# NSCG Estimation Issues when Using an ACS-based Sampling Frame

John Finamore[1], Steve Cohen[1], David Hall[2], Julie Walker[2], Donsig Jang[3]
[1]National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230
[2]U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233
[3]Mathematica Policy Research, 600 Maryland Ave., SW, Washington, DC 20024

## Abstract
The National Survey of College Graduates (NSCG) is the nation's only source of detailed statistics on the science and engineering labor force. Historically, the NSCG selected its sample once a decade from the decennial census long form respondents. In the 2010 NSCG survey cycle, the NSCG began using the American Community Survey (ACS) as the sampling frame for the NSCG. After considering numerous sample design options proposed by the NSCG survey sponsor, the National Science Foundation (NSF), and reviewed by the Committee on National Statistics (CNSTAT), the NSF approved the use of a rotating panel design for the 2010 decade of the NSCG. This rotating panel design allows the NSCG to address certain deficiencies of the previous long form-based design including the undercoverage of key interest groups. However, along with numerous improvements, the use of the ACS as a sampling frame for the NSCG and the implementation of the NSCG rotating panel design also introduced new challenges. This document summarizes the rotating panel design planned for the 2010 decade of the NSCG and discusses results from two research tasks related to NSCG estimation – The derivation of a simplified ACS final weight for use in the NSCG sampling effort and the recommended approach to derive NSCG estimates from the multiple panels included in each NSCG survey cycle.

**Key Words:** NSCG, ACS, Multiple Panel Estimation

## 1. Introduction and Background[3]

The National Survey of College Graduates (NSCG) is a longitudinal survey conducted every two to three years by the U.S. Census Bureau on behalf of the survey sponsor, the National Science Foundation (NSF). The NSCG is the largest of three surveys that combine to form the Scientists and Engineers Statistical Data System (SESTAT). The other two surveys are the National Survey of Recent College Graduates (NSRCG) and the Survey of Doctorate Recipients (SDR). SESTAT is a comprehensive and integrated system of information about the employment, educational, and demographic characteristics of the science and engineering population in the United States. The integrated data from these three surveys serve as the basis for the development of national estimates on the science and engineering (S&E) workforce.

[3] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the National Science Foundation, the U.S. Census Bureau, or Mathematica Policy Research, Inc.

Traditionally, the NSCG has selected its sample from the decennial census long form respondents. The long form was a large frame from which to select a sample (approximately 1 in 6 households in the United States, including Puerto Rico and the outlying areas[4]) and provided a wealth of information for sampling purposes that allowed the NSF to better identify the NSCG target population of college graduates residing in the U.S. However, the long form based-design for the NSCG did have certain drawbacks including the inability to efficiently sample the science and engineering (S&E) population and the increasing undercoverage of key interest groups throughout the decade.

In 2010, the Census Bureau discontinued the use of the decennial census long form. All U.S. residents received what had been known as the decennial census short form as part of the 2010 Census. Unfortunately for the NSCG, the short form does not collect information on educational attainment, occupation, and other variables used in the NSCG stratification and sample selection processing. As a result, the only practical alternative frame identified from which to select a new 2010 NSCG sample is the 2009 American Community Survey (ACS).

After considering numerous sample design options proposed by the NSF and reviewed by the Committee on National Statistics (CNSTAT), the NSF approved the use of a rotating panel design for the 2010 decade of the NSCG (National Research Council (2008)). The use of the ACS-based sampling frame within this rotating panel design will allow the NSCG to address the sampling efficiency and undercoverage issues that were associated with the long form-based design for the NSCG. However, along with these and other improvements, the use of the ACS as a sampling frame for the NSCG and the implementation of the NSCG rotating panel design also introduced new challenges.

This document summarizes the rotating panel design planned for the 2010 decade of the NSCG and discusses results from two research tasks related to NSCG estimation – the derivation of a simplified ACS final weight for use in the NSCG sampling effort and the recommended approach to derive NSCG estimates from the multiple panels included in each NSCG survey cycle. The incorporation of these two estimation research tasks will allow the NSCG to take advantage of the improvements introduced by the use of the ACS-based sampling frame.

## 2. The Impact of Using the ACS as a Sampling Frame

### 2.1 Increased Sampling Efficiency

The ACS is an ongoing survey conducted monthly by the U.S. Census Bureau. Similar to the decennial census long form, the ACS collects demographic, social, economic, and housing data from across the fifty states, the District of Columbia, and Puerto Rico.[5] It samples approximately three million households per year and completes an interview with nearly 2 million households. With the average U.S. household having about 2.6 persons, the ACS obtains complete interviews with over 4.5 million U.S. residents on an annual basis. When the NSCG degree, age, and residence eligibility criteria are taken into account, approximately 850,000 persons from each year of the ACS are eligible for

[4] The outlying areas include American Samoa, Guam, Northern Mariana Islands, and the U.S. Virgin Islands.

