, S&E bachelor's FOD	Mathematical scientists	-0.2065	0.5906	
Non-S&E low interest occupations, S&E bachelor's FOD	Mathematical scientists	-0.6946	0.7361	
Non-S&E occupations (high and low) non-S&E bachelor's FOD	Mathematical scientists	-0.555	1.153	
Not working, S&E bachelor's FOD	Mathematical scientists	-0.1039	0.8248	
Not working, non-S&E bachelor's FOD	Mathematical scientists	-14.62	1393	

		Model 2 - Recent Foreign Earned PhDs		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig**
Black	Hispanic	-1.952	1.107	.
Asian	Hispanic	-0.0365	0.3618	
AIAN	Hispanic	1.861	1.193	
NHPI	Hispanic	-15.18	4002	
White or other	Hispanic	0.488	0.3584	
Not Disabled	Disabled	-0.5234	1.058	
Female	Male	-0.0035	0.1876	
Born in PR or Outlying Areas; or Born Abroad of American Parents	US Citizen At Birth	-3.936	1.259	**
Not Citizen or Naturalized Citizen	US Citizen At Birth	-3.871	0.6712	***
Age at Entry into US		0.2777	0.02292	***

N = 47000, significance: *** < 0.0001, ** <0.001, * <0.05, . <0.1

Table C3 Coefficients and Standard Errors for Model 3A and 3B: Models of Current PhD Students

Description: This table shows the estimated coefficients and standard errors from two logistic model predicting whether 2010 NSCG respondent is a current S&E PhD student as a function of ACS frame variables and ACS Grade Attending variable

		Model 3A - Current PhD Students			Model 3B - Current PhD Students		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig**	**Estimate**	**Std. Error**	**Sig**
(Intercept)	-	0.428	0.3828		-2.334	0.4278	***
Age	-	-0.1218	0.00536	***	-0.06749	0.005548	***
Masters' Degree	Bachelor's or Professional Degree	0.9228	0.08502	***	0.8553	0.09143	***
Doctorate Degree	Bachelor's or Professional Degree	-1.985	0.2923	***	-1.132	0.3015	***
Computer and information scientists	Mathematical scientists	-1.193	0.354	***	-1.081	0.3654	**
Biological/medical students	Mathematical scientists	0.9655	0.2973	**	0.5662	0.3189	.
Agriculture and other life scientists	Mathematical scientists	-0.8528	0.7483		-0.6528	0.7662	
Chemists, except biochemists	Mathematical scientists	0.09398	0.3731		-0.1168	0.3935	
Physicists and other physical scientists	Mathematical scientists	1.63	0.2522	***	1.084	0.2691	***
Psychologists	Mathematical scientists	1.473	0.342	***	1.077	0.3726	**
Economics	Mathematical scientists	1.086	0.4077	**	0.8191	0.4474	.
Other social sciences	Mathematical scientists	-0.9669	0.4442	*	-1.002	0.4573	*
Chemical engineers	Mathematical scientists	-1.966	1.028	.	-1.912	1.039	.
Civil and architectural engineers	Mathematical scientists	-1.696	0.6232	**	-1.508	0.6366	*
Electrical and computer engineers	Mathematical scientists	-1.577	0.5526	**	-1.455	0.5635	**
Mechanical engineers	Mathematical scientists	-1.497	0.5526	**	-1.496	0.5656	**
Other engineers	Mathematical scientists	-0.6905	0.3718	.	-0.7103	0.3864	.
S&E-related health occupations	Mathematical scientists	-0.0966	0.3098		-0.379	0.3229	
S&E-related non-health occupations	Mathematical scientists	-0.4689	0.341		-0.6529	0.3545	.
Postsecondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	2.492	0.2373	***	1.258	0.2516	***
Postsecondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	0.5696	0.3306	.	-0.6973	0.3437	*

