Evaluating the American Community Survey as a Potential Sampling Frame for the National Survey of College Graduates

John Finamore and David Hall
U.S. Census Bureau, Washington, DC 20233

Abstract

The National Survey of College Graduates (NSCG) is a longitudinal survey that collects information on employment, educational, and demographic characteristics of scientists and engineers in the U.S. The current NSCG sample was selected from the 2000 decennial census long form.

In 2010, the NSCG plans to refresh its sample to address attrition and coverage concerns. With the American Community Survey (ACS) replacing the long form, the Census Bureau is evaluating using the ACS as a sampling frame for the NSCG. The evaluation will initially examine the change in available and effective sample if the ACS is used as the sampling frame under the current NSCG design. The evaluation will then examine the potential increase in sampling efficiency for the NSCG associated with the additional demographic information available on the ACS. This paper provides an overview of this evaluation and presents recommendations for the 2010 NSCG based on the evaluation results.

Key Words: ACS, NSCG, Sampling Frame, and Effective Sample

1. Introduction[1]

The National Survey of College Graduates (NSCG) is designed to collect information on employment, educational, and demographic characteristics of scientists and engineers in the United States. The NSCG is a longitudinal survey that derived its current sample from the 2000 decennial census long form. The NSCG survey contractor, the U.S. Census Bureau, will conduct interviews of the NSCG sample cases three times during the 2000 decade – 2003, 2006, and 2008.

In 2010, the NSCG plans to refresh its sample to address attrition and coverage concerns. With the American Community Survey (ACS) designed to collect decennial census long form type data, the Census Bureau is evaluating the advantages and disadvantages of using the ACS as a sampling frame for the NSCG. The evaluation initially examines the differences in available sample and effective sample if the ACS is used as the sampling frame under the current NSCG sample design[2]. The evaluation then examines the potential increase in NSCG sampling efficiency associated with the additional demographic information available on the ACS that was not available on the long form. Finally, the evaluation compares quality measures associated with the two sampling frames. This paper begins by describing the current NSCG sample design based on the 2000 long form sampling frame and a potential design based on the ACS sampling frame. Then, the paper provides an overview of the evaluation methodology and discusses findings from the evaluation.

[1] 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 U.S. Census Bureau.

[2] The NSCG survey sponsor, the National Science Foundation (NSF), with the assistance of the Committee on National Statistics (CNSTAT), evaluated the sample design options for the 2010 NSCG. Since the timing of the NSF/CNSTAT evaluation coincided with the research documented in this report, a final 2010 NSCG design had not been chosen in time for use in our research. As a result, for the findings presented in this report, we examined the use of the ACS as a sampling frame in the context of the current NSCG sample design. The current NSCG sample design is a total replacement design where an entirely new sample is selected from the sampling frame.

2. NSCG Sampling Frame Comparison: Current Frame vs. Future Frame

2.1 NSCG Target Population

The target population of the NSCG is the college-educated science and engineering (S&E) workforce residing in the United States. This target population includes:

- U.S. residents with a bachelor's degree or higher in an S&E degree field
- U.S. residents with a bachelor's degree or higher working in an S&E occupation.

2.2 Current NSCG Sampling Frame: The Decennial Census Long Form

To select a sample that would allow estimation of the NSCG target population, the 2003 NSCG used the 2000 decennial census long form as its sampling frame. In the 2000 census, the long form was sent to approximately one-sixth of the total housing units (about 18 million housing units), although sampling rates varied from 13 to 50 percent depending on the population size of the area (reference [1]).

The use of the 2000 decennial census long form as a sampling frame resulted in 6.4 million person-level records that were eligible for the NSCG sample. To prepare these 6.4 million records for sampling, the Census Bureau stratified the records into 565 sampling strata defined by race, ethnicity, citizenship, disability status, degree level, occupation, and gender information collected on the long form. The NSCG used this information for sampling purposes since most NSCG estimates are defined by these variables. From the set of 6.4 million eligible records, the Census Bureau selected 177,320 records for the 2003 NSCG sample.