[5] Unlike the decennial census, the ACS does not sample from the outlying areas.

selection into the NSCG. The questionnaire items included on the ACS are similar to what had been included on the decennial census long form. However, there is one noticeable questionnaire difference that has an immense impact on the NSCG sampling efficiency.

Beginning in the 2009 data collection year, an item was added to the ACS questionnaire asking for a respondent's field of degree for any *bachelor's degree* they may have earned (for those that reported having earned at least a bachelor's degree). The field of degree questionnaire item allowed for more efficient sampling of the S&E population from the ACS-based sampling frame. When using the long form-based sampling frame for the NSCG, the only information available to identify the S&E population was occupation. Since the NSF defines the S&E population by both their degree history and current occupation, the long form did not allow complete identification of the S&E population within the college graduates universe.

By collecting the field of degree information for any bachelor's degree, the ACS provides the NSCG with the ability to use a person's degree history (for any bachelor's degree) and occupation in order to determine whether they meet the S&E population definition. Since the ACS does not provide degree field information for advanced degrees[6] (e.g., master's or doctorate degrees), there is still room for sampling efficiency improvement associated with this desire to identify the S&E population. However, the ACS's ability to identify field of degree information for bachelor's degree is an improvement over the long form-based design for the NSCG and will likely result in the 2010 decade of the NSCG needing to sample almost 40,000 fewer cases than in the 2000 decade (130,000 vs. 177,000) while achieving similar estimation accuracy.

## 2.2 Improved Coverage

The decennial census long form occurred at the beginning of every decade. By using the long form as its sampling frame, the NSCG suffered from increasing undercoverage of certain population groups over the course of the decade, including recent college graduates and immigrants into the U.S. To provide coverage of the recent college graduates population, the NSRCG was included into the SESTAT design. This survey provided coverage of all recent college graduates from U.S. educational institutions that earned their degree after the long form reference date.

Unfortunately, in both the 1990 and 2000 decade, there was no sampling frame readily available to address the undercoverage of college graduates immigrating into the U.S. throughout the decade. In the 2010 decade, the ongoing design of the ACS monthly data collection in combination with the rotating panel design planned for the NSCG will allow the 2010 decade of the NSCG to provide coverage of college graduates immigrating into the U.S. throughout the decade.

[6] In previous NSCG survey cycles, we found the population of respondent with a non-S&E bachelor's degree and an S&E advanced degree to be relatively small. In the 2003 NSCG, we estimated this population to be around 5% of the total college educated population.

# 3 Derivation of Simplified ACS Final Weights for the NSCG Sampling

## 3.1 ACS Data Swapping

In the post-data collection processing of the ACS data, the Census Bureau incorporates data swapping techniques to ensure that individually identifiable data will not be released. The selection process for deciding which ACS households should be swapped is highly targeted to affect the records with the most disclosure risk. Pairs of households that are swapped match on a minimal set of demographic variables.

This ACS data swapping occurs after the ACS sample selection and determination of the household-level base weight, but prior to any additional ACS weighting adjustments. As a result of the timing of the data swapping, the ACS final weights are produced only for the swapped data file. It is this ACS swapped data file with ACS final weights that is used for all ACS data products - tables and microdata (U.S. Census Bureau (2009)).

From an NSCG perspective, the problem with this swapping approach for disclosure avoidance is that the swapped data file does not allow for the correct identification of the NSCG eligible sampling frame. As an example, assume that two households that match on county and number of residents are swapped. Also assume that one household includes four college graduates and the other household includes four residents that did not complete college.

If we were to use the ACS swapped file as the sampling frame for the NSCG, the household that does not have any graduates, but appears to have graduates on the ACS swapped file, would be eligible for the NSCG sampling frame. And the household that has graduates, but does not appear to have graduates on the ACS swapped file, would not be eligible. This incorrect identification of eligible sampling frame cases creates a problem in that cases could be chosen for sample who do not meet the target population of the survey.

## 3.2 Using the Unswapped ACS Data File and Producing ACS Final Weights

The Census Bureau maintains a file that allows identification of all swapped household pairs. However, the ACS data swapping cannot be undone by simply using this file to "unswap" all swapped pairs because the ACS weighting adjustments occur after the data swapping. This "swapping before weighting" approach is problematic because the final weight determined for each household (and each resident within a household) is based on the swapped characteristics applied to the household.

Since the correct identification of the eligible sampling frame is a top priority for the NSCG, the decision was made to use the unswapped ACS data file for the NSCG frame construction. However, since the ACS weighting occurs after the data swapping, the unswapped ACS data file only has ACS base weights that reflect the ACS sample selection and do not contain ACS final weights that reflect the ACS weighting adjustments, including the adjustments for the ACS subsampling and ACS nonresponse.

