Survey Description

1. Survey Overview (2015 Cycle)

a. *Purpose:* The Survey of Doctorate Recipients (SDR), sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation, provides data on the characteristics of science, engineering, and health (SEH) doctorate degree holders. It samples individuals who have earned an SEH research doctoral degree from a U.S. academic institution and are less than 76 years of age. By surveying SEH doctorate degree holders, the SDR provides data useful in assessing the characteristics and supply of the nation's SEH doctorates employed in educational institutions, private industry, and professional organizations, as well as federal, state, and local governments.

b. *Data collection authority:* The information collected in the SDR is solicited under the authority of the National Science Foundation Act of 1950, as amended, the Confidential Information Protection and Statistical Efficiency Act of 2002, and the America COMPETES Reauthorization Act of 2010. The Office of Management and Budget control number is 3145-0020 and expires on 31 August 2018.

c. *Major changes to the recent cycle:* The SDR significantly increased the size of the 2015 sample to 120,000 individuals from approximately 47,000 individuals in the 2013 cycle. This sample size increase was designed to significantly improve the survey's ability to examine employment characteristics for fields of study that aligns with degree fields reported in the Survey of Earned Doctorates (SED). The 2015 SDR expansion required the selection of a new sample drawn from the Doctorate Records File (DRF). The DRF includes information on all research doctorates earned from U.S. institutions and serves as the sampling frame for the SDR. The annual SED, in place since 1957, is used to update the DRF with new cohorts of U.S.-trained research doctoral graduates in all fields.

   The overarching 2015 SDR sample design objectives were twofold:

   - The expanded SDR is required to produce reliable estimates of employment outcomes by the fine field of degree (FFOD) taxonomy used in the SED; and

   - The expanded sample is expected to maintain the existing estimation capabilities associated with analytical domains defined by various demographic characteristics and currently used in NCSES publications such as Science and Engineering Indicators, Women, Minorities and People with Disabilities in Science and Engineering, and in the published SDR data tables.

   The eligibility rules remain unchanged since the 2010 cycle where the target population covers all SEH doctorates from U.S. academic institutions regardless of whether they reside in the U.S. or abroad.

   The new sample design improves the coverage of the SDR to include all internationally residing U.S.-trained SEH doctorate holders.

For additional information about the new estimation requirements for the 2015 SDR sample design, please contact the SDR Project Officer listed below.

There were no changes to the 2015 SDR survey content relative to the 2013 SDR.

2. Key Survey Information

a. *Frequency*: Biennial.
b. *Initial year of survey*: 1973.
c. *Reference period:* The week of February 1, 2015.
d. *Response unit*: Individuals with an SEH research doctorate degree.
e. *Sample or census*: Sample.
f. *Population size*: Approximately 1,047,900 individuals; 920,100 residing in the U.S. and 127,800 residing outside the U.S.
g. *Sample size*: 120,000 individuals.
h. Key Variables:

- Age
- Citizenship status
- Country of birth
- Country of citizenship
- Disability status
- Educational history (for each degree held: field, level, and when received)
- Employer information (e.g., type, size, geographic location)
- Employment benefits (e.g., health insurance, profit sharing)
- Employment status (e.g., unemployed, employed full time, or employed part time)
- Faculty rank for academically employed
- Immigration status (e.g., year of entry, type of current visa)
- Labor force status
- Marital status
- Number of children
- Occupation (current or past job)
- Postdoc status and reasons for taking a postdoc
- Primary work activity (e.g., teaching, basic research)
- Race and ethnicity
- Retirement status and year
- Salary
- Satisfaction and importance of various aspects of job
- School enrollment status
- Sector of employment (e.g., academia, industry, government)
- Sex
- Tenure status for academically employed
- Work-related training

3. Survey Design

a. *Target population:* The SDR target population includes individuals that meet the following criteria:
   - Earned an SEH research doctorate degree from a U.S. academic institution prior to 1 July 2013
   - Are not institutionalized or terminally ill on 1 February 2015
   - Are less than 76 years if age as of 1 February 2015

b. *Sample frame:* The Doctorate Records File (DRF) constructed from the annual Survey of Earned Doctorates (SED).

c. *Sample design:* The SDR uses a fixed panel design with a sample of new doctoral graduates added to the panel in each biennial survey cycle up until the 2013 SDR survey cycle. The refreshed sample for the 2015 SDR has a stratified sample design, where the strata are defined by 215 fields of study listed in the 2013 SED. The new SDR sample includes an oversample of the following groups:
   - Individuals included in the 2013 SDR
   - Underrepresented minorities in the doctorate population
   - Women

   The targeted oversampling was implemented to continue supporting researchers who use SDR data to conduct longitudinal studies, and to improve the precision of estimates for women and minorities within the new sampling strata.

