

Weighting Class versus Propensity Model Approaches to Nonresponse Adjustment: The SDR Experience

Y. Michael Yang¹, Yongyi Wang²

¹ICF International, 9300 Lee Highway, Fairfax, VA 22031

²NORC at the University of Chicago, 55 East Monroe Street, Chicago, IL 60603

Abstract

The Survey of Doctorate Recipients (SDR) is a longitudinal survey of U.S. residents with U.S.-granted doctoral degrees in science and engineering (S&E). There are two general approaches to nonresponse adjustments: the traditional weighting class approach and the relatively new propensity model approach. Since its inception, the SDR has used the weighting class approach to compensate for nonresponse. This approach incorporates a small set of variables in defining the weighting classes even though additional variables may also contribute to bias reduction. Changing to the propensity model approach would help overcome this limitation. Under the propensity model approach, weighting adjustments would be based on propensity scores predicted by multivariate models. Using the 2006 SDR data, this paper compares the two weighting approaches in terms of the distribution of the weights and the weighted estimates of key SDR variables. The distributions of the weights under the two approaches are almost identical, except that weights based on the propensity model tend to have slightly smaller variance. The weighted estimates of key variables also have similar distributions except that, under the propensity model approach, they tend to have slightly smaller design effects due to unequal weighting. We conclude that the propensity model approach may represent a viable alternative to the current weighting class approach. The propensity models may be further refined and future SDRs may consider adopting the new approach to nonresponse adjustments.

Key Words: Weighting class adjustment, propensity model adjustment, design effect

1. Introduction

The Survey of Doctorate Recipients (SDR) is a longitudinal survey of U.S. residents with U.S.-granted doctoral degrees in science and engineering (S&E). The SDR is one of three large-scale surveys sponsored by the National Science Foundation (NSF) that are integrated to create the Scientists and Engineers Statistical Data System (SESTAT). S&E doctorates are sampled shortly after degree receipt and followed until age 76 to collect demographic, education and career history information.

The target population of the 2006 SDR includes individuals who (1) received a doctoral degree in an S&E field from a U.S. institution, (2) was 75 years of age or younger on April 1, 2006, and (3) lived in the U.S. in a noninstitutionalized setting on April 1, 2006. As in prior cycles, the 2006 SDR sample frame is constructed as two separate databases, the new cohort frame and the old cohort frame. The new cohort frame includes 77,834 doctoral recipients who received their degrees between July 1, 2003 and June 30, 2006. The old cohort frame is developed from the 2003 SDR sample. Both respondents and nonrespondents of the 2003 SDR sample are included in the 2006 SDR old cohort frame insofar as they are not determined to be permanently ineligible (e.g., age ineligible, deceased, permanently incapacitated, etc.). Of the 40,000 cases in the 2003 SDR sample, 38,803 are included in the 2006 SDR old cohort frame. Therefore, the 2006 SDR frame contains a total of 116,637 cases.

The 2006 SDR used a stratified sample design. It contains 164 strata defined by a recoded demographic group variable, two degree field variables, and sex. The sample consists of 42,955 cases, including 4,928 new cohort cases and 38,027 old cohort cases. Within each stratum, the sample selection is conducted independently for the cohort substrata. The old cohort sample is selected systematically with probability proportional to size (PPS) where the measure of size is the 2003 base weight. The new cohort sample is selected systematically with equal probability within each stratum.

At the completion of data collection, every case in the sample received a disposition code to describe its final interview status. These disposition codes are classified into four response categories: eligible respondents (ER), eligible nonrespondents (EN), known ineligible cases (IN), and cases with unknown eligibility (UN). Table 1 presents the distribution of the 2006 SDR sample over the four response categories. In addition to the 42,955 cases, the sample also

includes 2,593 permanently ineligible cases that are considered part of the sample although they were not fielded. Table 1 shows that most of the SDR nonresponse is due to noncontact rather than noncooperation.

