Estimated Response Propensities as a Means to Evaluate Error Effects Due to Nonresponse

Leela M. Aertker[1] and William D. Kalsbeek[1]
University of North Carolina at Chapel Hill[1]

Abstract

In addition to providing a basis for the weighting adjustment for nonresponse, individual response propensities obtained from a logistic model may offer a means by which nonresponse bias effects can be routinely estimated in samples where useful auxiliary data are available for response propensity modeling. We present measures of nonresponse error effects and direct estimates of round-specific nonresponse bias using data from the multi-round National Longitudinal Study of Adolescent Health (Add Health). We further assess the utility of estimated response propensities as a direct mechanism to estimate nonresponse bias. This type of assessment method, based on estimated response propensities, accurately estimated non-response bias for health and risk behavior outcomes and less accurately estimated nonresponse bias for sexual behavior and substance use outcomes.

Keywords: unit nonresponse, response propensity, nonresponse bias

1. Introduction

The biasing effect of nonresponse is directly related to the extent of nonresponse and can greatly affect the quality of an otherwise well-designed survey. When respondents substantially and systematically differ from nonrespondents, an assessment of nonresponse bias becomes a necessity to assess data quality and potential effects on survey estimates, especially in light of the recent release of the Office of Management and Budget (OMB, 2006) guidelines. OMB guidelines state that if the expected response rates are less than eighty percent then a nonresponse bias plan should be implemented.

The recent OMB guidelines motivate this paper's investigation of two additional nonresponse bias assessment methods, which utilize response propensities, that are relatively easy and inexpensive to perform. Response propensities can serve as post-survey adjustment methods for nonresponse error reduction and may further be utilized as a means to evaluate nonresponse bias. Data from the National Longitudinal Study of Adolescent Health (Add Health) are used to illustrate our proposed assessment methods.

1.1 Utility of Estimated Response Propensities

There are two types of stochastic post-survey adjustment methods for nonresponse error reduction: weighting class and propensity modeling. These methods implicitly acknowledge the stochastic view of nonresponse since such adjustments are usually intended to reflect the likelihood of response. Each population (and thus sample) member has a response probability, or response propensity (p_i), which is unknown and must be estimated ($\hat{p}_i$). Response propensity modeling employs estimated response propensities, which are obtained from a weighted fitted logistic regression model, to adjust for nonresponse. In addition, these response propensities may be utilized to evaluate nonresponse bias in key survey estimates.

1.2 Overall Analytic Method

In order to convey the two assessment methods, it is convenient to define a survey sample as a collection of members randomly selected from the target population that can be divided into those who participate in the survey, sample responders, and those who do not participate in the survey, sample nonresponders. In the first approach to assess nonresponse bias, the *SR to S* approach, we will make inference from the sample responders (SR) to the survey sample (S). The *SR to S* approach considers the sample responders as an outcome of Poisson sampling with selection probabilities ($\bullet_i$) equivalent to estimated response propensities ($\hat{p}_i$). This approach estimates statistics for the entire sample, sample responders plus sample nonresponders, with data from the sample responders and their estimated response propensities. Nonresponse bias is then estimated by subtracting the estimate obtained using sample respondent data from the estimate using sample respondent data weighted by the inverse of the estimated response propensity. The estimated response propensities of the sample responders can be further utilized in the second approach, the *SRR to SR* approach, which randomly simulates two strata – the sample respondent responders and the sample respondent nonresponders – from the sample respondents based on their estimated response propensities. We will make inference from the sample respondent responders (SRR) to the sample responders (SR) in this approach. By subtracting the estimate obtained using the simulated sample respondent responder data from the estimate using data from the sample respondents, nonresponse bias can be estimated. Estimated nonresponse biases, obtained through the two alternative approaches, are then compared to the benchmark estimates of biases that serve as our evaluation criteria.

2. Methods

2.1 Estimation of Response Propensity

As previously discussed, response propensities may be available as a by-product of the process of computing nonresponse adjustments for sample weights (Lessler and Kalsbeek, 1992). They may also be estimated separately if they are not used in computing weights. The latter estimation approach was performed for purposes of this paper, since the weighting class adjustment strategy was used to produce the weights for the data that were used to illustrate the two methods.