		Model 3A - Current PhD Students			Model 3B - Current PhD Students		
Variable	**(Reference Category)**	**Estimate**	**Std. Error**	**Sig**	**Estimate**	**Std. Error**	**Sig**
Secondary teachers, S&E bachelor's field of degree (FOD)	Mathematical scientists	-1.154	0.4703	*	-1.797	0.4796	***
Secondary teachers, non-S&E bachelor's field of degree (FOD)	Mathematical scientists	-1.984	1.029	.	-2.52	1.036	*
Non-S&E high interest occupations, S&E bachelor's FOD	Mathematical scientists	-1.083	0.2844	***	-1.119	0.2955	***
Non-S&E low interest occupations, S&E bachelor's FOD	Mathematical scientists	-0.2817	0.2767		-0.5939	0.289	*
Non-S&E occupations (high and low) non-S&E bachelor's FOD	Mathematical scientists	-1.964	0.5514	***	-2.095	0.5589	***
Not working, S&E bachelor's FOD	Mathematical scientists	0.02001	0.3664		-0.4326	0.383	
Not working, non-S&E bachelor's FOD	Mathematical scientists	-12.78	236.3		-13.12	239.1	
Black	Hispanic	0.1504	0.1748		0.01055	0.1856	
Asian	Hispanic	-0.0856	0.1499		-0.08587	0.1605	
AIAN	Hispanic	0.1641	0.3731		0.0923	0.3962	
NHPI	Hispanic	-1.071	1.139		-0.8844	1.201	
White or other	Hispanic	-0.2191	0.1341		-0.2035	0.1441	
Not Disabled	Disabled	-0.3869	0.2142	.	-0.4031	0.2306	.
Female	Male	-0.4087	0.08167	***	-0.3107	0.08781	***
Born in PR or Outlying Areas; or Born Abroad of American Parents	US Citizen at Birth	-0.5296	0.3484		-0.6074	0.3651	.
Not Citizen or Naturalized Citizen	US Citizen at Birth	-0.0287	0.2069		0.002453	0.2179	
Age at Entry into US		0.0194	0.00788	*	0.01108	0.008389	
Attending Graduate or Professional Degree Program	Not Attending Graduate or Professional Degree Program	-	-		3.038	0.1137	***

N = 47000, significance: *** < 0.0001, ** <0.001, * <0.05, . <0.1

Appendix D. Comparison of NSC Data for GSS Eligible CIP Codes to the 2019 GSS Doctoral Counts

Table D1 Comparison of NSC data for GSS Eligible CIP Codes to the 2019 GSS Doctoral Counts

	GSS Data (from 2019 DT 4-1)		NSC Data					Comparison to GSS data			
			Original (All schools)		Updated (GSS schools only)		Difference	Original NSC vs GSS		Updated NSC vs GSS	
Field	Number	Percent	Number	Percent	Number	Percent	Number	Difference	% of GSS	Difference	% of GSS
All detailed fields	281,889	100.0	324,380	100.0	268,529	100.0	-55,851	42,491	115%	-13,360	95%
Science	193,896	68.8	192,058	59.2	167,317	62.3	-24,741	-1,838	99%	-26,579	86%
Agricultural sciences	3,889	1.4	3,364	1.0	3,364	1.3	0	-525	87%	-525	87%
Biological and biomedical sciences	53,915	19.1	43,337	13.4	42,932	16.0	-405	-10,578	80%	-10,983	80%
Biochemistry	4,534	1.6	3,427	1.1	3,406	1.3	-21	-1,107	76%	-1,128	75%
Biology	7,166	2.5	6,144	1.9	6,128	2.3	-16	-1,022	86%	-1,038	86%
Biomedical sciences	4,579	1.6	5,355	1.7	5,232	1.9	-123	776	117%	653	114%
Biophysics	890	0.3	950	0.3	950	0.4	0	60	107%	60	107%
Biostatistics and bioinformatics	3,192	1.1	2,478	0.8	2,452	0.9	-26	-714	78%	-740	77%
Biotechnology	98	*	101	0.0	101	0.0	0	3	103%	3	103%
Botany and plant biology	1,295	0.5	1,121	0.3	1,121	0.4	0	-174	87%	-174	87%
Cell, cellular biology, and anatomical sciences	4,975	1.8	3,344	1.0	3,327	1.2	-17	-1,631	67%	-1,648	67%
Ecology and population biology	2,571	0.9	2,119	0.7	2,119	0.8	0	-452	82%	-452	82%
Epidemiology	1,916	0.7	1,466	0.5	1,466	0.5	0	-450	77%	-450	77%
Genetics	2,082	0.7	1,392	0.4	1,392	0.5	0	-690	67%	-690	67%
Microbiological sciences and immunology	3,937	1.4	2,461	0.8	2,427	0.9	-34	-1,476	63%	-1,510	62%
Molecular biology	1,153	0.4	900	0.3	900	0.3	0	-253	78%	-253	78%
Neurobiology and neuroscience	5,138	1.8	3,556	1.1	3,550	1.3	-6	-1,582	69%	-1,588	69%
Nutrition science	948	0.3	738	0.2	730	0.3	-8	-210	78%	-218	77%
Pathology and experimental pathology	843	0.3	540	0.2	530	0.2	-10	-303	64%	-313	63%
Pharmacology and toxicology	2,151	0.8	1,418	0.4	1,393	0.5	-25	-733	66%	-758	65%
Physiology	2,703	1.0	1,732	0.5	1,613	0.6	-119	-971	64%	-1,090	60%
Zoology and animal biology	1,198	0.4	1,028	0.3	1,028	0.4	0	-170	86%	-170	86%
Biological and biomedical sciences nec	2,546	0.9	3,067	0.9	3,067	1.1	0	521	120%	521	120%
Computer and information sciences	17,192	6.1	20,570	6.3	14,667	5.5	-5,903	3,378	120%	-2,525	85%
Computer science	8,646	3.1	7,625	2.4	7,279	2.7	-346	-1,021	88%	-1,367	84%
Computer and information sciences, general	6,952	2.5	11,198	3.5	6,044	2.3	-5,154	4,246	161%	-908	87%
Computer and information sciences nec	1,594	0.6	1,747	0.5	1,344	0.5	-403	153	110%	-250	84%
Geosciences, atmospheric sciences, and ocean sciences	6,551	2.3	5,267	1.6	5,243	2.0	-24	-1,284	80%	-1,308	80%