The information available on the decennial census long form allows for efficient sampling of U.S. residents working in an S&E occupation. This efficiency results from the availability of the occupation information from the long form. But the use of the long form as a sampling frame does not allow maximum efficiency in the sampling of U.S. residents with a bachelor's degree or higher in an S&E degree field. Although the long form does collect educational attainment information (i.e., degree level), it does not collect degree field information related to any degrees earned. As a result, it is not possible to identify records in the long form-based sampling frame that have a bachelor's degree or higher in an S&E degree field. In response to this sampling frame shortcoming, the Census Bureau selected an oversample of cases from the long form-based sampling frame so that the 2003 NSCG sample would include an adequate number of U.S. residents with a bachelor's degree or higher in a S&E degree field.

Because of the long form-based sampling frame shortcoming and the resulting oversample, the 2003 NSCG sample included many cases that do not meet the NSCG target population criteria. Since the NSCG is a longitudinal survey, the subsequent survey cycles only include the cases that meet the NSCG target population criteria. As a result, when we sampled from the long form, approximately 40 percent of the respondents from the initial NSCG survey cycle were not followed in subsequent cycles because they did not meet the NSCG target population criteria.

2.3 Future NSCG Sampling Frame: The American Community Survey

As noted earlier, the ACS is designed to collect decennial census long form type data. However, instead of sampling approximately one-sixth of the nation's housing units once every ten years, the ACS samples approximately 250,000 housing units every month for an annual sample size of approximately 3,000,000 housing units. This annual sample allows for period estimation of the nation's characteristics – one year of ACS data will allow period estimates for areas with at least 65,000 people, three years of ACS data will allow period estimates for areas with at least 20,000 people, and five years of ACS data will allow period estimates for all areas including small governmental units.

While the ACS provides more timely information than the long form, a tradeoff is that there are fewer cases available when using the ACS as a sampling frame. In the 2005 ACS, for example, there were 818,216 respondents eligible for the NSCG. This value is about one-eighth the number available from the decennial census long form.

Since the ACS collects information similar to what is collected on the decennial census long form, we were able to use ACS data to create the same stratification variables used in the 2003 NSCG. To allow the examination of the differences in available sample between the long form-based and the ACS-based frames, the Census Bureau stratified the 818,216 cases from the 2005 ACS into sampling strata defined by the same stratification variables used with the long form-based sampling frame – race, ethnicity, citizenship, disability status, degree level, occupation, and gender.

In addition to having a smaller sample than the decennial census long form, the ACS, by design, has more weight variation among the sample cases. The weight variation in the ACS sample results partially from the subsampling of housing units prior to the field interviewing data collection phase. By having more weight variation built into the ACS design, it is likely that a sample selected from the ACS-based sampling frame will have a smaller effective sample size[3] than a sample of the same size selected from the long form-based sampling frame.

Another major difference associated with the ACS is the potential collection of information not available on the decennial census long form. As an example, the Office of Management and Budget recently approved the inclusion of a degree field question on the ACS beginning in 2009. Having this degree field information on the ACS-based sampling frame has the potential to drastically change the inefficient sampling of U.S. residents with a bachelor's degree or higher in an S&E degree field that occurred when sampling from the long form.

3. Sample Comparisons

A potential design for the 2010 NSCG involves using the 2009 ACS as the sampling frame to select an entirely new NSCG sample. However, numerous differences between the ACS-based sampling frame and the long form-based sampling frame raise the issue of whether one year of ACS data will provide enough sample to achieve the 2003 NSCG stratum-level sample sizes. This section investigates that issue by evaluating the ability of the ACS-based sampling frame to meet the 2003 NSCG actual sample sizes and the 2003 NSCG effective sample sizes derived from the long form-based sampling frame. To conduct this investigation, we used data collected on the 2005 ACS.

