New Procedures for Nonresponse Adjustments to the 2019 National Health Interview Survey Sampling Weights

Matthew D. Bramlett, James M. Dahlhamer,
Jonaki Bose, and Stephen J. Blumberg

Division of Health Interview Statistics
National Center for Health Statistics
Hyattsville, Maryland

Centers for Disease Control and Prevention
U.S. Department of Health and Human Services

September 2020

Abstract

The procedures for adjusting National Health Interview Survey sampling weights to account for nonresponse to the survey have been updated to coincide with the 2019 redesigned questionnaire. This paper documents the new weighting procedures for data users. The new procedures now include utilization of multilevel logistic regression models with variables from multiple sources to predict response propensities and raking procedures to include more variables for calibration to population control totals. The public data release for survey year 2019 will include sampling weights generated using the new procedures.

Suggested citation: Bramlett MD, Dahlhamer JM, Bose J and Blumberg SJ. New procedures for nonresponse adjustments to the 2019 National Health Interview Survey sampling weights. Published September, 2020.

Introduction

I. Problem Statement

The National Health Interview Survey (NHIS) has been in the field continuously since 1957 (1). There have been many advances in the field of survey sampling and operations in the decades since then, and periodically, the National Center for Health Statistics redesigns the survey questionnaire, sampling methods, and field operations to better reflect current best practices. A redesigned NHIS questionnaire was fielded in 2019, the first significant questionnaire redesign since 1997. In parallel, NHIS weighting procedures were updated in 2019. Updating weighting procedures was important, not only because of theoretical and methodological advances in the field, but also because household response rates for the NHIS have been steadily declining from roughly 92% in 1997 to 64% in 2018 (2). Lower response rates increase the potential for nonresponse bias to influence estimates, especially if the participants differ from the non-participants in the survey outcomes being measured. Weighting adjustments to correct for nonresponse bias that may have been adequate when the NHIS response rate was closer to 90% may no longer suffice when the response rate has dropped to about 60%.

The questionnaire redesign presented an opportunity to evaluate the weighting procedures and introduce enhancements or new techniques to better adjust for nonresponse. Many questions were deleted and many others were changed in various ways, including specific question wording, question order in the questionnaire, or the universe of respondents of whom the questions were asked (and there were new questions added as well). Any differences observed between estimates pre- and post-redesign may be due to real change in the population or partly attributable to the 2019 NHIS questionnaire redesign. Following a change in how survey data are weighted, similar uncertainty can result: any differences observed between estimates before and after the weighting change may be due to real change in the population or partly attributable to the updated weighting approach. NCHS decided to update the weighting procedures developed to adjust for nonresponse at the same time as the implementation of the redesigned questionnaire, to concentrate the uncertainty at one point in time rather than spreading it across multiple years. This paper documents the new weighting procedures developed to adjust for nonresponse in the 2019 NHIS data.

II. Background

The redesigned NHIS interview begins with a rostering of the household to identify all household members, to collect demographic data regarding each member, and to randomly select one sample adult and one sample child (if any children live in the household) for the detailed sample adult and sample child interviews. The adults typically respond for themselves, while an adult in the household who is knowledgeable about the child completes the interview for the child. All sampled housing units and individuals have a 'base' weight associated with them, which reflects their probability of selection. The sum of these base weights should be close to the total size of the number of households in the Unites States. In order to preserve this property, the weights of nonrespondents are distributed among respondents using different algorithms. This process is called nonresponse adjustment.

The nonresponse adjustment that was utilized through 2018 was based solely on geography. The weights for households, sample adults, and sample children were adjusted for nonresponse by multiplying the base weight by the inverse of the response rate within the sample segment (e.g., if the response rate in the geographic segment was 50%, the ratio adjustment factor was 2). Sample segments were defined by field operations as geographic units similar to Census block groups, generally consisting of approximately 25 sample addresses per quarter (approximately 100 sample addresses per year). The ratio adjustment factor was applied evenly to all housing units[1] in the geographic segment. The ratio adjustment was "capped" at 2, meaning that if the inverse of the response rate within the segment was larger than 2 (that is, if fewer than half of eligible households responded), the ratio adjustment for the respondents in the segment was 2 (i.e., the weight was doubled). If response in the segment was 0 (i.e., no households in the segment responded to the survey), or if the adjustment factor was capped at 2, a second-stage adjustment was performed at a higher level of geography: all the responding households in the county/county equivalent/county group had their weights inflated, again via simple ratio adjustment. The second-stage ratio adjustment factor was calculated as the sum of the base weights for all eligible households divided by the sum of the base weights for the responding households. The second-stage adjustment was capped at 1.5.