	GSS Data (from 2019 DT 4-1)		NSC Data					Comparison to GSS data			
			Original (All schools)		Updated (GSS schools only)		Difference	Original NSC vs GSS		Updated NSC vs GSS	
Field	Number	Percent	Number	Percent	Number	Percent	Number	Difference	% of GSS	Difference	% of GSS
Atmospheric sciences and meteorology	866	0.3	687	0.2	687	0.3	0	-179	79%	-179	79%
Geological and earth sciences	4,239	1.5	3,401	1.0	3,400	1.3	-1	-838	80%	-839	80%
Ocean and marine sciences	1,446	0.5	1,179	0.4	1,156	0.4	-23	-267	82%	-290	80%
Geosciences, atmospheric sciences, and ocean sciences nec	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Mathematics and statistics	13,565	4.8	11,426	3.5	11,426	4.3	0	-2,139	84%	-2,139	84%
Mathematics and applied mathematics	10,308	3.7	8,725	2.7	8,725	3.2	0	-1,583	85%	-1,583	85%
Statistics	3,257	1.2	2,701	0.8	2,701	1.0	0	-556	83%	-556	83%
Multidisciplinary and interdisciplinary studies	2,978	1.1	2,543	0.8	2,421	0.9	-122	-435	85%	-557	81%
Natural resources and conservation	3,677	1.3	3,262	1.0	3,189	1.2	-73	-415	89%	-488	87%
Environmental science and studies	1,738	0.6	1,640	0.5	1,568	0.6	-72	-98	94%	-170	90%
Forestry, natural resources, and conservation	1,939	0.7	1,622	0.5	1,621	0.6	-1	-317	84%	-318	84%
Physical sciences	36,506	13.0	30,262	9.3	30,257	11.3	-5	-6,244	83%	-6,249	83%
Astronomy and astrophysics	1,373	0.5	1,143	0.4	1,143	0.4	0	-230	83%	-230	83%
Chemistry	19,748	7.0	16,290	5.0	16,287	6.1	-3	-3,458	82%	-3,461	82%
Materials sciences	1,013	0.4	1,006	0.3	1,006	0.4	0	-7	99%	-7	99%
Physics	13,951	4.9	11,492	3.5	11,490	4.3	-2	-2,459	82%	-2,461	82%
Physical sciences nec	421	0.1	331	0.1	331	0.1	0	-90	79%	-90	79%
Psychology	20,231	7.2	40,675	12.5	23,822	8.9	-16,853	20,444	201%	3,591	118%
Clinical psychology	3,785	1.3	12,498	3.9	6,200	2.3	-6,298	8,713	330%	2,415	164%
Counseling and applied psychology	6,537	2.3	12,854	4.0	8,415	3.1	-4,439	6,317	197%	1,878	129%
Psychology, general	6,749	2.4	12,670	3.9	6,815	2.5	-5,855	5,921	188%	66	101%
Research and experimental psychology	3,160	1.1	2,653	0.8	2,392	0.9	-261	-507	84%	-768	76%
Social sciences	35,392	12.6	31,352	9.7	29,996	11.2	-1,356	-4,040	89%	-5,396	85%
Agricultural economics	806	0.3	734	0.2	734	0.3	0	-72	91%	-72	91%
Anthropology	4,365	1.5	3,634	1.1	3,634	1.4	0	-731	83%	-731	83%
Criminal justice and safety studies	900	0.3	1,241	0.4	860	0.3	-381	341	138%	-40	96%
Economics (except agricultural)	8,045	2.9	6,501	2.0	6,501	2.4	0	-1,544	81%	-1,544	81%
Geography and cartography	1,741	0.6	1,393	0.4	1,393	0.5	0	-348	80%	-348	80%
History and philosophy of science	257	0.1	241	0.1	241	0.1	0	-16	94%	-16	94%
Human development	731	0.3	581	0.2	581	0.2	0	-150	79%	-150	79%
International relations and national security studies	413	0.1	594	0.2	505	0.2	-89	181	144%	92	122%
Linguistics	1,616	0.6	1,435	0.4	1,435	0.5	0	-181	89%	-181	89%
Political science and government	5,488	1.9	4,646	1.4	4,646	1.7	0	-842	85%	-842	85%
Public policy analysis	2,414	0.9	1,811	0.6	1,790	0.7	-21	-603	75%	-624	74%
Sociology	5,070	1.8	4,259	1.3	4,259	1.6	0	-811	84%	-811	84%
Social sciences nec	3,546	1.3	4,282	1.3	3,417	1.3	-865	736	121%	-129	96%
Engineering	72,065	25.6	58,543	18.0	58,543	21.8	0	-13,522	81%	-13,522	81%

