

AN EVALUATION OF NONRESPONSE BIAS IN NHANES III (1988-91)

Meena Khare¹, Leyla K. Mohadjer², Trena M. Ezzati-Rice¹, Joseph Waksberg²

¹National Center for Health Statistics, ²WESTAT, Inc.

Meena Khare, National Center for Health Statistics, 6525 Belcrest Road, Rm 1000, Hyattsville, MD 20782

Key Words: Sample surveys, Weighted response rates, Unit and item nonresponse

Introduction

Unit and item nonresponse occur virtually in all population based surveys such as the National Health and Nutrition Examination Survey (NHANES), and affect the quality of the survey statistics in various ways. NHANES is a periodic survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention. It is designed to provide national estimates of health and nutritional status of the civilian noninstitutionalized population of the United States. In NHANES, medical history, and sociodemographic information are collected through household interviews, and standardized physical examinations are conducted in Mobile Examination Centers (MECs) to collect data on physical measurements, physiological tests, and biochemical measurements from blood and urine specimens.

The third NHANES or NHANES III (1988-94) is the seventh in a series of periodic health and nutrition surveys conducted by NCHS since the 1960s. NHANES III, with a sample of approximately 40,000 persons aged 2 months and older. The survey has been divided into two 3-year national samples (phase 1 and phase 2), so that national estimates can be obtained for the individual and combined phases. NHANES III is based on a complex, multistage area probability sample design and includes an oversample of children under 5 years of age, adults aged 60 years and older, Mexican Americans, and African Americans. Also, NHANES III is designed only to provide national estimates by age, gender and/or three race/ethnicity domains. Details of the sample design and data collection have been previously published (1-2).

NHANES III, like most sample surveys based on voluntary participation, experiences both unit and item nonresponse. Missing data result from unit nonresponse to household interviews, and/or physical examinations, and from item nonresponse to various examination components, and items within components or within interview sections. Nonresponse in NHANES III occurs primarily due to refusal, or inability to participate. Also, according to protocol, participants with preselected health conditions are excluded from selected examinations for medical safety. In NHANES III, phase 1 (1988-91), unit nonresponse to the household interview was 14%, and an additional 9% did not participate in the physical examination at the MEC. It is common survey practice to

compensate for unit nonresponse through weighting class adjustments (3-6). A two-stage nonresponse adjustment for NHANES III, phase 1 has been previously described (7). In addition to unit nonresponse, NHANES also experiences various levels of item nonresponse. In the household interview, item nonresponse to specific interview questions ranged from 1-5%. The component level nonresponse in the MEC ranged from 2-19%. Additional item nonresponse to various measurements within individual components ranged from 5-8% among examined persons, varying significantly by age. Item nonresponse is usually handled by some type of imputation-- single or multiple (8-12). Imputation methods substitute the missing items with one or more plausible values from similar units in the dataset or with predicted values obtained from a model, thus making it possible to use complete data methods for analyses. Special attention should be given to the imputation process, imputed values, and marginal and overall distributions of the data. If imputations are not done carefully, they can introduce large bias instead of reducing bias.

Potential for bias increases as the response rate decreases. To achieve higher response rates, NHANES III employs several field procedures including incentives to sample persons, such as cash, "certificate of appreciation", and "report of findings". The purpose of this paper is to provide a guideline to assess level of nonresponse and evaluate potential bias due to unit and item nonresponse in the NHANES III, phase 1. The methods section presents the evaluation process and significance of the weighted and unweighted response rates (13). Compensation methods to adjust for nonresponse and reduce nonresponse bias are also outlined in this section (14-18). The results section presents the nonresponse experienced in the NHANES III, phase 1. Finally, a summary of results and recommendations are presented including factors that should be considered when reporting results from a survey item with moderate or high nonresponse (13).

