Nonresponse Adjustment Methodology for NHIS-Medicare Linked Data

Michael D. Larsen[1], Michelle Roozeboom[2], and Kathy Schneider[2]

[1]Department of Statistics, The George Washington University,
Rockville, MD 20852, mlarsen@bsc.gwu.edu
[2]Buccaneer - a General Dynamics Company, 1401 50th St Suite 200, West Des Moines, Iowa 50266, Michelle.Roozeboom@vangent.com,
kathy.schneider@vangent.com

Abstract
National Health Interview Survey (NHIS) survey weights account for complex survey design, nonresponse, and post-stratification. The National Center for Health Statistics (NCHS) can link Medicare data to NHIS respondents. A self-selected subset of NHIS participants are not "eligible" for linkage to Medicare since they refused to provide a social security number (SSN), a Health Insurance Claim (HIC) number, or other key personal identifying information. In research with NCHS, we described how to assess potential linkage bias using NHIS 2005 with propensity score estimation as a weight adjustment methodology. This impact of adjustment on distributions of weights, point estimates, and estimated variance was quantified. Weight adjustments to fine groupings of subjects can reduce bias from nonresponse, but tend to produce more variable weights than coarser adjustments (a bias-variance trade-off). For most data users a single set of weights to reduce apparent nonresponse bias in survey variables would be sufficient, if weighting did not produce large increases in estimated variance. Other data users might wish to consider the nuances of developing adjusted weights. Recommended steps to be taken in forming adjusted weights are given.

Key Words: Bias-Variance trade-off; exact matching; propensity scores; record linkage; survey weights.

Acknowledgment and disclaimer: This work is based on *Nonresponse Adjustment Methodology for Linked Data*; a methodological guidance paper developed under contract GS-35F-0035L (Task 1) with the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS), September 2011. Thanks to Jennifer Park, Lisa Mirel, and others who participate in this project and provided comments. The authors assume sole responsibility for the contents of this proceedings paper. Opinions expressed are not necessarily shared by the CDC, NCHS, NHIS, NHANES, General Dynamics, Buccaneer, or Vangent. Substantial portions of the text are taken from the report *Nonresponse Adjustment Methodology for Linked Data*.

1. Introduction

The primary goal of this work is to appraise options for analyzing the National Health Interview Survey (NHIS)-Medicare linked data files. The primary goal is pursued through two aims. First, determine whether the NHIS survey weights alone are sufficient for analyzing the NHIS-Medicare linked data files. Second, evaluate and present options for weighting strategies to compensate for non response due to non-linkage, in order to avoid potential bias in estimates due to the linkage nonresponse. Section 2 briefly describes NHIS survey weighting. Section 3 discusses linkage of the NHIS survey to Centers for Medicare & Medicaid Services (CMS) Medicare databases. Section 4 comments on nonresponse bias. It outlines steps to be taken to address potential nonresponse bias through survey weighting. Section 5 concerns linkage probabilities in NHIS 2005. Section 6 lists some survey weight adjustment methods discussed in Roozeboom, Larsen, and Schneider (2011). Section 7 briefly describes application of ideas to the NHIS 2005-CMS linked file. Section 8 gives conclusions, a summary, and recommendations.

2. National Health Interview Survey Weighting

The National Health Interview Survey (NHIS) is a large-scale household interview survey of a statistically representative sample of the United States (U.S.) civilian, non-institutionalized population. Interviewers visit 35,000–40,000 households across the country and collect data regarding 75,000–100,000 individuals each year. Interviews consist of a broad range of health topics including health status, health care access, and use of health services. NHIS is an annual survey. The years of data available at the time of this study were 1994-2005.

NCHS surveys, such as the NHIS, identify samples for selection using **complex, multi-stage** probability sampling designs. The samples are selected so that participants are representative of the population of interest with **oversampling** performed on certain population subgroups to increase the reliability and precision of health status indicator estimates for these groups. **Weights** are created in NHIS to account for the complex ***survey design (including oversampling), survey nonresponse, and post-stratification***. When a sample is weighted in NHIS it is, therefore, representative of the U.S. civilian non-institutionalized Census population.

In NHIS, an individual is considered a non-respondent to the interview if he/she was selected to be in the sample, but **did not participate in the in-home interview.** Adjustments made for survey nonresponse account only for sample person interview non response, but not for component/item non response (i.e., a sample person declined to reveal his/her income but completed other interview components).