Field	GSS Data (from 2019 DT 4-1)		NSC Data					Comparison to GSS data			
			Original (All schools)		Updated (GSS schools only)		Difference	Original NSC vs GSS		Updated NSC vs GSS	
	Number	Percent	Number	Percent	Number	Percent	Number	Difference	% of GSS	Difference	% of GSS
Aerospace, aeronautical, and astronautical engineering	2,554	0.9	2,225	0.7	2,225	0.8	0	-329	87%	-329	87%
Agricultural engineering	662	0.2	574	0.2	574	0.2	0	-88	87%	-88	87%
Bioengineering and biomedical engineering	7,715	2.7	5,808	1.8	5,808	2.2	0	-1,907	75%	-1,907	75%
Biological and biosystems engineering	219	0.1	167	0.1	167	0.1	0	-52	76%	-52	76%
Chemical engineering	7,057	2.5	5,620	1.7	5,620	2.1	0	-1,437	80%	-1,437	80%
Civil engineering	7,752	2.8	6,094	1.9	6,094	2.3	0	-1,658	79%	-1,658	79%
Electrical, electronics, and communications engineering	18,577	6.6	15,090	4.7	15,090	5.6	0	-3,487	81%	-3,487	81%
Engineering mechanics, physics, and science	1,447	0.5	1,361	0.4	1,361	0.5	0	-86	94%	-86	94%
Industrial and manufacturing engineering	3,762	1.3	2,882	0.9	2,882	1.1	0	-880	77%	-880	77%
Mechanical engineering	11,247	4.0	9,257	2.9	9,257	3.4	0	-1,990	82%	-1,990	82%
Metallurgical and materials engineering	4,616	1.6	3,620	1.1	3,620	1.3	0	-996	78%	-996	78%
Mining engineering	201	0.1	187	0.1	187	0.1	0	-14	93%	-14	93%
Nanotechnology	146	0.1	122	0.0	122	0.0	0	-24	84%	-24	84%
Nuclear engineering	1,031	0.4	884	0.3	884	0.3	0	-147	86%	-147	86%
Petroleum engineering	607	0.2	469	0.1	469	0.2	0	-138	77%	-138	77%
Engineering nec	4,472	1.6	4,183	1.3	4,183	1.6	0	-289	94%	-289	94%
Health	15,928	5.7	73,779	22.7	42,669	15.9	-31,110	57,851	463%	26,741	268%
Clinical medicine[a]	4,571	1.6	6,091	1.9	4,273	1.6	-1,818	1,520	133%	-298	93%
Anesthesiology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Cardiology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Endocrinology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Gastroenterology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Hematology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Neurology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Obstetrics and gynecology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Oncology and cancer research	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Ophthalmology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Otorhinolaryngology	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Pediatrics	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Psychiatry	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Public health	4,191	1.5	5,445	1.7	3,681	1.4	-1,764	1,254	130%	-510	88%
Pulmonary disease	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Radiological sciences	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Surgery	ne	ne	0	0.0	0	0.0	0	0	--	0	--
Clinical medicine nec	380	0.1	646	0.2	592	0.2	-54	266	170%	212	156%
Other health	11,357	4.0	67,688	20.9	38,396	14.3	-29,292	56,331	596%	27,039	338%
Communication disorders sciences	911	0.3	2,814	0.9	2,532	0.9	-282	1,903	309%	1,621	278%