To produce final weights for the unswapped ACS data file, one consideration was to run the complete ACS production weighting methodology on the unswapped data. This approach would have resulted in final weights that could be used for NSCG sample selection purposes. However, since a goal of the ACS is to produce estimates of small geographic areas, the ACS production weighting methodology includes adjustments that

occur at very low levels of geography. The NSCG, on the other hand, provides national estimates for the science and engineering workforce. Since the NSCG emphasis is on national estimation (rather than county or subcounty estimation), the Census Bureau conducted research to explore whether a simplified set of weighting adjustments could be applied to the unswapped ACS data in place of the ACS production weighting methodology without adversely affecting the NSCG sample selection and the resulting NSCG estimates.

## 3.3 ACS Production Weighting and the Simplified Weighting Options

The ACS production weighting methodology includes separate steps for calculating ACS final weights depending on residence type. As listed below, sample persons residing in group quarters (GQ) receive an independent set of weighting adjustments as compared to the weighting adjustments for the sample persons residing in housing units (HU). For more information on the ACS weighting methodology, please see U.S. Census Bureau (2009).

Weighting Adjustments for ACS Sample Records Residing in GQs

- Initial base weight to reflect the probability of selection (BW is the abbreviation for the weight resulting from this adjustment)
- Adjustment to reflect the observed GQ population (FBW)
- Adjustment to reflect GQ nonresponse (WGQNIF)
- Applying the GQ person-level post-stratification factor (WGQPPSF)
- Rounding the final weight (PWT)

Weighting Adjustments for ACS Sample Records Residing in HUs

- Initial base weight to reflect the probability of selection (BW is the abbreviation)
- Adjustment to reflect the ACS subsampling (WSSF)
- Adjustment to reduce variability in the ACS monthly totals (WVMS)
- Adjustment #1 for HU nonresponse - by building type and tract (WNIF1)
- Adjustment #2 for HU nonresponse- by building type and month (WNIF2)
- Adjustment for data collection mode bias (WMBF)
- Rake #1 to person-level population totals (WPPSF – Rake #1)
- Rake #2 to person-level population totals (WPPSF – Rake #2)
- Rake #3 to person-level population totals (WPPSF – Rake #3)
- Rounding the final weight (PWT)

Simplified weighting options at the national level, state level, and division level were considered as part of this research. In addition, there were two national-level options that examined the impact associated with removing some of the ACS weighting steps. The five simplified weighting options include most of the same weighting adjustments included in the ACS production weighting methodology, but these adjustments were implemented at the national level, state level, or division level (which divides the U.S. into 10 areas, with Puerto Rico being its own area). Table 1 summarizes the ACS production weighting methodology and the five simplified options considered in this research.

In the table, a checkmark (✓) implies the full weighting adjustment is implemented as documented in the ACS production weighting methodology. A checkmark with the word national (✓ (national)) implies the weighting adjustment is implemented at the national level rather than at the ACS weighting area level. A checkmark with the word state (✓

(state)) implies the weighting adjustment is implemented at the state level rather than at the ACS weighting area level. A checkmark with the word division (✓ (division)) implies the weighting adjustment is implemented at the division level rather than at the ACS weighting area level. A shaded cell means the weighting adjustment is not included.

**Table 1.** ACS Production Weighting and the Simplified Weighting Options

| Resulting Weight | ACS Production Weighting | Simplified Weighting Options | | | | |
|---|---|---|---|---|---|---|
| | | **Option #1** (national) | **Option #2** (national) | **Option #3** (national) | **Option #4** (state) | **Option #5** (division) |
| **Weighting Adjustments for ACS Sample Records Residing in GQs** | | | | | | |
| BW | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| FBW | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| WGQNIF | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| WGQPPSF | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| **Weighting Adjustments for ACS Sample Records Residing in HUs** | | | | | | |
| BW | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| WSSF | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| WVMS | ✓ | ✓ (national) | ✓ (national) | | ✓ (state) | ✓ (division) |
| WNIF1 | ✓ | | | | | |
| WNIF2 | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| WMBF | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |
| WHPF | ✓ | | | | | |
| WPPSF – Rake #1 | ✓ | ✓ (national) | | | ✓ (state) | ✓ (division) |
| WPPSF – Rake #2 | ✓ | ✓ (national) | | | ✓ (state) | ✓ (division) |
| WPPSF – Rake #3 | ✓ | ✓ (national) | ✓ (national) | ✓ (national) | ✓ (state) | ✓ (division) |

## 3.4 Evaluation Methods

The current ACS weighting methodology and the five simplified weighting methodology options were compared using three main evaluation criteria.

### *3.4.1 Ease and Length of Implementation*

Each weighting option being considered as part of this research has a certain level of difficulty associated with implementation. In addition, each option requires a certain length of time for implementation. Since there was only a limited amount of time available to produce final weights for the 2009 ACS data as part of the 2010 NSCG

sample selection processing, the ease and length of implementation was used as an evaluation criterion. In addition, since the ACS replicate weight file is used in the production of any NSCG replicate weights, all weighting adjustments made in producing the ACS final weights need to be implemented in the production of the ACS final replicate weights. One issue that impacts the ease and length of implementation for all five methods is the collapsing of noninterview adjustment cells that occurs in response to cell size and adjustment factor limits.