The survey sample weights on the NCHS public use files for NHIS are created using the initial probability of selection (which accounts for oversampling), adjustment for non response, and post stratification to Census population control totals. **One should use weights in analysis.**

3. Record Linkage of NHIS to CMS

NCHS has developed a record linkage program designed to maximize the scientific value of the Center's population-based surveys. These data are not publicly released, but rather are available for approved users through the NCHS Research Data Centers (RDCs). Medicare enrollment and claims data are available for those NHIS respondents who ***agreed*** to provide personal identification data to NHIS and for whom NCHS was ***able to match*** with Medicare administrative records. CMS provided NCHS with Medicare benefit claims data for 1991 through 2007 for all successfully matched NCHS survey participants.

For the purposes of investigating the NCHS-Medicare linkage, refusal or inability to support data linkage is considered “nonresponse”. These are not a random sample of respondents. Instead, this is a self-selected subset of the initial survey respondents. Therefore there is p***otential for bias due to non random non linkage.***

4. Nonresponse Bias

In order for nonresponse bias to be present, two conditions must be satisfied. First, there must be different linkage rates across subgroups. Second, there must be differences in distributions of variables between linked and unlinked individuals. In the application considered here, survey response variables are available for **both** linked and unlinked individuals, whereas CMS variables are available only for linked individuals.

As part of the report Roozeboom, Larsen, and Schneider (2011), a literature review on non response bias measurement and adjustment was conducted. The focus of this literature review was the methods and impacts of nonresponse weight adjustment. The review is being edited and prepared as a NCHS Series 2 Report with Jennifer Parker and Lisa Mirel. It is anticipated that the literature review series report will be released in 2012.

In order to address potential nonresponse bias due to non linkage through weight adjustments, three steps are recommended. These can be described under the headings “Probabilities”, “Distributions”, and “Adjustment”.

- **Probabilities**: Determine whether there is variation among subgroups in terms of non response/linkage rates with NHIS to the CMS Medicare data files.
- **Distributions**: Compare distributions of variables observed for all cases between those that are linked and those that are not linked to determine if there is potential nonresponse bias.
- **Adjustment**: Explore methodological options for adjusting for nonresponse bias. The evaluation of methods should include consideration of bias, variance, and operational feasibility and usability.

5. Linkage Probabilities in NHIS 2005

The National Health Interview Survey (NHIS) in 2005 was linked to CMS Medicare databases. Roozeboom, Larsen, and Schneider (2011) reports on the characteristics of the groups that can and cannot be linked. There is variation in linkage rates by many demographic factors and by geographical regions. There also is variation in linkage rates by response categories to several survey variables.

Reporting study-related variables in the survey (versus don’t know or refusal on items) is associated with higher rate of linkage. Often the fact that someone reported or did not report a variable is more predictive of linkage than the value of the variable itself. For example, the propensity for linkage was much lower for not reporting income than it was for either income under $20,000 or income over $20,000. Five survey variables (US Born, Marital Status, Self-Reported Health Status, Private Health Insurance, and Activity Limitation) are missing at low rates, but if they are missing, then successful linkage to CMS is less likely. A count variable (# of out five missing) works well for predicting probability of linkage.

Logistic regression was used to predict the probability of linkage. Estimates, standard errors, and P-values are given in Table 1 below. One should note that the results in Table 1 are specific to the NHIS 2005-CMS Medicare linkage and might not apply to other years and surveys. Females are less likely to have successful linkage. This could be related to a higher tendency to not have or use a social security number for some survey respondents. There is some variation across geographic regions. Not missing any of the five variables, reporting income (either under or over $20,000), and reporting high school graduation status (less then high school, high school graduate, or more than high school graduate) are strongly associated with being linkable. Other variables also were useful in predicting linkage status.

In summary, the NHIS survey variables in addition to the demographic variable are useful for predicting linkage status. Some effort at model selection was made in Roozeboom, Larsen, and Schneider (2011), but a summary of that effort and detailed comparisons of models are not given in this proceedings paper.

Table 1: Estimated coefficients, standard errors, and p-values for predicting NHIS-Medicare linkage using logistic regression.

