

Third National Health and Nutrition Examination Survey  
(NHANES III), 1988-94

NHANES III SECOND EXAM FILE DOCUMENTATION

Series 11, No. 3A

July 1999

## Table of Contents

Introduction . . . . .	
Guidelines for Data Users. . . . .	
Survey Description . . . . .	
Sample Design and Analysis Guidelines. . . . .	
Data Preparation and Processing Procedures . . . . .	
General References . . . . .	
General Information . . . . .	
Data File Index . . . . .	
Data File Item Descriptions, Codes, Counts, and Notes . . . . .	

## Introduction

The National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) collects, analyzes, and disseminates data on the health status of U.S. residents. The results of surveys, analyses, and studies are made known through a number of data release mechanisms including publications, mainframe computer data files, CD-ROMs (Search and Retrieval Software, Statistical Export and Tabulation System (SETS)), and the Internet.

The National Health and Nutrition Examination Survey (NHANES) is a periodic survey conducted by NCHS. The third National Health and Nutrition Examination Survey (NHANES III), conducted from 1988 through 1994, was the seventh in a series of these surveys based on a complex, multi-stage sample design. It was designed to provide national estimates of the health and nutritional status of the United States' civilian, noninstitutionalized population aged two months and older.

The following table summarizes the NHANES III data which are currently available on CD-ROM, including this release.

Table 1. Available NHANES III CD-ROMs

CD-ROM Name	Release Date	Size in Megabytes	Data Files / Description
NHANES III, 1988-94, Series 11, No. 3A, ASCII Version (this release)	July 1999	33	Second exam sample files for dietary recall, examination, laboratory, additional laboratory analytes and documentation
NHANES III, 1988-94, Series 11, No. 2A, ASCII Version	April 1998	407	Dietary recall (replacement), electrocardiography, laboratory (additional analytes), and vitamins/medicines data files and documentation
NHANES III, 1988-94, Series 11, No. 1, Revised SETS Version 1.22a	October 1997	285	Adult and youth household questionnaire, examination, and laboratory data files and documentation, plan and operation, analytic and reporting guidelines, weighting and estimation methodology, field operations, non-response bias
NHANES III, 1988-94, Series 11, No. 1A, ASCII Version	July 1997	454	Adult and youth household questionnaire, dietary recall, examination, and laboratory data files and documentation
NHANES III, 1988-94, Series 11, No. 1, SETS Version 1.22a *	July 1997	285	Adult and youth household questionnaire, examination, and laboratory data files and documentation

NHANES III Reference Manuals and Reports October 1996	October 1996	152	Plan and operation, analytic and reporting guidelines, weighting and estimation methodology, field operations, non-response bias
---	-----------------	-----	---

\* Do not use this CD-ROM It had technical problems and has been superseded by the revised SETS version 1.22a, Series 11, No. 1, released in October 1997.

This CD-ROM, Series 11 No. 3A, contains data obtained from a second exam of selected survey participants who had a primary exam. This release does not replace the previous NHANES III data releases series 11 Nos. 1A and 2A).

Table 2. Location of the interview and examination components in the NHANES III public use data files

Topic	Data File							
	HA	HY	EXAM	LAB	DIET	VMS	ECG	
Sample weights	X	X	X	X	.	.	X	
Age/race/sex	X	X	X	X	.	.	X	
Ethnic background	X	X	.	.	.	.	.	
Household composition	X	X	.	.	.	.	.	
Individual characteristics	X	X	.	.	.	.	.	
Health insurance	X	X	.	.	.	.	.	
Family background	X	X	.	.	.	.	.	
Occupation of family head	X	X	.	.	.	.	.	
Housing characteristics	X	X	.	.	.	.	.	
Family characteristics	X	X	.	.	.	.	.	
Orientation	X	X	.	.	.	.	.	
Health services	X	X	.	.	.	.	.	
Selected health conditions	X	X	X	.	.	.	.	
Diabetes questions	X	.	.	.	.	.	.	
High blood pressure and cholesterol questions	X	.	.	.	.	.	.	
Cardiovascular disease questions	X	.	.	.	.	.	.	