An initial analysis was conducted to identify any auxiliary variables that were good predictors of response propensity in a weighted logistic regression model. Candidate predictors for the demographic model included those typically available in cross-sectional samples (i.e., region, degree of urbanization, and race/ethnicity) whereas predictors for the demographic-plus model included those typically available for cohort studies such as demographics (i.e., age, gender, and race/ethnicity), Wave I in-home respondent characteristics (i.e., standard verbal test score, self-rated intelligence, and self-rated attentiveness), school information (i.e., size, type, degree of urbanization, percentage of white students, and geographic region), and Wave I interviewer observations (i.e., whether respondent was bored or impatient, whether respondent was embarrassed, respondent's domicile, and respondent's neighborhood). Several demographic and demographic-plus models were fit – main effects only and main effects plus first-order interactions – with varying alpha levels of retention (i.e., alpha = 0.10, 0.05, 0.01, 0.001, 0.0001) using stepwise logistic regression. The best demographic and demographic-plus model at each alpha level, for both main effects only and main effects plus first-order interactions, was chosen based on -2 log likelihood, AIC, and SC statistics.

Several indicators of predictive strength were computed for all models – kappa, odds ratio, sensitivity, and specificity – in order to provide some quantifiable indication of how good the estimates of response propensities were. The estimated response propensities were categorized into two groups: propensity to respond (i.e., response propensity greater than or equal to the Wave III response rate) and propensity not to respond (i.e., response propensity less than the Wave III response rate). The kappa statistic measured the level of agreement between the response propensity group and the actual response status to Wave III. The odds ratio estimated the odds of being categorized in the propensity to respond group for the actual Wave III respondents versus the actual Wave III nonrespondents. Sensitivity was the probability of correctly classifying respondents in the propensity to respond group and specificity was the probability of correctly classifying nonrespondents in the propensity not to respond group. The models - demographic and demographic-plus - with the best combined values of the predictive strength indicators were selected for response propensity modeling.

The utility of the response propensities, under the stochastic view, was to evaluate error effects due to nonresponse, namely bias, for health or risk behavior characteristics. Nonresponse bias of these specific characteristics was estimated through population-directed and sample-limited perspectives. The population-directed perspective targeted the actual, underlying nonresponse bias in the population and attempted to estimate nonresponse bias for all members in the population. This perspective required stripping the post-stratification adjustment from the sample weights. In addition, a more simplistic indicator of nonresponse bias was estimated from a sample-limited perspective by comparing results from the respondents to the sample (not the population). This perspective did not involve stripping the post-stratification adjustment from the sample weights.

2.2 Estimation of Nonresponse Bias

2.2.1 SR → S Approach

Response propensity is denoted by p_i for the i^{th} member, where i = 1, 2, …, N, and is viewed as

$p_i = \Pr\{R_i = 1\}$

where R_i is 1 if the i^{th} population member responds and provides useful survey data when selected in the sample, and 0 if the i^{th} population member does not respond or does not provide useful data when selected in the sample. The inclusion probability in Poisson sampling is denoted by $\bullet_i$ for the i^{th} member, where i = 1, 2, …, N, and is viewed as

$\bullet_i = \Pr(I_i = 1)$

where I_i is 1 if the i^{th} population member is included in the sample, and 0 if the i^{th} population member is not included in the sample (Hajek, 1964; Brewer & Hanif, 1983).

In the setting where we wish to make inference from the sample respondents to the selected sample, we viewed the probability of whether or not a sample member decided to respond (p_i) as equivalent to the probability that a sample member became a respondent (•$_i$), which was a result of an independent Bernoulli trial. Both the stochastic view of nonresponse and Poisson sampling allow the respective response propensity or inclusion probability to be distinct for each member. When each response propensity (p_i) was substituted for each corresponding selection probability (•$_i$), the sample respondents were treated as an outcome of Poisson sampling. Since we did not know the actual response propensity of each member in the survey sample, it was estimated by $\hat{p}_i$. A weighted multiple logistic regression model was the mechanism by which estimated response propensities were determined in this paper.

We estimated each health or risk behavior characteristic (•) for all members in the survey sample and in the population through the *SR to S* approach. Sample-limited estimates of bias were obtained by the following equation

$$\hat{B}ias(\theta_{SR}) = [\theta_{SR}^{(UNWT)} - \hat{\theta}_S^{(PS)}]$$

where

$\theta_{SR}^{(UNWT)}$ = unweighted value of θ_{SR} based on the sample respondents (SR), and

$\hat{\theta}_S^{(PS)}$ = estimate of the survey sample (S) statistic, θ_S, based on all sample respondents weighted by the inverse of the estimated response propensity.

