Leveraging the American Community Survey (ACS) in Current Estimation for the National Survey of College Graduates (NSCG)

Michael D. Larsen[1]

[1]Department of Statistics, George Washington University
6110 Executive Blvd, Suite 750, Rockville, MD 20852
mlarsen@bsc.gwu.edu

Abstract
The National Survey of College Graduates (NSCG) has been conducted by the Census Bureau for the National Science Foundation (NSF) since the 1960s. It is the nation's only source of detailed statistics on the science and engineering (S&E) labor force. The NSCG uses a rotating panel design and selects its sample on a biennial basis from the American Community Survey to allow both cross-sectional and longitudinal analysis of education, employment, and demographic characteristics of the S&E labor force. Under this design, the NSCG data is collected and released on a biennial or triennial schedule. The 2010 survey cycle marked the initial use of the ACS as a sampling frame for the NSCG. The 2010 NSCG responses allow NSF the ability to produce estimates of the S&E labor force as of the 2010 calendar year. The next NSCG survey cycle is scheduled for 2013 and will allow NSF the ability to produce estimates of the S&E labor force as of the 2013 calendar year. Using the results from the 2010 NSCG survey cycle, we examined the feasibility of producing 2011 and 2012 S&E labor force estimates using 2011 and 2012 ACS data.

Key Words: Missing data; Model-assisted estimation; Panel survey; Prediction; Subsampling; Survey frame

1. Introduction

The National Survey of College Graduates (NSCG) has been conducted by the Census Bureau for the National Science Foundation (NSF) since the 1960s. It is the nation's only source of detailed statistics on the science and engineering (S&E) labor force. The NSCG uses a rotating panel design and selects its sample on a biennial basis from the American Community Survey to allow both cross-sectional and longitudinal analysis of education, employment, and demographic characteristics of the S&E labor force. Under this design, the NSCG data is collected and released on a biennial or triennial schedule.

Data are gathered in the survey on the number and characteristics of individuals with education and employment in S&E fields. Together the NSCG, the National Survey of Recent College Graduates (NSRCG), and the Survey of Doctorate Recipients (SDR) comprise the Scientists and Engineers Statistical Data System (SESTAT).

The Demographic Statistics Methods Division (DSMD) of the U.S. Census Bureau provides statistical support on the NSCG to the National Science Foundation (NSF).

2. NSCG Design Changes

The NSCG is undergoing design/frame changes. Prior to recent years, the frame for the NSCG was the Decennial Census. As illustrated in Table 1, the NSCG was selected from the Decennial Census. It is a longitudinal survey in the sense that subjects selected for one NSCG survey are again eligible for future NSCG surveys.

Table 1: NSCG design prior to recent changes

Survey year	Frame			
2001 NSCG	2000 Decennial			
2003 NSCG	2001 NSCG	2000 Decennial		
2005 NSCG	2003 NSCG	2001 NSCG	2000 Decennial	
2007 NSCG	2005 NSCG	2003 NSCG	2001 NSCG	2000 Decennial

In the near future, the frame for the NSCG will be the American Community Survey (ACS; see NRC 2008). Table 2 illustrates the future design. Each NSCG survey takes respondents from four years of the ACS. Once a ACS respondent is a respondent to the NSCG the respondent stays in the NSCG for four rounds. Several years of ACS data are not used in estimation, weighting, or design of the NSCG.

Table 2: NSCG future design.

NSCG Survey Year	NSCG interview round				ACS years unused in NSCG
	4th	3rd	2nd	1st	
	ACS year used by round				
NSCG 2017	2009	2011	2013	2015	'16, '14, '12, '10, etc.
NSCG 2019	2011	2013	2015	2017	'18, '16, '14, '12, etc.
NSCG 2021	2013	2015	2017	2019	'20, '18, '16, '14, etc.