### *3.4.2 Estimation Accuracy of NSCG Stratification Variables and Frame Totals*

The NSCG currently forms stratification variables using information derived from the ACS, including race, Hispanic origin, disability, citizenship, occupation, degree type, degree field, sex, and year of entry into the United States.

Estimates for each level of the NSCG stratification variables from the simplified weighting methodology options were compared with estimates from the ACS production weighting methodology. In addition, we cross-classified the NSCG stratification variables to produce the planned NSCG sampling strata and examined the change in the NSCG sampling strata totals using the simplified weighting methodology options.

### *3.4.3 Estimation Accuracy of ACS Weighting Variables*

The simplified weighting methodology options described in this document either reduce or remove some of the weighting adjustments used in the ACS production weighting methodology. Since most of the ACS weighting adjustments are correcting for some type of inaccuracy associated with certain variables, reducing or removing the adjustment may cause the inaccuracy to continue to exist. As an example, the weight after the variation in monthly response adjustment (*WVMS*) reduces variability in the ACS monthly totals. If this adjustment is not included into one of the simplified weighting methodology options, the variation in the monthly totals may continue to exist. As a result, an evaluation criteria included into this research examines the impact the change or omission of any adjustment has on the inaccuracies being corrected in the ACS production weighting methodology.

## 3.5 Selection of the Simplified Weighting Option for Use in the NSCG

After considering the evaluation methods discussed in the previous section, the division-level simplified weighting option was selected for use in the 2010 NSCG processing to produce final weights for all ACS cases.

Evaluation of the simplified weighting options began at the national level, since the NSCG produces national-level estimates. We eliminated simplified weighting option #3 from consideration because dropping the variation in monthly sample adjustment caused the sum of the weights to be incorrect for the subsequent intermediate weights. We then eliminated simplified weighting options #1 and #2 from consideration because these options (along with option #3) produced biased estimates of the Hispanic population.[5]

[5] The biased Hispanic population estimates are the result of the weighting adjustments. In the ACS production weighting methodology, all cases in Puerto Rico are assigned to be Hispanic for certain adjustments. When we did the weighting at the national level, cases from Puerto Rico were handled in the same way as every other state and not forced to be Hispanic. This resulted in an underestimate of the overall Hispanic population when compared to the ACS estimates.

The tables below display the Hispanic estimates being underestimated by the national-level simplified weighting options. Table 2 shows the weighted totals for Hispanic for all simplified weighting options in comparison to the ACS final production weight calculated through the ACS production weighting methodology. Table 3 shows the percent difference in the Hispanic estimates from the simplified weighting option compared to the ACS final production weight. For a comparison of the other demographic variables used as NSCG stratification variables, including age, sex, and race, please see Cohen, et al. (2010).

**Table 2.** Weighted Hispanic Totals

| Weighted Totals | | | | | |
|---|---|---|---|---|---|
| Production | Option #1 | Option #2 | Option #3 | Option #4 | Option #5 |
| 50,838,189 | 46,990,547 | 46,985,164 | 46,985,171 | 50,884,127 | 50,877,575 |

**Table 3.** Percent Difference of Hispanic Estimates Against ACS Production

| Percent Difference from ACS Production Final Weights | | | | |
|---|---|---|---|---|
| Option #1 | Option #2 | Option #3 | Option #4 | Option #5 |
| -7.57% | -7.58% | -7.58% | 0.09% | 0.08% |

Since the national-level simplified weighting options result in biased estimates of the Hispanic population, all three national-level simplified weighting options were deemed inappropriate for use in the NSCG processing. As a result of the national-level options not meeting the NSCG needs, the simplified weighting options with adjustments made at lower levels of geography were pursued.

When the comparison between the state-level and division-level option was made, the division-level option was chosen for use in the 2010 NSCG processing. Advantages of the division-level simplified weighting option are as follows:

*Ease and Length of Implementation* - The use of the division-level option and the national-level options do not require any collapsing of noninterview adjustment cells. The state-level option and the ACS production weighting methodology require collapsing of numerous noninterview adjustment cells.

*Estimation Accuracy of NSCG Stratification Variables and Sampling Frame Totals* - In comparing all the options, both the state-level and division-level options produce estimates of frame variables that are closer in value to the production ACS weighted estimates than the other three simplified weighting options. The three national options severely underestimate Hispanic estimates (see Tables 2 and 3).

*Estimation Accuracy of ACS Weighting Variables* - Removing the variation in monthly sampling adjustment, which was done in Simplified Option #3, shows negative consequences on the weighted totals of intermediate weights. For example, the sum of the weights for the intermediate weights produced in option #3 exceeds the sum of the same intermediate weights in the ACS production weighting methodology.

