Evaluating the Consistency between Responses to the 2010 NSCG and the 2009 ACS[1]

Phyllis Singer, Robyn Sirkis
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract
The purpose of this research was to examine the consistency of responses for similar questions asked in the American Community Survey (ACS) and the National Survey of College Graduates (NSCG) to better leverage the use of the ACS in the sample design and estimation of the NSCG. This research evaluated the ACS responses used in the NSCG sample design to determine if the current NSCG design was efficient for key statistics, and to identify new ACS responses that may strengthen the design. Additionally, this evaluation identified variables that could be used in model assisted and model-based estimation for the NSCG. The research consisted of two phases: response comparison and examination of response differences. We looked at individual responses to find patterns that lead to the inconsistency of occupation and education. This included constructing statistical models involving questionnaire responses and paradata to find covariates related to inconsistency.

Key words: consistency, ACS, NSCG, occupation, education

1. Introduction

The use of the American Community Survey (ACS) as a sampling frame for the National Survey of College Graduates (NSCG), and the desire to further leverage the ACS in the NSCG, rely on the consistency of responses for similar items between the surveys. The purpose of this evaluation was to determine the degree to which respondents gave consistent responses to similar items on the 2009 ACS and the 2010 NSCG. Another goal was to identify covariates and factors that may be related to occupation and degree field inconsistences. When consistency rate results were presented to internal sponsors and the National Center for Science and Engineering Statistics (NCSES), it was agreed to focus on inconsistencies in occupation and degree field. This consistency evaluation provides insight on the data collection differences between the ACS and NSCG, as well as dissimilarities in wording, time periods, coding rules, and proxy responses.

2. Background

The 2010 NSCG survey cycle marked the initial implementation of a rotating panel sample design that relied on biennial samples selected from the most recent ACS data file. ACS responses were used to determine the NSCG frame eligibility and to develop the NSCG sampling strata. Because 2010 marked the initial use of ACS responses for the NSCG, the National Science Foundation (NSF) requested that the U.S. Census Bureau conduct an evaluation to examine the effectiveness of the ACS responses in meeting the eligibility criteria and sampling efficiency needs of the NSCG.

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U. S. Census Bureau.

2.1 NSCG Survey Design

The NSCG is a longitudinal survey conducted every two to three years by the U. S. Census Bureau for the NSF (Finamore, 2013). Data from the NSCG, the National Survey of Recent College Graduates, and the Survey of Doctorate Recipients are used to develop national estimates on the science and engineering (S&E) workforce.

The target population for the 2010 NSCG survey cycle consisted of all individuals with the following characteristics:

- Under the age of 76 as of the survey reference date,
- Not institutionalized and living in the United States as of the survey reference date, and
- Recipient of a bachelor's degree or higher prior to the sampling frame construction date.

The 2010 NSCG incorporated a dual frame sample design. It selected about 65,000 persons from the 2009 ACS respondents who indicated they had a bachelor's degree or higher in any field of study. The majority of the remaining 35,000 sample persons of the 2010 NSCG sample were selected from respondents to the 2008 NSCG. Initial data collection used either a self-administered mail survey or a self-administered Web survey. Computer-assisted telephone interviewing (CATI) was used for nonresponse follow-up.

2.2 ACS Survey Design

The ACS (U. S. Census, 2009) is conducted every year to provide up-to-date social, economic, housing, and demographic estimates for the United States. The ACS sample is drawn from the Census Bureau’s Master Address File.

The ACS data collection operation consists of three modes: mail, telephone, and personal visit. The first phase for housing units included a questionnaire mailed to sample addresses, with a request that the household complete and return the questionnaire by mail. When a telephone number was available and no response was received, the U.S. Census Bureau followed up with CATI. If the U.S. Census Bureau was unable to reach an occupant using CATI, or if the household refused to participate, the address was eligible to be selected for computer-assisted personal interviewing (CAPI). Additionally, unmailable addresses were selected for the CAPI phase.