The design for NSCG 2010 is a hybrid that bridges the two designs. In 2010, there is a sample from the 2000 decennial census plus a supplement from the 2009 ACS. The 2000 decennial census was sub-sampled for the 2001, 2003, 2006, and 2008 NSCG surveys. A selection of individuals surveyed in those years is contacted again in 2010. Those individuals are referred to as the **Old Cohort**. The 2009 ACS was sub-sampled to add individuals to the 2010 NSCG. Those individuals are called the **New Cohort**. The new cohort has n=65,195 members. They are non-institutionalized, less than 76 years of age, and report having at least a bachelor's degree in ACS. The new cohort subsample from the 2009 ACS will be used in these analyses, because it has both ACS and NSG variables and will be similar to future scenario.

3. Project Overview

A project was funded by the U.S. Census Bureau (contract number YA-1323-SE-0066) to examine statistical issues in the NSCG. The four steps in the project are as follows.

1. Gather *documentation* on NSCG and ACS design and estimation;
2. Learn about the formation/use of *survey weights, estimation,* and *variance estimation* in NSCG (and ACS);
3. Investigate *models for data* in and between the NSCG and ACS; and
4. Conduct *analysis* on focal questions.

The first two steps have been accomplished and steps 3 and 4 are underway.

3.1 Gather documentation

Efforts to gather documentation yielded a variety and number of documents. Many documents were provided by staff at the U.S. Census Bureau. For NSCG and ACS, the documentation included the following classes of information. Some citations are listed by category in the References section.

- Survey questionnaires;
- Survey variable code books/dictionaries;
- Technical reports on sample selection and weighting (oversampling, non response adjustment, raking, post stratification) and other topics;
- Research reports and memos on survey design evaluation and study;
- Conference proceedings, such as those of the Joint Statistical Meetings (JSM) ; and
- Reports from the NSF to Congress and the public, including the NSF S&E Indicators (NSB 2012) and the NSF report on Women, Minorities, People with Disabilities (NSF 2013).

Generally not in existence are refereed journal articles on the survey design and estimation. Also generally not available are research articles based on data from the surveys. It appears that the primary use of the data is in reports from the NSF to Congress and the public.

3.2 Survey weighting

Survey weighting for the NSCG 2010 New Cohort involves several steps as described in several documents by Michael White (see references). The process of weighting the NSCG begins with the base weight, which comes from the ACS and sampling for the NSCG. After several steps, a final NSCG weight is produced. An outline of the steps based on the reports by Michael White is given below.

1. Base weight: sampling probability (ACS, NSCG)
2. Adjust weights of the sampling frame cases for duplication (8 pairs)
3. Development of the population controls – Female/Male; black/non black; young/old.

4. Weighting adjustment to account for the 2010 NSCG sample selection (adjust for non location of cases).
5. Identification of and adjustment for within-sample duplication (not needed for new cohort).
6. Weighting adjustment for unit nonresponse (propensity modeling).
7. Implementation of an iterative raking procedure: sex, race/ethnicity, S&E degree. Eight (8) iterations of raking were implemented to stabilize the weights
8. Identification of and adjustment for extreme weight values. This step was processed in the sampling cells and in three (3) primary analysis domains (PADs). Weights within cells were judged to be extreme if they were greater than the mean within the cell plus five (5) standard deviations within the cell: extreme > mean + 5 SD. Extreme weights are trimmed;
9. Reallocation of weights to address trimmed-weight bias: rake to sampling cell and secondary analysis domain SAD23 (ACS sex by broad occupation);
10. Construction of the 2010 NSCG final weights through an additional implementation of an iterative raking procedure.

Estimation in the 2010 NSCG uses the final weights for estimation and replicate variance estimation to estimate uncertainty associated with the estimates. Variance estimation was based on eighty (80) replicates with successive difference replication variance estimation (U.S. Census Bureau 2009; Fay and Train 1995). Issues concerning variance estimation were studied White and Opsomer (2011, 2012).

3.3. Modeling Overview

The modeling task of this project involves defining and estimating statistical models relating variables to one another within and across surveys. There are discrete and continuous variables in the ACS and NSCG.