The sum of weights for the intermediate weights produced in the other simplified weighting options does not vary from the sum of the same weights in the ACS production weighting methodology.

*Reduced Weight Variation* - All of the options produced weights that are less variable than the production ACS weights. Table 4 shows the weight distribution information for the housing unit final weight variable in the ACS production weighting methodology and in the five simplified weighting options.

**Table 4.** HU Final Weight Distribution Information

| HU Final Weight | Minimum | Median | Maximum | Coefficient of Variation | Design Effect of Weights |
|---|---|---|---|---|---|
| Final | 1 | 56 | 1,601 | 72.5377 | 1.52617 |
| Option #1 | 6 | 55 | 589 | 68.1849 | 1.46492 |
| Option #2 | 7 | 55 | 583 | 66.6938 | 1.44481 |
| Option #3 | 7 | 55 | 612 | 66.4665 | 1.44178 |
| Option #4 | 3 | 55 | 796 | 69.5050 | 1.48310 |
| Option #5 | 5 | 54 | 679 | 69.1279 | 1.47787 |

*Relationship Between Sampling and Weighting* - Division is being used as a sorting variable in the sample selection effort. Choosing the division-level weighting option allows for linkage between the weighting and the sample selection.

*Future Production of Subnational Estimates* - There has been discussion at NSF that a future goal may be to produce NSCG estimates below the national level. It is unlikely the estimates will be produced at the state level, but division-level estimation could be explored.

In conclusion, the division-level simplified weighting option was determined to be the best approach out of the five options for its ease of implementation and that it does not severely overestimate or underestimate any of the NSCG stratification variables and frame totals.

### 3.6 Limitations

The simplified weighting research described in this report was conducted on the 2008 ACS data files that were adjusted for disclosure avoidance (DA) purposes. When the simplified weighting methodology is implemented as part of the 2010 NSCG processing, it will be implemented on the 2009 ACS pre-disclosure avoidance data. The research described in this document assumes the findings derived from the 2008 ACS DA files would not differ dramatically had the research been conducted on the 2009 ACS pre-DA files. As a way to examine this assumption, versions of the tallies in Cohen, et al. (2010) were created when the simplified weighting methodology was implemented on the 2009 ACS pre-DA files during the 2010 NSCG processing. The tallies produced as part of this verification step were very similar to the tallies included in Cohen, et al. (2010).

## 4. Multiple Panel Estimation in the NSCG Rotating Panel Sample Design

In 2010, the NSCG began using the ACS sample as its sampling frame. After considering numerous sample design options proposed by the NSCG survey sponsor, the NSF, and reviewed by the CNSTAT, the NSF approved the use of a rotating panel design for the 2010 decade of the NSCG. This rotating panel design allows the NSCG to address certain deficiencies of the previous design, including the undercoverage of key interest groups.

This section describes how the rotating panel design will be implemented over the course of the 2010 decade.

## 4.1.Rotating Panel Design for the 2010 Decade of the NSCG

In the rotating panel design planned for the 2010 decade of the NSCG, the planned selected sample size will be 130,000 cases over the course of the four-panel design. Due to nonresponse and sample attrition, the expected realized responses per survey cycle (i.e., the completed number of interviews per cycle) will likely be in the 84,000 case range. This expected realized responses value is based on an assumed response rate of 75 percent in the initial survey cycle and 90 percent in the subsequent survey cycles.

## 4.2.Panel Rotation

In the 2010 NSCG survey cycle, 65,000 cases will be selected from the 2009 ACS (the remaining cases in the 2010 NSCG will be carried over from the 2008 NSCG and the 2008 NSRCG. In the 2012 NSCG survey cycle, 65,000 more cases will be selected from the 2011 ACS. The 2014 survey cycle will be the first cycle to implement the rotating in and rotating out feature of the rotating panel design with one-quarter of the on-going sample being replaced by new sample from the 2013 ACS.

To reach the four-panel design desired for the 2010 decade of the NSCG, we considered two main options for rotating panels to arrive at the four-panel design:

Option #1 – Drop 32,500 cases from the 2009 ACS prior to the 2014 NSCG and drop 32,500 cases from the 2011 ACS prior to the 2016 NSCG.

Option #2 – Drop 32,500 cases from the 2009 ACS prior to the 2014 NSCG and drop the remaining 32,500 cases from the 2009 ACS cases prior to the 2016 NSCG.

Each option has its advantages and disadvantages. Option #1 would provide more data for use in the NSCG longitudinal files[6]. And with this option, the complete four-panel rotation panel design with one panel from each ACS year will be in place starting from 2016. However, the inclusion of older sample tends to result in a higher attrition rate. At this point in our processing, we recommend Option #2 because it removes the oldest cases out of the survey sooner than Option #1 and, as a result, reduces the impact attrition will have on the NSCG survey estimates. This option, however, will delay the complete four-frame rotation panel design to 2018.

