Different Approaches for Non-Response Adjustments to Statistical Weights in the continuous NHANES (2003-04)

Lisa B. Mirel, Vicki Burt, Lester R. Curtin, and Cindy Zhang

Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Road, Room 4214 Hyattsville, MD 20782
LMirel@cdc.gov

Abstract

It is possible to use different methods to weight National Health and Nutrition Examination Survey (NHANES) data. Typically when weighting NHANES data we use a weighting class approach of predefined cells based on a limited number of variables. However, there are model based alternatives that allow for more flexibility in calculating the statistical weights. This flexibility is important because additional variables in non-response adjustments may be needed to better reduce the bias arising from the fact that non-respondents may be different from those who responded to the survey. In this paper, using NHANES data from 2003-04, we illustrate alternative methods to calculate statistical weights using both a weighting class approach and a model based approach to account for unit non-response. The WTADJUST procedure in SUDAAN version 10.0 was used to calculate weights using different approaches. Different weighting approaches were assessed using examples with the environmental subsample data and the dietary recall data from NHANES. Summary statistics derived from the different methods were compared. In addition, the distributions of the weights were compared and the influence on variance inflation was assessed by comparing the estimated design effects due to the differential weighting. We found that the comparison of the different weighting approaches varied by analysis variables with subsamples. Therefore, it is important to assess different weighting approaches for specific subsamples and specific variables when analyzing data subject to differential non-response.

Introduction

This paper explores an alternative approach to the unit non-response weight adjustment for the environmental subsamples and the dietary recall data from the National Health and Nutrition Examination Survey (NHANES). The purpose of non-response adjustments is to reduce the bias when non-respondents are different from those who responded to the survey and to maintain known marginal totals for population estimates. NHANES typically includes different sets of weights, each incorporating a non-response adjustment for its designed subsamples, such as the environmental subsamples and the fasting subsample. Also, several "full sample" components may have additional missing observations. Not all components have there own set of weights, but analysts are encouraged to examine the component missingness to determine the proper analytic strategy. Often, reweighting the data for component non-response has little effect on the estimates (Gregg 2004, 1593).

The public use data files for the environmental subsamples and the dietary recall data have unique statistical weights for use in analyses. The environmental subsample and dietary recall weights are calculated by creating weighting classes based on demographic variables (age, gender, and race/ethnicity) and the day of the week for the dietary recall data. These weights reflect adjustments to the examined participant's sample weight (exam weights). The exam weights are computed using the base probabilities of selection with non-response and post-stratification adjustments so that the weights sum to known totals of the targeted group in the population. The exam weights are adjusted for non-response based on age, sex, household education, household size, race/ethnicity, self-reported health status, and length of stay at current residence. They account for non-response in the sample of people who responded to the household component of the survey. For each of the three different environmental subsamples, demographic cells are defined subject to observed sample size and the exam weights within a weighting cell are then ratio adjusted for the non-responders in each specific subsample. For the dietary recall data, day of the week is an additional variable used to define the adjustment cells.

In this paper we explore alternatives to the non-response weight adjustment for both the environmental subsample weights and the dietary recall weights. The alternative approach attempts to use additional auxiliary information to, hopefully, better adjust the weights for factors related to non-response. Including more explanatory variables may help reduce bias in the estimates. We use a model based approach to adjust the weights for the environmental subsample while requiring that weights sum to the known totals of the targeted group in the population. The model based approach allows for more flexibility by permitting the use of continuous variables, more main effects, and lower order interactions in the model.

Because of the flexibility of the model it is not necessary to collapse levels of the variables due to small sample sizes (this can be a problem in the classic weighting cell approach). The model based approach is advantageous when there are good correlates with response propensity. Conversely, this method will not help to reduce bias if adding auxiliary data does not change the weights very much.

Background

The National Health and Nutrition Examination Survey (NHANES) is a multi-purpose survey designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements. Laboratory tests are conducted using biological specimens collected in the exam. Findings from this survey can be used to determine the prevalence of major diseases and risk factors for diseases. Information can be used to assess nutritional status and its association with health promotion and disease prevention. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure.