Table 3 illustrates the structure of the surveys over time. The ACS is a cross sectional survey, whereas the NSCG is a rotating panel (longitudinal) survey. One can model aggregates across time for cross sectional surveys. One can model aggregates across time for longitudinal surveys, but one also can model the longitudinal nature of the data; i.e., the same people are respondents in multiple survey years. The NSCG is a subsample from a previous year ACS. Therefore one can model aggregates or micro data across time between the two surveys.

In summary, one can consider models relating

- ACS in year t, ACS in year $t+1$, ACS in year $t+2$, ... (*aggregates, not longitudinal*);
- NSCG in year $t+1$ and NSCG in year $t+3$ (*aggregate and longitudinal*); and
- ACS in year t and NSCG in year $t+1$ (*aggregate, subsample*).

Table 3: Modeling overview – structure of the surveys over time

Year t	Year $t+1$	Year $t+2$	Year $t+3$	...
ACS	ACS	ACS	ACS	ACS
	NSCG		NSCG	

3.4. Topics for Analysis

Initial discussions with staff at the NSF and at Census identified four (4) possible topics for study. The discussion produced a rank ordering of the importance and hence timing of the planned studies. It is not anticipated that all four studies will be completed fully during the contract period. The four topics in rank order of importance are listed below. The rest of this paper concerns the first topic estimation for small domains.

1. Estimation for small domains (small area estimation) – do small area models lead to improvements in mean squared error and are they acceptable in practice?
2. Updating NSCG survey weights for intermediate year ACS – does this improve estimation?
3. Estimation for NSCG variables in intermediate years when an ACS is collected but not a NSCG sample – can this provide adequate estimates between survey years?
4. Question block rotation strategies – how should one reduce respondent burden and survey cost over time by rotating blocks of questions across time?

4. Small Area Estimation: Estimation for Small Domains

The NSCG is designed to give sufficient accuracy at the national level and at the level of large regions of the country. There is an interest in estimation in *small areas* (e.g., states) and *small domains* (e.g., subgroups by demographics, including female/male, race/ethnicity, age, and other factors). Estimation methods that "borrow strength" across areas/domains can produce reductions in mean square error (MSE), especially for small areas and domains. Estimation methods that utilize information from multiple surveys (NSCG, ACS) could also produce gains in MSE.

Small "areas" of interest to NSF, as suggested in conversations with the NSF staff, include Hispanics with U.S. bachelor's degrees by broad occupation group and American Indians Alaskan Natives (AIAN) and Native Hawaiians and Pacific Islanders (NHPI) by broad occupation group. In the 2010 NSCG, based on publicly available data, there were 7,533 respondents of Hispanic origin, representing 9.8% of the sample. In 2010, there are 317 AIAN respondents and 307 NHPI respondents, each representing 0.4% of the sample. A variable USCAB is a variable that indicates whether the bachelor's degree is predicted to be a U.S. bachelor's degree.

More generally, there are several race/ethnicity categories with low counts by broad occupation code. Table 4 reports counts of respondents with U.S. bachelor's degrees (USCAB) by broad occupation and race/ethnicity for a selection of race/ethnicity groups. Whites and Asians are not included, because they are larger in number. These data come from the public use web site for the NSCG.

Beyond the cross of race/ethnicity with broad occupation, there are many other variables that, if used in tabulating sample numbers, would produce small counts in many cells. Considering additional variables creates many more small areas of potential concern. In the 2010 NSCG, ACS_RACETH has 6 levels, ACS_SEX has 2 levels, ACS_DEMGROUP includes two age groups, ACS_SE has two levels (S&E versus not),

and ACS_HIDEG has 3 levels (BA/BS; MA/MS; PhD). Fully crossed, there are 12*6*2*2*2*3 = 1,728 cells. Of course, other variables might also be considered.

