Using Soft Refusal Status in the Cell-Phone Nonresponse Adjustment in the National Immunization Survey

Wei Zeng[1], David Yankey[2]
Nadarajasundaram Ganesh[1], Vicki Pineau[1], Phil Smith[2]
[1]NORC at the University of Chicago, 4350 East-West Highway, Bethesda, MD 20814
[2]Centers for Disease Control and Prevention, 1600 Clifton Road NE, A-19, Atlanta, GA 30333

Abstract
This paper seeks to evaluate the effectiveness of using metropolitan statistical area (MSA) geographic status as an adjustment variable in the cell-phone sample nonresponse adjustment for the National Immunization Survey (NIS) and investigates the potential to utilize the level-of-effort information collected during the telephone interview to improve the cell-phone sample nonresponse adjustment. We examined the viability of using two alternative measures of interviewing level of effort for cell-phone nonresponse adjustment: the number of call attempts to resolve a cell-phone number and soft refusal status which is an indicator of initial reluctance by the respondent to participate in the survey. The current approach of using MSA status in the NIS cell-phone sample nonresponse adjustment is compared to nonresponse adjustment utilizing the two alternative level-of-effort variables. We failed to find sufficient evidence to justify using the two new approaches or the current approach for the cell-phone nonresponse adjustment in the NIS. Consequently, eliminating the current MSA status cell-phone sample nonresponse adjustment is under consideration as we develop the future NIS weighting methodology

Key Words: weighting, nonresponse adjustment, cell-phone sample

1. Introduction

Random digit dial (RDD) telephone surveys have undergone some fundamental changes in recent years. Telephone surveys have been experiencing declining response rates, which raises concerns of an increase in potential nonresponse bias if respondents differ systematically from nonrespondents (Groves 2006). Moreover, one-third of American homes (34.0%) had only cell-phones during the second half of 2011 (Blumberg, Luke 2012). As a result, the landline frame is no longer able to provide sufficient coverage of the U.S. population, and in order to improve coverage, cell-phone sampling is increasingly used to supplement landline telephone surveys. These two fundamental changes jointly present a challenge for assessing and adjusting for potential nonresponse bias in cell-phone samples.

The NIS has been conducted quarterly since 1994 by the Center for Disease Control and Prevention (CDC) to estimate childhood vaccination coverage rates for the U.S., each state and select local areas. The NIS uses a two-phase survey design where the first phase is an RDD survey that identifies the households with age-eligible children age 19-35 months and collects information on vaccinations and vaccination providers of the eligible children. The second phase is a mail survey to the child's immunization provider. Prior to quarter 4 of 2010 (Q4/2010), the NIS consisted of only a RDD landline sample. However, beginning Q4/2010, a RDD cell-phone sample was added to the survey to take into account cell-phone only households. Cell-phone sampling in large-scale RDD surveys such as the NIS is a relatively new enhancement in recent years. As such, research into the best methods for weighting cell-phone samples to minimize potential bias due to nonresponse in the NIS is an important focus.

Effective auxiliary variables for use in nonresponse adjustments should have two properties: they should be predictive of the sampled unit's probability of responding to the survey, and also correlated with key survey variables of interest (Groves 2006). However, such variables are rare in RDD telephone surveys (West and Little 2012). Moreover, unlike the landline telephone sample, where the telephone exchange-level information are more readily available and reliable, auxiliary data available for the cell-phone sample nonresponse adjustments are often limited to geographic information, which is not always accurately assigned and can be unreliable due to the mobility of cell-phone users. Using error-prone auxiliary variables in nonresponse adjustment can be problematic since they are not necessarily measuring the true characteristics of the respondents and nonrespondents. For the current NIS cell-phone nonresponse adjustment, the innate assumptions are that metropolitan statistical area (MSA) status is correlated with the propensity for survey response and also correlated with vaccination coverage rates, thus using MSA status may help reduce nonresponse bias. To validate these assumptions and identify potential improvements to NIS cell-sample nonresponse weighting, we launched the evaluation described in this paper. We hypothesized that paradata, collected during the fielding of the NIS phone interviews measuring level-of-effort (LOE) may offer improvements in the cell-phone nonresponse adjustment for reducing nonresponse bias. LOE are known to be correlated with survey response rates (Groves and Couper, 1998) and previous studies have considered the use of LOE paradata, like the number of call-backs, to adjust for non-response bias (Biemer et al, 2012).