Please note that the 2005 ACS questionnaire did not include a field of degree question. As a result, the analysis included in this section determined the number of ACS months needed for the NSCG under the assumption that the field of degree information was not available from the ACS. Please note that the field of degree information will be available when the NSCG samples from the 2009 ACS. However, until the 2009 ACS data file with the field of degree information is available, conducting this analysis while assuming the field of degree information is not available allows a direct comparison of the ACS-based sampling efficiencies to the decennial census long form-based sampling efficiencies.

3.1 Comparison of Actual Sample Sizes

A major difference between the decennial census long form and the ACS is the number of cases eligible for the NSCG. As discussed earlier, the eligible cases available on the ACS-based sampling frame (818,216 cases in the 2005 ACS) is approximately one-eighth the number available on the long form-based sampling frame (6.4 million cases). To investigate the effect the smaller ACS-based sampling frame has on the NSCG sampling process, this section investigates how many months of ACS sample are needed to achieve the 2003 NSCG stratum-level actual sample sizes derived from the long form-based sampling frame.

3.1.1 Methodology

The 2000 decennial census long form was used as the sampling frame for the 2003 NSCG. As part of the 2003 NSCG sample selection process, we used the long form responses to create stratification variables and sampling strata for all 6.4 million eligible cases from the long form-based sampling frame. The 177,320 sample cases for the 2003 NSCG were then selected from within the sampling strata. We use the term "stratum-level actual sample size" to describe the number of cases selected for the 2003 NSCG sample from within a particular sampling stratum.

The ACS collects information that is similar to what was collected on the decennial census long form. As a result, we were able to use 2005 ACS data to create the stratification variables used in the 2003 NSCG. We then combined the stratification variables to form sampling strata and classified all 818,216 eligible cases from the ACS-based sampling frame into one of these sampling strata. Since the ACS is a monthly sample, we divided the stratum-level sampling frame counts by 12 to determine the monthly sampling frame cases available within a stratum.

[3] The effective sample size is the sample size from a simple random sample that would yield the same sampling variance as achieved by the actual design.

Knowing the 2003 NSCG actual sample sizes from the long form-based sampling frame and the number of cases available from the ACS-based sampling frame, it is possible to determine how many months of ACS data, on average, it would take to equal the 2003 NSCG actual sample size within each sampling stratum.

3.1.2 Findings

When using the 2003 NSCG sampling strata (565 strata), the months of ACS data needed to achieve the 2003 NSCG actual sample sizes varied widely by stratum – ranging from less than one month of ACS data to over two years worth. Overall, 468 of the 565 sampling strata (83%) required 12 months or less of ACS to populate the strata to a size equivalent to the 2003 NSCG actual sample size.

Table 1. Number of ACS Months Needed to Match the 2003 NSCG Actual Sample Sizes

Description	Number of ACS Months Needed			Percent Needing One Year or Less
	Maximum	Mean	Std. Dev.	
Actual 2003 NSCG Stratification (565 Strata)	28	7.44	4.86	83

3.2 Comparison of Effective Sample Sizes

While the number of ACS months needed to achieve the 2003 NSCG actual sample sizes provides some insight on the capability of the ACS-based sampling frame, it is the effective sample size investigation that gives a clear indication of the effect the weight variation in the ACS-based sampling frame has on the NSCG sampling process. To investigate this issue, this section examines the number of ACS months needed to achieve the 2003 NSCG stratum-level effective sample sizes derived from the long form-based sampling frame.

3.2.1 Methodology

For each of the 565 sampling strata used in the 2003 NSCG, we examined the distribution of weights, the design effect due to unequal weighting, and the effective sample size in the long form-based sample (177,320 cases) and the ACS-based sampling frame (818,216 cases). In our processing, we calculated the design effect due to unequal weighting as follows:

$$DEFF = \frac{n \times \sum_{i=1}^{n} w_i^2}{\left[\sum_{i=1}^{n} w_i\right]^2} \qquad (1)$$

where n is the unweighted sample size, w_i is the sample weight for sample case i, and DEFF is the design effect due to unequal weighting. We then calculated the effective sample size, $n_{effective}$, as the unweighted sample size divided by the corresponding design effect.