Table 4: Some counts of USCAB respondents by broad occupation, race/ethnicity

Broad occupation	AIAN non-Hispanic	Black, non-Hispanic	Hispanic	Non-Hispanic NHPI	Multiple race
Computer & mathematical scientists	16	551	454	16	126
Biological, agricultural, & other life scientists	10	133	249	10	53
Physical and related scientists	14	107	200	7	58
Social and related scientists	7	146	224	6	53
Engineers	23	521	692	32	136
S&E related occupations	48	960	1,160	61	197
Non-S&E related occupations	126	3,244	3,274	127	645
Logical Skip	73	1,418	1,280	48	293
Total	317	7,080	7,533	307	1,561

4.1. Small area models

The models that are being considered for these data are small area models for cross classified nominal variables. Area level models and unit level models can be considered. Area level models specify relationships of expected counts to characteristics describing areas. Respondents have variation within their respective areas. Unit level models specify relationships of probabilities of cell membership based on individual respondents. The former (area level models) in this application might use age group (2 levels) as a covariate. The latter (unit level models) might use age as a covariate for membership in other (non-age) categories.

In the case of area level models,

- **Data** have a multinomial distribution with proportion parameters
- The **Prior** distribution on the proportions is a Dirichlet distribution
- The **Posterior** distribution for proportions is a Dirichlet. In this case, means, variances, and simulated values of proportions are simple to produce
- The **Predictive** distribution for unknown data is a multinomial distribution with sample size 1: simulated cell entries are possible based on observed cell information and draws of proportions from the posterior distribution.

The number of variables involved in defining the contingency table and how variables are used in the model define a series of models. A **Large model** could involve a full cross classification of variables and produce a saturated log linear model. **Reduced models** are log linear model with some higher order interactions set to zero.

Unit level modeling ideas will be discussed below as potential model extensions.

4.2. Small area modeling issues

A number of modeling issues arise with small area estimation in general and in this application in particular.

- **How should one select models?** The set of available models include fully saturated or reduced area level models and unit level models with various uses of covariates. AIC or BIC criteria can be used to select models. One also can examine residuals: differences between model estimates and empirical (direct survey) estimates. One expects some shrinkage under the small area estimates, especially in cells with small sample sizes, but one probably does not want all the estimates to be nearly equal to the mean given the size of the data set.
- **Use of design and other variables**: Additional variables (e.g., detailed occupations crossed with demographics) were used for sampling cells. Should models be made bigger to account for the levels of these variables?
- **Use of survey weights**: A population size by cell is implied by the sum of survey weights. Posterior mean value for proportions for unobserved cases could be used in estimation. Then the weighted posterior means could be used to produce a population-based estimate of small area size. Methods of incorporating survey weights into the analysis need to be considered.
- **Replicate survey weights**: Replicate weights could be used in place of final survey weights in the weighting procedure described in the previous bullet; this would enable use of successive difference replication variance estimation on the modeled data. The question arises, however, whether one should do this. Instead, in the Bayesian context, one can simulate the posterior variance or posterior interval about an estimate.
- **Evaluation of models and their impact on MSE**: In the context of the example, how does one judge whether the modeling effort has produced improved estimates. Simulation studies can suggest the quality of performance that should be anticipated. Other anecdotal evidence, such assessments of patterns and reasonableness within the data itself, also can be considered.

4.3. Extensions to small area models

A number of extensions to the small area models can be considered. One idea is described briefly as follows.

- For each category (small area domain or cell), one could model the propensity of being in that category – this is multinomial (polytomous) logistic regression.
- Some variables (e.g., Highest degree, Sex, Age group) would then be used as predictors of cell membership in the logistic regression models. Models could have main effects and some interactions.

- A unit level model could use additional variables (such as design variables) for each person in the sample. If the ACS frame includes all the unit level variables, then predictions can be formed for all ACS sample members.
- Prior distributions would be placed on model regression parameters. This produces a hierarchical polytomous logistic regression model.