Methods

While higher response rates may result in less-biased data, there is no "magic" response rate at which data become free from bias. Sometimes an evaluation of bias may be possible if sufficient information for nonrespondents is available from other sources. When there is a substantially large nonresponse, an assessment of nonresponse bias becomes necessary to assess data quality and potential effects on survey estimates. This

section presents a systematic approach for evaluating the potential nonresponse bias.

Assessment of level of nonresponse

An assessment of the level of missing data due to nonresponse is the first step in evaluating potential for bias. The extent of missing data can greatly influence the survey estimates, their quality and interpretation. Nonresponse rates measure the magnitude of missing data, and can be derived in two ways: unweighted and weighted. Both rates are essential for a survey and serve different purposes. Unweighted response rates are used for administrative purposes as an indicator of how well the survey operations were carried out. They are particularly useful for quality control, and to measure the success of field operations. Also, unweighted nonresponse rates are useful for measuring the rate of missingness at various stages of the data collection.

Weighted response rates, on the other hand, are more appropriate for analytic purposes in examining the potential effect of nonresponse on survey estimates. Since survey estimates are based on weighted data, weighted response rates are better indicators of data quality. Sample size, relative errors, measurement errors, and weighted response rates should all be considered in drawing valid inferences from the survey estimates. In NHANES III, nonresponse occurs at several stages-- interview, examination, component, and items within interview and examination components. The combined effect of unit and item nonresponse should be considered in calculating the overall nonresponse rate.

The unweighted response rates are computed from the sample data. For example, the interview response rate is computed as the ratio of the number of interviewed persons to the total sample size and the examination rate as the number of examined persons to the total sample size. The nonresponse rate is then one minus the response rate or 100 minus the response rate expressed as percent.

The weighted response rates in NHANES III are computed in the following way. The weighted interview rate $WR_{(interview)j}$ is computed as the ratio of the sum of the post-stratified basic weight (PW) for interviewed persons to the total noninstitutionalized U.S. population aged 2 months and older. For a domain or subgroup j, the interview rate is defined as

$$WR_{(interview)j} = \frac{\sum_{i=1}^{n_j} PW_{ij}}{\sum_{i=1}^{nk_j} PW_{ij}}$$

where,

$WR_{(interview)j}$ = weighted interview rate in domain j
 PW_{ij} = post-stratified basic weight for

interviewed person i in domain j

nk_j = number of persons interviewed in domain j

n_j = number of persons screened in domain j

Similarly, the weighted examination response rate $WR_{(exam)j}$ is computed as the ratio of the sum of post-stratified basic weights (PW) for all examined persons to the sum of post-stratified basic weights for all interviewed persons in the sample. For persons in a domain or subgroup j, the examination rate is then

$$WR_{(exam)j} = \frac{\sum_{i=1}^{n_j} PW_{ij}}{\sum_{i=1}^{nk_j} PW_{ij}}$$

where,

$WR_{(exam)j}$ = weighted examination rate in domain j

PW_{ij} = post-stratified basic weight for examined person i in domain j

n_j = number of persons examined in domain j

nk_j = number of persons interviewed in domain j

The weighted item response rate $WR_{(item)j}$ is computed as the ratio of the sum of the post-stratified basic weights (PW) for all examined persons with usable data for the item to the sum of the post-stratified basic weights for all examined persons in the sample. Again, for persons in a domain or subgroup j, the item response rate is defined as

$$WR_{(item)j} = \frac{\sum_{i=1}^{n_j} PW_{ij}}{\sum_{i=1}^{nk_j} PW_{ij}}$$

where,

$WR_{(item)j}$ = weighted item response rate in domain j

PW_{ij} = post-stratified basic weight for person i with usable data in domain j

n_j = number of persons with usable data in domain j

n_j = number of persons examined in domain j

The overall weighted item nonresponse is then one minus the product of the weighted interview rate, the weighted examination rate among interviewed persons and the weighted item response rate among examined persons.