Against this backdrop, an evaluation was performed in 2019 to assess whether predicted response propensity could be used to improve the nonresponse adjustment. Various modeling

[1] For simplicity in description, the household is used here as the unit of response, but these procedures were applied to the household weights, the sample adult weights and the sample child weights in the same manner.

methods that could be used to predict response propensity for both respondents and nonrespondents were evaluated, including logistic regression, multilevel logistic regression, random forest prediction, and least absolute shrinkage and selection operator (LASSO) machine learning. Each modeling method included contextual data and paradata (i.e., data recorded by the interviewer in the process of recruiting sample households and contextual data measuring population characteristics in the area) in the set of predictor variables. The paradata considered included data from the household contact history instrument and the neighborhood observations instrument (two questionnaires that interviewers complete when attempting contact with a selected housing unit); the contextual data included medical population data from the county-level Area Health Resource File (AHRF) and demographic data from the tract-level Census Planning Database (CPD). The variables selected for modeling were available for respondents and nonrespondents alike and were correlated with response propensity as well as with key health outcomes from the NHIS. Models were estimated separately for households, for sample adults, and for sample children. The models were used to create weighting adjustment classes based on predicted response propensity. The assumption underlying this approach is that sample units that were less likely to respond but still completed the interview would be more similar to nonresponding units than were units with higher response propensity.

Once nonresponse weighting adjustments are complete, most surveys 'calibrate' the weights such that the sum of weights both overall and for specific subgroups either match or are close to known population totals from external sources. The two calibration methods examined were post-stratification and raking. Post-stratification was the process used by the NHIS through 2018, in which subgroup cells defined by simultaneous cross-classifications of age, sex and race/ethnicity had their sampling weights inflated such that the sum of each subgroup's weights matched the population counts of the same subpopulation, based on total counts provided by the Census Bureau. Raking is an iterative process in which weights are proportionally inflated within categories of demographic variables one at a time until marginals match or approximate population counts across all dimensions. Raking is typically able to handle more variables than post-stratification.

The evaluation included an examination of the following different options for final calibration to population control totals: keeping the post-stratification as it was through 2018 (post-stratifying within 100 adjustment cells formed by age, sex and race/ethnicity); raking to those same dimensions rather than post-stratifying; and raking to those dimensions while adding education, employment status, Metropolitan Statistical Area (MSA) status, and/or Census division. Including socioeconomic and/or geographic factors should improve the calibration over just including basic demographic factors.

Modeling and calibration alternatives were evaluated by how far the adjustment moved the estimates based on comparison with base-weighted estimates, as well as by how closely the estimates matched to estimates derived from the 2018 American Community Survey (ACS), the most timely available source of comparable sociodemographic estimates. The different modeling methods resulted in nonresponse adjustment factors for the low propensity weighting class that were often quite high. Weighting is always a trade-off between bias and variance: if no maximum value (a "cap") is imposed on the nonresponse adjustment factor, the heterogeneity introduced into the weights may be large, resulting in large variances and low statistical power. Capping reduces the variance increase at the expense of reducing the bias correction. Greater bias may result from capping. Therefore, various capping alternatives were evaluated based on their effect on the magnitude of the bias and the variance.

The remainder of this paper documents the methods that were chosen from this evaluation and applied to the 2019 sampling weights. A later publication will describe, in detail, all the various options assessed and the analyses that informed the final decision.

Solution

III. Summary

The nonresponse adjustment method that was chosen for the final 2019 sampling weights was an adjustment within quintiles formed from predicted response propensities generated from a multilevel logistic regression model. Households within a quintile received the following nonresponse adjustment factor: 1/(median response propensity for the quintile). The nonresponse adjustment was capped at 2.5.