Table 1: Classification of 2006 SDR Sample Cases

Response Category	Contact	Response	Case Count	Proportion
ER	1	1	30,817	67.66
EN	1	0	694	1.52
IN	1	.	4,788	10.51
UN	0	.	9,249	20.31
Total			45,548	100.00

2. The current weighting class approach

The 2006 SDR achieved a response rate of about 80 percent. There are two general approaches to nonresponse adjustments: the traditional weighting class approach and the relatively new propensity model approach. Since its inception, the SDR has used the weighting class approach to compensate for nonresponse. Under this approach, the final weight is calculated in three steps¹.

2.1 Base Weight

The first step is to calculate a base weight for all cases selected into the sample. The base weight accounts for the sample design, and it is defined as the reciprocal of the probability of selection under the sample design.

$$w_{i} = \frac{1}{p_i}$$

For all new cohort cases, the selection probability is the same for cases in the same strata. For old cohort cases, the selection probability is a conditional probability, i.e., conditional to the selection into the prior sample. Two adjustments are made to the base weight during nonresponse adjustments. The first adjustment accounts for cases with unknown eligibility, and the second compensates for nonrespondents that are known to be eligible. The 164 sampling strata are used as the building blocks for constructing adjustment classes. The sampling strata are collapsed to form adjustment classes to ensure that (1) there are at least 20 known eligibility cases per eligibility adjustment class and at least 20 complete cases per nonresponse adjustment class, and (2) the adjustment factor at both stages is less than 2 for all adjustment classes.

2.2 Adjustment for Unknown Eligibility

For a significant fraction of the sample members the eligibility status is unknown at the end of the survey. The adjustment for unknown eligibility distributes the base weights associated with such cases to cases for which the eligibility status is known, proportionately to the weighted distribution of known eligible and known ineligible cases. Through this adjustment, the base weights carried by the unknown eligibility cases are distributed to cases with known eligibility assuming that the eligibility rate among the two sets is the same. The adjustment factor is calculated as,

$$f_{1k} = \frac{ER_k + EN_k + IN_k + UN_k}{ER_k + EN_k + IN_k}$$

where k denotes the adjustment cell; ER_k represent the sum of base weights over all eligible respondents of cell k ; EN_k the sum of base weights over all eligible nonrespondents of cell k ; UN_k the sum of base weights over all unknown eligibility cases of cell k ; and IN_k the sum of base weights over all known ineligible cases of cell k .

2.3 Adjustment for Nonresponse

The base weight for the complete cases is adjusted to compensate for eligible nonrespondents. Using the same notation defined earlier, the nonresponse adjustment factor for eligible completes in cell l is calculated as,

¹ The SDR has used raking to adjust the weight in the final step, but raking is not part of the current weighting procedures.

$$f_{2l} = \frac{ER_l + EN_l}{ER_l}$$

We use a new index l for the nonresponse adjustment cells because they may be defined differently than k . Through this adjustment, the weights associated with eligible nonrespondents are transferred to the eligible respondents. The respondents thus represent all eligible doctorates in the frame.

2.4 Final Weight

The final weight for eligible complete cases is computed as the product of the base weight, the eligibility adjustment factor, and the nonresponse adjustment factor. That is,

$$w_{2i} = w_{1i} \times f_{1k} \times f_{2l}$$

The sum of w_{2i} over all eligible respondents represents an estimate of the number of eligible doctorates in the frame.

3. The alternative propensity model approach

Many researchers have proposed the use of logistic regression models to examine the nonresponse mechanism and calculate nonresponse adjustment. Examples of this can be found in Little (1986), Ekholm and Laaksonen (1991), Folsom (1991), Iannacchione, Milne, and Folsom (1991), and Eltinge and Yansaneh (1993). Numerous researchers have conducted empirical comparisons between the weighting class and propensity model approaches to nonresponse adjustments. Examples of this can be found in Lepkowski, Graham, and Kasprzyk (1989), Ezzati-Rice and Khare (1994), Carlson and Williams (2001), Diaz-Tena, et al (2002), and Siegel, Chromy, and Copello (2008).