Field	GSS Data (from 2019 DT 4-1)		NSC Data					Comparison to GSS data			
			Original (All schools)		Updated (GSS schools only)		Difference	Original NSC vs GSS		Updated NSC vs GSS	
	Number	Percent	Number	Percent	Number	Percent	Number	Difference	% of GSS	Difference	% of GSS
Dental sciences	208	0.1	3,510	1.1	1,789	0.7	-1,721	3,302	1688%	1,581	860%
Nursing science	3,439	1.2	12,080	3.7	8,174	3.0	-3,906	8,641	351%	4,735	238%
Pharmaceutical sciences	3,121	1.1	13,785	4.2	6,956	2.6	-6,829	10,664	442%	3,835	223%
Veterinary biomedical and clinical sciences	692	0.2	2,430	0.7	1,964	0.7	-466	1,738	351%	1,272	284%
Other health nec	2,986	1.1	33,069	10.2	16,981	6.3	-16,088	30,083	1107%	13,995	569%

* = value < 0.05%; ne = not eligible
nec = not elsewhere classified

Appendix E. Proposed Measures for a Longitudinal NSCG

UNIVERSE: Students in a PhD program and PhD recipients.

NOTE: Some measures require the collection of retrospective data or will include use of dependent interviewing. The methodologies for collection of retrospective data and dependent interviewing are under development and not fully reflected in the descriptions of the measures in this table or in the accompanying appendix flowcharts.

Table E1 Proposed Measures for a Longitudinal NSCG: **Time to Event**

Estimates/Measures	Topics	Calculation of the Measure
Time to Tenure-track job Time in months/years from receipt of the PhD to start of a tenure track job.	Employment	This measure is calculated as the number of months (or years) from receipt of the PhD to start of a tenure-track job. The measure is based on Date of Receipt of Doctorate Degree/NSCG-D11, and Start Month/Year of Current Job/NSCG-A20, if current job is a tenure track job and is the first tenure track job.
Time to Tenured Time in months/years from start date of tenure track job to date of receipt of tenured status.	Employment	This measure is calculated as the number of months (or years) from start date of tenure track job to receipt of tenured status. The measure is based on Start Month/Year of Current Job/NSCG-A20, if current job is a tenure track job, and Date of Receipt of Tenured Status/SDR-A18. This measure assumes that the job is a tenure-track professorship and is the same job across survey rounds. Status at the job would be collected and updated across survey rounds through dependent interviewing.