Population-directed estimates of bias were expressed by the following equation

$$\hat{B}ias(\hat{\theta}_{SR}) = [\hat{\theta}_{SR}^{(W1)} - \hat{\theta}_S^{(PS*W1)}]$$

where

$\hat{\theta}_{SR}^{(W1)}$ = estimate of θ_{SR} based on the sample respondents (SR) weighted by the pre-post-stratification adjusted (pre-PSA) Wave I weight, and

$\hat{\theta}_S^{(PS*W1)}$ = estimate of the survey sample (S) statistic, θ_S, based on all sample respondents weighted by the pre-PSA Wave I weight and the inverse of the estimated response propensity.

2.2.2 SRR → SR Approach

Each member's probability of response was estimated through a weighted multiple logistic regression model, which was denoted by $\hat{p}_i$ for the ith unit, i = 1, 2, …, n. Sample respondent responder and sample nonrespondent responder strata were then simulated from the sample respondents based on their estimated response propensity. Selection of the sample respondent responders from the sample respondents was accomplished through the result of a set of independent Bernoulli trials; that is, if a uniform random number (RN) was less than or equal to $\hat{p}_i$ the sample member was considered a sample respondent responder, and if a uniform RN was greater than $\hat{p}_i$ then the sample member was considered a sample respondent nonresponder. This was replicated 100 times to produce an average survey estimate for the sample respondent responders.

Each health or risk behavior characteristic (•) for all members in the survey sample and in the population were once again estimated, but this time through the *SRR to SR* approach. Sample-limited estimates of bias were expressed by the following equation

$$\hat{B}ias(\hat{\theta}_{SRR}) = [\bar{\hat{\theta}}_{SRR}^{(UNWT)} - \theta_{SR}^{(UNWT)}]$$

where

$\bar{\hat{\theta}}_{SRR}^{(UNWT)}$ = unweighted average of the 100 replicate estimates of θ_{SRR} based on the simulated subsets of sample respondent responders (SRR), and

$\theta_{SR}^{(UNWT)}$ = unweighted value of θ_{SR} based on all sample responders (SR), which includes the sample respondent responders plus the sample respondent nonresponders.

Population-directed estimates of bias were produced by the following equation

$$\hat{B}ias(\hat{\theta}_{SRR}) = [\bar{\hat{\theta}}_{SRR}^{(W1)} - \hat{\theta}_{SR}^{(W1)}]$$

where

$\bar{\hat{\theta}}_{SRR}^{(W1)}$ = average of the 100 replicate estimates of θ_{SRR} based on the simulated subset of sample respondent responders (SRR) weighted by the pre-post-stratification adjusted (pre-PSA) Wave I weight, and

$\hat{\theta}_{SR}^{(W1)}$ = unweighted estimate of θ_{SR} based on the sample responders (SR), which includes the sample respondent responders plus the sample respondent nonresponders, and weighted by the pre-PSA Wave I weight.

2.3 Add Health Background and Design

The National Longitudinal Study of Adolescent Health (Add Health) began as a school-based national survey of health-related behaviors among adolescents in grades 7 through 12 that had been randomly selected in a stratified multi-stage cluster sampling design (Harris et al., 2003). Adolescents attending these schools were eligible for selection into four waves: Wave I in-school survey during 1994-1995, Wave I in-home survey in 1995, Wave II in-home survey in 1996, and Wave III in-home survey in 2001. The Wave I in-home respondent sample represented the survey sample, the Wave III In-Home respondents represented the sample responders, and the nonrespondents to the Wave III in-home survey, who were eligible from the Wave I in-home sample, represented the sample respondent nonresponders.