This paper seeks to evaluate the effectiveness of using MSA geographic status in the cell-phone sample nonresponse adjustment, the original approach used in the NIS, and investigates the potential to utilize the LOE information collected during the household telephone interview to improve the cell-phone sample nonresponse adjustment. We examine whether the MSA and LOE variables meet the two criteria of being effective auxiliary variables, highly predictive of both the response propensity and key survey variables. Furthermore, we study the potential impact of the alternative nonresponse weighting adjustments on key NIS vaccination coverage rates.

2. Methods

The data we analyzed are from the Q4/2010 NIS cell-phone sample, which contains 1,926 household interview completes. The original NIS weighting methodology for the cell sample nonresponse adjustment utilizing weighting class methods (Little, 1986) includes the following steps: 1) applying the cell-phone number resolution nonresponse adjustment, compensating for some telephone numbers that are never determined to be active cell-phone numbers despite multiple call attempts; 2) applying the age screener nonresponse adjustment, adjusting for some households that fail to finish the age screener; and 3) applying the household interview nonresponse adjustment, adjusting for some households with age-eligible children that do not complete the household interview. Each of these adjustments are carried out within census region by forming weighting cells using MSA status (MSA, non-MSA), which is based on wire center[1] information associated with the cell-phone number.

The alternative approach we consider in this paper relies on LOE information captured during telephone interviewing. To quantify LOE, we considered two easy-to-access measures in our study. One is the number of call attempts to determine whether a cell-phone number is active or not (hereafter called “call attempts”), and the other is soft refusal status, an indicator of initial reluctance displayed by the respondent to participate in the survey. Soft refusal cases are different from hostile refusal cases that respond to survey contact by saying “Take me off your list” or “Don’t call me again”. Soft refusal cases are more likely to make statements such as “I am too busy” or ”I am not interested”, which are likely to be associated with differences in both data quality and substantive response (Couper 1997).

In order to examine whether the MSA and LOE variables are qualified as effective auxiliary variables in nonresponse adjustment, we first assess the power of MSA and the two LOE measures in predicting the age screener completion and household (HH) interview completion rates. We make a direct comparison of the age screener completion rates[2]and HH interview completion rates[3] by different levels of MSA status, soft refusal status and call attempts. In order to compare the three alternatives, we fit three logistic regression models predicting the two responses outcomes and use the area under the receiver operating characteristic (ROC)[4] curve, or AUC, to judge the predictive power of each model.

We also assess the second criteria for measuring the effectiveness of the auxiliary variables for use in nonresponse adjustment which is the association of the LOE variables with key NIS vaccination coverage rates.

Considering the above two assessments, we simulated nonresponse adjustments with the two new LOE measure(s) in both the cell-phone sample age screener nonresponse adjustment and the HH interview nonresponse adjustment. We compare the weighted demographic distribution and survey estimates using the original MSA based nonresponse adjustment, the two alternative approaches with the LOE nonresponse adjustment and the base weighted approach without nonresponse adjustment.

[1] A wire center is typically a general geographic area which is serviced by a defined set of exchanges. Each wire center has a set of area code exchanges with the same first six digits that are dedicated to providing wireless service.
[2] Defined as the number of cases who finished the age screener out of the total number of cell-phone numbers identified as active personal cell phone numbers.
[3] Defined as the number of cases who finished the household interview out of the total number of cases identified with at least one age eligible child in the household.
[4] A ROC curve plots the true positives (sensitivity) vs. false positives (1 − specificity) for a binary classifier system as its discrimination threshold is varied. Since, a random method describes a horizontal curve through the unit interval, it has an AUC of .5. Minimally, classifiers should perform better than this, and the extent to which they score higher than one another (meaning the area under the ROC curve is larger), they have better expected performance.

Only the age screener nonresponse adjustment and the HH interview nonresponse adjustment are considered here since we were not able to establish LOE measures at the cell-phone number resolution nonresponse adjustment stage.

3. Findings

Table 1 below presents three sets of comparisons for the age screener completion rates and interview completion rates by MSA status, call attempts, and soft refusal status. Cases from MSA areas and non-MSA areas have similar age screener completion rates and HH interview completion rates. Similar patterns were found for the number of call attempts to resolve a cell-phone number. With an increase in the number of call attempts, the age screener completion rates and HH interview completion rates did not decrease monotonically as one may expect. Only non-refusal cases and soft refusal cases show substantial differences in the age-screener completion rates (94.3% vs. 72.9%) and interview completion rates (76.5% vs. 42.3%). The completion rates comparison suggests that MSA status and number of call attempts are not very likely to be predictive of the response propensity while soft refusal status may have some potential.