Musculoskeletal conditions	X	.	.	.	.	.	.
Physical functioning questions	X	.	.	.	.	.	.
Gallbladder disease questions	X	.	.	.	.	.	.

Table 2. (continued) Location of the interview and examination components in the NHANES III public use data files

Data File

Topic	HA	HY	EXAM	LAB	DIET	VMS	ECG
Kidney conditions	X	.	.	.	.	.	.
Respiratory and allergy questions	X	X	.	.	.	.	.
Diet questions	X	.	.	.	.	.	.
Food frequency	X	.	X	.	.	.	.
Vision questions	X	X	.	.	.	.	.
Hearing questions	X	X	.	.	.	.	.
Dental care and status	X	X	.	.	.	.	.
Tobacco	X	.	X	.	.	.	.
Occupation	X	.	.	.	.	.	.
Language usage	X	X	.	.	.	.	.
Exercise	X	.	.	.	.	.	.
Social support/residence	X	.	.	.	.	.	.
Vitamin/mineral/medicine usage	X	X	X	.	.	.	.
Blood pressure measurement	X	.	X	.	.	.	.
Birth	.	X	X	.	.	.	.
Infant feeding practices/diet	.	X	.	.	.	.	.
Motor and social development	.	X	.	.	.	.	.

Functional impairment	X	X	.	.	.	.	.
School attendance	.	X	.	.	.	.	.
Cognitive function	.	X	X	.	.	.	.

Table 2. (continued) Location of the interview and examination components in the NHANES III public use data files

Data File

Topic	HA	HY	EXAM	LAB	DIET	VMS	ECG
Alcohol and drug use	.	.	X	.	.	.	.
Reproductive health	.	.	X	.	.	.	.
Diagnostic interview schedule	.	.	X	.	.	.	.
Activity	.	.	X	.	.	.	.
Physician's examination	.	.	X	.	.	.	.
Height and weight	.	.	X	.	.	.	.
Body measurements	.	.	X	.	.	.	.
Dental examination	.	.	X	.	.	.	.
Allergy skin test	.	.	X	.	.	.	.
Audiometry	.	.	X	.	.	.	.
Tympanometry	.	.	X	.	.	.	.
WISC and WRAT	.	.	X	.	.	.	.
Spirometry	.	.	X	.	.	.	.
Bone densitometry	.	.	X	.	.	.	.
Gallbladder ultrasonography	.	.	X	.	.	.	.
Central nervous system function evaluation	.	.	X	.	.	.	.
Fundus photography	.	.	X	.	.	.	.
Physical function evaluation	.	.	X	.	.	.	.
Fasting questions	.	.	.	X	.	.	.

Table 2. (continued) Location of the interview and examination components in the NHANES III public use data files

Topic	Data File							
	HA	HY	EXAM	LAB	DIET	VMS	ECG	
Laboratory tests on blood and urine	.	.	.	X	.	.	.	
Total nutrient intakes	.	.	X	.	.	.	.	
Individual foods	.	.	.	.	X	.	.	
Combination foods	.	.	.	.	X	.	.	
Ingredients	.	.	.	.	X	.	.	
Prescription Medicines	X	X	.	.	.	X	.	
Vitamins and Minerals	X	X	.	.	.	X	.	
Electrocardiography	.	.	.	.	.	.	X	

#### Data File Definitions

- HA - Household Adult Data File
- HY - Household Youth Data File
- EXAM - Examination Data File
- LAB - Laboratory Data File and Second Laboratory Data File
- DIET - Dietary Recall Data Files
- VMS - Vitamin Mineral Supplement Data File
- ECG - Electrocardiography Data File

This document includes the documentation for the NHANES III Second Exam Variable Ingredients File and also contains a general overview of the survey and the use of the data files. The general overview includes five sections. The first section, entitled "Guidelines for Data Users," contains important information about the use of the data files. The second section, "Survey Description," is a brief overview of the survey plan and operation. The third section, "Sample Design and Analysis Guidelines," describes some technical aspects of the sampling plan and discusses some analytic issues particularly related to the use of data from complex sample surveys. The "Data Preparation and Processing Procedures" section describes the editing conventions and the codes used to represent the data. The last and fifth section, "General References," includes a reference list for the survey overview sections of the document.