Variable	Estimate	Standard error	P-value
Female	-0.17	0.04	<0.0001
Not missing any of 5 variables	1.12	0.23	<0.0001
Region: Midwest	0.29	0.08	<0.001
Region: Northeast	0.10	0.08	0.21
Region: South	0.18	0.08	0.02
Income reported	~1.21	0.08	<0.0001
Limited physical activity	0.26	0.08	0.002
Married	0.20	0.05	<0.0001
Not U.S. born	-0.41	0.08	<0.0001
High school graduation status reported	~1.30	0.15	<0.0001

A simple strategy of imputing missingness as a ‘least desirable’ category was examined. This seemed like a possible alternative when missingness was low. For example, one might impute poor health status when self reported status is missing or serious activity limitation when activity limitation is not reported. This did not work well in models for the linkage probability and cannot be recommended. Instead, imputation should consider prediction models more seriously.

6. Survey Weight Adjustment Methods

There are four primary methods for addressing nonresponse in a sample survey. The first three listed here involve using survey weight adjustments (e.g., Kish 1990, Kalton and Flores-Cervantes 2003). They are raking, post stratification, and weighting based on propensity estimation. The fourth method for addressing nonresponse is through statistical modeling, which can lead to imputations for the missing data or statistical analysis that explicitly takes into consideration the cases with missing values.

Mirel et al. (2010) describe two weighting alternatives as a weighting class approach and a model-based weighting approach. The model-based weighting approach is encompassed by the propensity estimation weighting approach described below, whereas the weighting class approach is in the realm of raking and post stratification. The Research Triangle Institute (RTI 2008) implements a general methodology that enables raking, post stratification, and propensity estimation weight adjustment. Witt (2009) categorizes weight adjustments into nonresponse weight adjustments and post-stratification weight adjustments, also called benchmark adjustment factors. Särndal and Lundström (2005) discuss the use of auxiliary information, weighting adjustments, calibration weighting (which will be described below), and imputation of missing data. Little (1993) considers post stratification from a Bayesian perspective. Gelman (2007) discusses challenges involved in performing weight adjustments, post-stratification, and statistical modeling of survey data.

Before any measure to address nonresponse is taken in a complex scientific sample survey, a survey design weight typically is computed (Lohr 1999; Groves et al. 2004; Särndal, Swennson, and Wretman 2011; Kish 1990). This weight is often thought of as the inverse of the probability that a unit is selected into a sample. The sum of these weights then equals the population size. Weighted estimation of totals and means using these design weights then adjusts the usage of the sample values away from equal weighting so that they are representative of the target population. In reality, the weights can reflect a number of other considerations besides unequal probabilities of selection, such as control totals, benchmarks, and upper or lower bounds on weights (Groves et al. 2004, section 10.5). In the discussion that follows, the survey weights before adjustment for nonresponse will be called the design or sample weights. NCHS has implemented this step for its various surveys, but not additional weights for the CMS-linked subsample. Reviews of various methods were presented in Roozeboom, Larsen, and Schneider (2011). A NCHS Series 2 Report with Jennifer Parker and Lisa Mirel presenting the literature review is currently under review for publication.

Weighting-class adjustments are described in Lohr (1999; section 8.5.1). In this method, the sample is divided into classes based on variables known for all units in the sample. Variables typically eligible for such partitioning of the sample include strata and cluster information, variables from previous data collection (previous waves, phases, or stages) of the survey, and administrative information on the population. Within each class, the totals of the design weights overall and for the subset of respondents are computed. Then, within each class, the design weights for the respondents are multiplied (inflated) by the ratio of the weight total overall to the weight total for the subset of respondents (i.e., a subpopulation). The total of the respondents’ adjusted weights then equals the design weight total overall.

Raking is a procedure (also called iterative proportional fitting) that iteratively adjusts weight totals within sequences of weighting classes formed by a series of variable sets. A variable set is a collection of a few variables, typically two to three but possibly more. Examples include age-sex-race categories, education-sex categories, and race-owner/renter. Geographical controls, such as state or county, also could be crossed with these variables. Raking is used when one wants, in principle, to do weighting class adjustments but there is more than one set of weighting classes that cannot be computed simultaneously.