Estimates/Measures	Topics	Calculation of the Measure
Time to Full Professor Time in months/years from start of tenure track job to full professor.	Employment	This measure is calculated as the number of months (or years) from start date of tenure track job to receipt of status of full professor. The measure is based on Start Month/Year of Current Job/NSCG-A20, if current job is a tenure track job, and Date of Receipt of Full Professor Status/(new item based on SDR-A17 that collects year obtained faculty rank of professor), if the job is a professorship and is same job across survey rounds. Status at the job would be collected and updated across survey rounds through dependent interviewing.
Time on 1st Postdoc Time months/years between start date and end date of the first postdoc position.	Employment	This measure is calculated as the number of months (or years) between beginning and ending the first postdoc position. The measure is based on Start Date of Current Principal Job/NSCG-A20 and End Date of the Job (new item collected through dependent interviewing), if the current principal position is a Postdoc/SDR-A16 and is the First Postdoc (new item). Although this calculation is based on a question regarding principal job and assumes that the postdoc is the principal job, the revised survey would collect full employment history, not just principal job.
Time on 2nd Postdoc Time months/years between start date and end date of the second postdoc position.	Employment	This measure is calculated as the number of months (or years) between beginning and ending the second postdoc position. The measure is based on Start Date of Current Job/NSCG-A20 and End Date of the Job (new item collected through dependent interviewing), if the current position is a Postdoc/SDR-A16 and is the Second Postdoc (new item). Although this calculation is based on a question regarding principal job and assumes that the postdoc is the principal job, the revised survey would collect full employment history, not just principal job.
Time to 1st Retirement Time in years from first job to first retirement.	Employment	This measure is calculated as time in years from the first job (after receipt of the PhD) to the first retirement. The measure is based on Start Date of First Job (NSCG-A20) reported after Receipt of PhD (NSCG-D11), and Date of First Retirement/NSCG-A3 or NSCG-A8. New item will determine if this retirement is first retirement; if not first retirement, collect year of first retirement. These measures could likely be obtained through retrospective questioning on start dates of first job post-PhD and date of first retirement.

Estimates/Measures	Topics	Calculation of the Measure
Time to Leaving the U.S. Time in months/years between first arriving in the US and first leaving the US to reside in another country.	Work-Life Balance	This measure is calculated as the time in months/years between first arriving in the US for six months or more and first leaving the US to reside in another country. The measure is based on Year First Came to the US for Six Months or Longer (NSCG-E13) and Leaving the US to Reside in Another Country (new item).
Time to Naturalization Time in years between first arriving in the US to naturalization.	Work-Life Balance	This measure is calculated as the time in years between first arriving in the US for six months or longer and becoming a citizen by naturalization. The measure is based on Year First Came to the US for Six Months or Longer (NSCG-E13) and US Citizen by Naturalization (NSCG-E9 and Date of Naturalization (new item).
Persistence in Sector/Job Type Duration in months/years within same sector/job type.	Employment	This measure is calculated from the start month/year and end month/year of employment in the same sector. The measure is based on continuous employment within the same Sector (NSCG-A13). Although this calculation is based on the sector of the principal job, the revised survey would collect full employment history, not just principal job. Continuous employment within a sector could be calculated across any job held by the respondent.

Table E2 Proposed Measures for a Longitudinal NSCG: **Persistence**

Estimates/Measures	Topics	Calculation of the Measure
Persistence in S&E Employment Duration in months/years in S&E employment.	Employment	This measure is calculated from the start month/year and end month/year of employment with S&E. The measure is based on continuous employment in a S&E field based on Job Category (NSCG-A18). Persistence in S&E employment would be calculated across all jobs in the respondent's employment history. For individuals who are not continuously employed in S&E, duration of each spell of S&E employment could be calculated.
Spells/Duration of Employment Outside S&E Duration in years within non-S&E sector/job type.	Employment	This measure is calculated from the start month/year and end month/year of employment within S&E. The measure is based on continuous employment in a non-S&E field based on Job Category (NSCG-A18). Duration of employment outside of S&E would be calculated across all jobs in the respondent's employment history.
Duration of Current Principal Job Time in months/years spent in current principal job.	Employment	This measure is calculated based on the start month/year of the job reported as the current principal job. The measure is based on Start Month/Year of the current principal job (NSCG-A20) and the End Month/Year of the current principal job (new item collected through dependent interviewing).
Spells/Duration of Unemployment Duration in months/years of spells of unemployment.	Employment	This measure is calculated based on the dates of employment reported for all jobs held. Based on dates of employment, a period of months/years in which no job was held is a spell of unemployment, if the individual is in the labor force (NSCG-A3, or new item on whether individual seeks to work).
Spells/Duration of Not Working and Not Seeking Work Duration in months/years of spells of not in labor force (not working and not seeking work).	Employment	This measure is calculated based on the dates of employment reported for all jobs held. Based on dates of employment, a period of months/years in which no job was held is a spell of not working and not seeking work, if the individual is not in the labor force (NSCG-A3, or new item on whether individual seeks to work).