Under the response propensity modeling approach, one typically uses logistic regression and auxiliary data that are available for both respondents and nonrespondents to predict the response propensity of each sample member. The inverse of the predicted propensity score is used to calculate the weight adjustment factor. If the predicted propensity is unbiased, the potential nonresponse bias may be eliminated through this adjustment. In practice, however, the logistic regression model is only an approximation to the true response mechanism.

The propensity model approach starts with the same SDR base weight as calculated under the weighting class approach. We then build two logistic regression models for the unknown eligibility adjustment and the nonresponse adjustment, respectively.

3.1 The Contact Model

The contact model corresponds to the unknown eligibility adjustment under the weighting class adjustment approach. For the purpose of propensity modeling, cases with known eligibility status (ER, EN and IN) are considered "contact" whereas cases with unknown eligibility status (UN) are "non-contact". The logarithm of the odds of "contact" is modeled as a linear function of a set of covariates that are available for both contacts and non-contacts to predict the contact propensity of each sample member.

$$p_{1i}(\text{contact} = 1 | X) = \frac{1}{1 + e^{-f(X)}}$$

The inverse of the predicted contact propensity is the noncontact (or unknown eligibility) adjustment factor.

$$f_{1i} = \frac{1}{p_{1i}}$$

We used the SAS PROC LOGISTIC procedure to fit the contact model without including the survey weights. The set of predictors include race, sex, degree field, citizenship, disability status, and PhD Year.

3.2 The Response Model

The response model corresponds to the nonresponse adjustment under the weighting class adjustment approach. Among the contact cases, those in response category ER are "response" and those in EN are "non-response". The known ineligible cases (IN) and cases with unknown eligibility (UN) do not participate in the modeling because no

interviews were attempted with these cases. The logarithm of the odds of response is modeled as a linear function of the same set of variables as those in the contact model.

$$p_{2i} = \frac{1}{1 + e^{-g(X)}}$$

The inverse of the predicted response propensity is used as the nonresponse adjustment factor. That is,

$$f_{2i} = \frac{1}{p_{2i}}$$

3.3 Final Weight

The final weight under the propensity model approach is the product of the base weight, the noncontact adjustment factor, and the nonresponse adjustment factor. That is,

$$w_{2i} = w_{1i} f_{1i} f_{2i}$$

4. Comparisons of the weights and weighted estimates

The comparison between the two weighting approaches is conducted along two dimensions: the distribution of the final weights and the distribution of the weighted estimates.

4.1 Distribution of the Weights

The following statistics are used to describe the distribution of the weights: mean, standard deviation, minimum, maximum, coefficient of variation, and design effect due to unequal weighting. The distributions of the weights are compared overall as well as by population subgroups defined by gender, degree field, race and ethnicity, PhD year, cohort, disability status, citizenship, and demographics. Figures 1 through 3 below presents the comparisons for overall and by subgroups defined by gender, degree field, and race and ethnicity.

The overall mean weight is just over 23 under both weighting approaches. The variation of the mean across the subgroups reflects differential sampling rates. For example, white and male are undersampled relative to their counterparts. On the other hand, racial and ethnic minority groups are greatly oversampled. The means of the weights are almost identical under the two weighting approaches. This is true both overall and for the subgroups. The same pattern is observed for all the other subgroups comparisons that are not presented here.