The Add Health survey has a number of properties that make it advantageous for us to study nonresponse bias. The longitudinal nature of Add Health provides a unique source for comparison of a variety of health and risk behaviors, reported by both respondents and nonrespondents to the Wave III in-home survey, in the Wave I in-home questionnaire. The assigned Wave III in-home sample consisted of all Wave I in-home respondents, which allowed us to link Wave III respondents to their Wave I data. This enabled us to assess a true value, or benchmark estimate, of nonresponse bias and relative bias of selected key outcome variables using Wave I sample weights without post-stratification, regardless of Wave III response status, which provided our evaluation criteria for the estimated biases obtained through the *SR to S* and *SRR to SR* approaches. Furthermore, Wave I weights stripped of post-stratification adjustments were readily available on the dataset. These Wave I pre-post-stratification adjusted (pre-PSA) weights would then be utilized in the population-directed approach.

Key outcome variables (•*s)* were estimated in the survey sample and in the population from the following categories: health, sexual behaviors, substance use, and other risk behaviors. Health outcomes included the percentage with fair/poor self-rated health, percentage who were inactive, and the percentage with psychological distress (based on modified CES-D scale); sexual behavior outcomes included the percentage ever having sexual intercourse, percentage with more than two sexual partners, and the percentage with at least one STD among sexually active males and females; substance use outcomes included percentage ever having smoked cigarettes, drank alcohol, tried marijuana, tried cocaine, and tried other illegal drugs; and other risk behavior outcomes included the percentage having engaged in a serious physical fight in the past year, percentage who stole an item worth more than fifty dollars in the past year, and the percentage ever having seriously thought about suicide. The benchmark estimate of nonresponse bias for each of the key outcome variables was then computed, using Wave I in-home survey data, as the difference between the weighted estimate for the cases who responded in Wave III, sample responders, and the weighted estimate for all eligible cases from the Wave I in-home survey sample:

$$Bias\ (\hat{\theta}_{\mathrm{SR}}) = [\hat{\theta}_{\mathrm{SR}}^{\,W1} - \hat{\theta}_{\mathrm{S}}^{\,W1}].$$

The benchmark estimate of relative bias was computed by

$$RelBias\ \ (\hat{\theta}_{\mathrm{SR}}) = (Bias\ (\hat{\theta}_{\mathrm{SR}}) / \hat{\theta}_{\mathrm{S}})$$

where

$\hat{\theta}_{\mathrm{SR}}$ = weighted estimate of θ_{SR} based on the Wave III sample responders (SR), and

$\hat{\theta}_{\mathrm{S}}$ = estimated measure of θ_S based on all members of the Wave I respondent sample (S).

We used SAS-callable SUDAAN 9.0, for model construction, which incorporated the survey design characteristics into the computational formulas. A with-replacement method was assumed for design specification (Design = WR) since the primary sampling unit (PSU) selection probabilities were generally small. The Wave I sample weights without post-stratification, produced by the National Opinion Research Center (NORC), adjusted for the unequal selection probabilities, effect of nonresponse, and other imbalances, which occurred in Wave I. For nonresponse bias estimation, we used SAS 9.13 and employed the Wave I sample weights without post-stratification for the benchmark and population-directed estimates of nonresponse bias.

3. Findings

Logistic regression was utilized to examine the joint relationship between multiple predictors and a member's Wave III response outcome in both a fitted demographic and demographic-plus model using Wave I sample weights without post-stratification. The most predictive demographic and demographic-plus models of response propensity were determined by several indicators of predictive strength. The indicators measured the agreement between the estimated response propensity group and the actual response outcome. For the demographic main effects model at a significance level of 0.001, the kappa statistic was 0.08, the odds ratio was 1.56, and the sensitivity and specificity were 0.84 and 0.22, respectively. The demographic main effects with first-order interactions model at a significance level of 0.1 also had comparable values of the predictive strength indicators (kappa = 0.08, odds ratio = 1.51, sensitivity = 0.75, and specificity = 0.33); however, the demographic main effects model at a significance level of 0.001 was chosen due to its parsimony. There were no demographic predictors that were significant at 0.0001, and region alone was significant at 0.001 for both the models with main effects only and main effects plus first-order interactions. Therefore, the demographic model with region alone at a significance level of 0.001 was chosen for response propensity modeling using only those variables that would typically be available in a cross-sectional study.