In NHANES, a representative cross-sectional sample of the U.S. civilian, non-institutionalized population was selected by using a complex, multistage probability design. Design variables and statistical weights specific to participants who completed the interview and the examination components are provided for use in all analyses. Certain components have unique weights that are also provided. Examples of a component with unique weights are the environmental exposure subsamples. NHANES cannot collect environmental exposure data on all participants; therefore, random (typically one-third) subsamples are selected. Unique weights that adjust the Mobile Examination Center weights for non-response of the component are calculated for use in the analyses of the environmental subsamples. Similarly, unique weights are calculated for the dietary recall data. The dietary recall weights adjust the exam weights for non-response and the day of the week that the data were collected. Both sets of unique weights (environmental and dietary) are adjusted so that the sum of weights equals the population totals. The impact of reweighting the data on estimates of means and proportions is minimal. However, the sum of the adjusted weights for the specific component equals the population totals for the specific age, gender and race/ethnicity weight classes.

Methods

This paper examines data from NHANES 2003-04. For the environmental subsample component we analyzed specific chemical analytes for heavy metals. A 1/3 subsample of NHANES participants, ages 6 years and older, were randomly selected to be representative of the U.S. population. There were 2,621 respondents in the heavy metals laboratory subsample. A respondent had to have at least one analyte measured in the subsample to be assigned a weight. In this paper we assessed Cesium and Molybdenum. There were 2,558 participants in the subsample that had Cesium and Molybdenum measured. The levels of Cesium were fairly consistent across age, gender and race/ethnicity groups while the levels of Molybdenum were more variable across the groups (data not shown). We assessed the distribution of several unweighted demographic characteristics, comparing NHANES participants who completed at least one component in the Mobile Examination Center (MEC), participants in the environmental subsample and non-participants in the environmental subsample but who completed at least one exam component (Table 1).

There were some differences in demographic characteristics between participants and non-participants in this environmental subsample; all percentages are unweighted to illustrate the universe before weighting (Table 1). The Chi-square test of independence or Fishers Exact Test, when necessary, was used to assess statistically significant differences, two-sided p-value <0.05, between demographic characteristics, and response propensity. Significant associations were observed between participation in the environmental subsample component and the participants' age, gender, and race/ethnicity. In addition, other associations that we found between participants and non-participants included poverty index and marital status. There was no association between household size and response propensity. This led us to explore including additional variables to the weighting scheme to see if we could reduce bias while minimally increasing variance.

Table1. Characteristics of Sample Participants (SPs) for those aged 6 years and older in the heavy metals subsample

		Unweighted n SPs examined at the MEC	*Unweighted % of total at the MEC*	*Unweighted n SPs participation in Subsample*	*Unweighted % of total participants Subsample*	*Unweighted n non-participants in Subsample*	*Unweighted % of total non-participants in Subsample*
Age (years)	Total	7,982	100.0	2,621	100.0	52	100.0
	6-11	992	12.4	292	11.2	13	25.0
	12-19	2,248	28.2	741	28.2	16	30.8
	20+	4,742	59.4	1,588	60.6	23	44.2
Race/Ethnicity	Total	7,982	100.0	2,621	100.0	52	100.0
	Non Hispanic Black	2,090	26.2	724	27.6	9	17.4
	Mexican American	1,932	24.2	624	23.8	9	17.4
	White/ Other	3,960	49.6	1,273	48.6	34	65.4
Gender	Total	7,982	100.0	2,621	100.0	52	100.0
	Male	3,892	48.8	1,313	50.0	16	30.8
	Female	4,090	51.2	1,308	50.0	36	69.2
Poverty Index (imputed)	Total	7,982	100.0	2,621	100.0	52	100.0
	<1	1,999	25.0	663	25.2	17	32.6
	1-<2	2,139	26.8	722	27.6	7	13.4
	2-<4	2,115	26.4	698	26.6	22	42.4
	≥4	1,729	21.6	538	20.6	6	11.6
Marital Status	Total	7,982	100.0	2,621	100.0	52	100.0
	Age <14	1,560	19.6	471	18.0	15	28.8
	Married	3,045	38.2	1,011	38.6	10	19.2
	Not Married	3,377	42.4	1,139	43.4	27	52.0
Household Size	Total	7,982	100.0	2,621	100.0	52	100.0
	1-3	3,960	49.6	1,320	50.4	30	57.6
	4-6	3,315	41.6	1,078	41.2	19	36.6
	≥7	707	8.8	223	8.6	3	5.8

For the environmental components we compared a series of weights: the unadjusted exam weight provided in the demographic dataset, the subsample weight provided in the heavy metals dataset, and several new weights that were calculated using the WTADJUST procedure in SUDAAN, version 10. The WTADJUST procedure in SUDAAN computes non-response weight adjustments using a model based calibration approach similar to a logistic regression. The model is based on a generalized exponential model (Research Triangle Institute, 731).