Effect of nonresponse on estimates

The next step in the evaluation process is the assessment of the effect of nonresponse on survey estimates by investigating characteristics of respondents for over or under representation, and to compare survey estimates among respondents and nonrespondents. The difference D , defined as $(Y_r - Y_{nr})$, where Y_r is a mean or proportion among respondents and Y_{nr} is a mean or proportion among nonrespondents, is a useful summary statistic to evaluate the difference between respondents and nonrespondents for a given characteristic. A positive value of the difference D , implies an over representation among respondents, while a negative difference indicates an under representation. A large positive or negative difference suggests a potential for bias. After evaluation, if necessary, survey estimates should be adjusted for nonresponse in order to reduce bias.

Adjustment methods to compensate for nonresponse

After extensive evaluation, if respondents appear to be different from nonrespondents, sample estimates may be biased. There are several methods to adjust for unit and item nonresponse. Generally weighting class adjustment methods are used to compensate for unit nonresponse under the assumption that response of respondents and nonrespondents within weighting classes are similar. Item nonresponse is generally handled by imputation-- single or multiple, where missing values are substituted by one or more plausible values from similar units in the data or from a model(8-12). With imputations, complete data methods can be used for estimation. Other methods that have been used historically to adjust for item nonresponse are-- direct standardization or post-stratification, response propensity stratification to adjust cell weights, or sensitivity analysis(14-18).

Results

All persons selected to participate in NHANES III, phase 1 were screened for the basic demographic characteristics age, gender, and race/ethnicity. The screening rate in NHANES III, phase 1 was 100% with 6% of the information obtained from neighbors. All screened persons who completed key sections of the household questionnaire were defined as interviewed and all interviewed persons who completed one or more examination components in the MEC were defined as examined.

Unit nonresponse

Of the 20,277 persons selected and screened in NHANES III, phase 1, 14% did not participate in the interview portion and an additional 9% were not examined at the MECs. The primary reason for nonresponse was refusal or inability to participate. Characteristics of persons who refused the interview or examination varied substantially by age, race/ethnicity and geographic location. Interview and examination response decreased as age increased. To maximize

participation in the NHANES III, multiple persons were selected from a household based on their demographic characteristics. Interview and examination response rates were positively correlated with household size. Non-Hispanic white persons had the lowest participation in comparison to non-Hispanic black or Mexican Americans. Persons living in Northeast, urban metropolitan cities had lower participation in comparison to persons living in other locations. Non-Hispanic white women aged 80+ years and living alone had the highest nonresponse to the examination. Table 1 presents the weighted interview and examination response rates by selected sociodemographic characteristics. Chi-squares to test association of demographic characteristics with examination response were significant at 0.05 level of significance for all selected characteristics except for the family income.

Item nonresponse

In NHANES III, more than 20 different physical examinations, referred to as components, are conducted in the MECs. These components are assigned to persons based on their age (1-2). The component nonresponse varied substantially by the sociodemographic characteristics of the participants and the type of component. The primary reason for non-completion of an examination component was refusal or inability to participate. For medical safety, participants with preselected health conditions were excluded from selected examinations (e.g., women with confirmed or suspected pregnancy were excluded from the bone densitometry test). The component nonresponse ranged from 2-19% among examined persons and item nonresponse within individual components ranged from 5-8%, varying significantly by age.

The guideline outlined in the methods section for evaluating potential for nonresponse bias was applied to examination components with high nonresponse rates. The remainder of this section focuses on evaluation of item nonresponse in measurements from two selected components: hemoglobin (HGB) from phlebotomy (to draw blood for biochemistry), offered to all persons aged 1 year or older, and macular degeneration scores for early detection of diabetes from fundus photography of the eye, offered to adults 40 years and older (14). Table 2 presents the weighted and unweighted response rates for the selected components. Tables 3 and 4 compare the distribution of sociodemographic characteristics among respondents and nonrespondents for the two selected components. Only characteristics with large difference ($D \geq 3\%$) between respondents and nonrespondents are presented in these tables.