Table 1: Comparison of age-screener completion rates and household interview completion rates by MSA status, number of call attempts, and soft refusal status

	MSA Status		Number of Call Attempts					Soft Refusal Status	
	MSA	Non-MSA	1	2	3	4	5+	Soft Refusal	Non-Refusal
Age-Screener Completion Rate (%)	94.3	94.5	91.0	91.2	91.3	91.3	91.7	72.9	94.3
HH Interview Completion Rate (%)	70.5	72.3	72.7	71.0	69.1	68.8	69.7	42.3	76.5

To further examine the predictive power of auxiliary adjustment variables under study, we fit three different logistic regression models predicting age screener or HH interview completion, where the independent variable for the three models were defined by MSA status, call attempts, and soft refusal status, respectively. We used the areas under the ROC curve (AUC), interpretable as the degree the explanatory variable predicts the outcome, to help us to determine and compare the model performance. The performance is measured by the area under the ROC curve. An area of 1 represents a perfect prediction; an area of .5 represents no predictive power of a model. Figure 1 plots AUC for three logistic regression models predicting the age screener completion and Figure 2 shows AUC for models predicting the HH interview completion. They indicate almost no predictive power for MSA status and call attempts, both of which cover area less than 0.51 in both plots. It can be interpreted as both variables explain less than 1% of the variation in the age screener and interview completion outcomes. Soft refusal status performs better by covering an area of 0.6638 and 0.6179 for the age screener completion and the HH interview completion respectively. However, it still shows a weak association with the response propensity by only explaining about 16% and 11% variability in the age screener and HH interview completion respectively. Overall, the first assessment seems to reveal that the use of MSA status and the proposed call attempts auxiliary variables completely fail to predict the response propensity and the predictive power of soft refusal status is moderately better. When auxiliary variables lack power to predict the response propensity, they are not meaningful to use in the nonresponse adjustment.

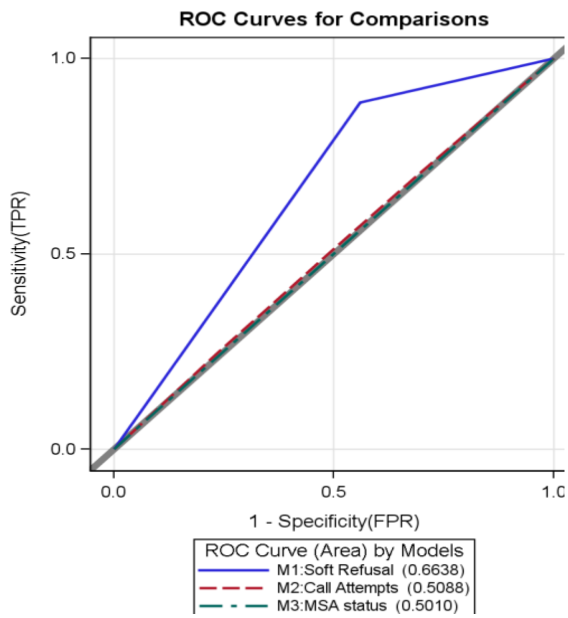

Figure 1: AUC for Logistic Regression Model Predicting the Age Screener Completion

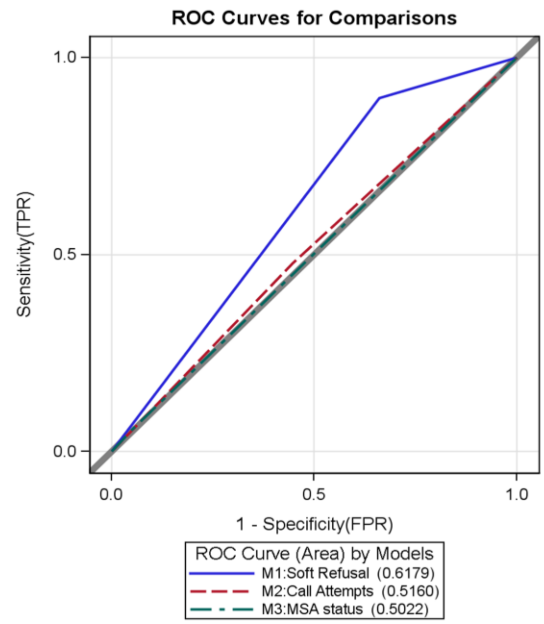

Figure 2: AUC for Logistic Regression Model Predicting the HH Interview Completion