Public Use Data Files for the third National Health and Nutrition Examination Survey will also be available from the National Technical Information Service (NTIS). A list of NCHS public use data tapes available for purchase from NTIS may be obtained from the Data Dissemination Branch at

NCHS. Information regarding a bibliography (on disk) of journal articles citing data from all the NHANES and the availability of NHANES III data in CD-ROM/SETS software format can be obtained from the Data Dissemination Branch at:

Data Dissemination Branch  
National Center for Health Statistics  
Room 1018  
6525 Belcrest Road  
Hyattsville, Maryland 20782

Phone: (301)458-4636

URL:<http://www.cdc.gov/nchswww>

NTIS can be contacted at:

NTIS - Computer Products Office  
5285 Port Royal Road  
Springfield, Virginia 22161  
(703) 487-4807

Copies of all NHANES III questionnaires and data collection forms are included in the Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94 (NCHS, 1994; U.S. DHHS, 1996). This publication, along with detailed information on NHANES procedures, interviewing, data collection, quality control techniques, survey design, nonresponse, and sample weighting can be found on the NHANES III Reference Manuals and Reports CD-ROM (U.S. DHHS, 1996). Information on how to order this CD-ROM is also available from the Data Dissemination Branch at NCHS at the address and telephone number given above.

#### NHANES III Second Exam Sample

The NHANES III Second Exam Sample was a sub-study of NHANES III, conducted for research purposes. These research files are intended to provide additional data for use with special statistical methods to improve estimates from the main survey data and for methodologic investigations. Following this description of the Second Exam Sample is information on the overall survey which is also relevant for the Second Exam Sample, including: general guidelines for data users, a description of the survey, sample design, analysis guidelines and a description of the data preparation and processing procedures.

#### Sample design and survey description

No statistical sampling design was applied for the second exam. However, a nonrandom sample of about five percent was obtained by selecting approximately 20

participants from the roughly 400 sample persons examined at each survey location. The following general guidelines were used by the MEC staff to select participants for the second exam:

- 1) select mainly adults, 2) half between the ages of 20-39 years, and half over 40 years; 3) select about half men and half women. The sample obtained consists of 2,603 persons, with 1,205 males (46 percent) and

1,398 females (54 percent).

Age group	2nd # of Exams	Percentage of 2nd Exams
< 12	212	8
12-19	231	9
20-39	809	31
40-59	578	22
> 60	773	30

The second exams were scheduled after the first or primary exams, when possible at the same time of day as the first exam. The second exams were conducted over the same time period as the primary exams for a particular survey location by the same MEC staff, although priority was given to scheduling and completing primary exams. The second exams were administered following the same protocols as for the primary exam, with the following exceptions: the food frequency questionnaire was not administered to adolescents 12-16 years; the WISC/WRAT was not administered to youths 6-16 years, and hand/knee x-rays were not re-administered on adults aged 60 and over.

#### Analytic Issues

Due to the research nature of these data, special caution should be used in analysis. All analyses should include thorough investigation of the potential selection bias of this small non-random sub-sample. Careful attention to identifying and evaluating differences in important characteristics (e.g., age and race-ethnicity) between the subsample and the main sample should be considered along with other issues.

The second exam data can be linked to the primary exam data and the household interview data using the unique identifier (SEQN). This is necessary to obtain the demographic data for the sample. NCHS recommends that the survey design variables (e.g., sample weights) not be linked with the second exam data, since the survey design variables were created for the full sample. There are no sample weights or other design variables specifically created for the second exam sample. There are weights labeled as "replicate...weight," but these are Fay's BRR Replicate Interview Weights. These weights are to be applied to the primary exam sample, with software which uses the balanced repeated replication (BRR) method. They should not be used with the Second Exam Sample.