For the demographic-plus main effects with first-order interactions model at a significance level of 0.1, the kappa statistic was 0.15, the odds ratio was 2.17, and the sensitivity and specificity were 0.62 and 0.57, respectively. The other demographic-plus models at varying significance levels had similar, albeit lower predictive strength values for all indicators except sensitivity. Therefore, the demographic-plus model with domicile, gender, standard verbal score, age, region, school size, bored, inattentive, neighborhood, school's degree of urbanization, school type, region × school size, gender × age, gender × bored, domicile × age, region × neighborhood, domicile × region, bored × standard verbal score, gender × school's degree of urbanization, school's degree of urbanization × neighborhood, school's degree of urbanization × school size, and gender × school size was the most predictive of response propensity, at a significance level of 0.1, and was selected for response propensity modeling using variables that would typically be available in cohort studies.

Benchmark estimates of nonresponse bias and relative bias were calculated using survey data from the Wave I in-home probability sample with the Wave I in-home sample weights without post-stratification and served as our evaluation criteria for the respective estimates from the *SR to S* and *SRR to SR* approaches. The majority of the benchmark estimates were negatively biased in Wave III and ranged from -1.73 to 0.14 percent. The estimate, ever had sexual intercourse, was negatively biased by approximately two percent and the following estimates - had multiple sexual partners, ever smoked cigarettes, ever drank alcohol, ever tried marijuana, and had a serious physical fight in past year - were all negatively biased by approximately one percent. The remaining variables (i.e., fair/poor self-rated health, inactive, psychological distress, sexually active with a STD, ever tried cocaine, ever tried other illegal drugs, and ever attempted suicide) had approximately zero percent nonresponse bias. The benchmark relative biases were in the range of -0.08 to 0.02 percent. The benchmark estimates indicated that nonresponse bias was an issue for a few key outcome estimates, although not to a considerable extent in this particular wave of the sample.

The *SR to S* and *SRR to SR* approaches, which employed estimated response propensities, were applied to estimate nonresponse bias. Tables I and II list the estimated nonresponse biases – for the sample-limited and population-directed analysis – obtained through each approach in both the demographic and demographic-plus fitted response propensity models.

In the sample-limited and population-directed *SR to S* approaches, the demographic and demographic-plus models accurately estimated nonresponse bias for fair/poor self-rated health, psychological distress, sexually active with an STD (excluding the population-directed demographic model), and ever considered suicide and somewhat accurately estimated nonresponse bias for inactivity, ever smoked (for the sample-limited demographic model only), ever tried cocaine, and stole an item within the past year. However, for the remaining sexual behavior estimates, substance use estimates, and engaged in a serious physical fight in the past year, the *SR to S* approach was less accurate in estimating nonresponse bias. For the sample-limited *SR to S* approach, the demographic model unexpectedly performed slightly better in estimating nonresponse bias compared to the demographic-plus model; for the population-directed *SR to S* approach, there was minimal difference between the estimated nonresponse bias for demographic and demographic-plus models in the.

In the sample-limited and population-directed *SRR to SR* approaches, the demographic and demographic-plus models accurately estimated nonresponse bias for fair/poor self-rated health, psychological distress, sexually active with an STD, ever tried cocaine, stole an item in the past year, engaged in a serious physical fight (for the sample-limited perspective only) and ever considered suicide and somewhat accurately estimated nonresponse bias for inactivity, ever smoked (for the sample-limited demographic model only), and ever tried other illegal drugs (for the population-directed perspective only). However, for the remaining sexual behavior and substance use estimates, the *SRR to SR* approach was less accurate in estimating nonresponse bias. Furthermore, the demographic and demographic-plus models in the sample-limited and population-directed *SRR to SR* approach tended to perform equally as well in estimating nonresponse bias.

4. Discussion

In this paper we have assessed the utility of two estimation strategies of determining the bias due to nonresponse where the inferences made to produce the bias estimates are limited to the initial sample for which indication of nonresponse bias is needed.

Table I: Sample-Directed Estimates of Nonresponse Bias[1] Based on the *SR to S* and *SRR to SR* Approaches for Selected Key Outcome Measure in Wave III Core Sample