The adjustment variables with low predictive power may still help to adjust for nonresponse bias when they are associated with the survey's outcome variables in practice (Kreuter et al 2010). We continue to test whether converted refusal cases, which are cases that initially refused but eventually completed the interview, differ from non-refusal cases in terms of vaccination coverage rates. We compared 15 childhood vaccines/vaccination series coverage rates estimated using the base weights, which only adjust for the selection probability, and observed some substantial differences in vaccination coverage rates (see Table 2) although they are not statistically significant. Directionally, vaccination coverage rates are always lower for converted refusal cases than for non-refusal cases. It suggests some potential association between soft refusal status and the key survey outcomes in the NIS. In other words, soft refusal status may meet the second criteria of being an effective adjustment variable.

Based on the findings above for soft refusal status, we tested the use of soft refusal status as a weighting cell variable in the cell sample nonresponse adjustments. Tables 3 and 4 compare the original MSA nonresponse adjustment and the proposed soft refusal nonresponse adjustment by examining the demographic distribution and vaccination coverage rates obtained using both methods. As far as demographic characteristics are concerned (see Table 3), both nonresponse adjustment methods produce very similar distributions. The differences are minor and not statistically significant.

Table 2: Base-weighted comparison of vaccination coverage rates between converted refusal cases and non-refusal cases: National Immunization Survey, Q4/2010

	Non-Refusal Cases		Converted Refusal Cases		
Dose Number and Vaccine/ Vaccine Series*	Mean (%)	Standard Error (%)	Mean (%)	Standard Error (%)	Difference **(%)**
3+ DTaP[5]	94.9	0.7	92.4	2.9	2.5
4+DTaP[6]	82.9	1.1	77.1	4.3	5.7
3+ Polio[7]	93.5	0.7	87.6	3.4	5.9
1+MMR[8]	90.2	0.9	82.9	4.0	7.4
3+ Hib[9]	92.5	0.8	89.5	3.2	3.0
3+ HepB[10]	89.9	0.9	84.8	3.9	5.1
1+ Var[11]	89.6	0.9	84.8	3.9	4.8
3+PCV[12]	92.3	0.8	90.5	3.1	1.8
4+ PCV[13]	83.5	1.1	78.1	4.2	5.4
4:3:1 Series[14]	80.5	1.2	71.4	4.6	9.1
4:3:1:3 Series[15]	79.3	1.2	71.4	4.6	7.9
4:3:1:3:3 Series[16]	76.0	1.2	67.6	4.8	8.4
4:3:1:3:3:1 Series[17]	74.6	1.3	66.7	4.8	7.9
4:3:1:3:3:1:3 Series[18]	73.7	1.3	65.7	4.8	8.0
4:3:1:3:3:1:4 Series[19]	71.5	1.3	63.8	4.9	7.7

Moreover, the two nonresponse adjustment methods do not result in substantial differences in vaccination coverage rates (See Table 4). Both methods produce almost identical survey estimates with most differences less than 0.2%. Little evidence was found to support the use of soft refusal status in the nonresponse adjustments for improving the demographic representation or key survey estimates compared to estimates derived using the original approach. Additionally, the resulting standard errors for the survey estimates associated with the two adjustment methods are almost identical indicating no advantage of one approach over the other in reducing survey variability.

[5] 3+ DTaP refers to 3 or more doses of any diphtheria and tetanus toxoids and pertussis vaccines including diphtheria and tetanus toxoids, and any acellular pertussis vaccine

[6] 4+ DTaP refers to 4 or more doses of any diphtheria and tetanus toxoids and pertussis vaccines including diphtheria and tetanus toxoids, and any acellular pertussis vaccine

[7] 3+ Polio refers to 3 or more doses of any poliovirus vaccine

[8] 1+MMR refers 1 or more doses of measles-mumps-rubella vaccine.

[9] 3+ Hib refers to 3 or more doses of Hib vaccine of any type or 2 or more doses of Hib of Merck type.