3. Methodology

This evaluation consisted of two phases. Phase 1 combined ACS and NSCG data, created comparison variables, and determined consistency rates. Phase 2 included an exploratory statistical analysis to determine potential covariates or factors related to occupation and degree field inconsistencies.

3.1 NSCG and ACS Data

The exploratory analysis consisted of using 2009 ACS and 2010 NSCG data and paradata. Minimal editing to correct for formatting differences across the different data collection instruments (mail, telephone, and web) was done to the NSCG data. The NSCG and ACS responses that were edited or allocated when creating the comparison variables were blanked when calculating the consistency rates. A comparison variable was defined as the combination of several NSCG or ACS response variables. After merging the ACS and NSCG data, retaining the overlap, and removing noninterviews, the analysis dataset contained 47,643 observations.

The data analyzed were a subset of the entire NSCG sample, but the purpose of this project was to determine the consistency of the NSCG cases that were sampled from the preceding ACS data collection year (2009). The data were not weighted; therefore one should not make inferences to the general population based on this analysis.

3.2 Comparison Variables and Consistency Rates

There were thirty-one ACS and NSCG comparison variables in the evaluation. Imputed or edited values were blanked. The comparison variables included the following topics:

Age	Likelihood of U.S. Earned Degree
Citizenship	Marital Status
Degree Level	Race
Degree Field	Race/Ethnicity
Disability	Occupation
Sex	Broad Occupation
Hispanic Indicator	Year of Entry

Some of these topics involved using 2-level or 3- or more level comparison variables. Consistency rates were calculated for the ACS and NSCG comparison variables. For this evaluation, the consistency rates are descriptive statistics for the population of 2010 NSCG cases selected from the 2009 ACS. The consistency rate was defined as follows.

$$Consistency\ Rate = \frac{Number\ of\ Consistent\ Responses}{Total\ Number\ of\ Eligible\ Cases} \times 100$$

Additionally, the coefficient of variation (CV) was calculated for each consistency rate. The CV is the standard deviation, divided by the mean.

3.3 Degree Field and Occupation Questions and Coding Instructions

The Phase 2 analysis involved constructing statistical models for the degree field and occupation topics. The second phase also involved conducting exploratory data analysis of write-in responses to identify any patterns to the occupation and degree field inconsistencies. Table 1 lists the degree field and occupation categories that correspond to the 2-level and 7-level degree field and occupation comparison variables.

Table 1. Degree Field and Occupation Comparison Variables

Category	Comparison Variable	Values or Additional Description
Degree Field	2-Level	Category 1 = At least one S&E bachelor's degree Category 2 = No S&E bachelor's degree
	7-Level	Category 1 = Computer and math sciences Category 2 = Life and related sciences Category 3 = Physical and related sciences Category 4 = Social and related sciences Category 5 = Engineering Category 6 = S&E Related fields Category 7 = Non-S&E
Occupation	2-Level	Category 1 = Science and Engineering (S&E) occupation Category 2 = Non-S&E occupation
	7-Level	Category 1 = Computer and math scientists Category 2 = Life scientists Category 3 = Physical scientists Category 4 = Social scientists Category 5 = Engineers Category 6 = S&E Related occupations Category 7 = Non-S&E occupations

Prior to conducting this analysis, it was important to become familiar with the wording of the degree field and occupation questions in the ACS and NSCG questionnaires and with the coding instructions for the ACS and NSCG occupation and degree field responses.

The degree field and occupation questions on the ACS and NSCG questionnaires were worded differently. There were also differences in question wording between modes. NSCG respondents with self-response paper questionnaires were instructed to enter codes found in the back of the questionnaire to complete the NSCG degree field and occupation questions after first supplying write-in responses. The self-response ACS questionnaire instructed the respondent to list the field of any bachelor's degrees received. The self-response NSCG questionnaires instructed respondents to enter their most recent degree, second most recent degree, and first bachelor's degree in a supplied degree grid. The NSCG respondents were then instructed to choose the education code from the back of the questionnaire that best described the majors. Additionally, the NSCG respondents entered their major field of study.