Variables	Benchmark Bias	Bias *SR to S* Approach		Bias *SRR to SR* Approach	
		Demographic	Demographic Plus	Demographic	Demographic Plus
Health					
Fair/poor self-rated health	0.06	-0.06	-0.01	-0.23	-0.23
Inactive	-0.38	-0.16	0.03	-0.05	-0.09
Psychological distress[(a)]	-0.04	-0.32	-0.18	-0.22	-0.31
Sexual Behaviors					
Had sexual intercourse	-1.70	-0.29	0.68	-0.31	-0.46
Had >2 sexual partners	-0.90	0.52	1.13	-0.04	-0.18
Had at least 1 STD[(b)]	0.00	0.08	0.21	-0.19	-0.16
Substance Use					
Ever smoked cigarettes	-0.83	-0.52	0.74	-0.24	-0.02
Ever drank alcohol	-1.07	-0.02	1.46	0.52	0.97
Ever tried marijuana	-1.07	0.06	0.96	-0.52	-0.25
Ever tried cocaine	-0.21	-0.02	0.17	-0.03	-0.02
Ever tried other illegal drugs	-0.45	-0.01	0.48	0.15	0.31
Other Risk Behaviors					
Had serious physical fight in past yr	-0.87	0.01	-0.13	-0.68	-0.87
Stole item worth >$50 in past yr	-0.32	-0.02	0.01	-0.18	-0.20
Ever considered suicide	-0.01	0.08	0.27	-0.08	0.03

[(1)]:Adjusted for stratified cluster sampling and unweighted.

[(a)]:Based on modified CES-D scale.

[(b)]:The subgroup of interest is sexually active males and females.

Table II: Population-Directed Estimates of Nonresponse Bias[1] Based on the *SR to S* and *SRR to SR* Approaches for Selected Key Outcome Measures in Wave III Core Sample

Variables	Benchmark Bias	Bias *SR to S* Approach		Bias *SRR to SR* Approach	
		Demographic	Demographic Plus	Demographic	Demographic Plus
Health					
Fair/poor self-rated health	-0.05	0.01	0.02	-0.01	-0.01
Inactive	-0.42	0.11	0.11	0.00	0.02
Psychological distress[a]	-0.17	-0.04	-0.11	0.00	-0.12
Sexual Behaviors					
Had sexual intercourse	-1.73	1.06	1.19	0.00	-0.12
Had >2 sexual partners	-0.85	0.20	1.46	0.01	-0.25
Had at least 1 STD[b]	0.14	1.41	0.27	0.01	0.06
Substance Use					
Ever smoked cigarettes	-0.58	0.61	1.05	-0.05	0.17
Ever drank alcohol	-0.94	1.10	1.79	-0.09	0.39
Ever tried marijuana	-1.03	0.73	1.10	-0.10	0.07
Ever tried cocaine	-0.13	0.14	0.21	-0.06	-0.02
Ever tried other illegal drugs	-0.45	0.25	0.53	-0.09	0.09
Other Risk Behaviors					
Had serious physical fight in past yr	-1.26	0.10	-0.16	-0.01	-0.30
Stole item worth >$50 in past yr	-0.30	0.03	0.00	-0.02	-0.05
Ever considered suicide	0.05	0.18	0.31	-0.01	0.10

[1]:Adjusted for stratified cluster sampling and using Wave I sample weights without post-stratification.

[a]:Based on modified CES-D scale.

[b]:The subgroup of interest is sexually active males and females.

Response propensities, which can be estimated for all members in a survey sample through a fitted logistic regression model, are often readily available from the process of computing nonresponse adjustments to sample weights and may be additionally used to evaluate nonresponse bias. One strategy uses the estimated propensities to infer to the initial sample from the set of respondents (the *SR to S* approach), whereas the other uses them to infer to the actual respondents from a contrived subset of respondents (*SRR to SR* approach).

The sample-limited and population-directed nonresponse bias estimates, based on the *SR to S* and the *SRR to SR* approaches, accurately estimated nonresponse bias for health estimates, sexually active with an STD, ever tried cocaine, stole an item in the past year, and ever considered suicide. The approaches were less accurate in estimating nonresponse bias for the majority of the sexual behavior and substance use estimates. The similar results in both the sample-limited and population-directed assessment methods justifies implementing the sample-limited method due to the relative simplicity (i.e., sample weights stripped of post-stratification adjustments are not necessary).

The sample-limited and population-directed nonresponse bias estimates tended to cluster around the horizontal line where the estimated bias equaled zero for the *SR to S* and the *SRR to SR* approaches. The underestimation of the magnitude of nonresponse bias could either be due to poor estimates of response propensities or the absence of any significant nonresponse bias in the key survey estimates. The former explanation is more reasonable due to the relative weakness of the predictive strength indicators (i.e., kappa, odds ratio, sensitivity, and specificity). The success of the *SR to S* and *SRR to SR* approaches in accurately estimating nonresponse bias depends on the predictive strength of the fitted models. Thus, models with low predictive strength may weaken the magnitude of estimates derived from these two approaches.