[10] 3+ HepB refers to 3 or more doses of hepatitis B vaccine.

[11] 1+ Var refers to 1 or more doses of varicella at or after child's first birthday, unadjusted for history of varicella illness.

[12] 3+PCV refers to 3 or more doses of pneumococcal conjugate vaccine.

[13] 4+ PCV refers to 4 or more doses of PCV.

[14] 4:3:1 Series refers to 4 or more doses of DTaP, 3 or more doses of poliovirus vaccine, and 1 or more doses of MMR.

[15] 4:3:1:3 Series refers to 4:3:1 plus 3 or more doses of Hib of any type or 2 or more doses of Hib of Merck type.

[16] 4:3:1:3:3 Series refers to 4:3:1:3 plus 3 or more doses of HepB.

[17] 4:3:1:3:3:1 Series refers to 4:3:1:3:3 plus 1 or more doses of varicella vaccine.

[18] 4:3:1:3:3:1:3 Series refers to 4:3:1:3:3:1 plus 3 or more doses of PCV.

[19] 4:3:1:3:3:1:4 Series refers to 4:3:1:3:3:1 plus 4 or more doses of PCV.

Estimates under both the original and LOE soft refusal status nonresponse adjustment are similar to the base weighted estimates with differences less than 1.5%. Since we failed to establish that MSA status and soft refusal status are effective in nonresponse adjustment, we were not able to conclude the differences observed here are the results of properly correcting the nonresponse bias. However, both adjustment methods lead to an increase in the standard errors (less than 0.2%) compared to the standard errors associated with base weighted estimates (that is, similar estimates generated using the base weights). Although the cost of increased variation is small, it still calls into question continued use of the original MSA based nonresponse adjustment.

Table 3: Nonresponse weighted demographic distribution for cell-phone only completes: National Immunization Survey, Q4/2010

Demographics	Under MSA Non-Response Adjustment	Under Soft Refusal Non-Response Adjustment
Child's Age		
19-23 months	28.8	28.7
24-29 months	36.9	37.1
30-35 months	34.3	34.2
Child's Gender		
Male	52.5	52.3
Female	47.5	47.7
Child's Race/Hispanic Ethnicity		
Hispanic	32.7	32.0
NH[20]-White	42.8	43.4
NH-Black	13.8	13.6
NH-Other	10.7	10.9
Mother's Education		
< 12 years	23.2	23.2
12 years	25.9	25.6
> 12 years	50.9	51.2
Mother's Age		
< 30	61.1	61.0
30+	38.9	39.0
Household Ratio of Income to Poverty		
< 1.0	46.3	45.7
1.0 - 1.99	21.4	21.6
2.0 - 3.99	18.9	19.4
4.0+	13.4	13.3
Housing Tenure		
Owner	37.7	37.7
Renter	57.4	57.4
Other Types of Arrangement	5.0	4.90
Household Number of Children		
1	31.0	30.5
2 or 3	55.8	56.4
4+	13.2	13.0

[20] NH refers to Non-Hispanic

Table 4: Nonresponse weighted vaccination coverage rates for cell-phone only sample: National Immunization Survey, Q4/2010

	Base Weight		Under MSA Non-Response Adjustment		Under Soft Refusal Non-Response Adjustment	
Dose Number and Vaccine/ Vaccine Series	Mean (%)	Standard Error (%)	Mean (%)	Standard Error (%)	Mean (%)	Standard Error (%)
3+ DTap	94.7	0.7	94.3	0.8	94.3	0.8
4+ DTap	82.4	1.1	81.1	1.2	80.9	1.3
3+ Polio	93.0	0.7	92.9	0.8	92.7	0.8
1+MMR	89.6	0.9	89.3	1.0	88.9	1.0
3+ HIB	92.2	0.8	91.5	0.9	91.5	0.9
3+ HepB	89.5	0.9	89.7	0.9	89.5	0.9
1+ Var	89.2	0.9	89.0	1.0	88.8	1.0
3+PCV	92.2	0.8	91.8	0.9	91.8	0.9
4+ PCV	83.0	1.1	81.8	1.2	81.6	1.2
4:3:1 Series	79.8	1.1	78.9	1.3	78.5	1.3
4:3:1:3 Series	78.6	1.2	77.5	1.3	77.3	1.3
4:3:1:3:3 Series	75.3	1.2	74.5	1.3	74.2	1.4
4:3:1:3:3:1 Series	74.0	1.2	73.0	1.4	72.7	1.4
4:3:1:3:3:1:3 Series	73.0	1.2	72.1	1.4	71.8	1.4
4:3:1:3:3:1:4 Series	70.9	1.3	69.8	1.4	69.6	1.4