The ACS occupation variable was based on a question that asked if the respondent had worked within the past 12 months, 1 to 5 years ago, or over 5 years ago or never worked. This was connected to the respondent's most recent occupation. Furthermore, respondents were asked to describe their kind of occupation and most important work activities and duties. The NSCG questionnaire asked respondents about their working status as of the week of October 1, 2010 to categorize occupation and to then choose the occupation code that best fit the occupation from the back of the questionnaire. Respondents were instructed to report the title, duties, and responsibilities for the title job.

It is important to note that in the NSCG the degree field and occupation codes could have been modified during the coding operations. On the NSCG, respondents supplied their own codes for occupation and degree field. Respondents did not supply their own codes for the ACS. The NSCG ran computer programs to code occupation and degree field. This is called auto-coding. Any cases that could not be coded by auto-coding were then sent to clerical coders. This is called manual coding. ACS occupation and degree field coding was done by manual coding in 2009.

3.4 Exclusion of Cases and Degrees

The examination of potential covariates related to degree field and occupation inconsistency focused on cases where the responses should have been the same on the ACS and NSCG questionnaires. Cases were excluded from the occupation analysis based on the following criteria.

- The occupation responses were imputed to create the ACS or NSCG comparison variables
- The respondent indicated their last employment was more than 1 year ago on the ACS questionnaire
- The respondent's answer to the NSCG question about the start year and month of their principal job was greater than their ACS questionnaire completion date

Degrees from the NSCG were excluded if the respondent indicated that they were earned after the ACS questionnaire completion date. In this situation, the 7-level comparison variable could change or stay the same, possibly changing the value of the 2-level degree comparison variable.

3.5 Statistical Models to Determine Covariates Related to Occupation and Degree Field Inconsistencies

Person-level logistic regression models were calculated to determine which of the questionnaire responses and paradata covariates were related to the consistency of response between the ACS and NSCG 2- and 7-level degree field and occupation comparison variables. Though the statistical models do not infer causation, they yield proximate determinants for research and analysis. Models were constructed for the four binary degree field and occupation response variable indicators (1=inconsistent response, 0=consistent response). The logistic model has the following form.

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \cdots \beta_k x_k + \varepsilon$$

where p is the propensity to be inconsistent and (x_1….. x_k) is the vector of covariates.

Separate statistical models were constructed using the ACS questionnaire responses and paradata only, and NSCG questionnaire responses and paradata only. More information about the covariates is in the next section. Finally, the ACS and NSCG covariates were used in the same statistical model. Both main effects and 2-way interaction terms were included in the statistical models.

There was a vast number of covariates to consider during the creation of the statistical models. Therefore, it was required to develop a variable reduction procedure. The method

chosen involved choosing the covariates with the largest absolute percent difference between those respondents who provided a consistent response (for either occupation or degree field) and a specific covariate value (in a category) and those respondents who provided an inconsistent response and the same covariate value. The covariates with the largest percent differences were used in the statistical models to determine which of these potentially had an effect on the degree field and occupation inconsistency.

Automatic selection methods were used to aid in choosing the initial parsimonious model. Issues relating to sample size, missing values, and collinearity among predictor variables (covariates) were taken into account when choosing an appropriate model. When two covariates were highly correlated, only one was included in the statistical model. The covariates were considered to be correlated for this analysis if the Pearson correlation coefficient was approximately 0.5 and higher. The covariate that was included in the statistical model was the one with the highest percent difference for the inconsistency flag response variable. Also, covariates with a vast number of missing values (e.g., over 95 percent) were excluded from the statistical model, as well as those that were heavily skewed (e.g., over 95 percent responses in one category). Covariates were recoded so that missing, don't know, and refusal values could be placed in a category and, therefore, not removed from the statistical model. Several covariates with more than two categories (not including missing values) were converted into binary covariates and used in the statistical models. For example, there were three binary variables created for the ACS mode (mail versus CATI/CAPI or CATI versus mail/CAPI, or CAPI versus mail/CAPI) covariate.