## 4.3.Coverage Scope in the NSCG Rotating Panel Design

Each new sample in the 2010 decade of the NSCG will provide coverage of the three S&E workforce population subgroups:

- S&E and S&E-related degree cases
- S&E and S&E-related occupation only cases
- Non-S&E Cases

[6] While the details of the longitudinal files are still in the planning stage, the general concept is that each survey cycle will include a longitudinal file for each panel that allows data users the opportunities to view the response history for sample cases.

In addition, the new sample for each NSCG survey cycle in the 2010 decade will provide coverage of the three coverage population subgroups:

- NSCG target population as of the previous survey cycle
- Immigrant cases since the previous survey cycle
- Recent college graduates since the previous survey cycle

## 4.4.Individual Panel Weights and Combined Panel Weights

For each NSCG survey cycle in the 2010 decade, the estimation goal is to produce both individual panel estimates and combined panel estimates. The individual panel estimates will allow evaluation of the NSCG target population characteristics using only the cases from one panel[7]. On the other hand, the combined panel estimates will allow evaluation of the NSCG target population using the cases from **all** the panels in sample.

The individual panel estimates will be derived by using the following weighting steps to determine the individual panel final weights:

- Determination of individual panel base weights through the sample selection procedure
- Weighting adjustment to account for unit nonresponse
- Implementation of an iterative raking procedure
- Identification of and adjustment for extreme weight values
- Construction of individual panel final weights through an additional implementation of an iterative raking procedure.

Once the individual panel final weights are derived, the combined panel weights for the particular survey cycle will be derived through a two-step weighting process:

- Determination of a combined panel weight adjustment factor
- Calculation of combined panel weights using the methods discussed in this section
- Implementation of an iterative raking procedure

## 4.5.Multiple Panel Estimation

This section discusses the topic of multiple panel estimation working under the recommendation to carry forward the individual panel weights across survey cycles. For the sections that follow, we use the following definitions and notation:

- Frame A is the 2008 NSCG and 2008 NSRCG frame.
- Frame B is the 2009 ACS frame.
- Frame C is the 2011 ACS frame.
- Frame D is the 2013 ACS frame.
- Frame E is the 2015 ACS frame.

[7] Please note that each panel within the rotating panel design covers a slightly different population. The older panels within the design are a subset of the more recent panels. To that end, while each panel will be able to produce individual panel estimates, the estimates for each individual panel will cover a slightly different population. Only the most recent panel within a given survey cycle will produce individual panel estimates that cover the same population as the combined panel estimates.

- Frame F is the 2017 ACS frame.
- $S(I)$ denotes the probability sample from frame $I$, $I = A,\ldots,F$.
- Unit $i$ in sample $S(I)$ has the base weight $w_i^I$, $I = A,\ldots,F$.

There is overlap among the sampling frames in each of the NSCG survey cycles. For notational simplicity, we define the overlapping domain by the number of frames included. For example, in the 2010 NSCG, domain *ab* denotes the overlapping of frame A and frame B; domain b denotes the component only covered in frame B, not overlapping frame A. Figure 1 below attempts to summarize the domain notation by survey year in the 2010 decade of the NSCG.