$$n_{effective} = \frac{n}{DEFF} \qquad (2)$$

3.2.2 Findings

Table 1 indicates the average 2003 NSCG sampling stratum required approximately 7.44 months of ACS-based sample to achieve the 2003 NSCG actual sample size and 83% of the strata needed one year or less of ACS-based sample. We present this information from Table 1 as a basis of comparison when evaluating the number of ACS months required to achieve the 2003 NSCG stratum-level effective sample sizes. Since the ACS, by design, has a large amount of weight variation among the sample cases, we expect the use of the ACS as a sampling frame will result in smaller stratum-level effective sample sizes when compared to the long form-based sampling frame.

To account for both the smaller number of sampling frame cases and the unequal weighting associated with the ACS design, the average 2003 NSCG sampling stratum required approximately 10.94 months of ACS-based sample to achieve the 2003 NSCG sampling stratum-level effective sample size. This average number of ACS months needed to achieve the effective sample sizes (10.94) is nearly 50% greater than the number of months needed to achieve the 2003 NSCG actual sample sizes (7.44 from Table 1). In other words, the unequal weighting built into the ACS design requires an additional 3.50 months of ACS-based sample to achieve the 2003 NSCG effective sample sizes.

Furthermore, only about 62% of the strata could match the 2003 NSCG effective sample sizes with one year of ACS sample as compared to 83% for the actual sample sizes. A summary of these findings is presented in Table 2.

Table 2. Number of ACS Months Needed to Match the 2003 NSCG Effective Sample Sizes

Description	Number of ACS Months Needed			Percent Needing One Year or Less
	Maximum	Mean	Std. Dev.	
Actual 2003 NSCG Stratification (565 Strata)	34	10.94	7.37	62

4. Sampling Efficiency Associated with the Field of Degree Question on the ACS

A major difference between the long form and the ACS is the availability of a field of degree question beginning in 2009. The long form asked sample cases to report their highest educational attainment level, but did not request information on the educational field of any reported degree. The Census Bureau recently evaluated the inclusion of a field of degree question on the ACS questionnaire (reference [2]). After reviewing the evaluation results, the Office of Management and Budget approved the inclusion of the field of degree question on the ACS beginning in 2009. The advantage of the field of degree question is that it may reduce the overall sample size from the value used in previous NSCG survey cycles that was needed to allow proper screening of the entire S&E target population.

It should be noted the field of degree question on the ACS would not remove the need for screening of cases to determine S&E eligibility. Instead, the field of degree question would allow for the identification of certain cases that meet the target population S&E criteria and would not need to be screened. As a result, fewer cases would need to be screened for S&E eligibility. Since less screening would be required, the overall sample size could decrease. The investigation of the decrease in the overall sample size (i.e., sampling efficiency) is the focus of this section.

As stated in the introduction, this report evaluates the use of the ACS as a sampling frame for the NSCG in the context of a total replacement design where an entirely new sample is selected from the sampling frame. Under the assumption of a total replacement design for an ACS-based sampling frame that **includes a field of degree question**, this section examines the amount of ACS-based sample needed to meet the effective sample sizes from the 2003 NSCG for key reporting domains. The key reporting domains are the important estimation levels used in NSCG publications. For this analysis, the key reporting domains are defined by a three-level demographic group variable (non-Hispanic White, non-Hispanic Asian, and Minority) and a 55-level degree field[4]/occupation variable. When combined, these two variables result in 165 unique key reporting domains.

Please note that while this section is determining ACS-based sample needs by examining the effective sample sizes, the analysis discussed in this section is slightly different than the analysis discussed in the previous section. In the previous section, we defined the 2003 NSCG effective sample sizes as *the number of sample cases in a 2003 NSCG sampling stratum* divided by the design effect due to unequal weighting for those cases. In this section, we define the effective sample size as *the number of survey respondents in a 2003 NSCG key reporting domain* divided by the design effect due to unequal weighting for those cases.