The Akaike Information Criterion (AIC) was used to compare the statistical models produced for each response variable. The model with the lowest AIC value was the preferred statistical model. See http://www.okstate.edu/sas/v8/saspdf/stat/chap39.pdf for the formulas.

The R-squared statistic, which refers to the fraction of variance explained by the model, gives information about the goodness of fit of the statistical model. An extremely low R-squared value provides an indication that the model does not fit the data well. Association statistics such as the *c* statistic were used in determining the effectiveness of the models. The *c* statistic represents the proportion of pairs with different observed outcomes for which the model predicts a higher probability for observations with the event outcome than the probability for nonevent observations. Additionally, the odds ratios were examined for those covariates left in the statistical model. The odds ratio is the probability of an occurrence of an event to that of nonoccurrence. It accesses the strength of association and the potential impact of confounding variables.

3.6 Covariates that May Have Impacted Occupation and Degree Field Inconsistency

As stated previously, separate statistical models were constructed using only the ACS questionnaire responses and paradata, and NSCG questionnaire responses and paradata. Additionally, statistical models were produced using both ACS and NSCG covariates. Table 2 gives descriptions pertaining to some of the variables or covariates that may have been related to the occupation and degree field consistency. Because of the wording and format of the ACS degree field question, it is possible that ACS respondents listed degrees other than bachelor's degrees or listed different degrees than on the NSCG. Because both surveys collected multiple degree fields, the covariates in Table 2 include

counts of distinct degree field levels (both ACS and NSCG), called "distinct fields", and indicators of overlap between the 7-level degree field categories on the NSCG and the ACS. The NSCG and ACS were said to have overlap between the 7-level degree field categories if the respondent listed degree fields from some of the same degree fields on both the NSCG and the ACS. Three covariates display whether the NSCG codes were converted to degrees or occupations by manual or auto-coding methods. Other ACS and NSCG covariates involve demographic characteristics, reasons for not working, data collection mode, educational attainment, time difference between the completion of the ACS and NSCG questionnaires, and whether the degree was related to a specific technical field expertise.

Table 2. Several ACS and NSCG Key Covariates
***indicates degree earned on or before ACS completion date**

Detailed Description	Category Description
ACS: Number of Distinct ACS Fields	1-4
Distinct ACS and NSCG Overlap ACS Fields with Distinct NSCG Bachelor's Fields*	no overlap / overlap
Distinct ACS and NSCG Overlap ACS Fields with Distinct NSCG Post Baccalaureate*	no overlap / overlap
Number of Distinct NSCG Bachelor's Fields*	0-3
Number of Distinct NSCG Higher Fields*	0-4
Number of Distinct NSCG Fields (Bachelor's and above)*	0-4
Degree Field Coding : Bachelor's Degree*	manually coded /auto-coded
Degree Field Coding: Bachelor's Degrees and above*	manually coded /auto-coded
Occupation Coding	manually coded/auto-coded

3.7 Verbatim Response Analyses

The occupation and degree field verbatim responses were analyzed to find patterns that lead to their inconsistency. Two occupation fields were examined for the ACS: the kind of work the respondent was doing and the respondent's most important activities or duties on the job. The two occupation fields analyzed for the NSCG were the title of the principal job and the kind of work on the job (duties and responsibilities). The ACS had a single response that listed the degree field for all bachelor's degrees. The NSCG had responses for the degree fields of the most recent degree, the second most recent degree, and the first bachelor's degree.

4. Findings

4.1 Consistency Rates

The consistency rates were lower than 90 percent for six of the 31 comparison variables. The CVs for the consistency rates were, overall, very small (<0.006). The 2- and 7-level occupation, year of entry, broad occupation, and 7-level degree field comparison variables had consistency rates lower than 90 percent. Table 3 lists the number of eligible cases, number of consistent cases, number of inconsistent cases, and consistency rate for the degree field, occupation, broad occupation, and year of entry topics. After calculating the consistency rates, it was decided not to focus on the year of entry and broad occupation comparison variables for this exploratory analysis. The year of entry had many levels (raw values) for year of entry. The broad occupation comparison variable had 16 levels.