Estimates/Measures	Topics	Calculation of the Measure
Persistence in Retirement Duration in years in retirement.	Employment	This measure is calculated based on time elapsed since date of first retirement, until the individual reports returning to the labor force. The measure is based on Employment Status—Not Working/NSCG-A1, being Retire/NSCG-A3 and Year of Retirement/NSCG-A34. Through dependent interviewing, follow-up data on employment status will be obtained.
Full-time to Part-time Transition from full-time position to part-time position since last interview.	Employment	This measure flags a change in hours worked per week from full time to part time. The measure is based on a change in response to Hours Worked in a Typical Week/NSCG-A32 from 35 or more hours per week in one round to less than 35 hour per week in the next round, if the respondent was employed in both rounds.

Table E3 Proposed Measures for a Longitudinal NSCG: **Conditional Transition (occurrence)**

Estimates/Measures	Topics	Calculation of the Measure
Full-time to Part-time Transition from full-time position to part-time position since last interview.	Employment	This measure flags a change in hours worked per week from full time to part time. The measure is based on a change in response to Hours Worked in a Typical Week /NSCG-A32 from 35 or more hours per week in one round to less than 35 hour per week in the next round, if the respondent was employed in both rounds.
Full-time to Retirement Transition from full-time employment to retirement since last interview.	Employment	This measure flags a change in employment status from employed full time to retired. The measure is based on a change in response across survey rounds from Employment Status— Working/NSCG-A1 and Hours Worked in a Typical Week /NSCG-A32 of 35 or more hours per week to a status of Employment Status—Not Working/NSCG-A1 and being Retired/NSCG-A3.
Part-time to Retirement Transition from part-time employment to retirement since last interview.	Employment	This measure flags a change in employment status from employed part time to retired. The measure is based on a change in response across survey rounds from Employment Status— Working/NSCG-A1 and Hours Worked in a Typical Week /NSCG-A32 of less than hours per week to a status of Employment Status—Not Working/NSCG-A1 and being Retired/NSCG-A3.

Estimates/Measures	Topics	Calculation of the Measure
Retirement to Full/Part-time Transition from retirement to full-time or part-time position since last interview.	Employment	This measure flags a change across survey rounds in employment status from retired to full-time or part-time employment. The measure is based on a change in response across survey rounds from Employment Status—Not Working/NSCG-A1 and being Retired/NSCG-A3 to Employment Status— Working/NSCG-A1.
Move to Different Industry Change in industry of primary job since last interview.	Employment	This measure flags a change across survey rounds in industry of employment. This measure is based on a change in response to Employment Sector/NSCG-A13.
Change of Citizenship Status Change in citizenship status since last interview.	Work-Life balance?	This measure flags a change across survey rounds in citizenship status. The measure is calculated based on a change in Citizenship Status/NSCG-E8 from non-US citizen to US citizen.
Number of Jobs Worked Number of employers since receipt of PhD.	Employment	This measure reflects the number of employers worked for since receiving the PhD. The measure is calculated by summing the number of jobs reported across survey rounds (based on the number of Employment Loops completed). Note that for respondents who have already received the PhD prior to the baseline survey, retrospective collection of data on jobs held prior to the baseline survey will be required in order to calculate a complete measure of jobs held since receipt of PhD.

Table E4 Proposed Measures for a Longitudinal NSCG: **Count of Transitions**

Estimates/Measures	Topics	Calculation of the Measure
Number of Jobs Worked Number of employers since receipt of PhD.	Employment	This measure is calculated as the number of employers worked for since receiving the PhD. The measure is calculated by summing the number of unique Employers/NSCG-A9 for all jobs reported across survey rounds. Dependent interviewing will be used to collect employment data between survey rounds. Note that for respondents who have already received the PhD prior to the baseline survey, retrospective collection of data on jobs held prior to the baseline survey will be required in order to calculate a complete measure of jobs held since receipt of PhD.
Number of Job Sectors Worked In Number of job sectors worked in since receipt of the PhD.	Employment	This measure reflects the number of job sections that individuals worked in. The measure is calculated based on summing the number of unique sectors worked in based on Sector (NSCG-A13) of the principal job. Dependent interviewing will be used to collect complete information on jobs worked between survey rounds.
Salary Change for Principal Job	Employment	This measure reflects a change in salary between rounds for the principal job. For individuals holding the same principal job across rounds, the measure is calculated based on changes across rounds in Basic Annual Salary/NSCG-A30, Weeks Worked in Year/NSCG-A31 and Hours Worked Per Week/NSCG-A32.

