COMPARISON OF NONRESPONSE ADJUSTMENT METHODS IN A HEALTH SURVEY

Todd L. Beck and Julia L. Bienias, Rush Institute For Healthy Aging
Todd L. Beck, Rush Institute For Healthy Aging, 1645 W. Jackson, Ste. 675, Chicago, IL 60612

KEY WORDS: missing data, weighting cells, propensity score methods

Abstract. The Chicago Health and Aging Project (CHAP) is a longitudinal community study examining risk factors for chronic health problems of older adults. The first wave of data collection, completed in 1997, comprised an extensive baseline population interview and a detailed clinical evaluation. To provide for more intensive data collection, a stratified random sample of people completing the population interview was chosen to participate in the clinical evaluation; the overall participation rate among non-decedents was 76%. Our goal was to examine different possible methods of adjusting for unit nonresponse in this sample. Because of the nature of a health study of older persons, one cannot assume the data are missing completely at random. A good case can be made for not making any nonresponse adjustments and simply generalizing the results to the population of participants. To more completely address this issue, we examined weighting class and propensity score methods to construct candidate nonresponse adjustment factors. We computed propensity classes based on several competing logistic regression models predicting an individual's propensity to participate in the study. We considered weighting class adjustments based on the survey design and on variables found significant in the logistic regression model. We compared these approaches (and making no adjustment) on the estimation of various quantities of interest.

I. Introduction

Nonresponse is an important consideration in data collection, particularly when it is suspected that the underlying mechanism determining response is related to the questions of interest. In health surveys, people may fail to respond to individual items or to the entire survey for a variety of reasons, including ill health. Nonresponse, and particularly unit nonresponse, can thus potentially be a serious problem when drawing inferences in health surveys. Hence it is important to both understand the possible mechanisms for nonresponse and to consider various methods of possible adjustment.

This paper examines the impact of unit nonresponse in the clinical evaluation sample of the first wave of a large community-based prospective epidemiologic study, the Chicago Health and Aging Project. The severity of the nonresponse is evaluated by examining the response rates, unweighted and weighted, across variables of interest. Next we examined weighting class and propensity-based methods to create nonresponse adjustment weights. We compared the different weights by examining a set of statistical analyses representing a range of scientific questions of interest, including estimation of population disease prevalence. Previous researchers have taken a similar approach (e.g., Khare, Mohadjer, Ezzati-Rice, & Waksberg, 1994; Mohadjer, Montaquila, Waksberg, Bell, James, Flores-Cervantes, & Montes, 1996)

II. The Chicago Health and Aging Project

The Chicago Health and Aging Project (CHAP) population was defined by a door-to-door census of all households in a geographically defined, biracial community on the south side of Chicago. All residents who were 65 years of age or older were then scheduled for an in-home interview that consisted of a combination of structured questions, performance tests, and physiological measures. The interview addressed a broad range of health and social factors relating to community-dwelling older adults. Of 8,501 age-eligible residents identified in the initial canvass, 7,813 were still living in the community at the time of the population interview, and 6,158 participated in the interview. A random sample of those interviewed, stratified by age (5 levels), gender (2), race (2), and cognitive performance (3) was then selected, using Poisson sampling, for a detailed clinical evaluation, which included further cognitive and physical assessments and neurological testing. Of the 1,055 subjects sampled for this first clinical evaluation, 86 died and 9 moved away before being able to participate, and 729 of the remaining 960 were evaluated. (See Wilson et al., 1999, for a further description of the study.) For purposes of the current analysis we considered 960 to be the selected sample. The study was approved by the Human Investigation Committee of Rush-Presbyterian-St. Luke's Medical Center.

III. Evaluation of Response Rates

The first step in determining the potential for bias in the survey is to assess the level of missing data. The potential for bias increases as the response rate decreases. We examined both the weighted and unweighted response rates for the first wave's clinical evaluation sample. The unweighted response rate is computed from the sample data as the ratio of the number of completed interviews to the total sample size. The weighted response rate is computed as the ratio of the sum of the sampling weights for those participating in the clinical evaluation to the sum of the weights for all those eligible for the clinical evaluation sample. The overall unweighted and weighted response rates were 75.9% and 73.6%, respectively. Figure 1 shows a much lower response rates among Blacks. Blacks have a weighted response rate of 65% compared to 87% for Non-Blacks. Figure 2 shows there was nearly equal participation among males and females. Figure 3 shows response rates in five year age groups tend to go up with age until the 85 and older category, where they decline slightly.