References on small area estimation are numerous. A selection of references that could be useful in the current investigations include Berg and Fuller (2012), Chattopadhyay et al. (1999), Ghosh and Maiti (2004), Jiang and Lahiri (2006), Larsen (2003), Pfefferman (2013), Rao (2003), and Xie, Raghunathan, and Lepkowski (2007).

5. Conclusions and Future Work

Conditions in the National Survey of College Graduates (NSCG) seem right for trying small area estimation (SAE). There are many domains and it is a large data set but with small sample counts in some places and subgroups. Bayesian log linear models can be one approach to try. Other approaches can be compared. The plan for the near future is to continue research on SAE for NSCG 2010 using the new cohort of NSCG 2010 drawn from the 2009 ACS.

The long term goal for this project is to address four research topics of interest to the NSF and Census for the National Survey of College Graduates (NSCG).

- Estimation for small domains/areas
- Update weighting for NSCG using new ACS
- Intermediate year estimation for the NSCG
- Design and evaluation for rotation of question blocks

Acknowledgements

References

References are listed by category:

- NSCG10-WT documents (ordered by document number);
- Other NSCG10 documents (ordered by document number); and
- Other references (ordered alphabetically).

NSCG10-WT Documents

Walker, J., and Hall, D. (2010). 2010 National Survey of College Graduates: Weighting Specification for the 2009 American Community Survey Group Quarters Simplified Weights.   Document # NSCG10-WT-1 (11-01-10). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Walker, J., and Hall, D. (2010). 2010 National Survey of College Graduates: Weighting Specification for the 2009 American Community Survey Housing Unit Simplified Weights. Document # NSCG10-WT-2 (11-01-10) U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Walker, J., and Hall, D. (2010). 2010 National Survey of College Graduates Research: Evaluating the Calculation of Final Weights for the American Community Survey. Document # NSCG10-WT-3 (11-17-10) U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). 2010 National Survey of College Graduates Weighting Plan. Document # NSCG10-WT-4 (4.17.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Weighting Specification for the 2010 National Survey of College Graduates New Cohort. Document # NSCG10-WT-5 (5.29.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Weighting Specification for the 2010 National Survey of College Graduates Old Cohort. Document # NSCG10-WT-6 (4.18.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Combined Cohort Weighting Specification for the 2010 National Survey of College Graduates. Document # NSCG10-WT-7 (4.19.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Duplicate Identification and Weighting Adjustments for the 2010 National Survey of College Graduates. Document # NSCG10-WT-8 (4.18.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Other NSCG 2010 documents (RES, RESEARCH, SAMP, VAR)

Walker, J.M. (2012). 2010 NSCG Propensity Model Recommendation. Document # NSCG10-RES-1 (3.27.12). National Survey of College Graduates, Demographic Statistical Methods Division, U.S. Census Bureau.

White, M., and Hall, D. (2012). 2010 Weight Trimming Research Results. Document # NSCG10-RES-2 (DRAFT, 1.24.12). National Survey of College Graduates, Demographic Statistical Methods Division, U.S. Census Bureau.

Hall, D., and White, M. (2011). Proposal for the 2010 National Survey of College Graduates (NSCG) Mode Effects Study. Document #NSCG10-RESEARCH-1 (1.26.11). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Hall, D., and Walker, J. (2011). Proposal for the 2010 National Survey of College Graduates (NSCG) Web Impact Study. Document #NSCG10-RESEARCH-2 (1.26.11). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Hall, D. (2010). Selecting the Sample for the 2010 National Survey of College Graduates New Cohort Methodological Studies. Document # NSCG10-RESEARCH-3 (10.13.10). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2011). Selecting the Sample for the 2010 National Survey of College Graduates New Cohort Incentive Study. Document # NSCG10-RESEARCH-4 (5.27.11). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Hall, D. (2010). Sampling Specification for the 2010 National Survey of College Graduates. Document # NSCG10-SAMP-1 (10.28.10). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Finamore, J., and Hall, D. (2010). Eligibility and Stratification Definitions for the 2010 National Survey of College Graduates Sample Selection. Document # NSCG10-SAMP-2. U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Hall, D. (2010). Evaluation of Sample Selection Techniques for the 2010 National Survey of College Graduates. Document # NSCG10-SAMP-3 (11.17.10). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Creating the Replicate Weights File for the 2010 National Survey of College Graduates Old Cohort. Document # NSCG10-VAR-2 (10.17.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Creating the Replicate Weights File for the 2010 National Survey of College Graduates New Cohort. Document # NSCG10-VAR-3 (11.1.12). U.S.

Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2012). Creating the Replicate Weights File for the 2010 National Survey of College Graduates Combined Cohort. Document # NSCG10-VAR-4 (11.26.12). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

White, M. (2013). 2010 National Survey of College Graduates Replicate Weight Data Files. Document # NSCG10-VAR-5 (2.6.13). U.S. Census Bureau, Demographic Statistical Methods Division, Health Surveys and Supplements Branch.

Other references

Berg, E.J., and Fuller, W.A. (2012). Estimators of error covariance matrices for small area prediction. *Computational Statistics and Data Analysis.* 56(10): 2949-2962.

Beyler, N., Vizcarra, B., Jang, D., and Lin, X. (2011). Imputation Specifications for 2010 SESTAT. Washington, D.C.: Mathematica Policy Research, Mathematica Reference Number: 06850.463

Chattopadhyay, M., Lahiri, P., Larsen, M.D., and Reimnitz, J. (1999). Composite estimation of drug prevalences for sub-state areas. *Survey Methodology*, 25, 81-86.

Fay, R., and Train, (1995). Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. Alexandria, VA: Proceedings of the American Statistical Association, Survey Research Methods Section.

Finamore, J., Cohen S., Hall, D., Walker, J., and Jang, D. (2011). NSCG Estimation Issues when Using an ACS-based Sampling Frame. Alexandria, VA: Proceedings of the American Statistical Association, Survey Research Methods Section.

Ghosh, M., and Maiti, T. (2004). SAE based on natural exponential family quadratic variance function models and survey weights. *Biometrika,* 91(1): 95-112.

Hall, D., Cohen, S., Finamore, J., and Lan, F. (2011). NSCG Sampling Issues When Using an ACS-Based Sampling Frame. Alexandria, VA: Proceedings of the American Statistical Association, Survey Research Methods Section.

Jiang, J., and Lahiri, P. (2006). Mixed model prediction and small area estimation (with discussion). *TEST* 15: 1-96.

Larsen, M.D. (2003). Estimation of small-area proportions using covariates and survey data. *Journal of Statistical Planning and Inference,* 112 (1-2), 89 – 98.

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.

National Science Board. (2012). *Science and Engineering Indicators 2012*. Arlington VA: National Science Foundation (NSB 12-01).

National Science Foundation, National Center for Science and Engineering Statistics. (2013). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013.* Special Report NSF 13-304. Arlington, VA. Available at http://www.nsf.gov/statistics/wmpd/.

Pfefferman, D. (2013). New important developments in small area estimation. *Statistical Science,* 28(1): 40-68.

Rao, J.N.K. (2003). *Small area estimation.* Wiley-Interscience

U.S. Census Bureau. (2009). *Design and Methodology.* American Community Survey. U.S. Government Printing Office, Washington, DC, 2009.

White, M., and Opsomer, J.D. (2011). Variance estimation issues in the 2010 NSCG two-phase sample design. Alexandria, VA: Proceedings of the American Statistical Association, Survey Research Methods Section.

White, M., and Opsomer, J.D. (2012). Replicate Variance Estimation in a Two-Phase Sample Design Setting – Simulation Study with 2010 National Survey of College Graduates Data. Alexandria, VA: Proceedings of the American Statistical Association, Survey Research Methods Section.

Xie, D., Raghunathan, T.E., and Lepkowski, J.M. (2007). Estimation of the proportion of overweight individuals in small areas – a robust extension of the Fay-Herriot model. *Statistics in Medicine*, 26: 2699–2715.