$$\theta_k = \gamma_k \alpha_k = \gamma_k \left[\frac{l_k(u_k - c_k) + u_k(c_k - l_k)\exp(A_k x'_k \beta)}{(u_k - c_k) + (c_k - l_k)\exp(A_k x'_k \beta)} \right]$$

k = Index corresponding to each record in the domain of interest.
Ω = Domain of interest. So $k \in \Omega$.
θ_k = Is the final weight adjustment for each record k in Ω.
γ_k = Is a weight trimming factor that will be computed before the β -parameters of the exponential model are estimated.
α_k = Is the non-response or post-stratification adjustment computed after the weight trimming step.
l_k = Lower bound imposed on the adjustment α_k.
u_k= Upper bound imposed on adjustment α_k.
c_k= Centering constant for the model.
$A_k = \frac{(u_k - l_k)}{(u_k - c_k)(c_k - l_k)}$
x_k= vector of model explanatory variables
β=Model parameters that will be estimated

The results from this model based approach will be equivalent to the weighting class approach if the highest order interaction of variables used to define the weighting classes is used with no main effects. Based on the assessment in Table 1 we recalculated subsample weights for the environmental components by 1) using age as continuous instead of categorical in a three way interaction with race/ethnicity and gender without main effects (Continuous Age); 2) using continuous age, race/ethnicity and gender as main effects and all two and three way interactions in the model (continuous Age and Main Effects); 3) using categorical age, race/ethnicity, gender, and a poverty index as main effects and all two and three way interactions of age, race/ethnicity and gender in the model (Add in Poverty Index); and 4) using categorical age, race/ethnicity, gender, a poverty index, marital status and household size as main effects and all two and three way interactions of age, race/ethnicity and gender in the model (All Variables). If the poverty index variable (ratio of family income to poverty) was missing we imputed the variable, only for the purpose of non-response adjustment, using a weighted sequential hotdeck imputation procedure in SUDAAN. Gender, age and race/ethnicity were used as the imputation classes. Marital status was missing for participants less than 14 years of age, a separate category was assigned for them. Three participants over age 14 were missing marital status and were assigned to "not married" for the weight adjustment model. The final results are also presented for the unadjusted exam weight and without any statistical weights.

While the environmental data are based on a designed subsample, the dietary data are based on the full sample but has additional non-response. Of those individuals considered part of the full exam sample, 89 percent completed the day one recall and, of those completing a day one recall, 92 percent completed a second day recall. The dietary data are unique in that the outcome variables differ by day of the week and the number of sample persons is highly variable by day of the week. To compensate for the day of the week effect and to adjust for the attrition in the dietary recall data, the dietary data have a specific weight that has been adjusted for the day of the week, race/ethnicity and age. In this paper we used the total nutrient file collected at the MEC. There is variability in consumption of nutrients by day of the week which is noted below for energy and caffeine. To illustrate the variability by day of the week we assessed the unweighted geometric mean (the underlying data is log-normal) and unweighted "n" by the day of the week (Table 2).

Table2. Unweighted sample size (n) and unweighted geometric mean by day of the week for Day 1 data, Caffeine (mg) and Energy (kcal)

		Overall	*Sunday*	*Monday*	*Tuesday*	*Wednesday*	*Thursday*	*Friday*	*Saturday*
Caffeine	Unweighted n	8,894	1,493	604	786	697	899	2,361	2,054
	Unweighted Geometric Mean	44.47	50.38	42.42	48.09	42.21	48.66	40.31	43.73
Energy	Unweighted n	8,894	1,493	604	786	697	899	2,361	2,054
	Unweighted Geometric Mean	1,847.43	1,844.49	1,770.86	1,757.30	1,845.67	1,752.62	1,854.98	1,944.94

Then we compared summary statistics for caffeine and energy using the adjusted Day 1 weight, the exam weights and without any weights.

For both the environmental component and the dietary recall data we assessed the differences in the various weighting schemes for certain variables by comparing the geometric mean, standard error of the geometric mean, percentiles, and the coefficient of variation (CV). Geometric means were reported because all variables were highly skewed. In addition, we assessed the summary statistics for the different weights by comparing the minimum and maximum weight values, the mean, the weighted totals, the CV of the weights and the estimated design effect due to differential weighting.

Given that it is difficult to measure bias because we do not actually know the true value in the population, our methods involved comparing the distributions of the weights and assessing the estimated design effects due to the differential weighting of the weighting class and model based methods. This paper thus provides different methods for assessing the weight adjustments and offers some diagnostics to assess which method is more appropriate.