Post stratification involves adjusting the sample weights within groups formed after the sample is selected ("post strata") to match totals in the population (e.g., for NHIS, it is civilian, non-institutionalized people in the U.S.). Post strata are typically defined by crossing two or more categorical variables measured on the sample and about which information is known in the population. The cross of several variables or variables with small categories can yield some cells with very small, even empty, counts. In these cases post stratification can run into problems. For empty cell counts in the sample, the adjustment is not defined. For small counts in the sample, the adjustment can be extreme. To deal with this aspect, cell collapsing is sometimes used. Cell collapsing combines post strata cells in order to increase sample counts in the combined cell.

In this NHIS survey-Medicare linkage study, there are extensive variables recorded on all original sample respondents that could be used in such post stratification adjustments for nonresponse due to non-linkage. Post stratification weight adjustments for nonresponse for linking to Medicare claims would make the respondents who link similar to the survey sample for classification variables of interest.

Calibration methods in survey sampling (Deville and Särndal 1992; Kim and Park 2010; Särndal 2007) allow one to adjust survey weights so that they are close to initial weights, such as the sampling design weights, but satisfy certain constraints. The connection to raking adjustment was demonstrated in Deville, Särndal, and Sautory (1993). The closeness of the weights is described by a distance function. Calibration can be implemented in a way to control the minimum and maximum value of weights and to match one or more control totals. It is therefore a very flexible methodology. Zhang (2000) describes how calibration can produce adjusted weights equivalent to those produced with post stratification.

Additional discussion of survey weight adjustment can be found in Gelman (2007) and Little (1993). Detailed adjustments to weights probably reduce bias more than gross adjustments, but highly variable weights can produce variable estimates. Weight trimming reduces range of weights and rescales to the total (by group/class/post stratum). Doing so tends to reduce variance, but can introduce bias.

7. Application to NHIS 2005-CMS Linkage

Four sets of adjusted weights were created to adjust the NHIS 2005 survey weights for nonresponse due to non linkage. Two sets used post stratification: one of these sets used age, race and gender cross classifications and the other set added cross classification by region. Two sets used propensity score estimation utilizing survey variables in addition to demographics: one of these sets used eight variables. The eight variable model was selected using the Akaike's Information Criterion (AIC). The other propensity score set used ten variables. The ten variable model added demographics not selected in the model with eight variables.

Table 2 summarizes the weight distribution after adjustment of public use NHIS final survey weights for nonresponse due to non linkage by each of the four methods. The two demographic methods produce substantially less variability in weights than do the methods based on propensity score estimation for weight adjustment.

Table 2: Summary of weight distribution after adjustment of public use NHIS final survey weights for nonreponse due to non linkage by four methods.

Weight model	Standard Deviation	Minimum	Maximum
Weights adjusted to match demographics (age, race, gender)	0.46	2.38	6.6
Weights adjusted to match demographics (age, race, gender) and regional segmentation	0.76	1.56	15.2
8 variable logistic regression estimation of propensity scores for weight adjustment	4.63	1.51	64.7
10 variable logistic regression estimation of propensity scores for weight adjustment	4.57	1.52	73.7

Table 3 presents estimates of BMI and height using weights from the four adjustment methods. Mean estimates and estimated standard errors (SE) are reported. As can be seen in the table, mean estimates are similar among the two demographic post-stratification and among the two propensity score weighting methods, but different across the two types of methods. One could speculate that the propensity score weighting methods are removing more nonresponse bias. Further research is needed to verify whether this is indeed the case. Standard errors also vary by type of method. Standard errors are smaller for the two demographic post-stratification results and higher for the two propensity score weighting results. The preference for one set of weight adjustment versus another is beyond the scope of this comparison. The differences in Table 3 illustrate the bias-variance trade-off that is at the center of the debate about weight adjustments for potential nonresponse bias.

Table 3: Estimates of BMI and Height using weights from four weight adjustment methods. Mean estimates and estimated standard errors (SE) are reported.