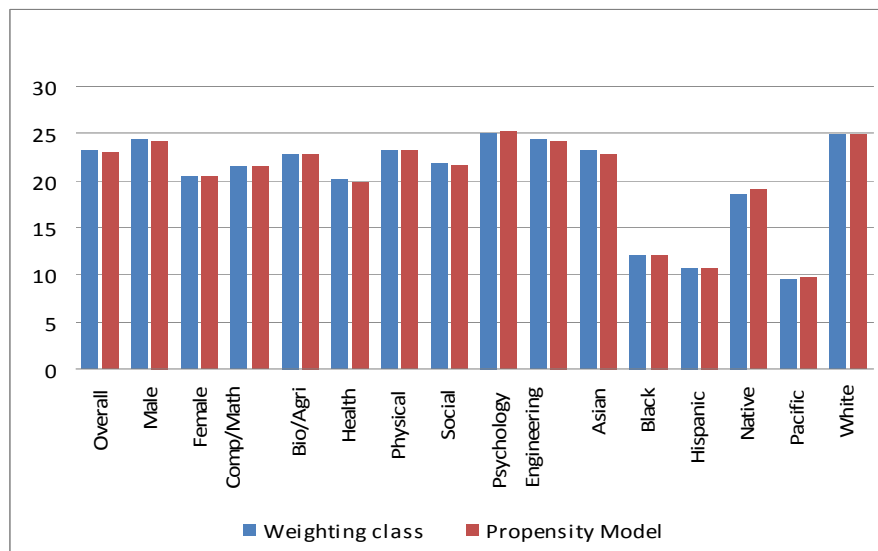


Figure 1: Comparison of the Distribution of the Weights: Mean weight

The variance of the weights differs slightly between the two sets of weights (Figure 2). With one or two exceptions, the propensity model weights tend to have smaller variance although the difference is not large. This pattern is also seen in Figure 3 which shows that, for most subgroups, the design effect due to weighting is slightly smaller under the propensity model approach. The same pattern is observed for those subgroups that are not presented here.