Phlebotomy

The overall unweighted and weighted phlebotomy rates were 66.5% and 66.8%, respectively. However, these rates differed substantially by age. For children aged 1-5 years, the weighted phlebotomy rate

was 56.6%, whereas for older adults aged 60 years and above, the rate was 58.4%. Among children, participation to the examination was very high (90% of those interviewed), however, refusal by parents to have blood drawn for their children resulted in a low phlebotomy rate (67% of those examined). A comparison of respondents and nonrespondents (Table 3), revealed that nonparticipants were non-Hispanic white, children weighing less than 29 lbs, and from family with poverty index below poverty level. Older persons, on the other hand, had low examination response (80% of those interviewed), but a high phlebotomy participation (93% of those examined), resulting in an overall component rate similar to children. Table 3 shows that among older adults aged 60+ years, nonrespondents to phlebotomy were non-Hispanic black, female, with family poverty index below poverty level, none or up to 8th grade education, never taken vitamins, and having physical activity less than or similar to others of their age.

These analyses show that nonresponse for this component varies by sociodemographic characteristics suggesting potential for bias in biochemistry measurements, especially among young children and older adults. This implies that the survey estimates should be properly adjusted for nonresponse in order to reduce bias. Also, the above mentioned factors should be taken into account when analyzing and interpreting the results.

Fundus photography

The overall unweighted and weighted response rates for fundus photography were 54.1% and 59.2%, respectively. An increase in the weighted rate indicates that perhaps adults who completed the test had larger sampling weights and oversampled adults with smaller sampling weights did not complete the test at the same rate as the remainder of the sample (note that black Americans, Mexican Americans and older persons were oversampled in the NHANES III). A comparison of weighted response rates for persons aged 40-59 years and 60+ years (68.4% and 54.0%, respectively) showed that older adults had higher item nonresponse. An investigation of reasons for nonresponse revealed that most of the older persons who could not complete this test, had a hard time holding their head or eye still for a good gradable photograph. Furthermore, a comparison of characteristics between respondents and nonrespondents in table 4 shows that nonparticipants were older adults, non-Hispanic whites, non-Hispanic blacks, with family poverty index below the poverty level, and persons with self reported fair or poor health status, diabetes, cataract, having trouble with vision, suggesting an under representation in the respondent sample.

The above analysis indicates potential for nonresponse bias in estimates, especially among older adults and suggests the need for an adjustment for this additional item nonresponse. Again, it is recommended to

take these facts into account when drawing inferences or reporting results from this component.

Summary

There is always a potential for nonresponse bias whenever characteristics of participants are different from nonparticipants in a survey. Generally, weighting class adjustment method is used to compensate for the unit nonresponse. In NHANES III, phase 1, a two-stage weighting class adjustment method was used to compensate for the examination nonresponse. Basic weights were first adjusted for the interview nonresponse by age, race/ethnicity, region, SMSA and household size. At the second stage, first stage weights for the interviewed persons were further adjusted for the examination nonresponse by age, race/ethnicity, household size, family income and self-reported health status (7). Weights from the first and second stages were then post-stratified to the total noninstitutionalized U.S. population aged 2 months and older to obtain final interview and examination weights, respectively. Since component nonresponse differed substantially by the type of component and individual measurement, no adjustment was made for the item nonresponse.

Nonresponse adjustments, if properly done, can serve to reduce bias and minimize sampling error. However, total elimination of such bias in a survey is not possible because in each weighting class, survey respondents will never fully represent the nonrespondents. To estimate nonresponse bias quantitatively, it is necessary to obtain similar information on respondents and nonrespondents. In NHANES III, phase 1, older adults, non-Hispanic whites, persons living alone, living in the Northeast region and in metropolitan cities had higher rates of missing data at each stage of the data collection. Analysts should be careful in analyzing data and interpreting results. Note that NHANES III is not designed to provide regional or any other domain estimates except for age, gender and/or race/ethnicity domains.