IV. Nonresponse Adjustment Methods

A. Weighting Classes

A common method of adjusting for survey nonresponse is the use of adjustment cells. Here individuals are placed into mutually exclusive cells. An example would be a combination of demographic variables such as race, gender and age. It is assumed that individuals within each group are approximately homogeneous with respect to their propensity to respond to the survey. Each individual in a given cell is then given an adjustment weight, which is the inverse of the mean response rate within the given cell. Note that these cells do not have to be the same as any sampling strata or other groupings. Variables that are available for both respondents and non-respondents can be used.

B. Propensity Stratification and Propensity Weighting

For some general discussion of propensity stratification and propensity weighting, see Rosenbaum and Rubin (1983) and Little (1986). In both methods one forms a set of weights to adjust for unit nonresponse. The calculation of weights for the two methods requires three main steps. The first is the calculation of each individual's estimated propensity of responding. This is done by fitting a logistic regression model to the respondents and non-respondents. The second step is the assignment of each individual into a propensity class. This is done by grouping the predicted propensities into a number of quantiles. Eltinge and Yansaneh (1997) gives a guide to forming these adjustment classes. The third step is the calculation of a weight for each propensity class.

The first method, propensity stratification, uses the individual propensity scores only in assigning an individual to a particular propensity class. Here, m classes can be defined by quantiles or some otherwise defined cut points of the propensity scores. The adjustment factor for propensity stratification can then be calculated as the inverse of the *mean number of individuals with complete data* from a given one of the m classes.

As with propensity stratification, propensity weighting uses quantiles or some otherwise defined cut points to divide the individuals into m propensity classes. The adjustment factors for propensity weighting are then calculated as the inverse of the *mean propensity* of all individuals falling into the mth class.

V. Computing Nonresponse Adjustments for the CHAP Clinical Evaluation Sample

For this project, the nonresponse to be studied is nonparticipation ("unit" nonresponse) in the CHAP baseline (first wave) clinical evaluation. We began the analysis with weights for the clinical evaluation data that account only for the stratified random sampling. We then applied the adjustment cell and propensity score methods to create additional weighting factors to adjust for possible nonresponse bias. Estimates based on using these methods are then compared to those based on using only the sampling weights and of using no weights.

To obtain unit response propensity estimates, we fit logistic regression models with the outcome being 1=Agreed to Participate in the clinical evaluation and 0=Did not agree to participate based on N=960 persons. Preliminary investigation of response rates by stratum indicated that race seemed to be the only stratum variable with differential participation, and that the effect was essentially constant across the other strata. Thus we included race in all models, and then examined the effects of several additional prediction variables. These variables generally fell into one of four categories: socioeconomic, cognitive function, physical function or health history. A best fitting model was then built up from this set of predictor variables. Table 1 reports the final logistic regression estimated odds ratios and Wald Chi-Square p-values. Variables included in the final model include indicators for race and marital status, years of education, an education by race interaction, Katz Activities of Daily Living scale (Katz & Akpom, 1976), and mini-mental status examination score (MMSE; Folstein, Folstein, & McHugh, 1975). The Katz scale, which is measure of physical function, ranges from 0 to 6 (excluding an item for grooming). The MMSE, which ranges from 0 to 30, is a measure of cognitive function.

Table 1. Parameter Estimates for Final Logistic Regression Model Predicting Participation in the CHAP Clinical Evaluation.

Factor	Odds Ratio	p-value for Wald Chi-Square
Intercept	----	0.0074
Race (Black/Non-Black)[a]	1.388	0.5980
Education	1.143	0.0025
Education*Race	0.891	0.0256
Katz	0.871	0.0380
Marital Status (Married/ Not Married)[a]	0.601	0.0017
MMSE	0.978	0.0269

[a]The first category is coded as "1," the second as "0"

The race by education interaction is illustrated in Figure 4. As can be seen in the figure, predicted response propensity was fairly constant for Blacks at all levels of education, but for Non-Blacks, the propensity to respond increased with increasing education.