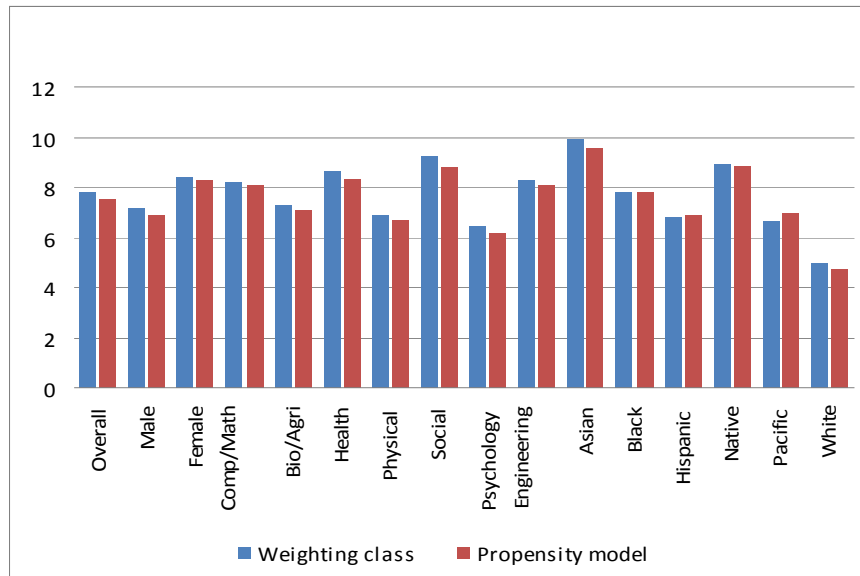


Figure 2: Comparison of the Distribution of the Weights: Standard deviation

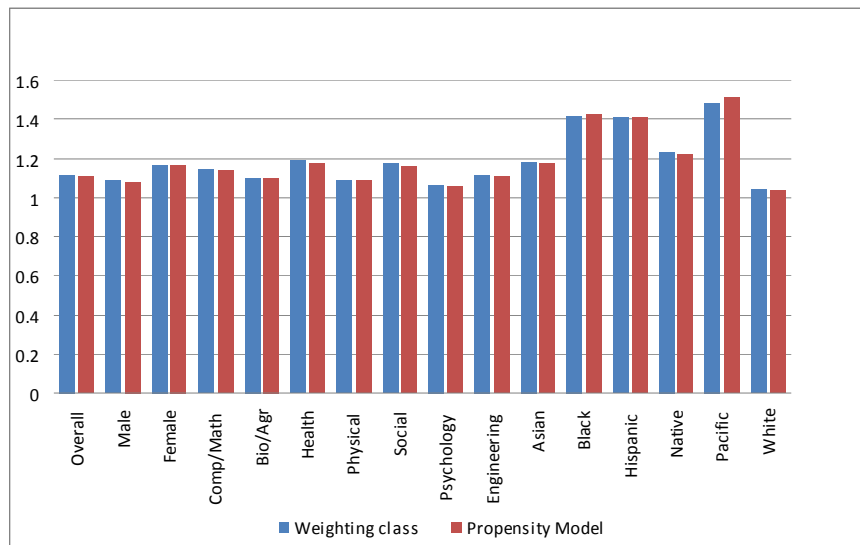


Figure 3: Comparison of the Distribution of the Weights: Design effect due to weighting

4.2 Distribution of Weighted Estimates

We compared the distribution of the weighted estimates for 30 key variables. The following statistics are used to describe the distribution of the weighted estimates: mean (total), standard error of the mean (total), and design effect of the mean (total). The SUDAAN PROC DESCRIPT procedure is used to estimate the standard errors and design effects. The distributions of the weighted estimates are compared overall as well as by population subgroups defined by gender, degree field, race and ethnicity, PhD year, cohort, disability status, citizenship, and demographics. For all these comparisons, the weighted means and totals are about the same under the two weighting approaches. The estimated standard errors are very similar except that the standard errors tend to be smaller under the propensity models. The difference is very small and insignificant though.

The following figures show the comparisons of the design effects of the weighted mean for six key variables, overall and by gender, degree field, and race and ethnicity. For all the key variables, the design effect is slightly above 1. Groups that are oversampled, i.e., racial and ethnic minorities, female, and certain degree fields, tend to have greater design effects than the other groups. Within the same group, however, the design effects are about the same under the two weighting approaches except that the design effects tend to be smaller with the propensity model adjustments.

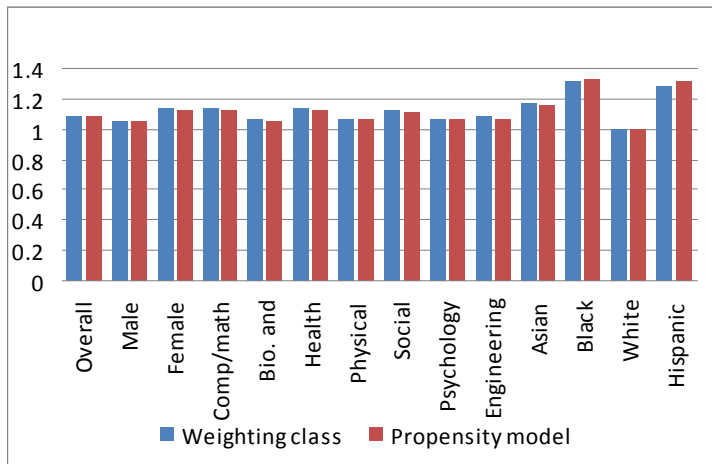


Figure 4: Comparison of the Design Effects of Weighted Estimates: Job change

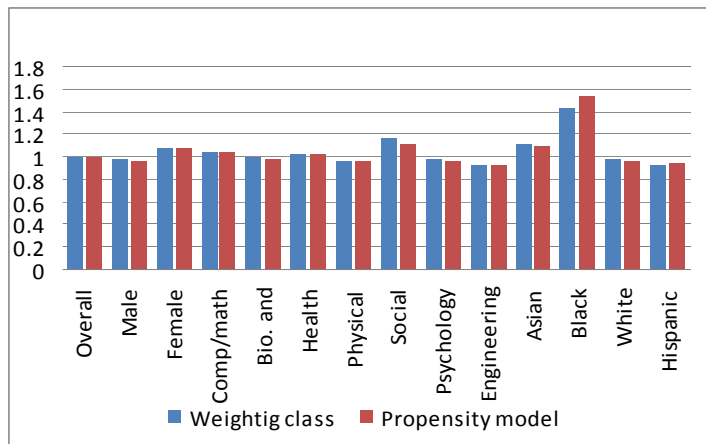


Figure 5: Comparison of the Design Effects of Weighted Estimates: Looked for work