Table 3. ACS and NSCG Comparison Variables with Consistency Rates

Category	Comparison Variable	Number of Eligible Cases	Number of Consistent Cases	Number of Inconsistent Cases	Consistency Rate (Percent)
Degree Field	2-Level	42,963	38,591	4,372	89.82
	7-Level	42,963	34,763	8,200	80.91
Occupation	2-Level	35,073	26,135	8,938	74.52
	7-Level	35,073	21,503	13,570	61.31
Broad Occupation	16-Level	34,695	18,462	16,233	53.21
Year of Entry	Continuous	10,809	8,191	2,618	75.78

Table 4 presents the number of eligible cases, consistent cases, and inconsistent cases, and consistency rates for the degree field and occupation comparison variable Phase 1 and 2 analyses. The consistency rates were similar between Phase 1 and Phase 2. The number of eligible cases decreased more in the Phase 2 for the occupation comparison variables than for the degree field comparison variables. The CVs for consistency rates in Table 4 were all less than 0.006.

Table 4. ACS and NSCG Comparison Variables Eligible Cases and Consistency Rates from Phase 1 and Phase 2

Category	Variable	Phase 1 Restriction		Phase 2 Restriction	
		Eligible Cases	Consistency Rate (%)	Eligible Cases	Consistency Rate (%)
Degree Field	2-Level	42,963	89.82	42,799	89.84
	7-Level	42,963	80.91	42,799	80.95
Occupation	2-Level	35,073	74.52	28,396	75.64
	7-Level	35,073	61.31	28,396	62.55

Statistical models were constructed for the 2- and 7-level degree field and occupation comparison variables.

4.2 Covariates Chosen in the Statistical Models

There were many covariates to consider during the creation of the statistical models. The covariates with the largest percent differences were used in the models to identify potential factors that could be related to the degree field and occupation inconsistency. This does not imply these were the causes of the inconsistency in the occupation and degree field variables, as the objective here was to reduce the number of variables when forming the models. Some of the covariates chosen by the procedure for the 2-level occupation inconsistency flag response variable were the ACS 2-level S&E degree field comparison variable, NSCG duties on a job requiring technical expertise of a bachelor's degree or higher (social sciences and other covariates), NSCG occupation coding, ACS educational attainment (master's degree), and the ACS gender comparison variable. Alternatively, some of the covariates used for the 7-level occupation inconsistency flag response variable were the NSCG duties on a job requiring technical expertise of a bachelor's degree or higher (social sciences, engineering, computer science, math or the natural sciences), ACS educational attainment (bachelor's degree), NSCG reasons for taking courses (to change academic or occupational field), ACS 2-level S&E degree field comparison variable, and the NSCG work on principal job related to highest degree.

Some of the covariates chosen by the procedure for the 2-level degree field inconsistency flag response variable were whether there was an overlap between the ACS and NSCG degrees among the seven categories used to create the degree field comparison variable (bachelor's degrees), NSCG degree field coding (bachelor's degree and higher degree covariates), NSCG duties on a job requiring technical expertise of a bachelor's degree or higher (engineering, computer science, math, or the natural sciences), ACS 2-level S&E occupation comparison variable, and the number of distinct ACS degrees (one and two distinct degrees covariates). Alternatively, some of the covariates used for the 7-level degree field inconsistency flag response variable were whether there was an overlap between the ACS and NSCG degrees among the seven categories used to create the degree field comparison variable (bachelor's degrees and post baccalaureate covariates), NSCG degree field coding (bachelor's degree and higher degree covariates), the number of distinct ACS degrees (one and two distinct degrees covariates), and the ACS data collection mode (mail).