4.1 Methodology

The Office of Management and Budget recently approved the inclusion of a field of degree question on the ACS questionnaire beginning in 2009. Unfortunately, since the field of degree question will not be added to the ACS until 2009, no ACS data is available that provides field of degree information for all the ACS sample cases.

To address the lack of data, the Statistical Research Division of the Census Bureau used classification trees to model the relationship between the field of degree information collected on the 2003 NSCG and the demographic information collected on the decennial census long form. Once the model was developed, it was used with the demographic information collected on the ACS to predict a field of degree response for each ACS respondent on the current ACS data file. Although there are limitations associated with this modeling approach to predict field of degree responses, the advantages of having an ACS file with field of degree for 2010 NSCG planning purposes outweigh the limitations.

[4] The 2003 NSCG collected degree field information as part of its data collection. As a result, the 2003 NSCG key reporting domains are defined by degree field.

Recall that the key reporting domains used in this evaluation were defined by a three-level demographic group variable (non-Hispanic White, non-Hispanic Asian, and Minority) and a 55-level degree/occupation variable. First, the 2003 NSCG survey respondents from the long form-based sampling frame were categorized into the 165 key reporting domains and effective sample sizes were calculated. Then, using the degree field value predicted for the ACS-based sampling frame cases, we determined the number of ACS-based cases needed to achieve the 2003 NSCG effective sample size for each domain.

Since the 2003 NSCG key reporting domains are for 2003 NSCG survey respondents, the number of ACS-based cases needed that was derived is the number of survey respondents needed in the 2010 NSCG. To determine the number of ACS-based sample cases needed in the 2010 NSCG, we need to inflate this number of survey respondents to account for nonresponse, ineligibility, screening, and misclassification.

4.2 Evaluation Assumptions and Limitations

It should be noted that numerous assumptions were made while evaluating the sampling efficiency associated with the degree field information. References [3], [4], and [5] provide an overview of some of the assumptions, briefly describe their derivation, and, when possible, discuss the implications of incorrect assumptions.

4.3 Findings

Using the key reporting domains defined by the three-level demographic group variable and the 55-level degree/occupation variable, the 2010 NSCG sample for the total replacement design would require approximately 120,000 sample cases from one year of ACS data to meet the target effective sample sizes from the 2003 NSCG key reporting domains. 84% of the domains had an adequate number of sampling frame cases available from one year of ACS data to meet the desired sample size for that domain. For the domains without an adequate number of sampling frame cases, slightly lowering the desired sample size would enable the sample demands to be met with one year of ACS data. As an example, the average desired sampling rate from the ACS-based sampling frame across the key reporting domains was approximately 1-in-400. For the domains without an adequate number of sampling frame cases, the desired sampling rate from the ACS-based sampling frame was approximately 1-in-100. Slightly reducing the oversampling in these domains would enable one year of ACS data to meet the 2010 NSCG sampling demands.

In terms of sampling efficiency, the use of an ACS-based sampling frame that **includes a field of degree question** requires about 30% less sample than the decennial census long form-based sampling frame to meet the key reporting domain effective sample sizes, even when taking into account the unequal weighting of the ACS-based sample. Please note that increased sampling efficiency is dependent upon the key reporting domains that are used within the sample selection process. The use of different key reporting domains may result in large changes in the sampling efficiency associated with an ACS-based sampling frame. Reference [5] documents the number of sample cases needed from the ACS-based sampling frame. References [6] and [7] document the specific process used to calculate the sample sizes.

5. Quality Measures Comparison

The previous sections evaluated the ability of an ACS-based sampling frame to provide an adequate actual sample. The next logical step is to evaluate the quality of the ACS data. To evaluate the ACS data quality, we investigated the imputation rate of key variables and the name quality for the long form and the ACS. In this evaluation, we used unweighted data to investigate the impact the imputation and name quality have on the actual number of cases eligible for the NSCG sampling frame.