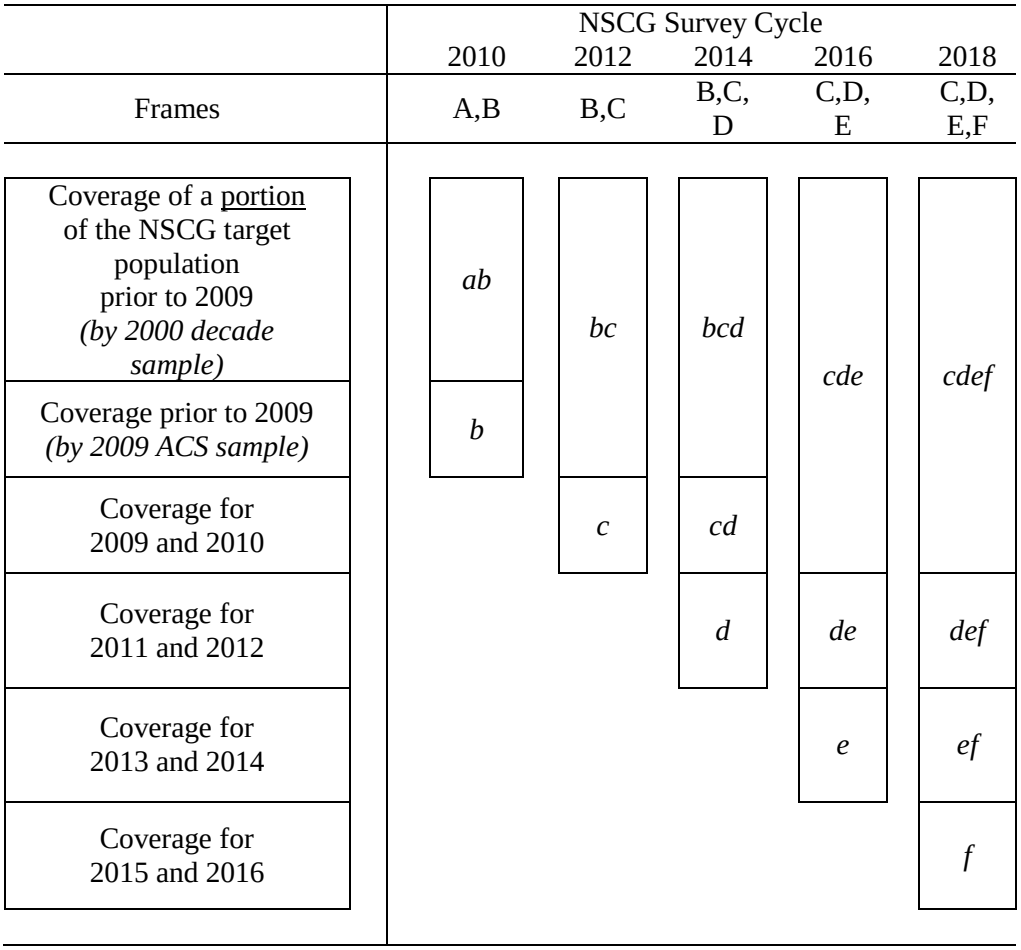


**Figure 1:** Frames and Domains for the 2010 Decade of the NSCG

## 4.6.Multiple Panel Estimation for the 2010 – 2012 NSCG

At the sample case level, the individual panel weight, $\text{ipw}_i^I$, I = A, B, …, F, reflects the inverse of probability selection and the other weighting adjustments discussed at the beginning of this section that could be applied either at or below the panel level.

We propose to use a single-weight estimation method to estimate the population total $Y$ as follows:

$$\hat{Y}^{2010} = \frac{1}{2}\hat{Y}^{A}_{ab} + \frac{1}{2}\hat{Y}^{B}_{ab} + \hat{Y}^{B}_{b} \qquad \text{(For the 2010 NSCG)}$$

$$\hat{Y}^{2012} = \frac{1}{2}\hat{Y}^{B}_{bc} + \frac{1}{2}\hat{Y}^{C}_{bc} + \hat{Y}^{C}_{c} \qquad \text{(For the 2012 NSCG)}$$

This is also the Mecatti estimator (Mecatti (2007)), where the individual panel weight totals for the 2010 NSCG are as follows:

$$\hat{Y}^{A}_{ab} = \sum_{i \in S(A) \cap ab} ipw_i^A \, y_i \,,\ \hat{Y}^{B}_{ab} = \sum_{i \in S(B) \cap ab} ipw_i^B \, y_i \,, \text{ and } \hat{Y}^{B}_{b} = \sum_{i \in S(B) \cap b} ipw_i^B \, y_i$$

And the individual panel base weight totals for the 2012 NSCG are:

$$\hat{Y}^{B}_{bc} = \sum_{i \in S(B) \cap bc} ipw_i^B \, y_i \,,\ \hat{Y}^{C}_{bc} = \sum_{i \in S(C) \cap bc} ipw_i^C \, y_i \,, \text{ and } \hat{Y}^{C}_{c} = \sum_{i \in S(C) \cap c} ipw_i^C \, y_i$$

This proposed estimation can be rewritten using combined panel weight ($cpw_i$) with the prerequisite of the individual panel weight ( $ipw_i^I$ ) as:

$$cpw_i^{2010} = \begin{cases} \dfrac{ipw_i^I}{2} \left( = \dfrac{ipw_i^A}{2} \text{ or } \dfrac{ipw_i^B}{2} \right) & \text{if } i \in (A \cup B) \cap ab \\ ipw_i^I = ipw_i^B & \text{if } i \in B \cap b \end{cases} \tag{1}$$

$$cpw_i^{2012} = \begin{cases} \dfrac{ipw_i^I}{2} \left( = \dfrac{ipw_i^B}{2} \text{ or } \dfrac{ipw_i^C}{2} \right) & \text{if } i \in (B \cup C) \cap bc \\ ipw_i^I = ipw_i^C & \text{if } i \in C \cap c \end{cases} \tag{2}$$

Then a population total estimator for variable y for 2012 NSCG can be expressed as:
$\hat{Y} = \sum_{i \in S(B) \cup S(C)} cpw_i^{2012} \, y_i$ .