Because the second exams were identical to the primary exams, with the exceptions noted above, the file structure for the second exams is the same as for the primary exam files. The variable nomenclature is the same with the following important distinction: the first or primary exam variable names have a 'p' in the third position while the second or "replicate" exam variable names have a 'r' in the third position (e.g., 'BMPWT' or 'BMRWT').

## GUIDELINES FOR DATA USERS

Please refer to the following important information before analyzing data.

### NHANES III Background Documents

- o The Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94, (NCHS, 1994; U.S. DHHS, 1996) provides an overview of the survey and includes copies of the survey forms.
- o The sample design, nonresponse, and analytic guidelines documents on the NHANES III Reference Manuals and Reports CD-ROM (U.S. DHHS, 1996) discuss the reasons that sample weights and the complex survey design should be taken into account when conducting any analysis.
- o Instruction manuals, laboratory procedures, and other NHANES III reference manuals on the NHANES III Reference Manuals and Reports CD-ROM (U.S. DHHS, 1996) are also available for further information on the details of the survey.

### Analytic Data Set Preparation

- o Most NHANES III survey design and demographic variables are found only on the Adult and Youth Household Data Files available on the first release. In preparing a data set for analysis, other data files must be merged with either or both of these files to obtain many important analytic variables.
- o All of the NHANES III public use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES III data files using this variable ensures that the appropriate information for each survey participant is linked correctly.
- o NHANES III public use data files do not have the same number of records on each file. The Household Questionnaire Files (divided into two files, Adult and Youth) contain more records than the Examination Data File because not everyone who was interviewed completed the examination. The Laboratory Data File contains data only for persons aged one year and older. The Individual Foods Data File based on the dietary recall has multiple records for each person rather than the one record per sample person contained in the other data files.
- o For each data file, SAS program code with standard variable names and labels is provided as separate text files on the CD-ROM that contains the data files. This SAS program code can be used to create a SAS data set from the data file.
- o Modifications were made to items in the questionnaires, laboratory, and examination components over the course of the survey; as a result, data may not be available for certain variables for the full six years. In addition, variables may differ by phase since some changes were implemented between phases. Users are encouraged to read the Notes

sections of this document carefully for information about changes.

- o Extremely high and low values have been verified whenever possible, and numerous consistency checks have been performed. Nonetheless, users should examine the range and frequency of values before analyzing data.
- o Some data were not ready for release at the time of this publication due to continued processing of the data or analysis of laboratory specimens. A listing of those data are available in the general information section of each data file.
- o Confidential and administrative data are not being released to the public. Additionally, some variables have been recoded to help protect the confidentiality of the survey participants. For example, all age-related variables were recoded to 90+ years for persons who were 90 years of age and older.
- o Some variable names may differ from those used in the Phase 1 NHANES III Provisional Data Release and some variables included in the Phase 1 provisional release may not appear on these files.
- o Although the data files have been edited carefully, errors may be detected. Please notify NCHS staff (301-458-4636) of any errors in the data file or the documentation.

#### Analytic Considerations

- o NHANES III (1988-94) was designed so that the survey's first three years, 1988-91, its last three years, 1991-94, and the entire six years were national probability samples. Analysts are encouraged to use all six years of survey results.
- o Sample weights are available for analyzing NHANES III data. One of the following three sample weights will be appropriate for nearly all analyses: interviewed sample final weight (WTPFQX6), examined sample final weight (WTPFEX6), and mobile examination center (MEC)- and home-examined sample final weight (WTPFH6). Choosing which of these sample weights to use in any analysis depends on the variables being used. A good rule of thumb is to use "the least common denominator" approach. In this approach, the user checks the variables of interest. The variable that was collected on the smallest number of persons is the "least common denominator," and the sample weight that applies to that variable is the appropriate one to use for that analysis. For more detailed information, see the Analytic and Reporting Guidelines for NHANES III (U.S. DHHS, 1996).