4. Data Collection and Data Processing

a. *Data collection:* The SDR uses a trimodal data collection approach: self-administered questionnaire (via mail), self-administered online survey, and computer-assisted telephone interview (CATI). Sample members are started in one mode depending on their past preference and their available contact information; most sample members were initially asked to participate in the online mode. At any time during data collection, sample members can choose to complete the survey using any of the three modes. Nonrespondents to the initial survey invitation were followed up with alternate survey modes. In total, the 2015 SDR data collection effort lasted approximately nine months.

b. *Data processing:* The data collected in the SDR are subject to both editing and imputation procedures. The SDR uses both logical imputation and statistical (hot deck) imputation as part of the data processing effort.

c. *Estimation techniques:* Since the SDR is based on a complex sampling design and subject to nonresponse bias, sampling weights are created for each respondent to support unbiased population estimates. The final analysis weights were calculated to account for:

- Differential sampling rates
- Adjustments for unknown eligibility
- Adjustments for nonresponse
- Adjustments to align the sample distribution with the DRF distribution with respect to gender, race and ethnicity, degree year, and degree field

The final sample weights enable data users to derive survey-based estimates of the SDR target population. The variable name on the SDR public use data files for the SDR final sample weight is WTSURVY.

5. Survey Quality Measures

a. *Sampling error:* The SDR is subject to sampling error. Estimates of sampling errors associated with this survey were calculated using the replicate weights included with the data file. Please contact the SDR Project Officer to obtain instructions for utilizing the replicate weights to calculate standard errors.

b. *Coverage error:* The concept of coverage in the survey sampling process is the extent to which the known population that is deemed eligible for sample selection (i.e., the sampling frame) "covers" the survey's target population. Any missed doctoral graduates within the DRF derived from the SED, which is a census survey of all research doctoral graduates awarded annually in the U.S., would create undercoverage in the SDR. Additional undercoverage errors may exist because of self-reporting errors in the SED that led to incorrect classification of individuals as not earning an SEH degree when in fact they held such a degree. The potential for overcoverage due to self-reporting errors in the SED is minimized by comparing and evaluating the SED SEH reported fields against the subsequent SDR reported information that indicates a non-SEH degree holder and thus ineligibility for the survey.

c. *Nonresponse error:* The weighted response rate for the 2015 SDR was 66%; the unweighted response rate was 68%. Results from the research and analysis of SDR nonresponse trends have been used in the development of the nonresponse weighting adjustments to minimize the potential for nonresponse bias in the SDR estimates.

   The SDR item nonresponse rate for key employment items, such as employment status, sector of employment, and primary work activity, ranged from 0.0% to 3.86%. Other variables, especially those involving sensitive information, had higher nonresponse rates. For example, salary and earned income had item nonresponse rates of approximately 12% to 19%. A hot deck imputation method was used to compensate for the item nonresponse.

d. *Measurement error:* The SDR is a survey of individuals and thus subject to reporting errors from differences in interpretation of questions. It is also true for any multimodal survey (web, mail, CATI) that some measurement errors will differ by modality. To reduce measurement errors, the SDR questionnaire items were pretested in focus groups and cognitive interviews.

   Please contact the SDR Project Officer to obtain further information on SDR survey evaluation research.

6. Data Availability and Comparability

a. *Data availability:* Data from 1993 to present are available through NCSES's SDR website at https: https://www.nsf.gov/statistics/srvydoctoratework/. Please contact the SDR Project Officer for more information on historical data.

b. *Data comparability:* Year-to-year comparisons can be made among the 1993 to 2015 SDR survey years because many of the core questions remained the same across these survey cycles. Small but notable differences exist across some survey years, however, such as the collection of occupation data being based on the different versions of the occupation taxonomy. Also, due to the reference month differences in some survey cycles, seasonal differences may occur when making comparisons across years and decades. As a result, use caution when interpreting across-year and across-decade comparisons. Also, the definition of the SDR survey target population has experienced the following changes over time:

- Surveys conducted before 1991 included individuals who received doctoral degrees in fields other than SEH and individuals who received their degrees from non-U.S. institutions;
- Surveys conducted prior to 2010 did not cover SEH doctorates residing outside of the U.S.
- Since 2010, coverage of SEH doctorates residing outside of the U.S. only included those having graduated since 2001.
- The 2015 SDR refreshed sample improved population coverage to include all SEH doctorates awarded by U.S. institutions regardless of the academic year of award or the graduate's post-graduation residency location. Thus, the 13% increase in the estimated total population of U.S.-trained doctorate recipients residing worldwide from 2013 to 2015 may be due to both an increase in the number of new doctoral recipients from academic years 2012 and 2013 compared to prior years, as well as the improved survey coverage.

Caution is recommended when considering any analysis of trends that span pre- and post-1991 surveys, pre- and post-2010 surveys, and pre-and post- 2015 surveys given the changes in the survey design and target population.

There is overlap in the cases included in the SDR which allows for the ability to conduct longitudinal analysis of the SDR sample. To link cases on the SDR public use data files across survey years, use the unique identification variable REFID.

7. Data Products

a. *Publications:* Data from the SDR are published in NCSES InfoBriefs and data tables, available at https://www.nsf.gov/statistics/srvydoctoratework/.

Information from this survey is also included in Science and Engineering Indicators and Women, Minorities, and Persons with Disabilities in Science and Engineering.

b. *Electronic access:* The SDR public use data are available in the SESTAT data tool and in public use downloadable files available through the NCSES data download page. The SDR restricted use data for researchers interested in analyzing microdata can be arranged through an NCSES restricted-use licensing agreement.

8. Contact Information

Daniel Foley
Project Officer
Human Resources Statistics Program
National Center for Science and Engineering Statistics
Phone: (703) 292-7811