5.1 Imputation Rates

For the variables used in the 2003 NSCG stratification, we compared the unweighted imputation rates from the ACS-based sampling frame against those from the decennial census long form-based sampling frame. These imputation rates were only calculated for the cases eligible for the NSCG sampling frame. The findings presented in Table 3 generally suggest the information collected on the ACS is of higher quality than the information collected on the long form[5]. This notion of higher data quality in the ACS is discussed in more detail in reference [1].

[5] All comparisons of the unweighted imputation rates presented in Table 3 are statistically significant at the 90 percent confidence level.

Table 3. Unweighted Imputation Rates for Key Variables – Long Form vs. ACS		
Variable	2000 Long Form Rate *(standard error)*	2005 ACS Rate *(standard error)*
Citizenship	3.44% *(0.04%)*	1.12% *(0.01%)*
Degree level	4.23% *(0.05%)*	1.74% *(0.02%)*
Disability status	9.36% *(0.07%)*	3.52% *(0.03%)*
Ethnicity	2.16% *(0.04%)*	1.11% *(0.01%)*
Gender	0.80% *(0.02%)*	1.03% *(0.01%)*
Occupation	9.41% *(0.07%)*	4.06% *(0.03%)*
Race	1.83% *(0.03%)*	1.10% *(0.01%)*

5.2 Name Quality

Since the NSCG is a person-based survey, it is important that the sampling frame provide adequate information to identify and locate the specific person selected for sample. To investigate the quality of name information, we evaluated the first name and last name variables. Any sample case that had missing information, a generic name[6], or a descriptor[7] rather than a name was flagged as a bad name record. Table 4 compares the unweighted bad name rates on the 2000 long form and the 2005 ACS. These bad name rates were only calculated for the cases eligible for the NSCG sampling frame. Table 4 provides the overall bad name rate and rates for the NSCG stratification variables. The results suggest the name information available on the long form-based sampling frame was of a higher quality than the name information available on the ACS-based sampling frame[8].

Table 4. Bad Name Rates for Key Variables – Long Form vs. ACS			
Variable	Level	2000 Long Form Rate *(standard error)*	2005 ACS Rate *(standard error)*
Citizenship	Non-U.S. Citizen	1.24% *(0.05%)*	3.19% *(0.07%)*
	U.S. Citizen	1.31% *(0.03%)*	2.48% *(0.02%)*
Degree level	Bachelor's degree	1.37% *(0.04%)*	2.59% *(0.03%)*
	Master's degree	1.05% *(0.05%)*	2.52% *(0.04%)*
	Doctorate degree	0.94% *(0.09%)*	2.68% *(0.11%)*
Disability Status	No	1.65% *(0.03%)*	2.51% *(0.02%)*
	Yes	1.20% *(0.08%)*	3.49% *(0.10%)*
Ethnicity	Non-Hispanic	1.42% *(0.03%)*	2.57% *(0.02%)*
	Hispanic	1.24% *(0.09%)*	2.64% *(0.10%)*
Gender	Male	1.19% *(0.04%)*	2.53% *(0.03%)*
	Female	1.35% *(0.04%)*	2.62% *(0.03%)*
Occupation	Physical Sciences	1.31% *(0.16%)*	1.95% *(0.22%)*
	Math and Computer Sciences	1.23% *(0.08%)*	2.22% *(0.11%)*
	Life Sciences	1.18% *(0.20%)*	2.49% *(0.30%)*
	Social Sciences	1.04% *(0.17%)*	2.15% *(0.23%)*
	Engineering	1.04% *(0.08%)*	2.24% *(0.12%)*
	S&E-Related	1.23% *(0.06%)*	2.37% *(0.05%)*
	Teachers	1.05% *(0.07%)*	2.23% *(0.10%)*
	Non-S&E Occupations	1.40% *(0.05%)*	2.68% *(0.03%)*
Race	White	1.09% *(0.03%)*	2.38% *(0.02%)*
	Black	1.88% *(0.10%)*	4.06% *(0.11%)*
	Asian	1.41% *(0.07%)*	3.14% *(0.08%)*
	American Indian/Alaska Native	1.82% *(0.27%)*	4.09% *(0.30%)*
	Native Hawaiian/Pacific Islander	2.43% *(0.71%)*	6.05% *(0.93%)*
	Other	1.75% *(0.19%)*	3.64% *(0.19%)*
Overall		**1.26% (*0.03%*)**	**2.57% *(0.02%)***

[6] Examples of generic names are John Doe and Jane Doe.