It is recommended to use both weighted and unweighted response rates to investigate for potential bias in survey estimates. Differences in the distribution of characteristics among respondents and nonrespondents should also be evaluated to determine correlates of nonresponse and for assessment of under or over representation of the respondents for a given characteristics. Weighting class adjustments or response propensity models are recommended to adjust for unit nonresponse. To adjust for item nonresponse, imputation is recommended to substitute missing data with one or more plausible values from similar units in the respondent sample or predicted values from a model. Additionally, if possible, sensitivity analysis can be done for evaluating the effect of various levels of nonresponse on selected statistics (16).

In addition to the above analyses, when reporting results from survey items with moderate or high nonresponse, the following factors should be considered: 1. How much margin of error can be tolerated in the estimates? 2. Are there systematic risks to the target population, if incorrect inferences are drawn? 3. Are there *a priori* reasons to believe that respondents are different from nonrespondents. If so, appropriate adjustments should be made to the final estimates. 4. Do the conclusions based on the statistics agree or conflict with the results from previous studies? With inconclusive results, analysts should evaluate the implications to the target population, and if necessary, propose further detailed studies. Finally, when reporting results with fairly high nonresponse, in addition to sampling errors and other limitations, analysts should point out the potential effects of nonresponse on survey estimates.

References

1. Ezzati, T., Massey, J., Waksberg, J., Chu, A., and Maurer, K. Sample Design: Third National Health and Nutrition Examination Survey. National Center for Health Statistics. Vital Health Statistics. Series 2, No. 113, 1992.
2. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. National Center for Health Statistics. Vital Health Statistics. Series 1, No. 32, 1994.
3. Madow, WG, Olkin, I., and Rubin, DB, eds. Incomplete Data in Sample Surveys, Volume 2, New York: Academic Press, 1983.
4. Cox, B. Weighting Survey Data for Analysis. Presentation for the ASA Continuing Education Program. 1991 Joint Statistical Meetings, August, 1991.
5. Kalton, G. and Kasprzyk, D. The Treatment of Missing Survey Data. Survey Methodology, Vol. 12, No. 1, 1-17, Statistics Canada, 1986.
6. Little, RJA. Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review 1986: 54:139-157.
7. Ezzati, T. and Khare, M. Nonresponse Adjustments in a National Health Survey. 1992 Proceedings of the Survey Research Methods Section of the American Statistical Association, pp.203-208, 1993.
8. Little, RJA and Rubin DB. "Statistical Analysis with Missing Data." John Wiley and Sons, New York, 1987.
9. Little, RJA and Rubin, DB. Final Report on NHANES Imputation Project. Datametrics Research, Inc., MA, December, 1992.
10. Khare, M, Little, RJA, Rubin, B, Schafer, JL. Multiple imputation of NHANES III. 1993 Proceedings of the Survey Research Methods Section of the American Statistical Association. In press.
11. Ezzati-Rice, TM, Fahimi, M, Judkins, D, Khare, M. Serial Imputation of NHANES III with Mixed Regression and Hot-Deck Techniques. 1993 Proceedings of the Survey Research Methods Section of the American Statistical Association. In press.
12. Ezzati-Rice, TM, Khare, M, Rubin, DB, Little, RJA, Schafer, JL. A Comparison of Imputation Techniques in the Third National Health and Nutrition Examination Survey. 1993 Proceedings of the Survey Research Methods Section of the American Statistical Association. In press.
13. Mohadjer, LK, Waksberg, J. Accounting for item nonresponse bias in NHANES III, unpublished internal report, Westat Inc., MD, 1994.
14. Chapman, DD. Reporting complete nonresponse in a sample survey. Proceedings of Section on Survey Research Methods of the American Statistical Association, 1991. pp. 266-271.
15. Kalton, G. Compensation for missing survey data. Institute for Social Research, University of Michigan, Anne Arbor, 1983.
16. Klein, R, Rowland, ML, Harris, MI. Racial/ethnic differences in age-related maculopathy: Third National Health and Nutrition Examination Survey. Ophthalmology, 1994. In press.
17. Lepkowski, JM, Landis, JR et. al.. Statistical methodologies for analyzing a complex sample survey. National Center for Health Statistics. National Medical Care Utilization and Expenditure Survey. Series A, Methodological Report No. 4. DHHS Pub. No. 88-20004. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1988.
18. Rowland, ML, Forthofer, RN: Adjusting for nonresponse bias in a health examination survey, Public Health Reports, 1993(3), 380-386.