Issues relating to correlated covariates were addressed before creating the models. Correlated covariates used in the statistical models included those pertaining to the number of distinct degrees, overlapping of ACS and NSCG degrees, data collection mode, race, and reasons for not working. Other correlated covariates included the ACS and NSCG comparison variables. Several covariates included a large portion of missing values or heavily skewed categories. These covariates were not included in the models. The majority of the NSCG reasons for not working covariates had over 95 percent missing values when examining the occupation inconsistency flag response variables. There were a few covariates related to the ACS reasons for not working topic that also had over 95 percent missing values for the same response variables.

4.3 Statistical Model Results

Several person-level logistic regression models were constructed to determine which characteristics of the respondents appear to be related to inconsistent responses between the ACS and NSCG degree field and occupation questions, for each inconsistency flag response variable. Models included ACS covariates only, NSCG covariates only, or both ACS and NSCG covariates.

The statistical models had a more appropriate fit for the 2-level and the 7-level degree field inconsistency flag response variables than occupation inconsistency flag response variables. Table 5 shows the model fit statistics. The c statistic was above 0.90 for the degree field models as opposed to below 0.70 for the occupation models. Possibly different covariates could lead to statistical models with improved model fit statistics for the occupation inconsistency response variables.

Table 5. Model Fit Statistics

	Lowest		Highest	
	R-squared	**c**	**R-squared**	**c**
2-Level Occupation	0.0441	0.631	0.0693	0.672
7-Level Occupation	0.0461	0.624	0.0856	0.676
2-Level Degree Field	0.0231	0.636	0.2457	0.934
7-level Degree Field	0.0270	0.614	0.4688	0.950

Table 6 shows the covariates that were kept in the 2-level degree field inconsistency flag model after the implementation of the backward elimination method using ACS and NSCG questionnaire responses and paradata covariates. The covariates that are shaded and in bold had a larger impact in the statistical model. The R-squared value was 0.2457 and the c statistic was 0.934. The odds of being in the inconsistent group for those reporting 2 or more distinct ACS degree field categories were about 7 times (7.328) the odds for those reporting 1 distinct ACS degree field category. The odds of being in the inconsistent group for those reporting no relationship between their degree and job were about 3 times (2.837) the odds for those reporting that there was a relationship between their degree and job.

Table 6. Covariates Remaining in the 2-Level Degree Field Inconsistency Flag Model

Detailed Description of Covariates
ACS Science and Engineering Occupation (S&E) Flag
ACS Number of Distinct ACS Fields (7 categories): 1
ACS Age Group
ACS Gender
NSCG Duties on Job Require Technical Expertise of a Bachelor's Degree or Higher in Engineering, Computer Science, Math or the Natural Sciences
NSCG Distinct ACS and NSCG Overlap ACS fields with Distinct NSCG Bachelor's Fields Earned on or Before ACS Completion Date
NSCG Duties on Job Require Technical Expertise of a Bachelor's Degree or Higher in Some Other Field (ex. Health, Business, or Education)
NSCG Duties on Job Require Technical Expertise of a Bachelor's Degree or Higher in Social Sciences
NSCG Mode (mail)
Distinct ACS and NSCG Overlap ACS Fields with Distinct NSCG Post Baccalaureate Earned on or Before ACS Completion Date
NSCG Degree Field Coding: Restricted Before ACS Completed and NSCG Bachelor's Degree
Interaction Effect: NSCG Mode (Mail) by NSCG and ACS Bachelor's Field Overlap

Table 7 shows the covariates that were kept in the 7-level degree field inconsistency flag model after the implementation of the forward selection method using ACS and NSCG questionnaire responses and paradata covariates. The covariates that are shaded and in bold had a larger impact in the statistical model. The R-squared value was 0.4688 and the c statistic was 0.950. The odds of being in the inconsistent group for those reporting 2 or more distinct ACS degree field categories were about 9 times (8.902) the odds for those reporting 1 ACS degree field category.