Consistent with Tucker (1987), we found that these estimated response propensities are not always as accurate as one would like. This underscores the need for improved predictors of response, which would strengthen the utility of these approaches and post-survey nonresponse weight adjustments in general. For example, the availability of more direct process indicators such as perceived security of the neighborhood and interviewer testing scores would most likely aid in the predictive power of the models. The key component to whether or not the *SR to S* and *SRR to SR* approaches succeed is if the auxiliary variables are highly correlated with both the likelihood of response and the key survey estimates.

It may also be of utility to target models to specific types of nonresponse. Groves and Couper (1998) suggest different functional forms for contact and refusal. Response propensities could be obtained from two separate logistic regression models: probability of contact (has been contacted / has not been contacted) and probability of refusal conditional on contact (refused when contacted / participated when contacted). Two-stage modeling of response outcomes may improve the estimated response propensities since there is a large and separate set of causes for survey contact and refusal.

The longitudinal nature of the Add Health survey provided an opportunity to directly assess a benchmark estimate of bias and relative bias of selected key outcome variables using Wave I in-home survey data, which served as our evaluation criteria for the two alternative approaches. It is plausible that certain estimates (e.g. sexual behavior and substance use estimates) that were correlated with response in the teen years (when the Wave I in-home survey occurred) are no longer correlated with response in early adulthood (when the Wave III in-home survey occurred). Thus, estimating nonresponse bias using survey data from the teen years (i.e., from the Wave I in-home survey) may be an inaccurate estimate for nonresponse bias in early adulthood (i.e., in the Wave III in-home survey). If our evaluation criteria (i.e., the benchmark estimates) of nonresponse bias were flawed then such comparisons are futile because the actual nonresponse bias is unknown.

Furthermore, this bias assessment is for round-specific (i.e., Wave III) nonresponse and did not account for the cumulative effects of nonresponse. Nonresponse bias is cumulative when there are multiple sources of attrition (i.e., multiple waves in a survey). The sample-limited estimates gave an indication of magnitude and direction but were not estimates of the actual bias in the population; the population-directed estimates did give an indication of the actual bias in the population but only round-specific bias. Therefore, for an actual assessment of nonresponse bias for the Wave III in-home sample, cumulative nonresponse bias should be evaluated.

The motivation for this paper is the recently released OMB (2006) guidelines regarding bias investigation in surveys. If expected nonresponse rates suggest the potential for bias to occur then a plan should be implemented, prior to the start of the survey, to study nonresponse bias. Assessing nonresponse bias using estimated response propensities, as illustrated in this paper, is an additional type of assessment method for researchers that may prove useful for estimating nonresponse bias of key outcome measures in survey samples if auxiliary variables are available that are highly correlated with both the likelihood of response and key survey estimates, there is separate modeling of response outcomes (e.g., contact and refusal), and the presented methods result in accurate estimates when applied to other surveys.

References

Brewer, KRW and M Hanif. (1983). *Lecture Notes in Statistics: Sampling With Unequal Probabilities*, New York, Springer-Verlag.

Groves, RM and J Montaquila. (2006). *Practical Tools for Nonresponse Bias Studies*, Odum Institute Short Course Notes.

Groves, RM and MP Couper. (1998). Nonresponse *in Household Interview Surveys*, New York, John Wiley and Sons.

Hajek, J. (1964). Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population, *The Annals of Mathematical Statistics*, 35, 1491-1523.

Harris, KM, F Florey, J Tabor, and JR Udry. (2003). *The National Longitudinal Study of Adolescent Health: Study Design [www document]*. URL: http://www.cpc.unc.edu/projects/addhealth/design

Lessler, JT and WD Kalsbeek (1992). *Nonsampling Error in Surveys*, New York, John Wiley and Sons.

Office of Management and Budget (2006). *Questions and Answers When Designing Surveys for Information Collections*, Office of Information and Regulatory Affairs.

Tucker, C. (1987). Toward the Development of an Ordered Set Data Quality Categories, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 558-563.