[7] A descriptor refers to a description of the sample person. The descriptor was recorded when a valid name was not provided. Common descriptors include "Gentleman of the House," "Lady of the House," "Son," and "Daughter."

[8] All comparisons presented in Table 4 are statistically significant at the 90 percent confidence level.

6. Summary and Future Research

6.1 Summary

For the 2010 cycle of the NSCG, the National Science Foundation (NSF) is examining the possibility of refreshing the NSCG sample to address attrition and coverage concerns. With the availability of the ACS as a source for decennial census long form type data, the Census Bureau is evaluating the strengths and weaknesses of using the ACS as a sampling frame for the NSCG. The strengths included:

- The data quality, as measured by the imputation rate, is better on the ACS than on the long form.
- The availability of degree field information on the ACS beginning in 2009 could lead to a reduction in the necessary sample for the initial NSCG survey cycle.
- Depending on the key reporting domains used in the sample selection process, the availability of degree field information on the ACS beginning in 2009 could allow one year of the ACS data to provide enough sample to meet the NSCG sample needs.

Weaknesses associated with the use of the ACS as a sampling frame for the NSCG included:

- The name information on the ACS appears to be of lower quality than on the long form.
- Even with the degree field information, it could take more than one year of ACS data to meet the NSCG sample needs for some key reporting domains (dependent on the domain definitions).

6.2 Future Research

From the information presented in this report, we list three areas that warrant further research.

- The determination of the 2010 NSCG sample size when selected from the ACS-based sampling frame is heavily dependent on the key reporting domains used within the sample selection process. As a result, resources should be allocated to identify the domains that best meet the needs of the NSCG and the NSF.
- Research should be conducted to investigate the effect the higher percentage of poor quality names on the ACS will have on NSCG estimation.
- The NSF, with the assistance of the Committee on National Statistics (CNSTAT), examined options for the 2010 NSCG sample design. From the options presented, CNSTAT recommended a rotating panel design (reference [8]). As a result, research should be conducted to examine the details associated with implementing this design in the 2010 NSCG.

References

[1] National Research Council. (2007). *Using the American Community Survey: Benefits and Challenges.* Panel on the Functionality and Usability of Data from the American Community Survey, Constance F. Citro and Graham Kalton, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[2] Raglin, David et al., "Testing a New Field of Degree Question for the American Community Survey," U.S. Census Bureau 2007 American Community Survey Content Test Report, May 2008.

[3] Finamore, John and David Hall, "2010 NSCG Sample Design Planning: Sample Size Information Using an ACS-Based Sampling Frame," Internal Census Bureau Memorandum for Landman from Killion, February 6, 2008.

[4] Finamore, John and David Hall, "2010 NSCG Sample Design Planning: Additional Sample Size Information Using an ACS-Based Sampling Frame," Internal Census Bureau Memorandum for Landman from Killion, February 29, 2008.

[5] Finamore, John and David Hall, "Evaluating the ACS as a Potential Sampling Frame for the NSCG – Full Report," Internal Census Bureau Memorandum for Landman from Killion, September 2008.

[6] Hall, David and John Finamore, "2010 NSCG – ACS-Based Sampling Feasibility Study Results for Goal #1A," Internal Census Bureau Memorandum for Landman from Killion, July 2, 2008.

[7] Hall, David and John Finamore, "2010 NSCG – ACS-Based Sampling Feasibility Study Results for Goal #1B," Internal Census Bureau Memorandum for Landman from Killion, July 2, 2008.

{8] 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 Academies Press.