This process was applied to the sample adult and sample child weights as well. Once the household base weight was adjusted for nonresponse, it was used as the starting point for the sample adult and sample child weights. For each sample adult, the nonresponse-adjusted household weight was multiplied by the inverse of the adult's probability of selection within the household. Then the sample adult nonresponse adjustment was applied. As with households, sample adults were split into quintiles based on the response propensities output from the sample adult multilevel response model. Adults within a quintile received the same nonresponse adjustment factor as for households: 1/(median response propensity for the quintile). The same process, again starting with the nonresponse-adjusted household weight, was then applied to the sample child weight as well. The nonresponse adjustments for the sample adult and sample child weights were each capped at 2.5.

The set of covariates included in the multilevel logistic regression models as predictors of response propensity were drawn from the contact history and neighborhood characteristics data compiled by the interviewers, as well as the external CPD and AHRF data, and are displayed in Table 1. The covariates that were selected for the household, sample adult and sample child models were chosen because they were related to both response propensity and key survey health outcomes in various domains: health status, health insurance coverage, health care access, health care utilization, health behaviors, and for the sample child model, stressful life events. The full set of covariates considered for the models, as well as a description of the covariate selection process, will be detailed in the forthcoming report.

The chosen method for calibrating the 2019 sample adult weights to population control totals was raking to dimensions formed from 18 categories of age by sex; 26 categories of age by race/ethnicity (Hispanic, non-Hispanic black, non-Hispanic Asian, non-Hispanic other); education (4 categories); and 23 categories of Census Division by MSA status (large MSA, small MSA, non-MSA). Employment status had also been considered for inclusion among the raking variables but ultimately was not chosen. The control totals for age by sex and age by race/ethnicity were provided to NCHS directly by the Census Bureau. The control totals for education, MSA status, and Census Division were calculated using ACS 2018 estimates. For the sample child weight, a smaller set of variables was used in the raking routine: age by sex (10 categories); age by race/ethnicity (15 categories); MSA status; and Census Division.

The final multilevel logistic regression models used to predict response propensity at the household, sample adult and sample child levels in 2019 are not necessarily the same models that will be used in subsequent years. The process of selecting the appropriate covariates to include in the model will be repeated for each new year, and thus the specific model may differ by year. When the model is updated and revised annually, the effects of the various capping alternatives will be reevaluated as well, and the specific cap level may thus be changed as a result of that evaluation. The dimensions used for final raking to population totals are also subject to annual re-evaluation and updating.

The conditional response rates for the sample adult and sample child interviews, given a household-level response, remain high as of 2019, approximately 90% for both the sample adult and sample child interviews, so the nonresponse adjustment at those stages doesn't have as large an impact on the survey estimates as does the household level nonresponse adjustment. However, this step was included in the weighting process in case these conditional response rates decrease over time. By including the adult and child nonresponse adjustment steps within the 2019 weight procedures, it allows for the procedures to be adapted to continue to account for nonresponse bias should there be decreases in conditional response.

IV. Conclusion

The new procedures for adjusting the sampling weights for nonresponse represent an improvement over the simple geography-based adjustments of the past. Differential adjustment by response propensity makes use of the assumption that respondents who show low response propensity are similar to nonrespondents with low response propensity and respondents with high response propensity are similar to nonrespondents with high response propensity. Including paradata in predictive models is also a more sophisticated procedure than was used previously, and the annual updating of the model will help ensure that the adjustment continues to perform well – the previous approach was static and could not be adjusted to account for changing patterns of nonresponse over time. And incorporating additional dimensions in a raking procedure is superior to post-stratifying to basic demographic totals only.

References

1. Parsons VL, Moriarity C, Jonas K, *et al*. Design and estimation for the National Health Interview Survey, 2006–2015. *Vital and Health Statistics* 2(165): National Center for Health Statistics, 2014.

2. National Center for Health Statistics. Survey Description, National Health Interview Survey, 2018. Published 2018. Accessed April 19, 2020. Available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2018/srvydesc.pdf.