The estimated propensities from the model presented in Table 1 were used to create classes for propensity weighting and propensity stratification. The classes were created by placing the propensities into six quantiles. Weights were then created as described in the previous section. For the weighting classes analysis, participants were placed into eight classes defined by two groups of race and four groups of education. We also generated adjustment weights using the sixty design strata as the weighting class cells.

VI. Results

A. Basic Descriptive Statistics

To compare some of the qualities of the different sets of weights, we first computed basic frequencies and means on some of the key demographic and health variables that describe the population. These weighted frequencies and means can then be compared back to those from the population (as represented by the sampling frame from which the clinical evaluation sample was drawn). This is simply allowing us to see how well each weighting method adjusts the sample back to the population. Table 2a lists the frequencies of race, gender, age, education and marital status for both the population and the unweighted and weighted sample. Here you can see that using only the sampling weights shows no evidence of upweighting the proportion of African Americans back up to the proportion found in the population, whereas each of the four nonresponse adjustment weights brings the sample proportion of African Americans up to reflect that of the population. Weighting classes, especially the one using the sixty strata to define the classes, appear to adjust well for gender and age. However, this is not surprising, as these are both stratification variables. As with race, the stratified sample weights alone also do a relatively poor job of adjusting for education, especially in the two higher education groupings. The propensity-based adjustments reflect marital status in the population better than weighting classes or the stratified sample weights. And as a general rule the propensity adjustments tend to give the most consistent estimators, regardless of whether or not the variable of interest was included in the logistic model from which the propensities were obtained.

There are only minimal differences among the weighted means listed in Table 2b. However, one can be clearly seen: In every case, using a nonresponse adjustment gives an estimate that is as close or closer to the population than the estimate using only the sampling weights.

B. Estimates of Disease Prevalence

A primary goal of the first wave of the CHAP study is the estimation of disease prevalence in the community as a whole and in important subgroups. Therefore, we examined nonresponse adjustment methods for their impact on these estimates. We illustrate with cerebrovascular disease (stroke).

In order to maximize power for prevalence estimation, a stratified design was chosen for CHAP that allowed for oversampling in the strata with the oldest people and those in poorer memory groups in order to deliberately enrich the sample with the highest expected prevalence of disease. Due to the disparity between the sample and the community under this design, estimating disease prevalence requires adjustments. However, conventional direct upweighting of the individual stratum prevalences may not perform well if the number of strata is large or the sample sizes in some strata are small. Therefore, we used the empirical Bayes approach of Beckett and Tancredi (in press) to estimate the prevalence of cerebrovascular disease in the community.

Each of the weights described in the Section V were used to estimate the prevalence of clinical diagnosis of probable stroke, first for the community as a whole and then in subgroups defined by race and gender. The overall prevalence rate of stroke in the community is estimated to be 10.52% using the standard sampling weights. The estimated prevalence is slightly smaller, 10.46%, using the eight group weighting class weights and a slightly higher, 10.58%, when the design strata are used as weighting classes. However, propensity weighing and propensity stratification give lower estimates of 10.08% and 10.10% respectively.

A similar pattern exists when looking at the prevalence rates by race and gender. These estimates are presented in Figures 5 and 6. Estimates using standard sampling weights and weighting classes show very similar estimates within each of the race and gender subgroups. However, a difference can be seen in the propensity methods. Here, the estimates for Non-Blacks and women remain similar to the methods that used the standard sampling weights or the weighting classes but the estimates of Blacks and men have both been reduced in comparison to the other methods. One difficulty here is that we no longer have frame values, so we have no way of knowing which estimates are "true."

VII. Conclusions

In this paper we presented an approach to the comparison of candidate adjustments for unit nonresponse and some preliminary results from the clinical evaluation component of the first wave of a longitudinal health survey. Our tentative conclusion is that either propensity-based method looks promising. However, we intend to examine more variables and models, including the impact of these methods on risk-factor regression analysis before making a recommendation. We also plan to expand the propensity model to incorporate nonresponse due to death. Eventually, we hope to expand this work to incidence analysis based on multiple waves of data.

VIII. References

Beckett, L. A., & Tancredi, D. J. (in press). Parametric empirical Bayes estimates of disease prevalence using stratified samples from community populations. *Statistics in Medicine.*

Eltinge, J.L., Yansaneh, I.S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey. *Survey Methodology*, **23**, 33-40.