Weight model	BMI		Height (in)	
	Mean	SE	Mean	SE
Weights adjusted to match demographics (age, race, gender)	30.16	0.21	68.27	0.125
Weights adjusted to match demographics (age, race, gender) and regional segmentation	30.16	0.22	68.31	0.131
8 variable logistic regression estimation of propensity scores for weight adjustment	31.08	0.38	68.60	0.198
10 variable logistic regression estimation of propensity scores for weight adjustment	31.14	0.39	68.62	0.198

9. Conclusion and Summary

NHIS survey weights account for complex survey design, nonresponse, and post-stratification. The NCHS link Medicare data to NHIS respondents who agree to linkage and provide sufficient information to enable linkage. Some NHIS participants are not "eligible" for linkage to Medicare. One option is to adjust survey weights to reduce potential nonresponse bias due to non linkage. This paper has summarized ideas for assessing potential linkage bias and performing weight adjustment. Ideas were illustrated with the 2005 NHIS-CMS Medicare linkage. Weight adjustments to fine groupings of subjects can reduce bias from nonresponse, but tend to produce more variable weights than coarser adjustments (a bias-variance trade-off).

This paper has summarized work from Roozeboom, Larsen, and Schneider (2011) that was a methodological guidance paper developed under contract GS-35F-0035L (Task 1) with the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS), September 2011. Substantial portions of the text are taken from the *Nonresponse Adjustment Methodology for Linked Data*. Further details can be found in that report.

Based on this research, it seems likely that it should be possible to create one set of weights adjusted for nonresponse due to non linkage for each survey-CMS linkage that has low variability. Such a set of weights would function much as survey weights for the original surveys do now, but they would be slightly adjusted for the non linkage. The user community should be able to use such weights in much the same manner as they use survey weights now.

In addition, it should be possible to describe the weight adjustment and evaluation procedures so that researchers can make weights for special purposes. That is, a researcher with sufficient technical background and interest in a specific topic and subgroup might be able to perform specialized weight adjustment in order to remove more potential nonresponse bias for a particular analysis. The researcher can be encouraged to report on the rationale and steps taken in the weighting and adjustment process. An alternative for researchers who do not think the original or general nonlinkage adjusted weights are sufficient is to use model-based analysis to adjust further for nonlinkage. In any case, part of the discussion could focus on the inherent bias-variance tradeoff and sensitivity (or lack thereof) of results to choices of methods.

Suggested steps for forming nonresponse due to nonlinkage weight adjustments are as follows.

1. Examine linkage rates by groups.
2. Examine distributions by groups.
3. Consider options for weight adjustments for non linkage.
4. Quantify weight distribution before and after adjustment.
5. Quantify impact of adjustments on estimates.
6. Quantify impact of adjustments on estimated standard errors.
7. Acknowledge weighting process in publications.
8. Report sensitivity of results to alternative weights.

References

Deville, J. C., and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87(418): 376-382.

Deville, J. C., and Särndal, C. E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88(423): 1013-1020.

Gelman, A. (2007). Struggles with survey weighting and regression modeling (with discussion). *Statistical Science*, 22(2): 153-164.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer E., and Tourangeau, R. (2004). *Survey Methodology*. John Wiley & Sons, Inc.: Hoboken, NJ.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19, 81-97.

Kim, J. K., and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, 78(1): 21-39.

Kish, L. (1990). Weighting: Why, when, and how? *Proceedings of the Section on Survey Research Methods of the American Statistical Association.* American Statistical Association: Alexandria, VA: 121-130.

Little, R. J. A. (1993). Post-stratification - A modeler's perspective. *Journal of the American Statistical Association*, 88(423): 1001-1012.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. Brooks/Cole Publishing: Pacific Grove, CA.

Mirel, L. B., Burt, V., Curtin, L. R., and Zhang, C. (2010). Different approaches for non-response adjustments to statistical weights in the continuous NHAHES (2003-04). Federal Committee on Statistical Methodology Research Conference.

Research Triangle Institute (2008). *SUDAAN Language Manual, Release 10.0*. Research Triangle Institute: Research Triangle Park, NC.

Roozeboom, M., Larsen, M.D., and Schneider, K. (2011). *Nonresponse Adjustment Methodology for Linked Data*, *September 2011*; a methodological guidance paper developed under contract GS-35F-0035L (Task 1) with the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS).

Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33(2): 99-119.

Särndal, C. E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, England.

Särndal, Swennson, and Wretman (2011). *Model Assisted Survey Sampling* (Springer Series in Statistics). Springer-Verlag**.**

Witt, M. B. (2009). Overview of software that will produce sample weight adjustments. *Proceedings of the Section on Survey Research Methods of the American Statistical Association.* American Statistical Association: Alexandria, VA, 3009-3023.

Zhang, L. C. (2000). Post-stratification and calibration - A synthesis. *American Statistician*, 54(3): 178-184.