Table 7. Covariates Remaining in the 7-Level Degree Field Inconsistency Flag Model

Detailed Description of Covariates
ACS Age Group
ACS Number of Distinct ACS Field (7 categories): 1
ACS Model (Mail)
NSCG Mode (Mail)
Distinct ACS and NSCG Overlap ACS fields with Distinct NSCG Bachelor's Fields Earned on or Before ACS Completion Date
Distinct ACS and NSCG Overlap ACS Fields with Distinct NSCG Post Baccalaureate Earned on or Before ACS Completion Date
NSCG Degree Field Coding: Restricted Before ACS Completed and NSCG Bachelor's Degree
Interaction Effect: NSCG Mode (Mail) by NSCG and ACS Bachelor's Field Overlap

Ultimately, the statistical models using the degree field inconsistency flag response variables appeared to be more informative, given the fit statistics, than the occupation degree field inconsistency flag response variables. Possibly different covariates could lead to statistical models with improved model fit statistics for the occupation inconsistency response variables.

4.4 Exploratory Verbatim Response Analysis Results

Occupation and degree field verbatim responses were (separately) compared between the ACS and NSCG questionnaires to find patterns that might have lead to the inconsistencies. There were many instances where the responses seemed similar, but the 2-level comparison variables did not have the same value for both surveys. There were differences in coding instructions for the ACS and NSCG occupations. For example, all post-secondary teachers on the ACS were coded as science and engineering related and all secondary teachers on the ACS were coded as non-science and engineering. For the NSCG, their status was determined by their field of expertise. Additionally, the respondents seemed to enter more detailed information when reporting their occupational duties for on the NSCG questionnaire.

The verbatim response analysis emphasized the definitional differences existed between the ACS and the NSCG coding operations. Secondary and post-secondary teachers are but one example. Additionally, pilots would be S&E for the ACS but non-S&E for the NSCG. Some items used to determine correct placement for problematic occupations (e.g., management) on the NSCG are not on the ACS.

5. Conclusions and Recommendations

The consistency rates were lower than 90 percent for six of the 31 comparison variables. The 2- and 7-level occupation, 2- and 7-level degree field, year of entry, and broad occupation comparison variables had consistency rates that were less than 90 percent.

The results from the verbatim analysis indicate that further investigation into the definitions for the coding operations between the ACS and NSCG may provide insight into the occupation and degree field inconsistencies. The statistical model analysis indicated that coding method (manual coding, auto coding), amount of overlap between

the ACS and NSCG 7-level degree categories, and the type of expertise for the job may be related to some of the inconsistencies.

It may be useful to reproduce the analyses in this paper with the final edited and imputed data from the ACS and the NSCG to determine if the results change. Additional recommendations are to further explore the ACS and NSCG coding operations and to conduct statistical modeling and correlation analysis using process data variables (e.g., contact attempts).

6. Limitations

The following were limitations for this analysis.

- Weights were not used because the findings were not projected to a larger population. Only the 2010 NSCG cases that were selected from the 2009 ACS were used in the analyses. Additionally, only minimal editing was done to the NSCG data and imputed ACS data were excluded from the analyses. The project consisted of exploratory analysis to determine potential covariates or factors related to occupation and degree field inconsistencies and therefore did not account for the complex survey design.
- It was not possible to run ACS data through the NSCG auto-coding programs because the ACS does not have all the variables needed to run the NSCG coding programs.
- Sample sizes were too small to put into models the variables that indicated reasons for not working.

Acknowledgements

We are grateful to John Finamore of the NSF and Benjamin Reist, David Warren Hall, Michael White, David Pysh, Greg Orlofsky, Kurt Bauman, Jessica Davis, and Melissa Chiu of the Census Bureau for their assistance throughout this project - supplying data, coding documentation, and insight into consistency issues between the ACS and the NSCG. We thank our reviewers for their helpful comments.

References

Finamore, J., *National Survey of College Graduates*, March 1, 2013, <http://www.nsf.gov/statistics/srvygrads>, accessed on March 22, 2013.

U. S. Census Bureau, *Design and Methodology: American Community Survey*, April 2009, <http://www.census.gov.acs/www/methodology/methodology main>, accessed March 22, 2013.