Table 1: Weighted Interview and Examination Response Rates by Selected Demographic Characteristics, NHANES III, phase 1, 1988-91

Demographic Characteristics	Sample Interview		Examination
	size	rate	rate
	n	%	%
Total	20277	82.8	73.8
Age			
2m-5 y	4647	93.9	85.0
6-19 y	3969	88.1	80.8
20-39 y	4249	81.4	73.5
40-59 y	3043	79.5	71.8
60+ y	4369	78.3	62.6
Gender			
Males	10131	81.7	72.8
Females	10146	84.1	74.8
Race/ethnicity			
NH-whites	8540	81.7	71.7
NH-blacks	5024	87.3	81.4
Mex-American	5979	87.9	81.4
Household Size			
1-2	5759	78.2	68.9
3-5	7949	83.1	76.1
≥5	6569	89.0	81.7
Region			
Northeast	2883	78.1	65.9
Midwest	3817	84.5	77.6
South	7438	84.3	75.8
West	6139	83.9	74.6
Urbanization			
not in SMSA	3828	88.4	82.1
SMSA, CC	7564	83.0	72.8
MA, not CC	8885	80.7	71.0
Interviewed persons only			
Marital Status			
Married	6088		84.9
Wid/Div/Sep	2277		79.5
Never married	2704		87.0
Education			
No school	4794		92.1
Elementary	4484		88.2
High school	5465		87.1
College+	2977		84.6
Family Income			
>\$10,000	3315		89.7
\$10,000-29,999	7083		88.7
\$30,000-49,999	3289		87.1
\$50,000+	2093		86.4
Poverty index			
<1	4478		92.8
≥1	11329		86.5

Table 2: Unweighted and Weighted Response Rates for Selected Examination Components, NHANES III, phase 1, 1988-91

Component	Age group	Unweighted	Weighted
Examination Rate	2 months+	77.1	73.8
Anthropometric measurements	2 months+	75.5	72.9
Phlebotomy	1 year+	66.8	66.5
Fundus photography	40 years+	54.1	59.2
Bone densitometry	20 years+	61.0	62.4

Table 3: Weighted Distribution of Selected Characteristics among Respondents and Nonrespondents of Phlebotomy, NHANES III, phase 1, 1988-91

Characteristics	Respondents	Nonrespondents
	%	%
Age: 1-5 years		
NH-white	67.1	73.0
Female	47.4	51.5
Poverty index ≥1	66.8	71.6
Weight <29 lbs	24.1	31.2
Age: 60+ years		
NH-black	8.2	14.8
Female	56.7	60.9
Poverty index <1	10.5	13.8
Education: 0-8 years	24.5	29.5
Ever taken vitamins: No	58.8	74.4
Physical activity: less than others	16.4	24.0

Table 4: Weighted Distribution of Selected Characteristics among Respondents and Nonrespondents of Fundus Photography, NHANES III, phase 1, 1988-91

Characteristics	Respondents	Nonrespondents
	%	%
Age: 60+ years	39.2	68.2
NH-white	82.5	79.0
NH-black	9.0	13.3
Income <\$10,000	12.0	20.0
Poverty Index <1	9.2	14.1
Health: fair/poor	21.6	34.4
Diabetes: yes	7.9	14.8
Cataract: yes	9.9	31.6
Trouble seeing: yes	14.0	25.7