Folstein, M. J., Folstein, S. E., & McHugh, P. R. (1975). Mini mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189-198.

Katz, S., & Akpom, C. A. (1976). A measure of primary sociobiological functions. *International Journal of Health Services*, **6**, 493-509.

Khare, M., Mohadjer, L.K., Ezzati-Rice, T. M., & Waksberg, J. (1994). An evaluation of nonresponse bias in NHANES III. *Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods*, 949-954. Alexandria, VA: American Statistical Association.

Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistics Review*, **54**, 139-157.

Mohadjer, L., Montaquila, J., Waksberg, J., Bell, B., James, P., Flores-Cervantes, I., & Montes, M. (1996, Feb.). National health and nutrition examination survey III: Weighting and estimation methodology. Technical report, available from Westat, Inc., or the National Center for Health Statistics.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in the observational studies for casual effects. *Biometrika*, **70**, 41-55.

Wilson, R. S., Bennett, D. A., Beckett, L. A., Morris, M. C., Gilley, D. W., Bienias, J. L., Scherr, P. A., & Evans, D. A. (1999). Cognitive activity in older persons from a geographically defined population. *Journal of Gerontology: Psychological Sciences*, **54B**, P155-P160.

Table 2. Basic Descriptive Statistics Computed in the Population and in the Sample, Using Different Methods for Computing Nonresponse Adjustments.

A. Frequencies (Percentage Estimates)

	Population (Sampling Frame)	No Weights	Sample Weights Only	w/Propensity Weighting	w/Propensity Stratification	w/Weighting Classes using Race x Educ.	w/Weighting Classes Using All 60 Strata
% Black	61.6	53.8	53.7	60.3	0.2	59.5	0.8
% Male	39.3	47.7	41.1	42.2	42.1	40.9	41.3
Age Group							
65-69	29.4	14.7	26.5	27.7	27.6	26.7	28.6
70-74	29.1	18.9	31.0	31.7	31.7	31.3	31.8
75-79	18.3	16.9	19.3	18.9	19.0	19.1	18.6
80-84	12.0	23.5	12.9	12.4	12.5	12.9	11.5
85+	11.2	26.1	10.2	9.2	9.9	9.9	9.6
Education (years)							
0-8	19.5	28.3	21.2	22.1	22.9	22.4	22.4
9-12	45.9	40.5	39.8	41.4	41.4	41.4	41.1
13+	34.6	31.3	39.1	35.7	35.7	36.1	36.5
% Married	47.8	38.7	44.1	47.1	47.1	43.9	44.5

B. Means

	Population (Sampling Frame)	No Weights	Sample Weights Only	w/Propensity Weighting	w/Propensity Stratification	w/Weighting Classes using Race x Educ.	w/Weighting Classes Using All 60 Strata
Education (Yrs.)	11.8	11.3	12.1	11.8	11.8	11.9	11.9
Occupation[a]	32.8	33.2	34.7	33.6	33.6	33.6	33.7
MMSE[b]	25.3	22.9	25.7	25.8	25.8	25.6	25.7
SDMT[c]	25.8	21.1	27.5	27.1	27.1	26.7	27.0
EBMT[d]	7.9	6.5	8.1	8.0	8.0	8.0	8.1
Katz[e]	5.5	5.2	5.6	5.6	5.6	5.6	5.6
SBP[f]	139.4	140.0	140.1	140.2	140.2	140.0	140.2
DBP[g]	76.9	77.1	77.7	77.4	77.4	77.2	77.4

a Occupational status code
b Mini Mental Status Examination score
c Symbol Digit Test
d East Boston Memory Test
e Katz Activities of Daily Living Scale, number able to do without difficulty
f Systolic Blood Pressure
g Diastolic Blood Pressure