4. Conclusion and discussion

We did not find sufficient evidence to show that either the current MSA based nonresponse adjustment or alternative methods using LOE measures are effective for the cell phone sample of the NIS. We found that MSA status failed to meet the criteria of being effective auxiliary variables in nonresponse adjustment due to the lack of power in predicting the age screener and interview completion outcomes. Similarly, the number of call attempts to resolve a cell-phone number as active lacked predictive power for use in nonresponse adjustment. One possible explanation is that cases requiring more phone calls to be resolved are harder to contact, but once they are contacted, they do not behave differently in terms of screener completion and interview completion compared to cases that are easier to contact. Soft refusal status performs slightly better but still shows only weak predictive power. Although converted refusal cases (i.e., soft refusal cases that completed the interview) show lower vaccination coverage rates than those who never refused, indicating some association of soft refusal status and key survey variables, the weak association of soft refusal status and the response propensity leads us to believe that soft refusal status will not be an effective auxiliary variable in nonresponse adjustments for the NIS cell-phone sample. When we used soft refusal status in the cell-phone sample nonresponse adjustments, we did not observe substantial and meaningful differences in demographic representation and vaccination coverage rates from estimates generated using the original approach (which uses MSA status in the nonresponse adjustment).

Most importantly, neither the current approach nor the alternative approach resulted in meaningful differences in vaccination coverage rates compared to estimates generated without any nonresponse adjustments at the age screener and interview completion stages. Nonresponse adjustments, effective or not, are usually at the cost of inflation of the standard errors. We also found that both the original MSA based approach and the alternative soft refusal nonresponse adjustment result in an increase in standard errors for the vaccination coverage rate estimates. As a result, one may argue implementing either the current or alternative nonresponse adjustment is not justified for the NIS cell sample, especially considering both approaches fail to show any benefit in effectively adjusting for the nonresponse bias. The unnecessary inflation of the standard errors, even with a small magnitude, should not be ignored since the data users will bear the consequences of data being less informative.

There is always a bias versus variance trade-off in survey estimates due to sample weighting. Ideally, we would like to reduce bias due to differential nonresponse, yet control variance by finding the most efficient weighting variables and methods. In our research, we did not find evidence that supports continuing the use of MSA status, though we are investigating whether it may be helpful for other estimates or subpopulations. Consequently, we are considering whether we should drop the current MSA status nonresponse adjustment in sample weighting of the 2013 NIS. The challenge for properly adjusting for potential nonresponse bias in the NIS cell-phone sample requires further exploration. as part of the development of the full RDD dual frame weighting approach for the 2013 NIS. We need to continue the search for more effective auxiliary variables used for nonresponse adjustments and conduct thorough a investigation before any implementation. If more informative adjustment variables are not identified, we will consider dropping the weighting class adjustment in the initial cell-phone sample nonresponse adjustment, relying more heavily the post-stratification and/or raking-ratio adjustment to reduce potential bias due to nonresponse.

References

Biemer, P.P., Chen, P. and Wang, K. (2012), Using level-of-effort paradata in non-response adjustments with application to field surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society). doi: 10.1111/j.1467-985X.2012.01058.x

Blumberg, S.J. and Luke, J.V. Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2011. National Center for Health Statistics. June 2012. Available from: http://www.cdc.gov/nchs/nhis.htm.

Couper, M.P. (1997), Survey introduction and data quality. Public Opinion Quarterly 61(2):317-338

Groves, R.M. (2006), Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly 70(5):646-75.

Kreuter, F., Olson, K.M., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. (2010), Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples from Multiple Surveys. Journal of The Royal Statistical Society Series A-statistics in Society 173 (2): 389-407

Little, R.J.A. (1986), Survey nonresponse adjustments for estimates of means.
International Statistical Review / Revue International De Statistique, 54(2), 139-157.

West, B.T. and Little, R.J.A. (2012), Nonresponse Adjustment of Survey Estimates Based on Auxiliary Variables Subject to Error. Journal of the Royal Statistical Society, Series C (Applied Statistics)