Table 1: Covariates used in 2019 Nonresponse Propensity Models, by Source and Sample Unit

Source *Variable*	*Sample Unit*		
	Household	*Sample Adult*	*Sample Child*
Contact History Instrument File			
Left materials such as advance letter, note/appointment card, informational packet	YES	YES	NO
Checked with neighbors, property manager/doorman, other family members whereabouts of householder(s)	YES	YES	NO
Called household/left message on answering machine	YES	YES	YES
Contact reluctance reason: expressed privacy/anti-government concerns, asked questions about survey content	YES	NO	YES
Contact reluctance reason: expressed time constraints	YES	YES	YES
Case reassigned to a different interviewer	YES	YES	YES
Scheduled an appointment	NO	YES	NO
Neighborhood Observation Instrument File			
How would you describe the condition of the sample unit or the building within which the sample unit resides?	YES	NO	YES
Based on your observation, does the sample unit have an adult-sized bicycle?	YES	NO	NO
Based on your observation, would you say at least one adult resident of the sample unit is employed?	YES	YES	NO
How old would you estimate the age of the residents to be?	YES	YES	NO
Would you judge this sample unit to have a household income in the bottom third, middle third, or top third?	YES	YES	NO
Would you say that the residents of the sample unit speak a language other than English?	YES	NO	NO
Does the sample unit have any indication that the residents of the sample unit are smokers?	YES	YES	NO
Does the sample unit or the building within which the sample unit resides have a well-tended yard or garden?	YES	YES	NO
Does the sample unit have a wheelchair ramp or other indicators that the residents of the sample unit are handicapped, disabled, or may have a chronic health condition (deaf, blind, use oxygen, etc.)?	YES	NO	NO
Census Planning Database File (Tract level)			
Average aggregate household income, American Community Survey (ACS)	YES	NO	NO
Average number of persons per ACS occupied housing unit	YES	YES	NO
Percentage of ACS population aged 25 years and over that have a college degree or higher	YES	NO	NO
Percentage of ACS occupied housing units that have more than 1.01 persons per room (quintiles)	YES	YES	NO
Percentage of addresses for which the first Mailout/Mailback form mailed was completed and returned	YES	NO	YES
Percentage of ACS housing units that are considered mobile homes (quartiles)	YES	NO	NO
Percentage of ACS occupied housing units where a householder lives alone or with nonrelatives only	YES	NO	NO
Percentage of ACS population Mexican, Puerto Rican, Cuban, or other Hispanic/Latino/Spanish origin (quintiles)	YES	YES	NO
Percentage of ACS population non-Hispanic Asian Indian, Chinese, Filipino, Korean, Japanese, Vietnamese, or other Asian (quintiles)	YES	NO	NO

Source *Variable*	*Sample Unit*		
	Household	*Sample Adult*	*Sample Child*
Census Planning Database File (Tract level) -- continued			
Percentage of ACS population aged 5 years and over that speaks a language other than English at home	YES	YES	NO
Percentage of ACS population that is 65 years old or over	YES	NO	NO
Percentage of 2010 ACS family-occupied housing units with a related child under 6 years old	YES	NO	NO
Percentage of 2010 Census total population that lives outside of an Urbanized Area or Urban Cluster (tertiles)	YES	YES	NO
Percentage of ACS housing units classified as the usual place of residence of the individual/group living in it	YES	YES	NO
Percentage of ACS population that have two or more types of health insurance	YES	NO	NO
Percentage of ACS housing units that do not have complete plumbing facilities (quartiles)	NO	YES	NO
Percentage of ACS population that is between 5 and 17 years old	NO	YES	NO
Percentage of ACS housing units that are in a building that was constructed in 2010 or later (tertiles)	NO	YES	NO
Percentage of addresses in a 2010 Census mailback area that were confirmed as vacant housing units (quintiles)	YES	NO	NO
Percentage of addresses in a 2010 Census mailback area deleted (did not correspond to a valid housing unit)	NO	YES	NO
Percentage of valid addresses where a 2010 Census form was expected to be delivered returned to Census	NO	YES	NO
Area Health Resource File (County level)			
Natural log of number of hospitals per 100,000 county residents	YES	NO	YES
Number of medical doctors per 100,000 county residents	YES	YES	NO
Number of deaths in 2018 per 100,000 county residents	NO	YES	NO
County included 1 or more registered toxic waste sites	NO	YES	NO