## 4.7.Multiple Panel Estimation for the 2014 – 2018 NSCG

As we stated, the individual panel weight from the previous survey cycle will be carried forward and used as the initial weight in subsequent survey cycles. Assume the individual panel weight is panel weight, $ipw_i^I$ , *I = A, B, …, F*. Under this assumption, and using rotating panel option #2 from section 4.2, the combined panel weights for the 2014 NSCG, 2016 NSCG, and 2018 NSCG are as follows:

$$cpw_i^{2014} = \begin{cases} \frac{ipw_i^I}{4} \quad \left( = \frac{ipw_i^B}{4} \text{ or } \frac{ipw_i^D}{4} \right) & \text{if } i \in (B \cup D) \cap bcd \\ \frac{2 \times ipw_i^I}{4} = \frac{ipw_i^I}{2} = \frac{ipw_i^C}{2} & \text{if } i \in C \cap bcd \\ \frac{2 \times ipw_i^I}{3} = \frac{2 \times ipw_i^C}{3} & \text{if } i \in C \cap cd \\ \frac{ipw_i^I}{3} = \frac{ipw_i^D}{3} & \text{if } i \in D \cap cd \\ ipw_i^I = ipw_i^D & \text{if } i \in D \cap d \end{cases} \tag{3}$$

$$cpw_i^{2016} = \begin{cases} \frac{ipw_i^I}{4} \quad \left( = \frac{ipw_i^D}{4} \text{ or } \frac{ipw_i^E}{4} \right) & \text{if } i \in (D \cup E) \cap cde \\ \frac{2 \times ipw_i^I}{4} = \frac{ipw_i^I}{2} = \frac{ipw_i^C}{2} & \text{if } i \in C \cap cde \\ \frac{ipw_i^I}{2} \quad \left( = \frac{ipw_i^D}{2} \text{ or } \frac{ipw_i^E}{2} \right) & \text{if } i \in (D \cup E) \cap de \\ ipw_i^I = ipw_i^E & \text{if } i \in E \cap e \end{cases} \tag{4}$$

$$cpw_i^{2018} = \begin{cases} \frac{ipw_{i,cdef}}{4} \quad \left( = \frac{w_i^C}{4} \text{ or } \frac{w_i^D}{4} \text{ or } \frac{w_i^E}{4} \text{ or } \frac{w_i^F}{4} \right) & \text{if } i \in (CUD \cup EUF) \cap cdef \\ \frac{ipw_{i,def}}{3} \quad \left( = \frac{w_i^D}{3} \text{ or } \frac{w_i^E}{3} \text{ or } \frac{w_i^F}{3} \right) & \text{if } i \in (D \cup E \cup F) \cap def \\ \frac{ipw_{i,ef}}{2} \quad \left( = \frac{w_i^E}{2} \text{ or } \frac{w_i^F}{2} \right) & \text{if } i \in (E \cup F) \cap ef \\ ipw_{i,f} = w_i^F & \text{if } i \in F \cap f \end{cases} \tag{5}$$

Please note that the calculation of the combined panel weights for the NSCG survey cycles after 2018 will use the same formula presented for the 2018 NSCG with changes in the notation to account for the different sampling frames.

## 4.8.Additional Weight Adjustment Considerations

In the previous sections, we described the weighting steps tentatively planned for use in deriving the individual panel weights and the combined panel weights. One issue to keep in mind when considering these weighting steps is that the noninterview adjustment discussed as part of the individual panel weight calculation could be performed within each panel or within each cohort. The cohort level adjustment would likely reduce the number of small sample cells, but the drawback is that it could ignore different response propensities across panels. To investigate this issue, we plan to include the panel identifier as an independent variable in a model examining response propensity. We will determine how to implement the noninterview adjustment based on the model results. The 2014 NSCG is the first survey cycle that will need to address this issue. As a result, when 2012 NSCG data is available, we will begin research to examine this issue.

A second issue to keep in mind is that we tentatively plan to perform raking/benchmarking by adjusting weights to the ACS population control total for key demographics. However, this plan is under the assumption that we will be confident with the concordance between ACS and NSCG responses for the key demographics. Research to examine the ACS and NSCG concordance and to determine the level at which the ACS controls should be used (panel, survey cycle, etc) will be conducted in the future. As part of this research, the research will examine ACS item nonresponse rates and ACS response error issues.

## References

Cohen, Steve, et al. (2010). "Multiple Panel Estimation in the National Survey of College Graduates," 2010 NSCG Statistical Evaluation Committee.

Hall, David and Julie Walker. (2010). "2010 NSCG Research – Evaluating the Calculation of Final Weights for the American Community Survey (Document # NSCG10-WT-3)," Internal Census Bureau Memorandum.

Mecatti, Fulvia (2007) "A Single Frame Multiplicity Estimator for Multiple Frame Surveys," *Survey Methodology Vol22, No 2. pp151-157.*

National Research Council. (2008). *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs*. Panel on Assessing the Benefits of the American Community Survey for the NSF Division of Science Resources Statistics, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

U.S. Census Bureau. (2009). *Design and Methodology, American Community Survey*. Washington, DC: U.S. Government Printing Office. (http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf)