NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY III

ACCOUNTING FOR ITEM NONRESPONSE BIAS

Prepared for:

National Center for Health Statistics
Hyattsville, Maryland

Prepared by:

Westat, Inc.
1650 Research Boulevard
Rockville, Maryland 20850

Leyla Mohadjer
Bridgett Bell
Joseph Waksberg

November 4, 1994

TABLE OF CONTENTS

LIST OF TABLES

ACCOUNTING FOR ITEM NONRESPONSE BIAS IN NHANES III

1. INTRODUCTION

The National Health and Nutrition Examination Survey (NHANES) is one of the major programs in the series of health-related studies conducted by the National Center for Health Statistics (NCHS). The first survey, started in 1956, concentrated on the adult population and the prevalence of chronic disease. The next two phases of the survey were conducted from 1963 to 1970 and were largely devoted to the growth and development of children. In 1971 the survey was expanded to include the population from 1 to 74 years of age and a battery of questions and measurements pertaining to nutrition was added. This survey is referred to as the first National Health and Nutrition Examination Survey or NHANES I and was in the field between 1971 and 1975. NHANES II was conducted from 1976 to 1980; and a special survey of Hispanics was conducted from July 1982 through December 1984 (HHANES).

NHANES III is the latest in the series of health surveys. It began in 1988 and will continue through 1994. The NHANES III sample was split into two random halves, so that each half (referred to as "phase" in this report) was a nationally representative sample of the target population. Phase 1 of the survey was conducted from 1988 to 1991 and some selected data items are now available for analysis.

Properly weighted estimates based on the data obtained from a health survey, such as NHANES, would be approximately unbiased (from a sampling perspective) if every sampled household agreed to participate in the survey and if every selected person was responsive to the interview and was examined. However, nonresponse is always present in any survey operation, even where participation is not voluntary. The best guard against potential nonresponse bias is to utilize field procedures that maintain high cooperation rates. For example, the payment of a $50 incentive and repeated callbacks for refusal conversion were very effective in reducing nonresponse and thus the potential for nonresponse bias.

Achieving high response rates has always been a major concern. In the early national health examination surveys the rates were high, averaging about 90 percent. However, the picture began to change with NHANES I. For NHANES I, the overall response rate was 74 percent; for NHANES II it was 73 percent; and for HHANES the rate was 73 percent. In phase 1 of NHANES III, the overall examination response rate increased to 77 percent, presumably because of the efforts

to achieve high response rates (i.e., a $50 incentive and extensive callbacks and refusal conversion procedures). The interview response rate in phase 1 of NHANES III was much greater, 86 percent.

NCHS carried out a two-stage procedure for nonresponse adjustment and poststratification to known population totals to adjust for unit (interview and examination) nonresponse for phase 1 of NHANES III. The practical consequence of this adjustment was that the distributions of characteristics of the pool of nonrespondents within an adjustment class were implicitly assumed to be the same, on average, as those of the respondents within the same adjustment class. If this implicit assumption holds, the estimates are effectively unbiased. Recent analyses (Ezzati, et al. 1992) of weighting adjustments for unit nonresponse confirmed this assumption. Differences between respondents and nonrespondents for the examined samples in phase 1 of NHANES III appeared to be minor, as measured by information reported in the interview phase.

Unfortunately, unit nonresponse is not the only response problem with the survey. NHANES, like other surveys which perform multiple observations on the same person, is subject to incompleteness of data not only through failure to interview and examine all sample persons (unit nonresponse), but also from the failure to obtain and record all items of information for the examined persons (item nonresponse). This report concentrates on the issues related to item nonresponse.

NCHS is concerned about the levels of potential bias that may be present in estimates derived from data with high item nonresponse rates, and has conducted a number of studies since 1956, when the national health examination surveys began, in an effort to improve the response rates and to assess nonresponse bias. The occurrence of missing data, either unit or item, creates potential for bias in estimates derived from survey data. For certain important characteristics, the respondents may differ significantly from the nonrespondents. In such instances, survey data may not adequately reflect or characterize the nonrespondents. Projections of individual characteristics of the nonrespondents that are based on responses of the participants may be incorrect. If such differences exist and are not adjusted for in the analyses, then any estimates or inferences made to the target population may be misleading. The potential for bias is particularly great when nonresponse rates are high. As a result, assessment of nonresponse bias is very important in analyses of data from a probability sample such as NHANES.

As indicated by Chapman (1991), the assessment should include the following sections that focus on item nonresponse:

- A definition of item nonresponse and the level of nonresponse;
- A description of the effect of item nonresponse on statistics of interest;
- A description of the item nonresponse adjustment;
- Assumptions used in the adjustment methodology; and
- An assessment of the imputation procedure and the impact of nonresponse adjustment on survey estimates.

Section 2 provides a description of item nonresponse and an approach for measuring and reporting nonresponse rates in NHANES III. An approach for the evaluation of item nonresponse bias is given in section 3. Section 4 provides a summary of approaches available for item level nonresponse adjustments. A description of the criteria used for selection of a compensation procedure is given in section 5. Section 6 includes a discussion of the methods available for assessment of the imputed data. Section 7 provides guidelines on reporting results of analyses of items with high nonresponse rates. A bibliography is included at the end of this report.

In this report, we examine the potential for item nonresponse bias after adjustments are done to correct for unit nonresponse. Three components of the NHANES III Mobile Examination Center (MEC) file with high item nonresponse rates were selected for nonresponse bias analysis: (1) phlebotomy, (2) fundus photography of the eye, and (3) bone densitometry. Phlebotomy (drawing blood for biochemistry) is offered to all persons aged 1 year or older, fundus photography is offered to adults 40 years and older, and bone densitometry to persons aged 20 years and older (pregnant women or those expected to be pregnant were excluded from this component). Hemoglobin (HGB) from phlebotomy, macular degeneration scores for early detection of diabetes from fundus photography, and total bone mineral density measurement (total_BMD) from bone densitometry were selected for evaluation of item nonresponse.

2. MEASURING AND REPORTING NONRESPONSE IN NHANES III

An assessment of the extent of missing data is critical in evaluating potential nonresponse bias and how to compensate for missing data. For any statistical analysis, a data analyst must decide how to treat variables containing missing data. The extent of missing data can influence these decisions, the interpretation of the results, and the strength of the confidence in the interpretations. Including the rates of missing data in published reports is helpful to the reader for interpreting the results and conclusions of the study.

As indicated in the previous section, the first step is to provide a definition of item nonresponse in the documentation. The following provides a general definition for item nonresponse in NHANES III.

Item Nonresponse. Missing data for items can result from the following factors: the answers are classified as unusable; the respondent does not have the information to answer a particular item or refuses to answer a specific question or undergo a particular test; laboratory equipment fails; test results are faulty; specimens are lost in shipping; or some items of information fail to be recorded on the examination record.

Nonresponse rates can be derived in two ways: unweighted rates or weighted nonresponse rates. The two serve different purposes and both should be calculated. Unweighted nonresponse rates are indicators of how well the survey operations were carried out. They are useful during the course of the survey as part of the quality control process, and at the completion of fieldwork as a measure of success. A comparison of unweighted nonresponse rates in NHANES III with similar rates in earlier phases of NHANES is illuminating and provides useful information on whether the quality is continuing to improve.

However, weighted nonresponse rates are more appropriate in examining the potential effect of nonresponse on statistics. Since the estimates are based on weighted data, weighted nonresponse rates are better clues to potential quality problems. The magnitude of the weighted response rates should be considered when drawing conclusions from the sample estimates. Section 2.1 provides a brief description of the methodology used for computation of weighted response rates.

2.1 Computation of Weighted Response Rates

Nonresponse can occur at several stages of the survey: interview, examination, and steps within the examination. The effects of all three need to be taken into account in calculating nonresponse rates.

The weighted response rates can be computed in the following way. The weighted interview response rate for subgroup j, where j denotes the domain under study, (e.g., persons 40 years and older, females 20 years and older, etc.) can be calculated as the sum of the poststratified base weights for persons who completed the interview divided by the sum of the poststratified base weights for all screened persons in subgroup j. That is,

$$WR_{(interview)\,j} = \frac{\sum_{i=1}^{NK_j} PW_{ij}}{\sum_{i=1}^{NL_j} PW_{ij}}$$

where

$WR_{(interview)j}$ = weighted interview response rate for subgroup j,

PW_{ij} = poststratified base weights for respondent i in subgroup j,

NKj = total number of persons interviewed in subgroup j, and

NLj = total number of persons screened in subgroup j.

Poststratified weights that include nonresponse adjustments should not be used in this calculation since those weights bring respondents up to the level of the total population including the contributions expected from nonrespondent. However, the poststratified base weights do not correct for nonresponse. There was a virtually 100 percent response rate at the screener level. The poststratified base weights mainly force the base weights to add up to the same total population as the interview and examination weights.

The weighted examination response rate can be calculated in a similar way using the interview weights, that is,

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

where

$WR_{(examination)j}$ = weighted examination response rate for subgroup j,

PW_{ij} = poststratified base weights for respondent i in subgroup j,

NJ_j = total number of persons examined in subgroup j, and

NK = total number of persons interviewed in subgroup j.

The weighted item nonresponse for subgroup j is equal to

$$WR_{(item)\,j} = \frac{\sum_{i=1}^{NI_j} PW_{ij}}{\sum_{i=1}^{NJ_j} PW_{ij}}$$

where

$WR_{(item)j}$ = weighted item response rate for subgroup j,

PW_{ij} = poststratified base weights for respondent i in subgroup j,

NI_j = total number of persons with usable data for the item in subgroup j, and

NJ_j = total number of persons examined in subgroup j.

The overall weighted response rate is computed as the product of the three response rates.

As indicated in section 1, three variables were selected from the MEC file for the nonresponse analyses in this report. Tables 1 through 4 contain the weighted and unweighted response rates for these variables. Table 1 shows the weighted and unweighted response rates for the macular degeneration item. The macular degeneration score was collected only from persons 40

years and older. Out of 7,412 persons eligible for the macular degeneration examination, 5,897 persons were interviewed and 4,933 persons underwent the examination. Although 4,540 persons took the macular degeneration examination, only 4,007 ended up with usable data for analysis. A comparison of the response rates shows a fair-sized difference between weighted and unweighted item response rates, indicating that persons with no usable macular degeneration data had, on average, lower sampling weights than those with usable data. In other words, there is a potential for underrepresentation of the oversampled subgroups (i.e., Mexican-Americans, black persons, and persons 60 years and older) in the fundus sample. Furthermore, the interview and examination response rates for the population eligible for the macular degeneration examination are much lower than for the entire NHANES III sample, reflecting the relatively low response rates for the population eligible for macular degeneration examination. In addition, Table 1 shows much lower response rates for persons 60 years and older when compared to persons 40 to 59 years of age. The higher nonresponse rates among older people were partially due to health problems that prevented them from undergoing the macular degeneration examination. A more detailed comparison of respondents and nonrespondents for the macular degeneration item is given in section 3.

Table 2 shows the weighted and unweighted response rates for the bone density item (total bone mineral density measurement). The total bone mineral density measurement examination was collected only from persons 20 years and older. Out of 11,661 persons 20 years and older, 9,488 persons were interviewed, and 8,213 persons had the examination. A total of 7,116 ended up with usable data for analysis. Females show higher weighted and unweighted response rates for interview and examination stages. However, pregnant women, or those who thought they were pregnant, did not have (or undergo) the bone density examinations. This resulted in much lower bone density item response rates for females. The unweighted item response rate for females was about 83 percent compared to about 70 percent for males. The corresponding weighted response rates were about 85 percent for females and 92 percent for males. The difference between the weighted and unweighted response rates for the total bone mineral density measurement examination are less than the associated differences for the macular degeneration examination.

Interview response rates were higher for persons 20 to 39 years of age compared to those 40 years and older. The examination response rate was lower among persons 60 years and older partially due to higher rate of health problems preventing older persons from undergoing the examination. This age group had the lowest overall response rate when compared to other age groups with an unweighted overall response rate of about 53.5 percent

Table 1. Weighted and unweighted response rates for the macular degeneration examination

Sample	Sample size	Unweighted response rates (%)		Weighted response rates (%)	
		Percent at each level[1]	Percent overall levels[2]	Percent at each level[1]	Percent overall levels[2]
Total sample screened	7,412	100.00	100.00	100.00	100.00
Interviewed	5,897	79.56	79.56	78.95	78.95
40-59 yrs	2,429	79.82	79.82	79.45	79.45
60+ yrs	3,468	79.40	79.40	78.29	78.29
Examined	4,933	83.65	66.55	85.95	67.86
40-59 yrs	2,220	91.40	72.95	90.42	71.84
60+ yrs	2,713	78.23	62.11	79.97	62.61
Macular degeneration respondents (with usable data)	4,007	81.23	54.06	87.28	59.23
40-59 yrs	1,249	91.04	66.41	92.52	66.47
60+ yrs	2,758	73.20	45.46	79.36	49.69

[1]This column shows the response rate for each row of the table, e.g., the unweighted response rate for the examined sample is equal to 83.65 and the associated weighted value is equal to 85.96.

[2]This column shows the product of response rates for each row of the table with the preceding rows, e.g., the unweighted overall response rate for the examined sample is equal to 66.55, and is computed by the product of screened sample response rate, the interview sample response rate, and the examined sample response rate. This is usually referred to as the overall response rate at the associated level.

and a weighted overall response rate of about 56 percent. The total bone mineral density item response rate was highest among persons 40 to 59 years of age. This age group also had the highest overall response rate (about 66 %).

Tables 3 and 4 show the weighted and unweighted response rates for the hemoglobin item. The response rates are provided separately for age groups: 1 to 5, 6 to 19, 20 to 59, and 60+ years of age. For 1 to 5 year-olds the total screened sample was equal to 3,473, of which 3,278 were interviewed and 3,043 were examined. Only 2,056 children out of the examined sample of 3,043 had usable hemoglobin data for analysis. As can be seen in Table 3, although the response rate for interview and examination was high among young children, the response rate for the phlebotomy item was low resulting in an overall weighted response rate of about 56.6 percent for the sample. The

Table 2. Weighted and unweighted response rates for the total bone mineral density examination

Sample	Sample size	Unweighted response rates (%)		Weighted response rates (%)	
		Percent at each level[1]	Percent overall levels[2]	Percent at each level[1]	Percent overall levels[2]
Total sample screened	11,661	100.00	100.00	100.00	100.00
Interviewed	9,488	81.37	81.37	80.08	80.08
Sex					
Male	4,737	80.33	80.33	78.03	78.03
Female	4,751	82.43	82.43	81.94	81.94
Age					
20-39 yrs	3,591	84.51	84.51	81.37	81.37
40-59 yrs	2,429	79.82	79.82	79.45	79.45
60+ yrs	3,468	79.38	79.38	78.29	78.29
Examined	8,213	86.56	70.43	88.02	70.49
Sex					
Male	4,122	87.02	69.90	88.10	68.74
Female	4,091	86.11	70.97	87.96	72.07
Age					
20-39 yrs	3,280	91.34	77.19	90.33	73.50
40-59 yrs	2,220	91.40	72.96	90.42	71.84
60+ yrs	2,713	78.23	62.10	79.97	62.61
Total bone mineral density respondents (with usable data)	7,116	86.64	61.02	88.58	62.44
Sex					
Male	3,707	89.93	62.86	92.39	63.51
Female	3,409	83.33	59.14	85.27	61.45
Age					
20-39 yrs	2,775	84.60	65.30	86.03	63.24
40-59 yrs	2,003	90.23	65.82	91.99	66.09
60+ yrs	2,338	86.18	53.51	89.45	56.00

[1] Same as Table 1.
[2] Same as Table 1.

main reason for the high rate of item nonresponse for this age group was refusal. A comparison of the weighted and unweighted response rates shows that children with no usable hemoglobin data have, on average, higher sampling weights than those who responded and have usable data. That is, those who were oversampled in NHANES III are overrepresented in the sample.

The response rate for the hemoglobin item was higher for 6 to 19 year-olds, as shown in the lower part of Table 3. Of the 3,969 children and young adults screened for the phlebotomy sample, 3,575 were interviewed and 3,348 were examined. A total of 2,876 persons had usable hemoglobin data. The pattern of nonresponse among sampled persons is the same as that for young children. That is, a comparison of the weighted and unweighted response rates shows that those with no usable hemoglobin data have, on average, higher sampling weights than those who responded and have usable data. That is, those who were oversampled in NHANES III are overrepresented in the sample.

Table 4 shows similar data separately for persons aged 20 to 59 and those 60 years and older. For those aged 20 to 59, the interview response rate is considerably lower than that of the younger age groups. However, the response rate for the hemoglobin item is much higher than that of the younger age groups. As a result, the overall response rates for persons 20 to 59 years of age is very close to those aged 6 to 19. The sample of persons 60 years and older have the lowest interview and examination response rates. The hemoglobin item response rate is around 93 percent for the oldest age group, resulting in an overall response rate of about 58 percent.

In addition to reporting nonresponse rates, to the extent possible, any analysis of nonresponse should include information on reasons for missing data (e.g., unable to contact, medical reasons, refusals, etc). Analysts should study the reasons for nonresponse at each stage of sampling. This information is valuable for diagnostic purposes in the evaluation of nonresponse bias.

The following section provides some general guidelines for computing response rates using the NHANES III data file.

Table 3. Weighted and unweighted response rates for the hemoglobin item for children 1 to 5, and 6 to 19 years old

Sample	Sample size	Unweighted response rates (%)		Weighted response rates (%)	
		Percent at each level[1]	Percent overall levels[2]	Percent at each level[1]	Percent overall levels[2]
Age = 1 to 5 Years					
Total sample screened	3,473	100.00	100.00	100.00	100.00
Interviewed	3,278	94.39	94.39	93.52	93.52
Examined	3,043	92.83	87.62	90.44	84.60
Hemoglobin respondents (with usable data)	2,056	67.57	59.20	66.97	56.64
Age = 6 to 19 Years					
Total sample screened	3,969	100.00	100.00	100.00	100.00
Interviewed	3,575	90.07	90.07	88.07	88.07
Examined	3,348	93.65	84.35	91.73	80.79
hemoglobin respondents (with usable data)	2,876	85.90	72.46	85.73	69.26

[1]Same as Table 1

[2]Same as Table 1

Table 4. Weighted and unweighted response rates for the hemoglobin item for 20 to 59, and 60 years old and older

Sample	Sample size	Unweighted response rates (%)		Weighted response rates (%)	
		Percent at each level[1]	Percent overall levels[2]	Percent at each level[1]	Percent overall levels[2]
Age = 20 to 59 Years					
Total sample screened	7,292	100.00	100.00	100.00	100.00
Interviewed	6,020	82.56	82.56	80.61	80.61
Examined	5,500	91.36	75.43	90.37	72.85
Hemoglobin respondents (with usable data)	5,128	93.24	70.32	93.63	68.21
Age = 60+ Years					
Total sample screened	4,369	100.00	100.00	100.00	100.00
Interviewed	3,468	79.38	79.38	78.29	78.29
Examined	2,713	78.23	62.10	79.97	62.61
Hemoglobin respondents (with usable data)	2,517	92.78	57.61	93.33	58.43

[1]Same as Table 1

[2]Same as Table 1

2.2 General Guidelines for Computing Response Rates

Response rates are computed based on the interview status. The interview status is used to identify the respondents to the interview, the examination, and the item. The status is defined as follows:

- A person identified as a sampled person (SP) who was not interviewed has a status of 0;
- An SP who has completed the interview but not the examination has a status of 1;
- An SP who has completed the interview and any component of the mobile examination center (MEC) examination has a status of 2;
- An SP who has completed the interview and a home examination has a status of 3.
- A person not eligible for the study has a missing value for the response status.

The following steps should be taken to compute the unweighted and weighted interview, examination, and item response rates.

Fundus Photography Comp onent -- Macular Degeneration Scores for Early Detection of Diabetes

Unweighted and weighted interview response rates:

1. Subset the main file to sampled persons (SPs) who are 40 years of age and older. The main file refers to the file of all screened persons. Identify the SPs by their assigned interview status. (Note, all SPs should have a nonmissing value for interview status).

2. The subset of SPs identified in step 1 represent all persons eligible for the interview. The interview response rate to the macular degeneration item is computed as the ratio of the total number of SPs that actually completed the interview to the total number of SPs eligible to complete the interview. Eligible SPs have an interview status of 0,1,2, or 3. The SPs that have completed the interview have an interview status of 1,2, or 3.

3. Repeat step 1 for computing the weighted response rate.

4. Use the poststratified base weight to compute the weighted interview response rate. Compute the weighted response rate as the sum of the poststratified base weights for SPs who completed the interview divided by the sum of the poststratified base weights for SPs eligible to complete the interview.

Unweighted and weighted examination response rates:

1. As for the interview response rate, the first step is to identify and subset the data file to the SPs that are eligible for examination. For the macular degeneration item, create a file containing SPs who are age 40 and older with an interview status of 1,2, or 3.

2. The next step is to compute the examination response rate. Since the focus is on a component administered at the MEC, the response rate is computed based on SPs who completed some portion of the MEC examination. The respondents include SPs who have an interview status of 2. Compute the unweighted response rate as the ratio of the total number of MEC-examined SPs to the total number of eligible SPs (those with an interview status of 1,2, or 3).

3. For the completion of weighted response rates, create a file containing eligible SPs. Follow the instructions given in step 1.

4. Use the poststratified base weight to compute the weighted examination response rate. Compute the weighted response rate as the sum of the poststratified base weights for SPs who completed the MEC examination divided by the sum of the poststratified base weights of SPs eligible to complete the MEC examination.

Unweighted and weighted item response rates:

1. Subset the file used for computing the examination response rates to persons that completed any given component of the MEC. The SPs should be age 40 and older with an interview status of 2.

2. The eligible SP to the macular degeneration item is defined as the SP with or without a macular degeneration score. The respondents to the examination are defined as SPs with a gradable (or usable) macular degeneration score. A respondent's score is gradable if: there is no degeneration detected, early degeneration detected, or late degeneration detected. Persons who do not have a gradable score are considered eligible nonrespondents. Compute the unweighted response rate to the fundus item as the ratio of the total number of SPs with gradable macular degeneration scores to the total number of SPs eligible to complete the macular degeneration examination.

3. For the computation of weighted response rates, create a file containing eligible SPs. Follow the instructions given in step 1.

4. Use the poststratified base weight to compute the weighted macular degeneration item response rate. Compute the weighted response rate as the sum of the poststratified base weights for SPs who have gradable scores for the macular degeneration item divided by the sum of the poststratified base weights for all SPs eligible to complete the macular degeneration examination.

Unweighted and weighted overall response rates:

1. Compute the overall unweighted response rate as the product of the unweighted interview, examination, and item response rates for the macular degeneration item.

2. Compute the overall weighted response rate as the product of the weighted interview, examination, and item response rates for the fundus photography component.

Bone Density Component -- Total Bone Mineral Density Measurement

Unweighted and weighted interview response rates:

1. Subset the main file to persons with an interview status of 0, 1, 2, or 3 who are age 20 and older.

2. Repeat steps 2 through 4 described for computing the interview response rate for the fundus component. Note that the file should include SPs age 20 and older.

Unweighted and weighted examination response rates:

1. Subset the file used for computing the interview response rate for the total bone mineral density measurement item to SPs with an interview status of 1, 2, or 3.

2. Repeat steps 2 through 4 described for computing the examination response rate for the fundus component. (Include SPs age 20 and older).

Unweighted and weighted item response rates:

1. Subset the file used for computing the examination response rate for the total bone mineral density measurement item to SPs with an interview status of 2.

2. The unweighted item response rate is based on the MEC bone scan test results. The SPs eligible for the total bone mineral density item consist of those persons with a good bone scan, bad bone scan, or no scan. Respondents to the total bone mineral density item have good bone scan (or usable) test result. Compute the unweighted response rate as the ratio of the total number of SPs with good bone scans to the total number of SPs eligible to complete the bone scan.

3. For the computation of weighted response rates, create a file of eligible SPs following the instructions given in step 1.

4. Use the poststratified base weight to compute the weighted item response rate. Compute the weighted response rate as the sum of the poststratified base weights for SPs who have good scans divided by the sum of the poststratified base weights for SPs eligible to complete the total bone mineral examination.

Overall response rate to the total bone mineral density item:

1. Repeat steps 1 and 2 for computing the overall response rate to the macular degeneration item.

Phlebotomy Component -- Hemoglobin Test

Unweighted and weighted interview response rates:

1. Subset the main file to persons with an interview status of 0, 1, 2, or 3 who are age 1 and older.

2. Repeat steps 2 through 4 described for computing the interview response rates for the fundus component. Note that the file should include SPs age 1 and older.

Unweighted and weighted examination response rates:

1. Subset the file used for computing the interview response rate for the hemoglobin item to SPs with an interview status of 1, 2, or 3.

2. Repeat steps 2 through 4 described for computing the examination response rates for the fundus component. (Include SPs age 1 and older).

Unweighted and weighted item response rates:

1. Subset the file used for computing the examination response rate for the hemoglobin item to SPs with and interview status of 2.

2. The unweighted item response rate is based on the hemoglobin test results (HGB) taken at the MEC. The eligible SPs to the hemoglobin item consist of those persons with a HGB test result greater than or equal to zero and persons with a missing HGB test result who are age 1 and older. Respondents to the hemoglobin item have a HGB test result greater than zero. Compute the unweighted response rate as the ratio of the total number of SPs with a HGB result greater than zero to the total number of SPs eligible to take the examination.

3. For the creation of weighted response rates, create a file of eligible SPs following the instructions given in step 1.

4. Use the poststratified base weight to compute the weighted item response rate. Compute the weighted response rate as the sum of the poststratified base weights for SPs who have HGB results greater than zero divided by the sum of the poststratified base weights for SPs eligible for the hemoglobin item.

Overall response rates to the hemoglobin item:

1. Repeat steps 1 and 2 for computing the overall response rates to the fundus item.

3. EVALUATION OF NONRESPONSE BIAS

There is always a potential for nonresponse bias whenever sample persons who did not participate in the survey have somewhat different characteristics than those who did. Adjustment methods for unit nonresponse involve making adjustments to the weights of the responding units. The adjustments are typically made for different subgroups of the sample. The outcome is a data set containing data for respondents whose weights have been adjusted to compensate for the data missing for nonrespondents. As mentioned in section 1, NCHS has conducted a two-stage procedure for nonresponse adjustment and poststratification to known population totals to adjust for unit nonresponse. For the first step age, race/ethnicity, household size, region of the country, and urbanization status were used to classify respondents and nonrespondents into adjustment classes for interview nonresponse adjustment. Adjustments for examination nonresponse focused on the use of health history data as well as demographic information (Ezzati, et al. 1992). However, the current version of phase 1 of the NHANES III data file does not include any adjustments for item nonresponse.

Nonresponse adjustment methods can serve to reduce nonresponse bias. However, the total elimination of such bias is not possible, since within any weighting class the respondents ordinarily will not be fully representative of the nonrespondents. To estimate the nonresponse bias quantitatively, it is necessary to obtain external validating data which are not available for incorporation into the survey estimation procedure. Thus, it is not possible to quantify the extent of nonresponse bias remaining in the survey estimates after nonresponse adjustment. However, methods are available to assist in evaluating the likelihood that nonresponse bias remains a serious problem, even after making weighting adjustments. The same types of methods can be used to assess nonresponse bias for phase 1 of NHANES III, where adjustments are made for unit nonresponse but not for item nonresponse.

The most common method of nonresponse bias assessment is to compare the distribution of respondents and nonrespondents from the same survey with respect to characteristics that are known for both groups. Another common method is to compare the distribution of respondents in the survey with associated distributions coming from other independent surveys. Other methods include the use of independent sources of data to gain information about the nonrespondents. Kammerman (1987) provides a review of various methods of nonresponse bias analysis and an extensive literature review of studies related to nonresponse bias assessment in earlier NHANES surveys. The most recent assessment of nonresponse bias in NHANES was carried out by

Rowland and Forthofer (1993). In this report, the authors used the chi-square automatic interaction detection (CHAID) technique to summarize the data for assessment of nonresponse bias in HHANES. CHAID is a descriptive procedure used to derive estimates of relationships between the dependent variable (the response status) and the predictor variables (other classifications or descriptive variables) by calculating the chi-square measure of association between the dependent and each independent variable. The predictor variable with the highest significant level for the chi-square test is used to split the sample into groups. This process is repeated for each of the new groups until there are too few observations for further splitting. The result is a tree-like structure that suggests which predictor variables may be important and need further investigation.

The investigation of potential bias in HHANES consisted of four parts. Part 1 of the investigation included a study of the interview status. This investigation was limited to the screener interview variables: age, season of the year, gender, family size, language of the screener interview, and mobile examination center location. Part 2 of the investigation provided a study of the examination status. In this stage, the weighted interview examination response rate was studied in relation to demographic and screener variables from the screener and family questionnaires and medical history variables from child and adult medical history questionnaires. Part 3 of the investigation compared the estimated population proportions of various characteristics for Mexican-Americans, Cubans, and Puerto Ricans in HHANES and the 1982-1984 NHIS. Part 4 of the investigation estimated possible bias in disease prevalences. An estimate of possible nonresponse bias was computed for selected variables. For each variable, a biased adjusted estimate was compared with a survey estimate based on the analytic sample.

Groves, et al. (1992) used the information given in the decennial Census to obtain information on survey nonrespondents from Census records. Addresses from seven different personal visit surveys conducted by various agencies and organizations were matched to decennial records.

Most of the analysis of nonresponse bias is limited to investigations of the bivariate relationship between demographic characteristics and response rates. The disadvantage of this approach is the lack of models that examine nonresponse in a multivariate context. Groves has compared the bivariate patterns of nonresponse to multivariate patterns of nonresponse using a logistic regression model. Variables in the analysis included sociodemographic characteristics and health status indicators of sample persons, information about their family situation, and geographical and administrative variables.

Other recent articles related to the issue of nonresponse include Kott (1994) in which he compares two different models for handling nonresponse in survey sampling theory. In a response model, the propensity of survey responses is modeled as a random process, as an additional phase of sample selection. In a parametric model, the survey data are themselves modeled. He indicates that these two models can be used simultaneously in the estimation process so that one provides some compensation for the possible failure of the other. Other related research includes: Hidiroglou, et al. (1993); Tennant (1991); and Melton et al. (1993).

In this report, we provide a rather simple and straightforward method of assessing nonresponse bias when unit nonresponse adjustments are conducted but no adjustments are made for item nonresponse. This approach is simple to implement and does not require access to any specialized software. However, analysts are encouraged to review the literature and conduct other bivariate or multivariate approaches if appropriate computer softwares are available to them.

In general, analysts are advised to develop tabulations that focus on the remaining differences between respondents and nonrespondents after unit nonresponse adjustments are conducted. In effect, they will consist of the differences that remain, after discounting the effect of unit nonresponse adjustments. An examination of the gross differences between respondents and nonrespondents, without reflecting the unit nonresponse adjustments, may substantially overstate the potential biases. The format given in Tables 5 through 10 provides an example of a proposed method of nonresponse assessment after adjustments are made for interview and examination unit nonresponse for phase 1 of NHANES III.

3.1 Analysis of Nonresponse Bias

As mentioned earlier, one variable was selected from the fundus photography, bone density, and phlebotomy components in the MEC file. The first step of any nonresponse bias assessment is to identify the covariates that are highly correlated with the variable of interest and the response status. Analysts are advised to review the literature and talk to experts to gather information about the correlates before conducting any nonresponse bias analysis. In the analysis reported in this section, the information given in Klein (1994) was used to identify potential correlates for the macular degeneration measurement item. The information given in Looker, et al. (1992) was used to derive correlates for the total bone mineral density item, and general background information was used to derive the correlates for the hemoglobin item.

Tables 5 through 10 compare the weighted distribution of respondents and nonrespondents for the macular degeneration item, bone mineral density measurement item, and the hemoglobin item using unit nonresponse adjusted sampling weights. The following five sample distributions are compared in these tables.

Sample (1) The sample distribution of the original sample selected for the item of interest.

The poststratified base weights were used to produce weighted distribution for sample (1), the selected sample. Since the screener sample had about 100 percent response rate, and the sample size is quite large, sample (1) distribution is expected to be close to the true distribution in the population.

Sample (2) The sample distribution of the interviewed sample for the item.

The interview weight was used to compute the weighted distribution for the interview sample. The interview weight incorporates nonresponse adjustments for demographic variables available in the screener. Therefore, sample (2) demographic distribution is exactly the same as that for sample (1) for variables with the same categories as used in the nonresponse and poststratification stages. To the extent that the health variables listed in Table 5 are correlated with the demographic variables used in unit nonresponse adjustments, the distribution of the health-related variables are also expected to be close to the true values in the population.

Sample (3) The sample distribution of the examined sample persons who were expected to take the item.

The examination weights were used to compute the distributions for this sample. The demographic distribution of the examined persons, as shown in Table 2, is almost the

Table 5. Weighted distributions of respondents and nonrespondents to the macular degeneration the interviewed sample, and the examined sample

Characteristics	Selected Sam (1)		Examined Respon (3)		Respondents to mac degeneration exami (4)		Nonrespondents to degeneration exami (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	ample Siz	%
Total	7,412		4,933		4,007		926	
Age								
40 - 59	3,043	56.9	2,220	56.9	2,021	60.8	199	31.8
60+	4,369	43.1	2,713	43.1	1,986	39.2	727	68.2
Race/ethnicity								
White, non-Hispar	4,084	82.4	2,563	82.1	2,085	82.5	478	79.0
Black, non-Hispar	1,600	9.6	1,154	9.6	917	9.0	237	13.3
Mexican Americar	1,529	3.0	1,071	3.0	883	3.0	188	3.1
Other	199	5.0	145	5.4	122	5.5	23	4.7
Sex								
Male	3,700	49.9	2,502	46.2	2,038	46.2	464	46.0
Female	3,712	50.1	2,431	53.8	1,969	53.8	462	54.0
	Interviewed sample (2)							
Total	5,897	100.0						
Poverty Index Level								
Below poverty lev	953	9.9	815	9.9	625	9.2	190	14.1
At or above pover	4,157	80.6	3,535	81.4	2,933	82.7	602	73.0
Missing	787	9.5	583	8.7	449	8.1	134	13.0
Health Status								
Good/very good/e	3,968	76.5	3,381	76.6	2,822	78.3	559	65.2
Poor/Fair	1,919	23.3	1,546	23.3	1,180	21.6	366	34.4
Missing	10	0.2	6	0.1	5	0.1	1	0.4
Smoke now								
Yes	1,324	25.7	1,141	24.8	958	25.4	183	20.9
No	4,561	74.1	3,791	75.2	3,048	74.6	743	79.1
Missing	12	0.1	1	0.0	1	0.0		
Ever had diabetes								
Yes	734	8.9	603	8.8	442	7.9	161	14.8
No	5,150	90.9	4,321	91.1	3,560	92.0	761	84.9
Missing	13	0.2	9	0.1	5	0.1	4	0.3
Ever had cataracts								
Yes	1,180	12.7	889	12.8	548	9.9	341	31.6
No	4,715	87.3	4,042	87.2	3,458	90.1	584	68.3
Missing	2	0.0	2	0.0	1	0.0	1	0.0
Ever had trouble seeing								
Yes	1,072	15.2	870	15.6	615	14.0	255	25.7
No	4,791	84.4	4,046	84.1	3,377	85.7	669	74.2
Missing	34	0.5	16	0.3	15	0.3	2	0.1

*Includes SPs with usable data

same as the screened sample because of the nonresponse adjustments carried out at the interview level. Furthermore, the examination weight incorporates nonresponse adjustments for self-reported health status as well as some demographic variables available from the interview (Ezzati, et al. 1992).

Sample (4) The sample distribution of the cases with usable data for the item.

Sample (5) The sample distribution of the cases with no usable data for the item.

The examination weights were used to compute the weighted distributions for both samples (4) and (5). As noted earlier, the examination weight includes adjustments for interview and examination nonresponse, but does not include any adjustments for item nonresponse. The sample distributions are compared for a group of variables thought to be related to response status, and to the three items under investigation.

Table 5 provides the weighted distributions of respondents and nonrespondents to macular degeneration examination in the fundus photography component compared to the selected sample. The variables are divided into two groups, those obtained through the screener, and those provided in the questionnaire. Age, race/ethnicity, and sex variables are collected in the screener, and thus are available for the entire selected sample. The health status, smoking status, ever had diabetes, ever had cataracts, and ever had trouble seeing variables come from the interview and are available only for persons with completed interviews.

A sample of 7,412 was originally selected for the macular degeneration item, of which a total of 5,897 persons responded to the interview (sample (2)). Sample (3) shows the distribution of the examined persons for the macular degeneration sample. A total of 4,933 persons were examined.

A profile of nonrespondents is usually the first step in understanding their potential effects on the statistics. The actual profiles, however, may shed little or no light on the biases because the weighting system is designed to reduce the effects arising from differences between the demographic characteristics of respondents and nonrespondents, as shown in Table 5. Although the weighting cannot take all differences between respondents and nonrespondents into account, the weighting cells that were used appear to have largely eliminated any potential differences between respondents and nonrespondents in distributions of health measures that are likely to be related to the results of macular degeneration examination. The biases that may arise from the residual differences that remain in the sample appear to be quite minor for the examined sample.

Sample (4) presents the distribution of the 4,007 cases with macular degeneration scores that were gradable for age-related maculopathy. Sample (5) provides the same distribution for persons who completed the examination but did not have gradable macular degeneration scores for the fundus. As reported by Klein et al. (1994), of the 926 SPs with no usable data, 159 were not photographed because of subject refusal, equipment problem, physical inability, insufficient time, or ocular problems (eye movement, pupil dilation, corneal opacity, etc.). The photographs were not gradable in another 767 because of camera malfunctions, presence of media opacities, or poor pupillary dilation.

A comparison of the demographic and health-related distributions of samples (1), (2), (4) and (5) in Table 5 shows the following patterns:

- Persons aged 60 and older are underrepresented in the macular degeneration sample;
- Persons with "good/very good/excellent" self-reported health status responded at a higher rate than the rest;
- Persons at or above poverty are overrepresented in the macular degeneration sample;
- Persons with diabetes and persons who have trouble seeing are underrepresented in the macular degeneration sample; and
- Persons with cataracts have the highest underrepresentation rate in the macular degeneration sample.

Table 6 provides similar information for the total bone mineral density item. A sample of 11,661 was originally selected for the total bone mineral density examination, of which, a total of 9,488 persons responded to the interview (sample (2)).

Sample (3) shows the distribution of the examined persons for the total bone mineral density sample. A total of 8,213 persons were examined. Sample (4) presents the distribution of the 7,116 cases with usable data for the total bone mineral density item. Sample (5) provides the same distribution for persons who completed the examination but did not have usable data for the total bone mineral density item.

Table 6. Weighted distributions of respondents to the total bone mineral density item compared to the the interviewed sample, and the examined sample

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to total-BM (4)		Nonrespondents to total-B (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	Sample Siz	%
Total	11,661		8,213		7,116		1,097	
Age								
20 - 49	5,940	64.3	4,546	64.3	3,922	63.9	624	67.6
50+	5,721	35.7	3,667	35.7	3,194	36.1	473	32.4
Race/ethnicity								
White, non-Hispanic	5,510	78.8	3,586	78.6	3,217	79.4	369	72.9
Black, non-Hispanic	2,792	10.9	2,137	10.9	1,831	10.5	306	14.3
Mexican American	3,004	4.4	2,231	4.4	1,840	4.1	391	6.4
Other	355	6.0	259	6.1	228	6.1	31	6.4
Sex								
Male	5,897	47.7	4,122	47.7	3,707	49.6	415	32.4
Female	5,764	52.3	4,091	52.3	3,409	50.4	682	67.6
	Interviewed sample (2)							
Total	9,488	100.0						
Marital status								
Not Married	2,239	18.5	1,804	18.2	1,538	18.0	266	19.8
Married	5,723	64.8	5,089	65.5	4,431	65.5	658	65.5
Never Married	1,478	16.5	1,307	16.2	1,135	16.3	172	14.7
Missing	15	0.2	11	0.1	10	0.1	1	0.0
Education Level								
0 - 8 years	2,296	10.8	1,946	11.0	1,628	10.6	318	13.5
9 - 11 years	1,610	13.9	1,383	14.1	1,194	14.1	189	13.9
12 years	2,704	33.3	2,393	33.3	2,100	33.5	293	31.1
13+ years	2,766	41.2	2,430	41.1	2,144	41.2	286	40.6
Missing	79	0.8	59	0.6	48	0.5	11	0.8
Poverty Index Level								
Below poverty level	1,770	11.3	1,587	11.4	1,325	11.3	262	12.4
At or above poverty l	6,562	80.2	5,722	80.9	5,019	81.1	703	79.4
Missing	1,156	8.5	904	7.7	772	7.6	132	8.2
Weight								
< 70 lbs.	3	0.0	1	0.0	-	0.0	1	0.0
70 - 139 lbs.	2,390	27.1	1,996	26.7	1,637	25.5	359	36.7
140 - 209 lbs.	5,658	59.1	4,951	59.6	4,392	60.7	559	50.8
210 - 269 lbs.	840	10.1	757	10.3	670	10.6	87	8.4
270+ lbs.	123	1.4	111	1.3	84	1.2	27	2.8
Missing	360	2.3	300	2.1	248	2.2	52	1.4
Health status								
Good/very good/exce	6,980	82.9	6,121	83.0	5,361	83.8	760	76.4
Poor/Fair	2,498	17.0	2,086	17.0	1,750	16.1	336	23.6
Missing	10	0.0	6	0.0	5	0.0	1	0.0
Physical activity								
More active than othe	2,917	31.8	2,540	32.1	2,269	32.9	271	25.7
Less active than other	2,073	21.5	1,752	21.3	1,450	20.6	302	26.8
About the same as ot	4,215	43.8	3,704	44.0	3,216	43.9	488	45.0
Missing	283	2.9	217	2.6	181	2.6	36	2.5

Table 6. Weighted distributions of respondents to the total bone mineral density item compared to the the interviewed sample, and the examined sample (continued)

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to total-BM (4)		Nonrespondents to total-B (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	Sample Siz	%
Smoke now								
Yes	2,568	31.6	2,285	31.1	2,020	31.1	265	30.6
No	6,905	68.4	5,925	68.9	5,094	68.8	831	69.4
Missing	15	0.1	3	0.0	2	0.0	1	0.0
Mother had osteoporosis								
Yes	217	3.4	200	3.6	169	3.5	31	4.7
No	8,760	92.1	7,601	91.9	6,603	92.1	998	90.2
Missing	511	4.5	412	4.5	344	4.4	68	5.1
Mother broke hip								
Yes	431	4.9	368	5.0	318	4.9	50	5.7
No	8,741	92.7	7,586	92.6	6,583	92.8	1,003	91.6
Missing	316	2.4	259	2.4	215	2.3	44	2.7
Mother's age when hip broken								
< 50 years	44	0.5	37	0.5	31	0.5	6	0.6
50+ years	372	4.2	318	4.3	277	4.3	41	4.8
NA	8,741	92.7	7,586	92.6	6,583	92.8	1,003	91.6
Missing	331	2.5	272	2.5	225	2.4	47	3.0
Ever had osteoporosis								
Yes	223	2.3	184	2.4	152	2.3	32	3.4
No	9,211	97.2	7,993	97.2	6,938	97.3	1,055	96.2
Missing	54	0.5	36	0.3	26	0.4	10	0.3
Number of times fallen								
None	2,586	17.4	2,063	17.5	1,810	17.9	253	13.8
1 - 5 times	805	5.2	597	5.2	485	4.9	112	7.5
6+ times	52	0.3	37	0.3	30	0.3	7	0.5
NA	6,020	76.9	5,500	76.9	4,778	76.8	722	78.0
Missing	25	0.1	16	0.1	13	0.1	3	0.1
Ever had broken hip								
Yes	205	1.4	157	1.5	110	1.2	47	4.0
No	9,274	98.4	8,051	98.5	7,002	98.8	1,049	96.0
Missing	9	0.2	5	0.1	4	0.1	1	0.0
Ever had broken wrist								
Yes	694	8.2	591	8.4	529	8.7	62	6.0
No	8,784	91.7	7,617	91.5	6,583	91.2	1,034	94.0
Missing	10	0.2	5	0.1	4	0.1	1	0.0
Ever had broken spine								
Yes	165	1.8	147	2.0	125	1.9	22	2.6
No	9,313	98.0	8,060	97.9	6,986	98.0	1,074	97.3
Missing	10	0.2	6	0.1	5	0.1	1	0.0
Had prescribed medication in the last month								
Yes	4,484	44.8	3,764	45.2	3,264	44.9	500	47.2
No	4,976	55.0	4,434	54.7	3,839	54.9	595	52.7
Missing	28	0.2	15	0.1	13	0.2	2	0.1
Ever take vitamins								
Yes	3,159	37.5	2,721	37.7	2,319	36.7	402	45.1
No	6,305	62.3	5,481	62.2	4,787	63.2	694	54.8
Missing	24	0.1	11	0.1	10	0.1	1	0.1

*Includes SPs with usable data

A demographic comparison of the total bone mineral density item sample shows that females (and to some extent persons aged 20 to 49) are underrepresented in the sample. This is partially because women who were pregnant or thought they were pregnant did not take the total bone mineral density test for safety reasons. Throughout the NHANES medical examination components, various groups of respondents were excluded for medical or safety reasons. However, the exclusion of pregnant women from the bone density item constituted the largest exclusion of this type. There are no reasons to believe that the bone density of a woman who was pregnant at the time of the NHANES examination was any different from a nonpregnant woman with the same sociodemographic characteristics. This hypothesis can be tested by comparing various characteristics (i.e., demographic and health characteristics) of the pregnant and nonpregnant women.

Another approach is to look at differences between women who had had a child less than 6 months or a year before taking the NHANES bone density examination and other women with similar demographic characteristics who took the total bone mineral density examination.

Table 6 shows that persons who weighed less than 140 pounds responded at a lower rate than the rest of the SPs. Persons with poor or fair health status are also somewhat underrepresented in the bone density sample. The second page of Table 6 includes a group of variables thought to be related to the total bone mineral density item. These health-related variables did not show any noticeable correlation with response status.

The only health-related variable that is somewhat related to response status is the "ever take vitamins" variable. It seems that persons who take vitamins are somewhat underrepresented in the sample.

Tables 7 through 10 show the weighted distribution of respondents and nonrespondents to the hemoglobin item for persons aged 1 to 5 years, 6 to 19 years, 20 to 29 years, and 60 years and older, respectively. Table 7 shows that a sample of 3,473 children aged 1 to 5 years was originally selected for the hemoglobin item. Out of the screened sample of 3,473 children aged 1 to 5 years, 3,278 children responded to the interview, and 3,043 responded to the examination. A total of 2,056 children had usable data for the phlebotomy (hemoglobin) sample. The resulting sample shows a slight underrepresentation of females and white non-Hispanic children. Children below the poverty level are

Table 7. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare the interviewed sample, and the examined sample for 1 to 5 year-olds

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglobin (4)		Nonrespondents to hemoglobin (5)	
	Sample Size	%	Sample Size	%	Sample Size	%	Sample Size	%
Total	3,473		3,043		2,056		987	
Race/ethnicity								
White, non-Hispanic	1,111	69.3	927	69.0	603	67.1	324	73.0
Black, non-Hispanic	989	15.2	906	15.2	626	15.9	280	13.8
Mexican American	1,238	8.1	1,092	8.1	738	8.6	354	7.1
Other	135	7.4	118	7.6	89	8.4	29	6.1
Sex								
Male	1,674	51.3	1,469	51.3	1,017	52.6	452	48.5
Female	1,799	48.7	1,574	48.7	1,039	47.4	535	51.5
	Interviewed sample (2)							
Total	3,278	100.0						
Poverty Index Level								
Below poverty level	1,174	23.8	1,141	24.3	804	27.2	337	18.4
At or above poverty l	1,726	67.7	1,573	68.4	1,056	66.8	517	71.6
Missing	378	8.5	329	7.3	196	6.0	133	10.0
Health Status								
Good/very good/exce	3,009	95.9	2,785	95.9	1,876	95.3	909	97.1
Poor/Fair	267	4.0	257	4.0	180	4.7	77	2.7
Missing	2	0.1	1	0.1	-	0.0	1	0.2
Weight in pounds								
< 29 lbs	829	26.2	773	26.4	468	24.1	305	31.2
30 - 69 lbs	1,638	57.5	1,494	56.4	1,041	57.4	453	54.4
70 - 120 lbs	9	0.4	8	0.4	6	0.4	2	0.2
Missing	802	15.9	768	16.8	541	18.1	227	14.2
Had prescribed medication in the last month								
Yes	688	21.1	639	20.7	411	20.4	228	21.3
No	2,549	78.8	2,365	79.2	1,616	79.6	749	78.4
Missing	2	0.1	1	0.1	-	0.0	1	0.3
Ever take vitamins								
Yes	1,261	44.4	1,143	44.2	778	43.8	365	44.9
No	1,990	55.3	1,873	55.6	1,258	56.0	615	54.8
Missing	3	0.2	3	0.2	2	0.2	1	0.3
Last visit to the doctors								
Less than two years	3,183	98.0	2,954	97.9	1,987	97.7	967	98.2
Two years or more	67	1.6	65	1.7	49	1.8	16	1.7
Never	24	0.3	21	0.3	18	0.4	3	0.1
Missing	4	0.1	3	0.1	2	0.1	1	0.0
Weight at birth								
Greater than 5.5 lbs	49	0.9	43	0.9	34	1.0	9	0.7
Less than 5.5 lbs	8	0.1	8	0.2	5	0.1	3	0.3
na	3,076	98.4	2,859	98.4	1,929	98.2	930	98.5
Missing	45	0.6	40	0.5	26	0.7	14	0.4

Table 7. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare
the interviewed sample, and the examined sample for 1 to 5 year-olds (continued)

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglo (4)		Nonrespondents to hemogl (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	Sample Siz	%
Ever had anemia								
Yes	339	8.4	326	8.7	225	8.9	101	8.2
No	2,919	91.0	2,698	90.7	1,818	90.5	880	91.2
Missing	20	0.6	19	0.6	13	0.6	6	0.6

*Includes SPs with usable data

Table 8. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare the interviewed sample, and the examined sample for 6 to 19 year-olds

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglobin (4)		Nonrespondents to hemoglobin (5)	
	Sample Size	%	Sample Size	%	Sample Size	%	Sample Size	%
Total	3,969		3,348		2,876		472	
Race/ethnicity								
White, non-Hispanic	1,255	69.7	1,003	69.5	860	69.7	143	68.5
Black, non-Hispanic	1,050	15.4	906	15.4	735	14.7	171	19.9
Mexican American	1,514	7.4	1,310	7.4	1,165	7.6	145	5.6
Other	150	7.5	129	7.7	116	8.0	13	6.0
Sex								
Male	1,960	51.0	1,642	51.0	1,413	51.1	229	50.5
Female	2,009	49.0	1,706	49.0	1,463	48.9	243	49.5
	Interviewed sample (2)							
Total	3,575	100.0						
Poverty Index Level								
Below poverty level	1,164	20.7	1,125	20.6	971	20.7	154	20.0
At or above poverty l	2,025	70.9	1,880	72.1	1,613	72.2	267	71.8
Missing	386	8.4	343	7.3	292	7.1	51	8.2
Marital Status								
Not married	5	0.1	5	0.1	4	0.1	1	0.5
Married	86	2.9	78	2.7	72	2.9	6	1.8
Never married	3,449	96.1	3,241	96.3	2,778	96.2	463	96.5
Missing	22	0.8	21	0.9	19	0.8	2	1.2
Education Level								
0 - 8 years	2,500	61.4	2,357	61.6	1,990	60.0	367	71.0
9 - 11 years	664	22.6	624	22.7	552	23.4	72	19.1
12 years	270	10.3	247	10.3	224	10.8	23	7.0
13+ years	94	4.6	89	4.7	79	5.0	10	2.9
Missing	34	1.0	28	0.8	28	0.9	-	0.0
Health Status								
Good/very good/exce	3,231	95.3	3,023	95.5	2,590	95.3	433	96.5
Poor/Fair	343	4.7	324	4.5	285	4.7	39	3.5
Missing	1	0.0	1	0.0	1	0.0	-	0.0
Ever had hypertension								
Yes	47	1.6	43	1.4	37	1.4	6	1.1
No	3,526	98.3	3,303	98.5	2,838	98.5	465	98.7
Missing	2	0.1	2	0.1	1	0.1	1	0.2
Ever had high cholesterol								
Yes	30	1.3	29	1.3	23	1.3	6	1.7
No	3,503	98.7	3,281	98.7	2,820	98.7	461	98.3
NA	-	0.0	-	0.0	-	0.0	-	0.0
Missing	-	0	-	0	-	0.0	-	0.0

Table 8. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare the interviewed sample, and the examined sample for 6 to 19 year-olds (continued)

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglobin (4)		Nonrespondents to hemoglobin (5)	
	Sample Size	%	Sample Size	%	Sample Size	%	Sample Size	%
Ever had diabetes								
Yes	4	0.2	4	0.2	4	0.2	-	0.0
No	1,177	76.0	1,110	77.0	992	78.6	118	65.4
Missing	451	23.8	405	22.8	337	21.2	68	34.6
Weight in pounds								
< 70 lbs	708	30.9	656	30.4	537	29.4	119	36.6
70 - 139 lbs	860	36.8	799	36.5	691	37.4	108	31.1
140 - 209 lbs	330	17.3	312	17.7	274	18.2	38	14.0
210 - 269 lbs	21	1.6	19	1.6	15	1.3	4	2.6
270+ lbs	2	0.1	2	0.1	2	0.2	-	0.0
Missing	650	13.2	630	13.7	552	13.3	78	15.7
Had prescribed medication in the last month								
Yes	527	18.3	499	18.5	442	19.1	57	14.9
No	3,024	81.6	2,826	81.4	2,412	80.8	414	85.1
Missing	2	0.1	1	0.1	1	0.1	-	0.0
Ever take vitamins								
Yes	884	27.8	804	26.9	686	26.5	118	29.5
No	2,668	72.0	2,522	72.9	2,170	73.4	352	69.9
Missing	4	0.2	3	0.2	2	0.1	1	0.6
Physical activity								
More active than othe	168	31.4	162	31.8	146	32.0	16	30.6
Less active than other	138	21.4	130	20.8	119	21.2	11	17.0
About the same as oth	322	46.3	297	46.8	263	46.2	34	52.4
Missing	4	0.9	3	0.5	3	0.6	-	0.0
Ever had chest pains								
Yes	147	23.1	141	24.2	124	24.2	17	24.3
No	485	76.9	451	75.8	407	75.8	44	75.7
Missing	-	0.0	-	0.0	-	0.0	-	0.0
Last visit to the doctors								
Less than two years	2,522	87.2	2,364	87.5	2,007	86.8	357	90.9
Two years or more	369	11.5	349	11.8	300	12.5	49	8.5
Never	51	1.2	42	0.7	37	0.7	5	0.6
Missing	1	0.0	1	0.0	1	0.0	-	0.0
Weight at birth								
Greater than 5.5 lbs	70	1.7	64	1.5	54	1.4	10	2.2
Less than 5.5 lbs	10	0.2	9	0.2	6	0.2	3	0.7
NA	2,729	97.2	2,553	97.3	2,164	97.4	389	96.8
Missing	49	0.9	66	0.9	62	1.0	4	0.3
Ever had anemia								
Yes	191	5.8	184	6.1	168	6.6	16	3.5
No	2,726	93.4	2,546	93.1	2,153	92.6	393	95.5
Missing	26	0.7	26	0.8	24	0.8	2	1.0

*Includes SPs with usable data

Table 9. Weighted distributions of respondents and nonrespondents to the hemoglobin item compared the interviewed sample, and the examined sample for 20 to 59 year-olds

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglobin (4)		Nonrespondents to hemoglobin (5)	
	Sample Size	%	Sample Size	%	Sample Size	%	Sample Size	%
Total	7,292		5,500		5,128		372	
Race/ethnicity								
White, non-Hispanic	2,798	76.4	1,983	76.3	1,875	76.6	108	71.4
Black, non-Hispanic	1,962	11.6	1,575	11.6	1,419	11.1	156	18.6
Mexican American	2,257	5.1	1,734	5.1	1,636	5.1	98	4.4
Other	275	7.0	208	7.1	198	7.2	10	5.6
Sex								
Male	3,790	49.1	2,754	49.1	2,585	49.6	169	40.7
Female	3,502	50.9	2,746	50.9	2,543	50.4	203	59.3
	Interviewed sample (2)							
Total	6,020	100.0						
Poverty Index Level								
Below poverty level	1,193	11.4	1,127	11.6	1,049	11.5	78	14.1
At or above poverty l	4,211	81.3	3,829	81.8	3,590	82.3	239	72.9
Missing	616	7.3	544	6.6	489	6.2	55	13.0
Marital Status								
Not married	865	13.4	795	13.3	733	13.2	62	14.9
Married	3,809	66.3	3,503	66.9	3,286	67.1	217	64.1
Never married	1,317	20.1	1,192	19.7	1,100	19.6	92	20.9
Missing	11	0.2	8	0.1	7	0.1	1	0.1
Education Level								
0 - 8 years	967	6.9	904	6.8	844	6.8	60	6.8
9 - 11 years	966	12.5	895	12.8	833	12.7	62	13.8
12 years	1,960	34.7	1,799	34.8	1,679	34.9	120	32.6
13+ years	2,065	45.3	1,863	45.1	1,740	45.0	123	45.5
Missing	44	0.7	37	0.6	30	0.6	7	1.3
Health Status								
Good/very good/exce	4,845	87.3	4,410	87.3	4,116	87.3	294	87.4
Poor/Fair	1,173	12.7	1,089	12.7	1,011	12.7	78	12.6
Missing	2	0.0	1	0.0	1	0.0	-	0.0
Ever had hypertension								
Yes	1,116	18.2	1,013	18.4	936	18.4	77	19.2
No	4,891	81.6	4,476	81.4	4,181	81.5	295	80.8
Missing	13	0.1	11	0.1	11	0.1	-	0.0
Ever had high cholesterol								
Yes	725	14.7	666	15.0	632	15.2	34	12.1
No	5,073	85.3	4,633	85.0	4,312	84.7	321	87.9
na	-	0.0	-	0.0	-	0.0	-	0.0
Missing	5	0.0	2	0.0	2	0.0	-	0.0

Table 9. Weighted distributions of respondents and nonrespondents to the hemoglobin item compared the interviewed sample, and the examined sample for 20 to 59 year-olds (continued)

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglobin (4)		Nonrespondents to hemoglobin (5)	
	Sample Size	%	Sample Size	%	Sample Size	%	Sample Size	%
Ever had diabetes								
Yes	276	3.3	253	3.4	241	3.4	12	3.1
No	5,727	96.4	5,232	96.4	4,874	96.4	358	96.0
Missing	17	0.3	15	0.3	13	0.2	2	0.9
Weight in pounds								
< 70 lbs	-	0.0	-	0.0	-	0.0	-	0.0
70 - 139 lbs	1,416	26.6	1,280	26.3	1,186	26.1	94	30.7
140 - 209 lbs	3,581	58.8	3,282	59.1	3,073	59.2	209	57.0
210 - 269 lbs	618	10.8	564	10.9	525	11.0	39	9.0
270+ lbs	105	1.7	98	1.7	90	1.7	8	0.9
Missing	205	2.1	192	2.0	178	2.0	14	2.4
Had prescribed medication in the last month								
Yes	2,066	37.3	1,895	37.7	1,761	37.5	134	39.4
No	3,944	62.6	3,597	62.2	3,359	62.3	238	60.6
Missing	10	0.1	8	0.2	8	0.2	-	0.0
Ever take vitamins								
Yes	1,935	37.0	1,759	37.0	1,650	37.0	109	36.4
No	4,077	62.9	3,735	62.9	3,472	62.9	263	63.6
Missing	8	0.1	6	0.1	6	0.1	-	0.0
Physical activity								
More active than othe	1,693	29.9	1,536	30.0	1,430	30.1	106	28.6
Less active than other	1,348	22.5	1,226	22.6	1,150	22.9	76	17.2
About the same as oth	2,849	45.1	2,624	45.1	2,446	44.9	178	48.3
Missing	130	2.5	114	2.4	112	2.1	12	5.9
Ever had chest pains								
Yes	1,640	27.8	1,537	28.9	1,444	29.2	93	23.6
No	4,379	72.2	3,962	71.1	3,683	70.7	279	76.4
Missing	1	0.0	1	0.1	1	0.1	-	0.0

*Includes SPs with usable data

Table 10. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare the interviewed sample, and the examined sample for persons 60 years and older

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglo (4)		Nonrespondents to hemogl (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	Sample Siz	%
Total	4,369		2,713		2,517		196	
Race/ethnicity								
White, non-Hispanic	2,712	86.7	1,603	86.4	1,506	86.8	97	79.8
Black, non-Hispanic	830	8.6	562	8.6	499	8.2	63	14.8
Mexican American	747	2.0	497	2.0	466	2.0	31	1.7
Other	80	2.6	51	3.0	46	3.0	5	3.7
Sex								
Male	2,107	43.0	1,368	43.0	1,277	43.3	91	39.1
Female	2,262	57.0	1,345	57.0	1,240	56.7	105	60.9
	Interviewed sample (2)							
Total	3,468	100.0						
Poverty Index Level								
Below poverty level	577	10.8	460	10.7	421	10.5	39	13.8
At or above poverty l	2,351	76.6	1,893	78.3	1,762	78.5	131	76.3
Missing	540	12.6	360	11.0	334	11.0	26	9.9
Marital Status								
Not married	1,374	35.5	1,009	34.6	934	34.7	75	33.2
Married	1,914	60.0	1,586	61.1	1,475	60.9	111	63.3
Never married	161	4.4	115	4.2	105	4.3	10	3.5
Missing	4	0.1	3	0.1	3	0.1	-	0.0
Education Level								
0 - 8 years	1,329	24.1	1,042	24.8	957	24.5	85	29.5
9 - 11 years	644	18.9	488	18.6	451	18.6	37	18.1
12 years	744	28.7	594	28.1	553	28.1	41	28.8
13+ years	701	27.7	567	27.9	536	28.2	31	23.3
Missing	35	0.6	22	0.5	20	0.6	2	0.3
Health Status								
Good/very good/exce	2,135	68.5	1,711	68.4	1,595	68.5	116	67.5
Poor/Fair	1,325	31.3	997	31.3	917	31.3	80	32.5
Missing	8	0.2	5	0.3	5	0.2	-	0.0
Ever had hypertension								
Yes	1,539	44.8	1,187	43.9	1,098	43.7	89	46.3
No	1,912	54.9	1,514	55.7	1,407	55.9	107	53.7
Missing	17	0.3	12	0.4	12	0.4	-	0.0
Ever had high cholesterol								
Yes	737	28.1	612	29.1	575	29.2	37	27.4
No	2,467	71.6	1,920	70.6	1,777	70.5	143	72.6
na	-	0.0	-	0.0	-	-	-	0.0
Missing	10	0.3	6	0.3	6	0.3	-	0.0

Table 10. Weighted distributions of respondents and nonrespondents to the hemoglobin item compare the interviewed sample, and the examined sample for persons 60 years and older (continued)

Characteristics	Selected Sample (1)		Examined Respondents (3)		Respondents to hemoglo (4)		Nonrespondents to hemogl (5)	
	Sample Siz	%	ample Siz	%	ample Siz	%	Sample Siz	%
Ever had diabetes								
Yes	529	13.2	418	12.9	387	13.0	31	12.0
No	1,912	86.7	2,291	87.1	2,128	87.0	163	87.8
Missing	17	0.1	4	0.0	2	0.0	2	0.2
Weight in pounds								
< 70 lbs	3	0.0	1	0.0	1	0.0	-	0.0
70 - 139 lbs	974	28.4	716	27.8	660	27.8	56	27.9
140 - 209 lbs	2,077	60.0	1,669	61.1	1,557	61.1	112	60.6
210 - 269 lbs	222	8.0	193	8.3	181	8.5	12	5.8
270+ lbs	18	0.4	13	0.3	12	0.3	1	0.3
Missing	155	3.0	108	2.5	93	2.3	15	5.4
Had prescribed medication in the last month								
Yes	2,418	69.8	1,869	70.3	1,738	70.3	131	70.4
No	1,032	29.7	837	29.5	773	29.6	64	28.3
Missing	18	0.4	7	0.2	6	0.1	1	1.3
Ever take vitamins								
Yes	1,224	39.4	962	40.1	910	41.1	52	25.6
No	2,228	60.2	1,746	59.8	1,602	58.8	144	74.4
Missing	16	0.3	5	0.1	5	0.1	-	0.0
Physical activity								
More active than othe	1,224	38.0	1,004	39.1	950	39.9	54	27.9
Less active than other	725	18.2	526	16.9	472	16.4	54	24.0
About the same as oth	1,366	39.6	1,080	40.4	1,001	40.3	79	42.8
Missing	144	4.3	103	3.5	94	3.4	9	5.3
Ever had chest pains								
Yes	1,090	33.8	895	36.2	834	36.1	61	37.4
No	2,375	66.2	1,818	63.8	1,683	63.9	135	62.6
Missing	3	0.0	-	0.0	-	0.0	-	0.0

*Includes SPs with usable data

overrepresented in the sample. Furthermore, children under 29 pounds (the younger children, in general) are underrepresented in the phlebotomy sample.

Table 8 shows the weighted distribution of the hemoglobin sample for persons aged 6 to 19 years. Black non-Hispanics show a slight underrepresentation in this sample. Other groups that are also slightly underrepresented in the sample are those who: (1) have fewer than 9 years of education, (2) weigh less than 70 pounds, (3) have taken vitamins, (4) had no chest pains, and (5) did not have anemia.

Table 9 shows the same distribution for persons aged 20 to 59 years. The distribution of the hemoglobin item with respect to the variables included in the table, is very similar to the selected sample, except for one variable. It seems that persons who had chest pains are overrepresented in this sample. The same overrepresentation is apparent for persons aged 60 years and older, as given in Table 10. In addition, Table 10 shows that persons with high cholesterol, those who take vitamins, and those who are more active than others seem to be overrepresented in the hemoglobin sample for persons aged 60 years and older. Persons with hypertension are somewhat underrepresented in this sample.

Furthermore, the analysis described above did not include any tests of significance for differences between respondents and nonrespondents in the samples. Analysts are encouraged to conduct chi-squared tests of significance to assess the differences in population distribution among respondents and nonrespondents. Section 3.2 contains general guidelines for the construction of Tables 5 through 10.

As mentioned earlier, the kinds of analyses provided in Tables 5 through 10 can only present approximations to the effects of nonresponse biases. Extrapolating from differences between respondents and nonrespondents in interview characteristics to items obtained in examinations can only be done by assuming that the health patterns for nonrespondents are similar to those of respondents, within each category analyzed. It is unlikely that this assumption holds exactly. Even as approximations, however, these types of analyses do provide a useful guide to the potential effects of nonresponse on the statistics under study.

3.2 General Guidelines for Computing Weighted Sample Distributions

The sample distributions for the macular degeneration, total bone mineral density, and hemoglobin items are computed based on the interview status and the item result code. The interview status is used to identify the sampled persons selected in the original sample. It is also used to identify the respondents in the interview and examination samples. For the examination item sample, the interview status along with the item result code are used to identify the respondents and nonrespondents. Refer to Section 2.1 on the guidelines for computing response rates for the definition of interview status. The item result code, for each item, is determined according to a specific test administered during the examination. The SP is assigned an item result code based on whether the test was completed or not. In order to compute the weighted distributions for the original sample, interview sample, examination sample, and item sample, follow the steps described below.

Fundus Photography Component -- Macular Degeneration Scores for Early Detection of Diabetes

Weighted sample distributions for the selected sample:

1. Subset the main file to sampled persons (SPs) who are age 40 years and older (retain all variables associated with the SP). The main file refers to the file of all screened persons. Identify the SPs by their assigned interview status. Note, all SPs should have a nonmissing value for interview status.

2. Next, define specific categories for the demographic and health-related variables. Before defining categories, it would be helpful to produce frequencies for the variables of interest. For some variables the reported value may be out of range or missing. An out-of-range value is reported when the response of the SP is “unknown” or is a “don't know” response. A missing value, on the other hand, appears for two reasons: when there is a skip pattern in the questionnaire to avoid asking non applicable questions and when the SP does not answer the question when prompted. Both the out-of-range response and the missing value due to the SP's nonresponse should be categorized as missing when computing estimates. In the case of the skip pattern, the missing value can be eliminated by recoding the missing value of the variable of interest to a valid answer. This requires examining the questions preceding the question (or variable) of interest to determine the appropriate recode value. For example, the preceding question may ask if the SP ever smoked cigarettes and the answer is no. The question of interest asks if the SP smokes now; however, a skip pattern appears for this question and the SP is not asked the question. The value is left missing for this SP. Change the missing value to the assigned value for "no, the SP does not smoke now".

3. Use the poststratified base weight to compute the weighted distribution for the original sample of SPs identified. For each demographic or health-related subgroup defined in step 2 , calculate the weighted distribution as the sum of the

poststratified base weights for SPs in a given category divided by the sum of the poststratified base weights for all SPs in the original sample.

Weighted distributions for the interview sample:

1. Identify the SPs who responded to the interview and subset the main data set to those people. For the macular degeneration item, create a file containing SPs who are age 40 and older and with an interview status of 1, 2, or 3.

2. Repeat step 2 described above for the original sample distribution computations.

3. Use the final interview weight to compute the weighted interview sample distribution. For each demographic or health-related subgroup defined, compute the weighted sample distribution as the sum of the final interview weights for SPs in a given category divided by the sum of the final interview weights for all SPs in the interview sample.

Weighted distributions for the examination sample:

1. The next step is to compute the examination sample distribution. Since the focus is on a component administered at the MEC, the sample distributions are computed based on SPs that completed some portion of the MEC examination. The sample includes SPs who have an interview status of 2 and are age 40 or older. Create a data set containing these SPs.

2. Repeat step 2 described above for the original sample distribution computations.

3. Use the final MEC examination weight to compute the weighted examination sample distributions. For each demographic or health related subgroup defined, compute the weighted sample distributions as the sum of the final MEC weights for SPs in a given category divided by the sum of the final MEC weights for all SPs in the examination sample.

Weighted distribution of respondents and nonrespondents for the item sample:

1. Start with the data set used for computing the examination sample distributions. The SPs should be age 40 and older with an interview status of 2. Split this data set into two files, one for the respondents to the item and the other for the nonrespondents. The respondents to the item are defined as SPs with a gradable (or usable) macular degeneration score and an item result code of complete. A respondent's score is gradable if: there is no degeneration detected, early degeneration detected, or late degeneration detected. Persons with a nongradable (or missing) score and an item result code of incomplete are classified as nonrespondents. Before splitting the file, run a cross tabulation between the two variables used to define respondents and nonrespondents to check for discrepancies. An example of a discrepancy would be if the SP has a gradable

score for the item and an item result code of incomplete. In this case, the SP is treated as a respondent. The opposite is true also. If the SP has a nongradable (or missing) score and an item result code of complete, the SP is treated as a nonrespondent.

2. Repeat step 2 described above for the original sample distribution computations.

3. To compute the respondent distributions, use the file created in step 1 for respondents to the macular degeneration item. Use the final MEC weight to compute the weighted macular degeneration item sample distributions for the respondents. For each demographic or health-related subgroup defined, compute the weighted sample distribution as the sum of the final MEC weights for SPs in a given category divided by the sum of the final MEC weights for all SPs in the respondent item sample.

4. To compute the nonrespondent distributions, use the file created in step 1 for the nonrespondents. Use the final MEC weight to compute the weighted macular degeneration item sample distributions for the nonrespondents. For each demographic or health related subgroup defined, compute the weighted sample distributions as the sum of the final MEC weights for SPs in a given category divided by the sum of the final MEC weights for all SPs in the nonrespondent item sample.

Bone Density Component -- Total Bone Mineral Density Measurement

Weighted original sample distributions:

1. Subset the main file to persons with an interview status of 0, 1, 2, or 3 who are age 20 and older.

2. Repeat steps 2 and 3 described for computing the original sample distributions for the fundus component. Note that the file for bone density should include SPs age 20 and older.

Weighted interview sample distributions:

1. Subset the file used for computing the interview sample distributions for the bone density file to SPs with an interview status of 1, 2, or 3.

2. Repeat steps 2 and 3 described for computing the interview sample distributions for the fundus component. (Include SPs age 20 and older).

Weighted examination sample distributions:

1. Subset the file used for computing the interview sample distributions for the bone density item to SPs with an interview status of 2.

2. Repeat steps 2 and 3 described for computing the examination sample distributions for the fundus component. (The file should contain SPs age 20 and older).

Weighted item sample distributions for respondents and nonrespondents:

1. Start with the data set used for computing the examination sample distributions. The SPs should be age 20 and older with an interview status of 2. Split this data set into two files one for the respondents to the item and the other for the nonrespondents to the item. The weighted item sample distributions are based on the MEC bone scan test results. Respondents to the total bone mineral density item have a good bone scan (or usable) test result and an item result code of complete. Persons who do not have a good scan (including SPs with no scan) and an item result code of incomplete are classified as nonrespondents. Similar to the fundus component, it would be helpful to check for discrepancies and make any necessary changes before splitting the file.

2. Repeat step 2 described above for the original sample distribution computations for the fundus component.

3. Repeat steps 3 and 4 for computing the respondent and nonrespondent distributions for the fundus component.

Phlebotomy Component -- Hemoglobin Test

Weighted original sample distributions:

1. Subset the main file to persons with an interview status of 0, 1, 2, or 3 who are age 1 and older.

2. Repeat steps 2 and 3 described for computing the original sample distributions for the fundus component. Note that the file for phlebotomy should include SPs age 1 and older. Also when choosing variables of interest, remember that the adult and youth questionnaires are slightly different. That is, there are some questions that pertain to adults only or youth only.

Weighted interview sample distributions:

1. Subset the file used for computing the interview sample distributions for the phlebotomy component to SPs with an interview status of 1, 2, or 3.

2. Repeat steps 2 and 3 described for computing the interview sample distributions for the fundus component. (Include SPs age 1 and older).

Weighted examination sample distributions:

1. Subset the file used for computing the interview sample distributions for the phlebotomy component to SPs with an interview status of 2.

2. Repeat steps 2 and 3 described for computing the examination sample distributions for the fundus component. (The file should contain SPs age 1 and older).

Weighted item sample distributions for respondents and nonrespondents:

1. Start with the data set used for computing the examination sample distributions. The SPs should be age 1 and older with an interview status of 2. Split this data set into two files one for the respondents to the item and the other for the nonrespondents. The weighted item sample distribution is based on the MEC hemoglobin (HGB) test results. Respondents to the item have a HGB test result greater than zero and an item result code of complete. Persons who have a HGB less than or equal to zero (including SPs with no test result) and an item result code of incomplete are classified as nonrespondents. Like the fundus component, check for discrepancies and make any necessary changes before splitting the file.

2. Repeat step 2 described above for the original sample distribution computations for the fundus component.

3. Repeat steps 3 and 4 for computing the respondent and nonrespondent distributions for the fundus component.

4. COMPENSATING FOR MISSING DATA ITEMS

The results of the nonresponse bias analysis provided in the previous section can help in developing and applying techniques for compensating for missing data items. This can be accomplished by incorporating the information obtained in Tables 5 through 10 into an item nonresponse adjustment procedure to reduce the bias. This section provides brief descriptions of various item nonresponse compensation procedures.

Items in the components of the medical examination are subject to potential bias from missing data since no item nonresponse adjustment was carried out for phase 1 of the NHANES III sample. As with unit nonresponse, this bias generally cannot be eliminated by nonresponse adjustments. However, adjustments for nonresponse can be performed in such a way that the associated nonresponse bias is reduced without sharply increasing the variance of a survey estimate.

A basic approach in compensation for missing data develops relationships for various survey measurements among the respondents that can be used to extrapolate measurements to the nonrespondents. Auxiliary variables exist for many items in NHANES III data, either in the examination or interview component, that can assist in the nonresponse adjustments procedures. However, when considering the magnitude of the total nonresponse for certain variables, such as the biochemical variables, questions must be raised concerning the effect of compensating for missing data elements and the most appropriate adjustment procedures to be employed.

In general, there are two strategies available to overcome the problem of missing items in sample surveys: (1) nonresponse weighting adjustment, and (2) imputation. In weighting adjustments for nonresponse, missing or incomplete units (or items) in the sample are ignored and the sampling weights of the responding units (or items) are inflated (or referred to as adjusted) to account for the nonresponding units. In the imputation approach incomplete or missing data are included in the sample with missing values replaced by imputed values.

The second strategy is more popular among survey analysts. A discussion of the advantages and disadvantages of imputing for missing values is given by Kalton (1983). The benefits include reduction in nonresponse bias in survey estimates when using an appropriate imputation technique. Furthermore, the result of imputation is a complete data set which makes analyses easier, since the analyst need not make special provisions for handling of observations containing missing data only. The disadvantage of working with data sets including missing items is that a person's

record will only be included in an analysis if the record does not have any missing values with respect to the variables included in the analysis. Thus, different sets of respondents will be included in different analyses.

A disadvantage of imputation is the possibility of ending up with a data file after imputation that is more biased than if no imputation had been performed. Bias reductions depend on the suitability of the assumptions made in the imputation. When imputations are performed separately on different variables, the bias may be reduced for univariate statistics based on the variables containing the imputed data, but multivariate relationships among variables could become distorted. Also, researchers may treat the resulting data set as if it were complete, thus affecting the variances of the estimates. In some instances, it is argued that when a small proportion of observations are imputed, the effect of imputation is relatively minor. However, in such cases, analyses performed on subgroups may contain a high proportion of imputed values.

Data files containing imputed data should also contain certain information regarding imputation. In particular, imputed values should be flagged. Flagging imputed values on a data set allows an analyst to explore the impact of nonresponse on the results from an analysis. Even when the imputed values are flagged, assessing their impact on the results may still be difficult. Other recommendations include coding the number of times a particular record is used as a donor of imputed values, coding the number of attempts to achieve a successful imputation, and identifying the donors and their donated values. Furthermore, no one imputation is good for all types of analyses. Analysts may wish to choose imputations for a particular variable based on the type of analysis to be performed.

Almost all nonresponse adjustments involve imputations within classes of the sample. These adjustment classes are analogous to the weighting classes in unit nonresponse adjustments and are used to bring together sample persons with similar characteristics. The variables used to define the adjustment classes are typically used as matching or control variables. Using these matching variables, sample persons with data on a particular item of interest are matched with sample persons with data missing for the item of interest. Then the adjustment methods are implemented within the classes. The guidelines for defining the adjustment classes are the same as those used for defining weighting classes and will not be explored further here. Refer to Trena, et al. (1993) for a discussion of weighting class adjustments for unit nonresponse adjustments in NHANES III.

In summary, item nonresponse adjustment is most conveniently carried out through imputation techniques. Imputation has several advantages:

- If the compensation is done well, it will usually reduce bias in survey estimates. It is almost always preferable to impute for missing items rather than treating them as randomly missing data at the analysis stage. Although imputation obviously will not eliminate all nonresponse biases, it can be expected to dampen their effects considerably;

- Imputation makes analysis easier and the results simpler to present; and

- Compensation for missing data will lead to consistent results from different analyses if imputation is done centrally at NCHS.

The ability of imputation techniques to eliminate or reduce bias depends on the extent to which, for nonparametric techniques, classes can be defined that are relatively homogeneous with regard to the variable of interest or, for parametric techniques, covariates can be defined that are highly correlated with the variable of interest. Their effectiveness also depends on the extent to which classes can be defined, or covariates identified, for which the response mechanism or probability of response is approximately the same within class and the extent to which response rates differ between classes.

As mentioned earlier, all imputed items should be flagged. Item nonresponse rates should be calculated the same way whether or not imputation is used. However, caution is necessary in using adjusted or imputed data. The magnitude of bias that still remains after imputation should be investigated. A particular concern is that some subgroups of the sample may contain sizable proportions of missing data and estimates for the subgroups may significantly depend on adjusted or imputed values. The effect of compensation for missing data needs to be assessed for these subgroups. In addition, the guidelines described in section 7 apply with or without imputation.

Unit nonresponse was handled by weighting-class adjustments, with little or no compensation made for item nonresponse, in previous NHANES surveys. NHANES III applied similar weighting-class adjustments for unit nonresponse (Ezzati and Khare, 1993). In 1992, NCHS initiated a project to investigate alternatives approaches to current NHANES nonresponse adjustment methodology. Two single and one multiple imputation procedures were investigated in this research project. In addition, Klein et al. (1994) presents another method of compensating for missing data in NHANES III. Brief descriptions of the methods, and the results of the evaluation carried out for NHANES III Research on imputation techniques is provided in section 4.1. Section 4.2 provides a list of other compensation procedures described in the literature.

4.1 Evaluation of Missing Data Compensation Procedures for NHANES III

As indicated earlier, two single and one multiple imputation procedures were investigated for NHANES III. The two single imputation methods applied two closely related regression techniques. The first one involved predictive mean matching, and the second one was based on a hot-deck procedure in which empirical residuals were added to the predicted values. A multiple imputation procedure based on a multivariate model for mixed normal and categorical data was the third method investigated under this research. The following provides descriptions of the underlying compensation procedures and a summary of the conclusions of the investigation.

Hot-deck. A common technique for imputation that randomly selects a donor with a complete set of data whose values are then assigned to a recipient with missing data. The donor and recipients typically are matched according to certain characteristics such as sex, race, age, etc. The hot-deck method selects donors internally, that is from the data file containing the recipient.

The method of hot-deck imputation is applied within imputation classes. Within a class, a respondent with information on the variable of interest is designated as a donor. The donor is selected either at random or in some determined manner. The donor's value is then assigned to a recipient whose variable of interest contains a missing response. In general, the hot-deck procedure is a duplication technique imputation. Refer to Ford (1983) for a review of other research that explore sequential hot-deck procedures.

Regression Imputation. The imputation technique of regression does not use imputation classes. Instead, a regression model is developed by using the variable requiring imputation as the dependent variable, and control variables as the independent variables. Once the model is developed, the recipient's values for the control variables are inserted into the model and a predicted value is obtained. The imputed value can take on the value either of the predicted value or of the predicted value plus a randomized residual.

Kalton (1983) describes the relationship between regression imputation and imputation based on imputation classes. Essentially, estimation of a value through the use of imputation classes defined by qualitative variables reduces to an analysis of variance in a linear model, where no assumptions are made for the residuals. A suitable estimator is provided by selecting at random a donor within an imputation class. The lack of assumptions about the distribution of the residuals reflects that no assumptions are made when using imputation classes. Kalton points out that better

estimators can be obtained if assumptions are made for the model. Such assumptions could concern the distribution of the residuals or linear relationships between the dependent variables and the independent variables.

Regression models can be used to impute values for certain continuous variables containing missing data. The models, either single or multiple, can be fit to the subgroup of records containing valid values. Once an adequate model is obtained, the variance of the residuals will be obtained through PROC UNIVARIATE; the residuals, by definition will have a mean of zero. To obtain an imputed value for the item of interest, the values corresponding to the independent variables in the fitted regression model will be substituted from the record containing the missing data into the equation, resulting in a predicted value. A randomly chosen error term will be added to the predicted value to give the imputed value. The residual will be selected from the distribution of the residuals, determined from PROC UNIVARIATE.

Two separate imputation methods were investigated using independent imputation programs specifically created for NHANES (Ezzati, et al. 1993). The first method, WESMATCH, is capable of performing various forms of statistical matching. This SAS macro has been used to perform a version of predictive mean matching imputation. For the second method, Hane-Deck, a customized imputation program was developed to perform a hybrid hot-deck imputation.

In both methods, a model was built by regressing the target variable on a carefully selected set of covariates. For this purpose, the missing values of all categorical covariates were presented as separate categories. With the regression coefficients estimated, a predicted value was computed for the target variable in question. All missing values of the continuous covariates were replaced by average value within sex, age, and race categories.

The first step was the same for both methods. For the second of the first method, a variation on the conventional regression method of imputation was used. All missing values following the computation of predicted values for a target variable, were imputed via WESMATCH. For this purpose, donors were selected based on the proximity of their predicted values. Specifically, all observations were first listed in ascending order of the computed predicted values. Then, when a record with a missing value was encountered, the observed value corresponding to the record with the closest predicted value with the least number of prior donations was selected. Each donor could donate three times before it was banned from further donation.

With a complete set of predicted values, the second step of the second method consisted of imputation of all missing residuals via hot-deck. For this purpose, all records were sorted by sex, age, and race; and pools of donors and missing values were created within the resulting pools. Next, all missing residual values were imputed by selecting donors from the corresponding donor pools of residuals. The missing values of the target variable were then imputed by adding the imputed residuals to the corresponding predicted value.

Multiple Imputations. Rubin (1977, 1978), and Rubin and Little (1978) considered a Baysian approach using auxiliary variables to correct the respondent mean for the bias caused by nonignorable nonresponse. Other recent articles in this area include: *Multiple Imputation after 15 Years* (Rubin, 1993). *Evaluation of Imputation Methods* (Schaefer, 1992), and *Multiple Imputation* (Khare et al., 1993).

Khare, et al. (1993) describe the preliminary efforts to multiply impute a portion of the data from phase 1 of NHANES III. A data set consisting of 27 key variables for 12,392 sampled adults was multiply imputed, for both item and unit nonresponse, using techniques of iterative Bayesian simulation via Markov chains described by Schaefer (1991). Exploratory analysis of the imputed values suggests that both the marginal distributions of variables and important relationships between them were accurately preserved. For further research, refer to Schaefer (1993) on multiple imputation in NHANES III.

The model was a special case for mixed continuous and categorical data introduced for discriminate analysis by Krzanowski ((1982) and applied to incomplete multivariate data by Little and Schluchter (1985). Multinomial and multivariate normal distributions were assumed for the discrete and continuous variables. A full description of the approach is given in Khare, et al. (1993).

Comparison of the Approaches. Ezzati, et al. (1993) describe the research conducted to compare the alternative missing data adjustment methods for selected survey components in phase 1 of NHANES III based on single and multiple imputation methodology.

The results focus on single and multiple imputation for item nonresponse among examined persons only. The two singly imputed values resulted in extremely small differences even when imputing for 14 percent missing data. Thus, from a methodological point of view, one method was not superior to another. Fahimi et al. (1992), however, has pointed to some computational advantages to the empirical residual method.

For each variable, the relative differences were all less than 1 percent when data sets with imputations generated by both single and multiple imputation methods were compared.

For the subset of evaluated data, values generated from two single imputation techniques showed nearly identical distributions. In addition, single and multiple imputation methods resulted in similar point estimates. Both methods preserved the marginal distribution of the variables and the relationship between them.

As Ezzati, et al. (1993) pointed out, a number of important issues must be addressed and additional research should be carried out before specific recommendations on imputation strategies for NHANES III can be made. Although easy-to-use SAS software for single imputation was developed, methods like hot-deck imputation are complex and difficult to implement in multivariate data sets. On the other hand, the model-based multiple imputation technique works well in the multivariate survey setting, but it requires specialized computing software. There is also a need to do further research on the validity of each approach and the derivation of better methods of variance estimation such as model-based variances.

4.2 Other Missing Item Compensation Methods

The following provides brief descriptions of other compensation procedures for item nonresponse.

Logical Imputation. This type of imputation is useful when the logic of the case suggests a value for the missing data. An example is a record for an observation that does not give the person's sex, but it would be logical to assign a sex based on the person's name.

Mean-value Imputation. The mean-value technique is easy to comprehend and to implement. Within imputation classes the mean-value for the respondents is assigned to the records with missing data. Intuitively, this imputation method may seem to be the best, in that it is a good procedure for estimating the sample mean. However, the distribution of the variable for which the imputation is being performed becomes distorted, since the imputed values for the nonrespondents become centered at the mean for the respondents. A result is that the variability of the distribution is dampened. Furthermore, if the resulting complete set of data is treated as a set containing full response and the imputation is not taken into account, the variance for the mean will be underestimated.

Advantages of this technique are that it is easy to implement and that the end result is a complete set of data. The resulting data are appropriate for producing estimates of aggregates. Imputing the cell-mean, however, distorts the distribution by causing a reduction or dampening of the variance.

Weighting Class. When estimating aggregate statistics, weighting class adjustments yield an estimate that is identical to an estimate based on cell-mean imputations. Weighting class adjustments are appropriate when it is not necessary to produce a data set containing a complete set of values for the item of interest.

Cold-Deck Imputation. Cold-deck imputation requires the use of data from a previous survey or from other information about the population under study. Once the data are collected into a cold deck, values are assigned to nonrespondents in the current survey. An advantage of a hot-deck imputation procedure, as compared with the cold-deck, is that information from the survey -- not external data -- is used for the imputation.

A major disadvantage of this technique is that past survey data may not be comparable to the database requiring imputation. Differences in the two data sets may arise simply because of obsolescence or because different definitions were used.

Random Imputation. As described by Kalton (1983), random imputation techniques are made up of procedures that use a probability mechanism to select donors within an imputation class. The donors could be selected, for example, by simple random sampling (SRS) with or without replacement from within an imputation class.

When applying random imputation, the number of donors and the probability sampling scheme must be decided. Kalton (1983) discusses several options for situations in which the number of nonrespondents exceeds or is less than the number of respondents. The techniques can involve stratifying the potential donors prior to sampling or selecting a systematic sample of donors ordered by responses to items. One specific approach imputes multiple values instead of a single value, thereby reducing the increase in variance associated with using a sample of respondents as donors. A variant of this approach is to select a sample of donors for each recipient and to assign the mean value from the sample of donors to the recipient. Since imputed means are used, the distribution of the resulting imputed values and the respondents' values combined becomes distorted.

Flexible Matching Imputation. Kalton (1983) uses the phrase “flexible matching imputation” to identify a modified hot-deck imputation procedure that was used by the Census Bureau for one of their studies. In this modification, potential donors and recipients are matched on many control variables that are prioritized. If no donors are matched on the variables for a particular recipient, variables are dropped in reverse order of priority until a match is obtained. Advantages of this technique include obtaining closer matches and reducing the frequency of using multiple donors.

Distance Function Matching. Kalton (1983) describes an imputation technique, "distance function matching," that allows for quantitative control variables, thus avoiding difficulties associated with categorizing continuous variables. A function measuring the distance between potential donors and recipients is used to select a donor.

In the case of a single quantitative control variable, for example, the potential donors and recipients could be ordered by the quantitative control variable and the donor selected as the one nearest to the recipient. A distance function would provide a measure of nearness, and could take into account the number of times a record had been used as a donor. Alternatively, the recipient could be assigned a value that was the mean of the potential donors within a certain distance of the recipient. In the case of one quantitative control variable and more than one qualitative control variable, the matching on the quantitative variable could be performed within classes defined by the qualitative variables.

When more than one quantitative control variable is to be used, the distance function could be generalized to accommodate more than one quantitative variable. Care must be taken when defining the distance function, since variables with long tails could have a large impact on the distance metric. One suggestion is to use ranks instead of the actual value of the variable.

Zero Substitution. The zero substitution method for imputation simply ignores missing data due to nonrespondents by substituting a value of zero at the estimation stage of analysis. While such a method may intuitively appear inappropriate, Platek and Gray (1986) describe conditions when zero substitution may be called for. One instance is when there is a very high response rate which allows the use of the Horvitz-Thompson estimator without adjustments for nonresponse. The result is that the underestimate may result in a lower mean squared error, and may be less costly to produce than by other imputation methods.

A second instance of application can occur when the estimate of interest is a mean per unit using only information from respondents and making no other adjustments for missing data. A

third application produces an estimate through the use of ratio estimation. If the numerator and denominator are each underestimates produced by using zero substitution, the ratio may not necessarily be an underestimate. Also, this type of ratio estimate may be less costly than using other imputation techniques leading to estimates of the numerator and denominator, and could result in smaller increases in mean squared error. Platek and Gray (1983) present derivations and discussions of the estimate, and its expected value and variance.

Historical Substitution. Platek and Gray (1986) discuss the use of historical substitution, substituting data from historical or external sources for a unit with missing data. The authors discuss two types of historical substitution methods. One is in which historical or external data are available for all units with nonresponse. The other is where historical or external data are available for some, but not all units, and other imputation methods must be used.

Boostrap Methods. A recent paper by Efron (1994) focuses on missing data, imputation, and the bootstrap. Three main topics are discussed in this paper: bootstrap methods for missing data, the relationship of these methods to the theory of multiple imputation, and computationally efficient ways of executing them. This paper is accompanied by responses from Rubin.

It would be more efficient for imputation to be done centrally by NCHS. Common methods can be used for different sets of items. Although the quality of imputation will usually be better if it is done centrally, there may be time and budget constraints that makes it impractical for NCHS to conduct the imputation. If imputation is not done in advance of analysis by NCHS, analysts are advised to perform imputation as part of their overall activities. Frequently, an analyst's intimate knowledge of a particular subject will produce improved imputation for an item, or a related group of items.

Section 5 provides a discussion of the types of criteria used for selecting a compensation procedure and section 6 contains a description of the methods used for assessment of the imputed data. The compensation procedures discussed in section 5 and 6 are mainly imputation techniques rather than an item nonresponse adjustment approach. However, similar guidelines on the criteria for selecting an approach and assessment of adjusted values should be followed for nonresponse adjustment methods.

5. CRITERIA FOR SELECTING A COMPENSATION PROCEDURE

Kalton (1983) provides a discussion on criteria for selecting a procedure to compensate for missing data. The criteria include the precision of the resulting estimates, the estimation of the standard error of the estimates, and the suitability of the final data set for producing the desired estimates. Brief descriptions of the criteria follows:

Precision of Estimators. One consideration in choosing a compensation technique is the resulting precision of the estimator. Kalton (1983) provides a discussion on this subject under the assumption that population means for the nonrespondent and the respondents are equal. Another common assumption is that the data are missing at random. These can be translated to the constraints that $\bar{y}_m = \bar{y}_r$, or that the expected values of $\bar{y}_m$ over repeated application of the compensation method $\bar{y}_r$, i.e., $E_2(\bar{y}_m) = \bar{y}_r$ where E_2 is the conditional expectation over the random imputation scheme.

The cell-mean method is an example of a compensation method that imposes the constraint that $y_r = y_m$. Relaxing the constraint to $E_2(y_r) = y_m$ allows for other procedures such as a hot-deck method.

Two aspects involved in selecting an adjustment procedure are the sample design to be used and the size of the sample selected. The results from Kalton (1983) are based on the assumption that the population contains a fixed number of nonrespondents and a fixed number of respondents. The variances are derived conditional on a fixed number of respondents in the sample.

The amount of increase depends on the compensation method used. Calculations could be constructed for different schemes under consideration, and compared.

Estimation of Standard Errors. Computation of the standard errors is another consideration when choosing a missing data compensation technique. For a SRS, estimation is simpler for weighting adjustments than for imputation. This is because only the respondents are included in the data set used for analysis and the weights and responses are known for each unit. On the other hand, appropriate use of the imputation procedures requires flagging the respondents and indicating the number of times they are used as donors. Because every variable must be handled

separately, computing variances for variables containing imputed values can be complex and time-consuming.

Kalton cites some work showing that when nonresponse is extensive, treating a SRS data set with imputations as if the set were complete, results in a substantial overestimation of the precision for the estimate.

Suitability of Procedure for the Data Set. A major consideration when choosing a compensation procedure is the objective of an analysis. The assumption that the population means for respondents and nonrespondents are identical is reasonable when it is desired to estimate the population mean or total. However, it is not appropriate in other instances where distributional properties are important.

Kalton (1983) investigates the effect of different schemes on the distribution of a variable and its variance and also investigates the covariance between two variables. For example, he demonstrates that for cell mean imputation, the resulting covariance is an underestimate, and that in a SRS subsampling scheme such as hot-deck, the resulting distribution is unbiased for the respondent population distribution. In the case where two variables are missing or both contain data, and the nonrespondent receives both values from a single donor, the resulting covariance is unbiased for the population distribution. In the case where two variables are missing or both contain data, and the nonrespondent receives both values from a single donor, the resulting covariance is unbiased for the population covariance for respondents. If the two values are assigned independently of the variable with complete data, the covariance underestimates that for the population respondents. If both variables are subject to missing data, and the imputations are made independently, the resulting covariance is again underestimated.

Kalton (1983) points out that simple weighting procedures for unit nonresponse retains all properties of the respondent sample and thus leads to good estimates of univariate and multivariate statistics. The use of weighting adjustments is preferred to imputation techniques wherever possible. The use of imputation creates difficulty concerning the construction of sampling errors and the relationships between variables. The use of subsampling, such as by hot-deck methods, adds another sampling error component that can be reduced through multiple imputations or choice of subsampling scheme.

6. ASSESSMENT OF THE IMPUTED DATA

Once an imputation technique is chosen and applied to the data, the results must be examined to assess the performance of the technique. The assessment should include an examination of the distribution of the imputed values for detection of outliers. It should also include an assessment of bias and estimates of variance and mean squared error for the resulting estimates. Section 6.1 presents a number of approaches that can be used in the detection of outliers. Section 6.2 summarizes the criteria used for assessing the compensation procedures.

6.1 Detection of Outliers and Unusual Imputed Data

One of the primary evaluations of the imputation technique is the detection of outliers in the imputed data. Kammerman, et al. (1987) describe and recommend procedures that can be used to detect outliers resulting from imputation and provide methods of accommodating outliers in analysis. They discuss some steps that can be taken to avoid the imputation of outliers. These steps are general guidelines rather than detailed specifications because each imputation task has its own special characteristics that make it a unique problem. However, with regard to the problem of imputing outliers, there are two important aspects in the design of any imputation procedure that need to be emphasized. First, the presence of outliers in the data will increase the chances of creating outliers through imputation. Thus, to the extent feasible, these outliers should be identified and excluded from the imputation process, if not from the data files themselves. Second, it is important to use as many relevant covariates as possible in designing the imputation procedure.

The need to review the reported data for extreme values is illustrated by the following two examples. Flegel, et al. (1986) performed an analysis of subscapular and triceps skinfold measurements taken during NHANES II. Their analysis focused on the development of appropriate models for use in detecting extreme values, and on the use of these models for imputation. The paper demonstrates the usefulness of such models for outlier detection and other purposes, but also points out the consequences of including extreme values in the imputation process. In the paper the authors cite an example using NHANES II data where the unweighted mean subscapular skinfold for white males age 15 was reduced from 10.6 mm to 9.9 mm when an “obvious outlier” and two imputed values were replaced by values generated from their model. What is interesting is the fact that the outlier and two imputed values each had the same value of 65 mm. If the two imputed values were in fact determined by the outlier, then removing it from the imputation process would have avoided

imputing values that were too large. On the other hand, if the imputed values were determined from other records (e.g., in a hot-deck process), then it is clear that irrelevant classing variables were used in the imputation process. In either case, some analysis of the relationship between the variable to be imputed and the potential covariates might have led to improved procedures for imputation.

A second example comes from a paper by Kovar (1984). In this paper, Kovar discusses the effect of imputations on estimates from National Medical Care Utilization and Expenditure Survey for small domains. The imputations were made by a weighted hot-deck procedure. Again, some extreme values were apparently not challenged and permitted to remain in the data files for imputation purposes. It turns out that of the six hospitalizations with charges of more than $90,000, three were imputed. Moreover, the three cases that were not imputed were later discovered to be in error. Thus the impact of the erroneous data on the survey estimates was magnified by the resulting imputations. Even if the data for the largest hospitalizations were correct, the effect of the imputation would have been to double the frequency of largest hospitalizations in the sample, in effect giving the extreme values an unduly large weight in the estimates.

The above examples point out some of the pitfalls in imputation, in particular with regard to the problem of imputing outliers. However, they also indicate steps that can be taken to minimize this possibility. One is to screen the data for outliers before actually carrying out the imputation process. Often this review will lead to the identification of erroneous data that can be corrected. In other cases, an extreme or unusual value will be identified that cannot be discarded on the grounds that it is in error. Here, it is less clear what the proper action should be. One approach, as discussed earlier, would be to retain the value in the data fields, but to exclude it from the imputation process.

The basic idea behind most outlier detection techniques is that individual values in a data set can be compared against the overall “pattern” of observations. A measure of the “unusualness” of particular value can then be computed in the form of an appropriate test statistics, or alternatively, significance probability. Extreme values are declared to be “outliers” deserving further examination or review if the computed value of the test statistics is larger than would be expected under the assumption that all of the observations are from the same population. Sets of data that are thought to be homogeneous but which actually contain observations from different populations are sometimes said to be “contaminated.”

Barnett and Lewis (1984) refer to such tests as “tests for discordancy.” Originally these tests were derived under the assumption that the data followed a normal distribution, but later the

tests were modified to handle certain non-normal distributions, most notably the exponential and gamma-type distributions. We note that the two general classes of distributions mentioned above cover a wide variety of situations, and thus should be applicable to many of the numeric variables in the NHANES data files. Since the significance probabilities associated with the tests depend on the underlying distribution of the sample observations, it is important to initially examine the distribution of the data to determine the appropriate test(s) to use. Alternatively, certain nonparametric tests may be used (e.g., see Naus, 1975).

Graphical Methods. Simple plots or charts of the observed data that are used to identify certain patterns in the data can often expose extreme and potentially contaminated values. A histogram not only displays the general shape of the distribution, but also the frequency of extreme values. From the histogram and associated frequency distribution, it is then possible to establish a rule in which all of the values exceeding or falling below specified cutoffs would automatically be reviewed for validity. For example, the cutoffs may be the 95th and 5th percentiles, respectively, or other cutoffs chosen on the basis of considerations of costs and expected payoff.

In the case of variables that are correlated, a graphical display provides a powerful tool for identifying outliers. Note that in this case, an "extreme" observation is not necessarily one that carries a large value, or is unusual when examined by univariate histogram but rather one that departs from the overall pattern of the data. Software packages that produce three-dimensional plots can similarly be used to examine triplets of variables. To discover important relationships between variables, a substantial analytic effort may be involved. Practical considerations will dictate the extent to which these analyses can be carried out, and users of data files such as the numerous NHANES files should be aware of these constraints when interpreting imputed values.

The advantage of the graphical methods is that gross outliers can often be easily detected by a simple visual inspection. In a sense, the more formal techniques described later are simply refinements of this "crude" visual inspection. The principal disadvantage of the graphical methods is that they depend to some extent on judgments concerning the unusualness of observations, especially those that are considered "borderline."

Graphical methods (e.g., scatter plots) are useful for identifying patterns in multivariate data, and hence observations that disturb or deviate from that general pattern. A limitation of graphical methods is that only up to three dimensions can be plotted. However, in the case of continuous variables, the dimension of observed multivariate data can sometimes be reduced with an appropriate transformation This was the case of the skinfold example of Flegel, et al. mentioned

earlier. In that example, the original data were expressed in terms of three variables, height, weight, and skinfold measurements. Considerable simplification was achieved by reducing the problem to a bivariate one involving only skinfold and Quetelet index, Q, defined by $Q = \text{weight}/(\text{height})^2$.

Univariate Tests. Univariate tests can be used as an initial step in testing for the presence of outliers. For example, normal, exponential or gamma are distributions used in univariate cases in Kammerman, et al. (1987).

Assumptions on the underlying distribution of the variables under study are necessary for the calculation of significance probabilities, i.e., the probability that a particular observation comes from the same population as the rest of the sample. From the point of view of data editing and verification, it is not so important to know the exact values of these probabilities, as it is to identify cases in the sample which are potentially contaminated, i.e., for which these probabilities are low. On the other hand, some extreme values may be "borderline" cases in the sense that they are not so extreme as to warrant deletion from the data files, but unusual enough to be suspicious. In this case, it may be useful for analysts to have some sort of measure of the unusualness of a particular observation so that a decision can be made regarding whether or not the observation should be included in the analysis. Use of these measures also flags an observation as having been examined, but for which no obvious errors (e.g., in coding or keypunching) were detected. Of course, physically impossible values should be discarded, or re-measured and corrected if feasible.

If the underlying distributions of the variable under study is normal or gamma, the tests described earlier are appropriate, and significance probabilities can be computed using the methods described in Barnett and Lewis (1984). However, these tests are generally sensitive to the assumptions made, and where the usual assumptions about the underlying distribution of the data are not appropriate, nonparametric tests can be used in outlier detection, and approximate significance probabilities can be obtained for these.

Nonparametric Tests. An example of a nonparametric test procedure is the "simple editing rule" described in Naus (1975). This procedure is relatively simple to apply and has the advantage of being able to be implemented in batches. Moreover, it assigns to each record a "measure of suspicion" that can be used by the analysts to decide whether to exclude a particular observation from analysis. For more detail on nonparametric tests, refer to Kammerman, et al. (1987).

Multivariate Tests. It was indicated earlier that univariate methods tend to be less effective when applied to variables that are correlated. The reason for this is that such procedures do not take account of the intrinsic relationship between variables that may offer additional clues about the potential discordancy of a particular observation. Thus, a particular value of X may not be considered "extreme" when the variable is examined alone, and similarly a particular value of Y may not be considered unusual, but when taken together, the pair (X, Y) may be discordant in the sense that the (bivariate) observation departs "significantly" from the overall pattern exhibited by all other (X, Y)'s.

Many of the items collected in NHANES (e.g., serum cholesterol, body measurements, blood, and biochemistry data) depend on factors such as age, sex, race, health status, and others, and failure to take account of these when applying univariate tests can lead to misleading conclusions. This points out the essential difficulty in outlier analysis. The techniques described in the univariate case, while valid, apply only to an oversimplified situation in which a homogeneous subset of the sample can be identified for which (in the absence of outliers) the variable under consideration behaves in some known fashion. The goal then is to identify these subsets or factors that are associated with the outcome of the variable.

The most common approach for testing multivariate data for outliers involves determining an appropriate transformation of the (multivariate) observations to reduce the original multivariate problem to a univariate one. Additional details, including approximate significance tests for determining discordancy, are given in Barnett and Lewis (1984).

For the procedures discussed above to be effective in detecting outliers, it is important to have good information about the underlying structure of data. Thus, these methods require either substantial advance knowledge of the underlying relationships governing the response to a particular survey item, or a significant expenditure of time and resources to determine these relationships through extensive analytic work. Given this information, implementation of these procedures for detecting outliers is straightforward. The difficult part is the preliminary work needed to establish the appropriate methods to use.

In general, it will not be possible to completely avoid the imputation of outliers, nor for that matter to identify all outliers in the data file. The simplest approach is to ignore the extreme values in the calculation of means and ratios and in modeling work such as regression.

Other accommodation procedures are discussed in detail in the book by Barnett and Lewis (1984). Basically, they all involve the dampening of the effect of extreme values in analysis. The reader is referred to that book for specific details not mentioned here.

In general, Kammerman, et al. (1987) list the following as recommendations for dealing with extreme values:

- Remove extreme values from the imputation process. They should not be used as donors, nor should they be used in development of models for imputation;
- Subject all imputations to prescribed edit checks. Review imputations for consistency, and make any adjustments as necessary;
- Maintain an audit trail of the imputations and review results for reasonableness and consistency; and
- Use models whenever possible to verify the "reasonableness" of imputed values.

6.2 Criteria for Assessment of the Imputed Data

As noted earlier, the criteria for assessment of any imputation procedure should include an assessment of bias, variances, and mean square error. The following provides brief descriptions of each approach.

Bias. Estimating the bias for an estimate requires the knowledge of the true population values. To estimate the nonresponse bias quantitatively, it is necessary to obtain external validating data that are not available for incorporation into the survey estimation procedure. Thus, it is not possible to quantify the extent of nonresponse bias remaining in the survey estimates after nonresponse adjustment. However, methods are available to assist in evaluating the likelihood that nonresponse bias remains a serious problem, even after making the indicated types of weighting adjustments. For example, refer to Rowland, et al. (1993) and Klein, et al. (1994) for nonresponse bias analysis techniques.

Rancourt, et al. (1992) provides bias correction methods for survey estimates from data with imputed values for nonignorable nonresponse. Rancourt uses simple correction factors for the bias problem in cases where multiple imputation by regression is used.

Variances. A second approach is to construct the variances of the estimates of the population parameters of interest. Such variances must account for the impact of the imputation procedure used and should not treat the data as if they were a complete set of responses. It has been common practice to treat the imputed data as if they were true values and then compute the variances using standard formula. This approach could lead to serious underestimation of true variance when the proportion of missing items for a variable is relatively large.

Rao and Shao (1992) provide a method of jackknife variance estimation with survey data when a hot-deck imputation method is used for imputing the missing data. They demonstrate that the procedure is consistent as the sample size increases. Two other papers can also be used as a reference to the computation of variances for imputed data include *Linearized Variance Estimators under Imputation: An Empirical Investigation* (1993) and *Jackknife Variance Estimation under Imputation for Missing Data*, by Rao and Sitter (1992).

Mean Square Error. Once estimates of bias and variance are obtained, they can be combined into an estimate of the mean square error (MSE) of the population parameter of interest. Comparing the MSEs will provide some insight into the evaluation of the imputation procedures. The use of MSE criteria as the basis of choosing one imputation procedure over another provides a trade-off using only bias or variance as the basis for the choice.

7. REPORTING RESULTS OF ANALYSES OF ITEMS WITH HIGH NONRESPONSE RATES

The occurrence of missing data, either unit or item, creates a potential for bias. For certain important characteristics, the respondents may differ significantly from nonrespondents. In such instances, survey data may not adequately reflect or characterize the nonrespondents. The potential for bias is particularly great in the presence of high nonresponse rates. An evaluation of nonresponse bias is thus an important aspect of the analyses of data from a probability sample such as the NHANES III. It is also very important to guide the analysts on ways to deal with items with very high nonresponse rates.

When reporting results of analyses based on items with high nonresponse rates, analysts should point out obvious pitfalls in using these data and illustrate the kinds of conclusions that might be appropriate. The following provides a general set of guidelines for analysis of data with high nonresponse rates.

Taking Nonresponse into Account in Data Analyses

1. Inferences drawn from data with moderate or high nonresponse rates should depend on a number of factors. (Moderate or high nonresponse rates are defined in the context of potential uses of the data and the definitions of "high" may vary from survey to survey or among items being studied, as can be seen in the examples below. There is no purpose in a classification of levels of nonresponse that is independent of particular uses. It is the analyst's responsibility to determine whether the level of nonresponse for an item is serious, relative to the accuracy needed for a particular inference.) The main considerations are listed below:

    - How much margin of error can be tolerated? Obviously, if there is a need for extreme accuracy, an analyst should hesitate about drawing conclusions with a moderate amount of nonresponse. Fairly high errors can be tolerated for other kinds of statistics. For example, there are occasions in which nothing is known about a subject and even ball-park estimates are useful. In contrast, there are time series where the direction of change is important, and even small errors could reverse the true trend; a moderate amount of nonresponse should inhibit one from drawing conclusions in such cases. (However, if a very large increase, or decrease is shown in the time series, nonresponse would not be of much concern.)

- A special case of the situation described above is when comparisons are made between two domains. When there is a large or dramatic difference, an analyst can be reasonably confident in assuming the direction of the difference is real, even if a high level of nonresponse makes the amount of the difference somewhat uncertain. When the difference is small (even if outside of sampling error), an analyst should draw the conclusion that it is not clear whether there is a real difference between the domains.

- Are there symmetric risks to the U.S. population in making inferences that may be wrong? For example, in a case in which the hypothesis appears to be confirmed by the data and is accepted, the result would lead to preventative measures that could save lives. If it turns out that the hypothesis was wrong, no lives would be lost but a moderate amount of money would have been spent unnecessarily. In such a case, it would be reasonable for an analyst to lean in the direction of accepting such hypotheses, even if the nonresponse rates were high. (An illustration of such nonsymmetric risks is the conclusion drawn in the analyses of earlier phases of NHANES that poor black children had much higher levels of lead in their blood than other groups.)

- Are there *a priori* reasons to believe that nonrespondents may be very different from respondents? Have there been previous studies on the subject, or could analysis of the NHANES data file shed some light on the subject? For example, in determining whether to implement a survey on HIV infection several years ago, an important consideration was the possibility that a high proportion of drug abusers and the gay population would refuse to cooperate, possibly because of uncertainty about confidentiality assurances or for other reasons. This would have introduced major biases in the results.

- Do the conclusions implied by the statistics agree with results from other studies or contradict them? Somewhat lower standards are usually permissible when the results are used to confirm other studies.

2. Even if results are so uncertain that an analyst, or NCHS, would hesitate to draw any clear conclusion, would they have important implications if they were true? If so, the analyst may want to consider the possibility of describing tentative conclusions and recommend that other, more tightly controlled, studies be carried out.

3. Whenever the data are used, the analyst has the responsibility of pointing out the caveats from nonresponse in the reports, just as limitations from sampling error are described. Although possible nonresponse effects cannot be quantified in the same way as sampling errors, analysts should point out potential effects on the estimates when nonresponse is fairly high.

Bibliography

Barnett, V. and Lewis, T., 1984, Outliers in Statistical Data, John Wiley & Sons, New York.

Chapman, D., 1991, ÒReporting Complete Nonresponse in a Sample SurveyÓ, Proceedings of the Section on Survey Research Methods, American Statistical Association, 266-271.

Efron, B., 1994, ÒMissing Data, Imputation, and the Bootstrap,Ó Journal of the American Statistical Association, Vol. 89, No. 426, 463-474.

Ezzati, T., Khare, M., 1992. ÒNonresponse Adjustment in a National Health SurveyÓ, Proceedings of the Section on Survey Research Methods, American Statistical Association, 339-344.

Ezzati, T., et al., 1993, A Comparison of Imputation Techniques in the Third National Health and Nutrition Examination Survey, National Center for Health Statistics, Hyattsville, Maryland.

Fahimi, M., et al., 1992, ÒSerial Imputation of NHANES III with Mixed Regression and Hot-Deck Techniques,Ó Westat, Inc., Rockville, Maryland.

Flegel, K., et al., 1986, ÒPredictive Models for Skinfold Measurements in National Survey Data. Unpublished Manuscript.

Ford, B., ÒAn Overview of Hot-Deck Procedures,Ó Incomplete Data in Sample Surveys, Vol 3: Procedures of the Symposium, Chapter 14, Madow WG and Olkin I, eds., Academic Press, New York.

Groves, R. and Couper, M., 1992. ÒCorrelates of Nonresponse in Personal Visit Survey,Ó Proceedings of the Section on Survey Research Methods, American Statistical Association, 102-111.

Hidiroglou, M. et al., 1993, ÒA Framework for Measuring and Reducing Nonresponse in Surveys,Ó Survey Methodology, Statistics Canada, Vol 19, No.1, 81-84.

Kalton, G., 1983, Compensation For Missing Survey Data Institute for Social Research, The University of Michigan, Ann Arbor, Michigan.

Kammerman, L.A., Nisselson, H., and Schreiber, G., 1987, ÒEvaluation of HANES Nonresponse BiasÓ, Contract No. 282-85-2104, Social & Scientific Systems, Inc., Rockville, Maryland, and Westat, Inc., Rockville, Maryland.

Khare, M. et al., 1993, ÒMultiple Imputation of NHANES III,Ó National Center for Health Statistics, Hyattsville, Maryland.

Klein, R. et al. 1994, ÒRacial/Ethnic Differences in Age-Related Maculopathy: The Third National Health and Nutrition Examination Survey,Ó Contract No. 200-89-7004, National Center for Health Statistics, Hyattsville, Maryland.

Kott, P., 1994, ÒA Note on Handling Nonresponse in Sample Surveys,Ó Journal of the American Statistical Association, Vol. 89, No. 426, 693-696.

Kovar, M., 1984, ÒImputation in Sample Surveys: The Effect on Small Domain Estimates,Ó Proceedings of the Section on Survey Research Methods, American Statistical Association.

Krzanowski, W., 1982, ÒMixtures of Continuous and Categorical Variables In Discriminant Analysis: A Hypothesis Testing Approach,Ó Biometrics 38, 991-1002.

Little, R. and Schluchter, M., 1985, ÒMaximum-likelihood Estimation for Mixed Continuous and Categorical Data with Missing Values,Ó Biometrika, 72, 492-512.

Looker, A., et al., 1992, ÒAssessing Hip Fracture Risk in a Population-based Health Survey: The NHANES III Osteoporosis Component,Ó Aging Clin. Exp. Res., Vol. 4, 53-60.

Melton, L., et al., 1993, ÒNonresponse Bias in Studies of Diabetic Complications: The Rochester Diabetic Neuropathy Study,Ó Journal of Clinical Epidemiology, Vol. 46, No. 4, 341-348.

Naus, J., 1975, Data Quality Control and Editing, Marcel Dekker, Inc., New York.

Platek, R. and Gray, G., 1986, ÒOn the Definitions of Response Rates,Ó Survey Methodology, Vol. 12, No.1, 17-27.

Rancourt, E., et al. 1992, ÒBias Corrections for Survey Estimates From Data With Imputed Values for Nonignorable Nonresponses,Ó Statistics Canada.

Rao, J. and Shao, J., 1992, ÒJackknife Variance Estimation With Survey Data Under Hot Deck Imputation,Ó Biometrika, Vol. 79, No. 4, 811-822.

Rao, J. 1993, ÒJackknife Variance Estimation with Imputed Survey Data,Ó Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada.

Rao, J., 1993, ÒLinearized Variance Estimates Under Imputation: An Empirical Investigation,Ó Technical Report, Business Survey Methods Division, Statistics Canada, Ottowa, Ontario, Canada.

Rao, J. and Sitter, R., 1992, ÒJackknife Variance Estimation Under Imputation for Missing Survey Data,Ó Technical Report, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa.

Rowland, M., Forthofer, R., 1993, ÒInvestigation of Nonresponse Bias: Hispanic Health and Nutrition Examination Survey,Ó Vital and Health Statistics, Series 2: Data Evaluation and Methods Research, No. 119, National Center for Health Statistics, Hyattsville, Maryland 7-8.

Rubin, D., 1977, ÒFormalizing Subjective Nations About the Effect of Nonrespondents in Sample Survey,Ó The Journal of the American Statistical Association, Vol. 72, No. 359, 538-543.

Rubin, D., 1978, ÒMultiple Imputations in Sample Surveys- A Phenomenological Bayesian Approach to Nonresponse,Ó Proceedings of the Section on Survey Research Methods, American Statistical Association, 20-34.

Rubin, D., 1993, ÒMultiple Imputation After Fifteen Years,Ó Harvard University, Cambridge, Massachusetts.

Schafer, J., 1992, ÒEvaluation of Imputation Methods,Ó Memorandum: Recommendations on Model for NHANES III Imputation Project, Datametrics Research, Inc., Waban, Massachusetts.

Schafer, J., 1991, ÒAlgorithms for Multiple Imputation and Posterior Simulation From Incomplete Multivariate Data With Ignorable Nonresponse,Ó Ph.D. Dissertation, Department of Statistics, Harvard University, Cambridge, Massachusetts.

Schafer, J. et al., 1993, ÒMultiple Imputation of Missing Data in NHANES III,Ó Proceedings of the 1993 Annual Research Conference, United States Bureau of the Census, Washington, DC.

Tennant, A. and Bradley, E., 1991, ÒInvestigating Nonresponse Bias in a Survey of Disablement in the Community: Implications for Survey Methodology,Ó Journal of Epidemiology and Community Health, 45: 247-25.