Table D5 Proposed Measures for a Longitudinal NSCG: **Growth & Experiences**

Estimates/Measures	Topics	Calculation of the Measure
Change of Total Earned Income	Employment	This measure reflects a change in total earned income between survey rounds. The measure is calculated based on changes in Total Earned Income/NSCG-A38 between survey rounds.

Note: Survey item numbers were drawn from the NSCG 2019 survey for new respondents and the 2019 SDR.

Appendix F. Employment History Flowcharts

Figure F1 Baseline Flowchart

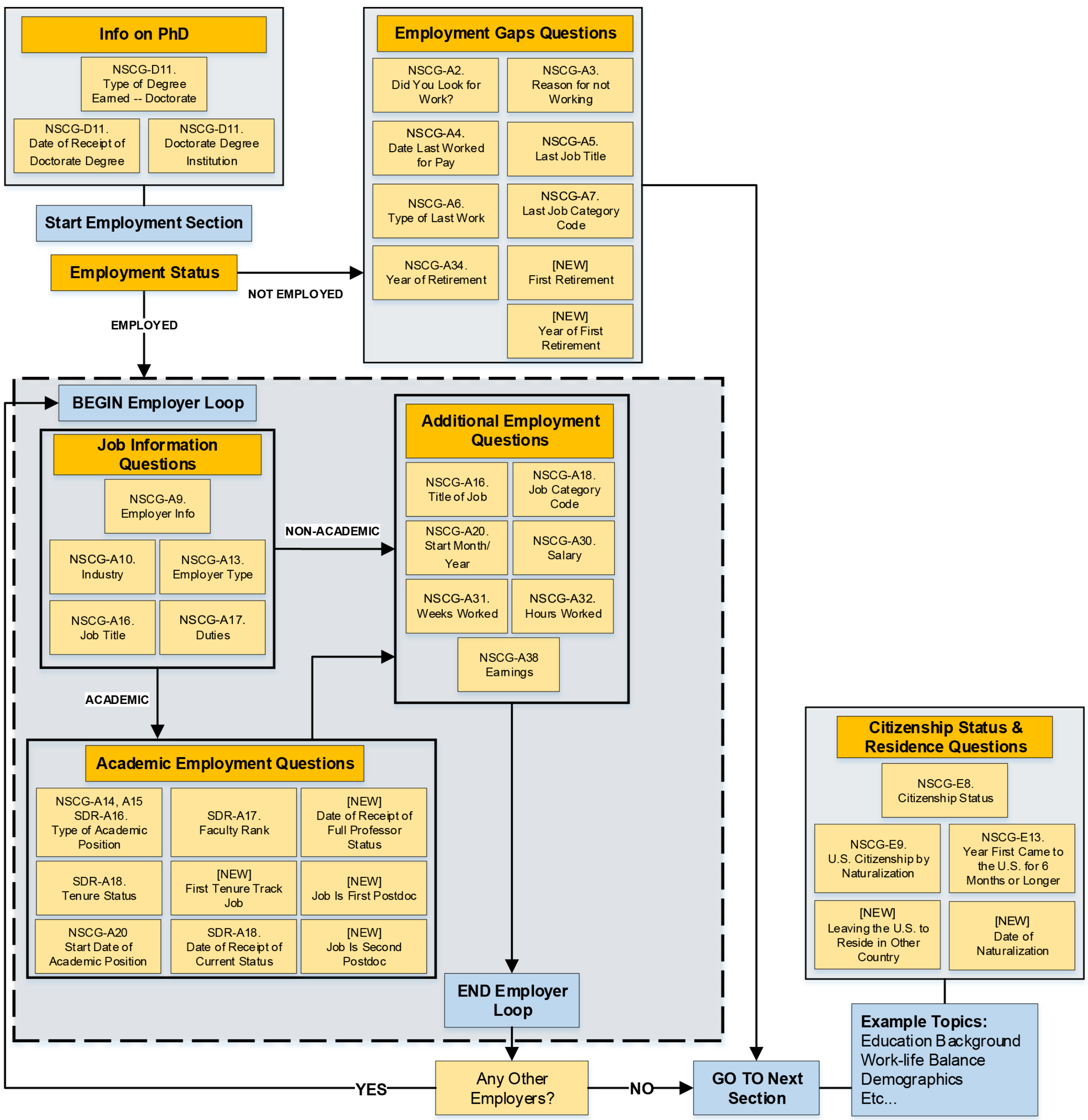


Figure F2 Follow-up Flowchart