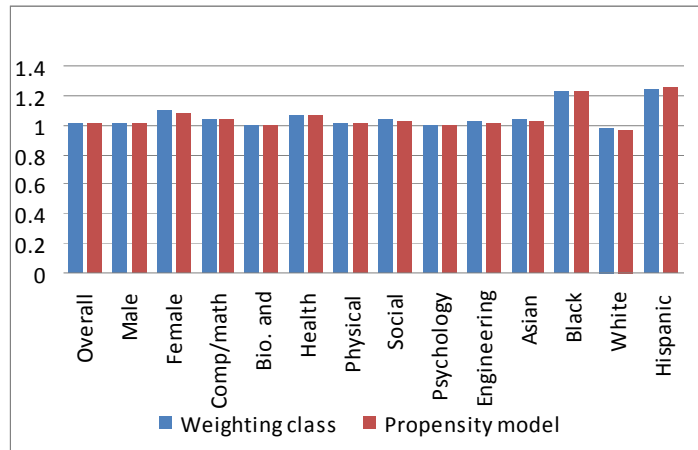


Figure 6: Comparison of the Design Effects of Weighted Estimates: Married

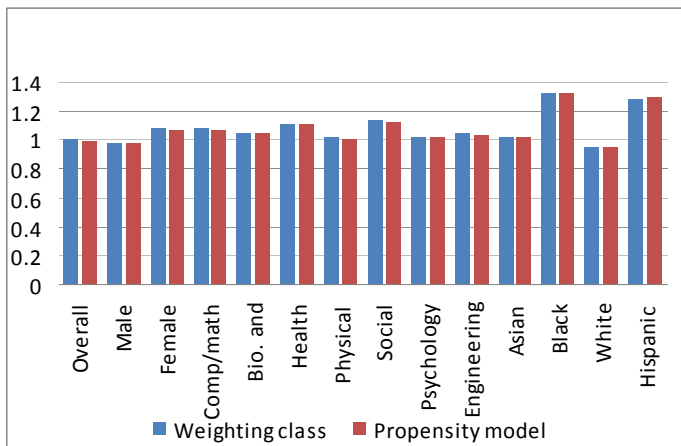


Figure 7: Comparison of the Design Effects of Weighted Estimates: Worked for educational institution

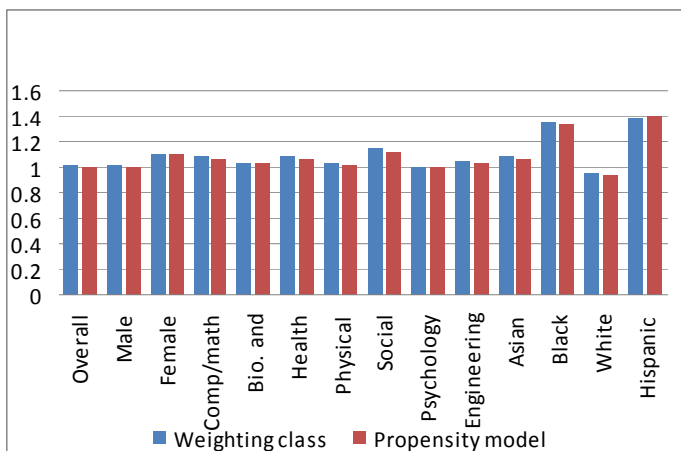


Figure 8: Comparison of the Design Effects of Weighted Estimates: Salary over \$100k