We analyzed all NHANES data using SUDAAN Release 10.0 to incorporate the survey design and the statistical weights.

Results

Overall changing the weight for the environmental subsample data does not appear to make significant differences in the results of the summary statistics for Cesium and Molybdenum. Similarly, using the Day 1 dietary weight in analyses produces similar summary statistics to using the exam weight for both Caffeine and Energy. However, the differences reside in the sum of the weights and slight changes in variance. By design, the subsample and the adjusted weights both maintain the correct population totals for the particular sub groups. The exam weight, however, does not. This difference is expected because the exam weight is not adjusted to account for the 1/3 subsample and non-response. The dietary data are similar. Their weights have been adjusted for non-response from the full exam sample. The exam weight that is used in the tables below does not account for these adjustments.

Environmental Variables

The summary statistics for Cesium and Molybdenum (Table 3) illustrate the results of an unweighted analysis, where the CV is small but the point estimate and variance are not accounting for the sample design and therefore may be incorrect. The other weighting schemes have some shifts in the distributions, the percentiles, compared to the unweighted estimates. The CVs for the procedures that incorporate weights for Cesium range from 2.84 to 2.90 and for Molybdenum range from 2.32 to 2.38. The increase in variance between the alternative weighting methods is minimal.

Table 3. Comparing different weighting approaches in for Cesium (ug/L) and Molybdenum (ug/L) for geometric mean, standard error of geometric mean, coefficient of variation (CV), and select percentiles for Sample Participants with non-missing data

		Sample Size	*Geom. Mean*	*SE Geometric Mean*	*CV (%)*	*10th*	*25th*	*50th*	*75th*	*95th*
Environmental Variable	*Weight procedure*									
Cesium	*Unweighted*	2,558	4.73	0.09	1.97	1.89	3.18	5.18	7.41	11.74
	Exam	2,558	4.67	0.13	2.84	1.76	2.92	5.13	7.66	12.63
	Subsample	2,558	4.67	0.13	2.88	1.76	2.91	5.13	7.67	12.69
	Continuous Age	2,558	4.68	0.13	2.84	1.76	2.93	5.16	7.67	12.64
	Cont. Age and Main Effects	2,558	4.68	0.13	2.87	1.76	2.93	5.16	7.68	12.67
	Add in Poverty Index	2,558	4.67	0.14	2.90	1.75	2.91	5.13	7.67	12.70
	All Variables	2,558	4.67	0.14	2.90	1.75	2.91	5.13	7.67	12.70
Molybdenum	*Unweighted*	2,558	44.84	0.95	2.13	14.15	26.23	49.20	82.78	150.23
	Exam	2,558	39.46	0.93	2.37	11.50	22.34	44.30	78.13	137.78
	Subsample	2,558	39.68	0.94	2.36	11.50	22.38	44.41	78.40	137.75
	Continuous Age	2,558	39.82	0.93	2.32	11.56	22.57	44.56	78.46	138.09
	Cont. Age and Main Effects	2,558	39.75	0.93	2.33	11.55	22.51	44.49	78.37	137.85
	Add in Poverty Index	2,558	39.61	0.94	2.38	11.49	22.35	44.37	78.36	137.67
	All Variables	2,558	39.63	0.94	2.38	11.49	22.36	44.38	78.36	137.68

The comparison of summary statistics of the different weights (Table 4) indicate a narrow range of values for both the estimated design effect due to differential weighting ($1+CV^2$), 1.82 to 1.88, and for the CV of the weights, 90 to 94. However, the exam weights do not sum to the correct population totals for the 1/3 subsample. The alternative weighting procedures do sum to the correct population totals. The exam weights sum to the correct population totals for the full exam sample.

Table 4. Summary statistics for the different weights and the estimated design effect due to differential weighting in the Heavy Metals Subsample. Number of respondents in Table 4 is different than the sample size in Table 3 since Table 4 includes all respondents in the Heavy Metals Subsample and Table 3 includes only those Sample Participants with a non-missing value for Cesium and Molybdenum.