FIGURE 1: Weighted Participation by Race

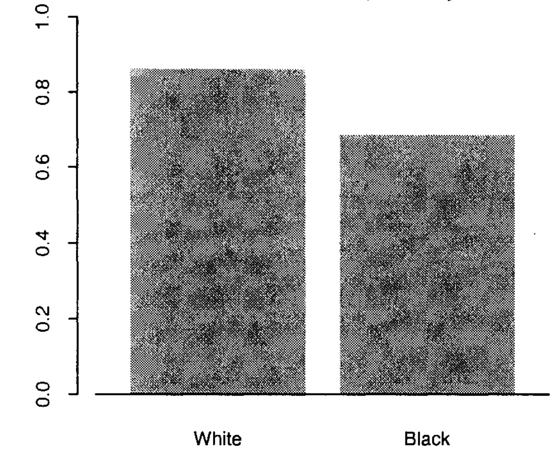

FIGURE 2: Weighted Participation by Gender

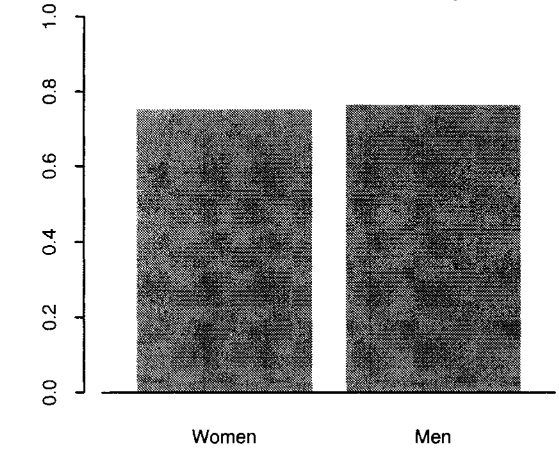

FIGURE 3: Weighted Participation by Age Group

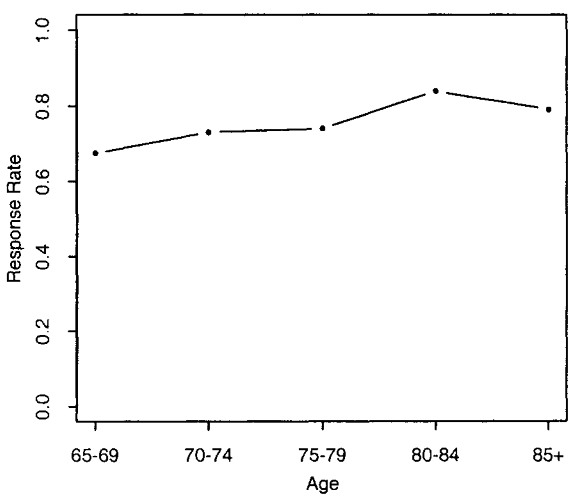

FIGURE 4: Predicted Propensity by Education

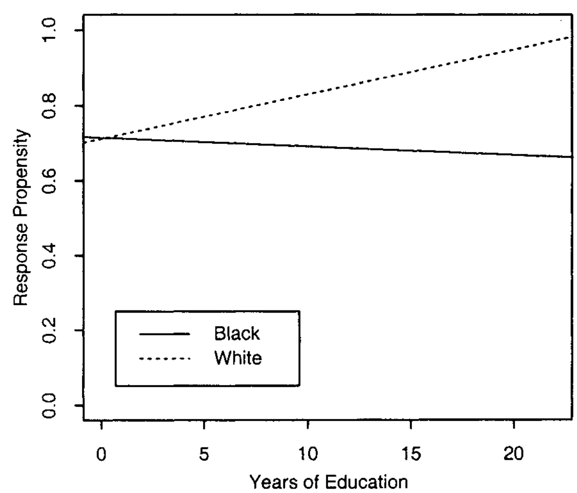

FIGURE 5: Estimated Prevalence of Stroke by Race

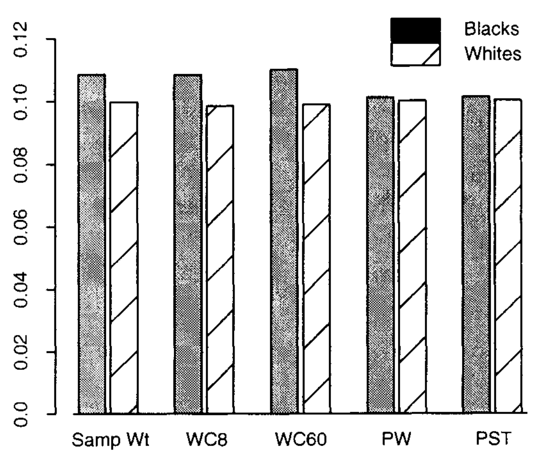

FIGURE 6: Estimated Prevalence of Stroke by Gender