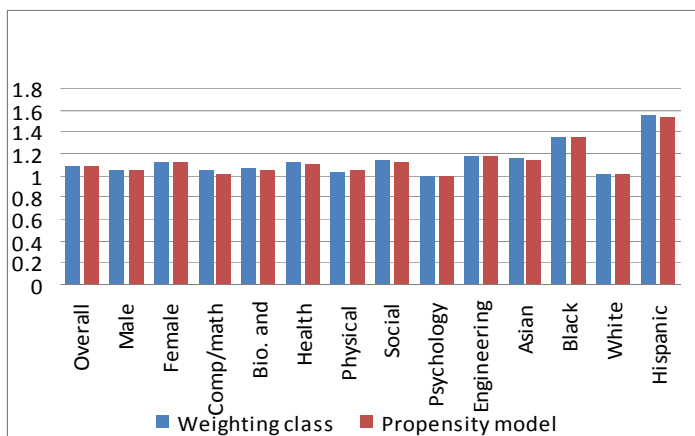


Figure 9: Comparison of the Design Effects of Weighted Estimates: Employed

5. Discussions

We found little difference in the distribution of the weights and the weighted estimates between the two weighting approaches. The propensity weights tend to have smaller variance than the weighting class weights, but the difference tends to be small. The similarity holds both overall and for various subpopulations that are of analytical interest to the SDR. Empirical investigations in the past, including those cited earlier, typically conclude that different weighting approaches produce similar weights and weighted estimates, so this finding is not surprising. On the one hand, the SDR contact rate and cooperation rate are high, which leaves little variation in the adjustment factors regardless which weighting approach is used. On the other hand, the contact and response mechanisms assumed under the two weighting approaches are basically the same because the variables used to define the weighting classes are also used as predictors in the propensity models.

Overall, the propensity modeling approach seems to offer a viable alternative to the current weighting class approach. The contact and response models may be further refined or improved by incorporating additional variables into the models. For example, response status of the previous round may prove to be good predictors of the response status of the current round. Substantive survey variables, such as income, may also be excellent predictors. However, including these variables in the models will require creative coding because past response status is not available for new cohort sample members. Similarly, variables like income are not available for nonrespondents. In addition, as more variables become potential predictors, careful model selection will be necessary. These will be the subjects for further research.

References

- Carlson, B.L. and Williams, S. (2001). "A comparison of two methods to adjust weights for nonresponse: propensity modeling and weighting class adjustments," Proceedings of the Section on Survey Research Methods, American Statistical Association.
- Diaz-Tena, N., Potter, F., Sinclair, M., and Williams, S. (2002). "Logistic propensity models to adjust for nonresponse in physician surveys," Proceedings of the Section on Survey Research Methods, American Statistical Association, 776-781.
- Eltinge, J.L. and Yansaneh, L.S. (1993). "Weighting adjustments for income nonresponse in the U.S. Consumer Expenditure Survey," Technical Report No. 202, Department of Statistics, Texas A&M University, College Station, Texas.
- Ekholm, A. and Laaksonen, S. (1991). "Weighting via response modeling in the Finnish Household Budget Survey," *Journal of Official Statistics*, 7, 325-337.
- Ezzati-Rice, T.M. and Khare, M. (1994). "Modeling of response propensity in the third National Health and Nutrition Examination Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association, 955-959.
- Folsom, R.E. (1991). "Exponential and logistic weight adjustments for sampling and nonresponse error reduction," Proceedings of the Social Statistics Section of the American Statistical Association, 197-202.
- Iannacchione, V.G., Milne, J.G., and Folsom, R.E. (1991). "Response probability weight adjustments using logistic regression," Proceedings of the Section on Survey Research Methods, American Statistical Association, 637-642.
- Lepkowski, J., Graham, G., and Kasprzyk D. (1989). "Weighting adjustments for partial nonresponse in the 1984 SIPP Panel," Proceedings of the Section on Survey Research Methods, American Statistical Association, 296-301.
- Little, R.J.A. (1986). "Survey nonresponse adjustments for estimates of means," *International Statistical Review*, 54(2):139-157.
- Siegel, P., Chromy, J.R., and Copello, E. (2008). "Propensity models versus weighting class approaches to nonresponse adjustment: a methodological comparison," Paper presented at the American Association for Public Opinion Research conference.