Weight procedure	*Number of respondents*	*Minimum weight*	*Maximum weight*	*Mean weight*	*Weighted total*	*CV of the weights (%)*	*Estimated design effect due to differential weighting*
Exam	2,621	2,145	138,287	32,481	85,132,757	90	1.82
Subsample	2,621	5,567	455,772	100,144	262,476,798	94	1.88
Continuous Age	2,621	5,482	418,318	100,144	262,476,798	91	1.83

Weight procedure	*Number of respondents*	*Minimum weight*	*Maximum weight*	*Mean weight*	*Weighted total*	*CV of the weights (%)*	*Estimated design effect due to differential weighting*
Cont. Age and Main Effects	2,621	6,463	423,866	100,144	262,476,798	92	1.85
Add in Poverty Index	2,621	5,386	474,211	100,144	262,476,799	94	1.88
All variables	2,621	5,319	474,002	100,144	262,476,799	94	1.88

Dietary Variables

We note some differences with the summary statistics for Caffeine and Energy when we compare the Day 1 weights with the exam weight (Table 5). This is probably due to the day of the week variation on total nutrient intake. For Caffeine the variance is smaller with the Day 1 weight compared to the exam weight while the variance is larger with the Day 1 weight for Energy. For Caffeine the weighted distribution with the Day 1 weights is shifted slightly left compared to the weighted distribution with the exam weight. While the distribution of Energy is shifted slightly right with the Day 1 weight compared to the exam weight.

Table 5. Comparing summary statistics, geometric mean, standard error of geometric mean, coefficient of variation (CV), and select percentiles for Sample Participants with non-missing data for Caffeine (mg) and Energy (kcal) unweighted, with the exam weights and with the Day 1 weight.

		Sample Size	*Geom. Mean*	*SE Geometric Mean*	*CV (%)*	*10th*	*25th*	*50th*	*75th*	*95th*
Environmental Variable	*Weight procedure*									
Caffeine	*Unweighted*	8,894	44.47	1.95	4.39	.	0.04	29.12	124.19	402.57
	Exam	8,894	73.47	3.58	4.87	.	5.15	76.43	204.81	531.77
	Day1	8,894	72.22	3.27	4.53	.	4.81	73.51	202.59	524.42
Energy	*Unweighted*	8,894	1,847.43	14.33	0.78	993.40	1,379.50	1,894.33	2,557.50	3,913.90
	Exam	8,894	1,969.92	16.82	0.85	1,100.61	1,476.87	2,002.68	2,706.84	4,108.59
	Day1	8,894	1,954.29	17.67	0.90	1,091.44	1,467.84	1,999.22	2,682.58	4,033.42

"." Could not estimate the 10^{th} percentile because of ties in the data

We note that the CV of the weights is smaller for the exam weights than the Day 1 weights. Similar to the environmental variables the weighted totals differ between the exam and Day 1 weights. The Day 1 weights include adjustments to the exam weights that account for the non-response to the dietary recall section (Table 6).

Table 6. Summary statistics for the different weights and the estimated design effect due to differential weighting for exam weights and Day 1 weights

Weight procedure	*Number of respondents*	*Minimum weight*	*Maximum weight*	*Mean weight*	*Weighted total*	*CV of the weights (%)*	*Estimated design effect due to differential weighting*
Exam	9,034	1,533	145,844	29,761	268,856,842	95	1.90
Day1	9,034	887	293,829	31,683	286,222,757	121	2.46

Conclusion

Given the different reasons and propensity for non-response in the different components of NHANES, there may not be just one method that can be applied to the weight adjustment, and each component may require a different type of adjustment. Adjusting weights for non-response is important to help reduce bias in the estimates. In our examples this importance is illustrated by the slightly different estimates in the total nutrient data. This might be explained by differences in food consumption on certain days of the week. However, when the non-response appears to be random, reweighting does not seem to make a difference in the estimates.

It is important to assess the different components of NHANES before proceeding with an analysis. As detailed in our methods, many tools can be used with success to adjust weights from complex surveys.

Acknowledgements

We would like to thank Jeffery Hughes and Brittany Mellinger for their helpful suggestions and comments.

References

Gregg, Edward W., Paul Sorlie, Ryne Paulose-Ram, Quiping Gu, Mark S. Eberhardt, Michael Wolz, Vicki Burt, Lester Curtin, Michael Engelgau, Linda Geiss. 2004. Prevalence of Lower-Extremity Disease in the U.S. Adult Population ≥40 Years of Age with and Without Diabetes. *Diabetes Care 27* (7): 1591-1597.

Research Triangle Institute (2008). *SUDAAN Language Manual, Release 10.0* Research Triangle Park, NC: Research Triangle Institute.

Lohr S: *Sampling Design and Analysis*. Albany, NY, Duxbury Press, 1999.

Kish L: *Survey Sampling*. New York, NY, John Wiley & Sons, Inc., 1965.