#### Referencing or Citing NHANES III Data

- o In publications, please acknowledge NCHS as the original data source. For instance, the reference for the NHANES III Laboratory Data File on this CD-ROM is:

U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. Third National Health and Nutrition

Examination Survey, 1988-1994, NHANES III Second Laboratory Data File (CD-ROM, Series 11, No. 3A). Hyattsville, MD.: Centers for Disease Control and Prevention, 1999.

- o Please place the acronym "NHANES III" in the titles or abstracts of journal articles and other publications in order to facilitate the retrieval of such materials in bibliographic searches.

## SURVEY DESCRIPTION

The third National Health and Nutrition Examination Survey (NHANES III) was the seventh in a series of large health examination surveys conducted in the United States beginning in 1960. Three of these surveys, the National Health Examination Surveys (NHES), were conducted in the 1960's (NCHS, 1965; NCHS, 1967; NCHS, 1969). In 1970, an expanded nutrition component was added to provide data with which to assess nutritional status and dietary practices, and the name was changed to the National Health and Nutrition Examination Survey (Miller, 1973; Engel, 1978; McDowell, 1981). A special survey of Hispanic populations in the United States was conducted during 1982-1984 (NCHS, 1985).

The general structure of the NHANES III sample design was similar to that of the previous NHANES. All of the surveys used complex, multi-stage, stratified, clustered samples of civilian, noninstitutionalized populations. NHANES III was the first NHANES without an upper age limit; in fact, the age range for the survey was two months and older. A home examination option was employed for the first time in order to obtain examination data for very young children and for elderly persons who were unable to visit the mobile examination center (MEC). The home examination included only a subset of the components used in the full MEC examination since it would have been difficult to collect some types of data in a home setting. A detailed description of design specifications and copies of the data collection forms can be found in the Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-1994 (NCHS, 1994; U.S. DHHS, 1996).

NHANES III was conducted from October 1988 through October 1994 in two phases, each of which comprised a national probability sample. The first phase was conducted from October 18, 1988, through October 24, 1991, at 44 locations. The second phase was conducted from September 20, 1991, through October 15, 1994, at 45 different locations. In NHANES III, 39,695 persons were selected over the six years; of those, 33,994 (86%) were interviewed in their homes. All interviewed persons were invited to the MEC for a medical examination. Seventy-eight percent (30,818) of the selected persons were examined in the MEC, and an additional 493 persons were given a special, limited examination in their homes.

Data collection began with a household interview. Several questionnaires were administered in the household: Household Screener Questionnaire, Family Questionnaire, Household Adult Questionnaire, and Household Youth Questionnaire.

At the MEC, an examination was performed, and five automated questionnaires or interviews were administered: MEC Adult Questionnaire, MEC Youth Questionnaire, MEC Proxy Questionnaire, 24-Hour Dietary Recall, and Dietary Food Frequency (ages 12-16 years). The health examination component included a variety of tests and procedures. The examinee's age at the time of the interview and other factors determined which procedures were administered. Blood and urine specimens were obtained, and a number of tests and measurements were performed including body measurements, spirometry, fundus photography, x-rays, electrocardiography, allergy and glucose tolerance tests, and ultrasonography. Measurements were taken of bone

density, hearing, and physical, cognitive, and central nervous system

functions. A physician performed a limited standardized medical examination and a dentist performed a standardized dental examination. While some of the blood and urine analyses were performed in the MEC laboratory, most analyses were conducted elsewhere by contract laboratories.

A home examination was conducted for those sample persons aged 2-11 months and aged 20 years or older who were unable to visit the mobile examination center. The home examination consisted of an abbreviated version of the tests and interviews performed in the MEC. Depending on age of the sample person, the components included body measurements, blood pressure, spirometry, venipuncture, physical function evaluation, and a questionnaire to inquire about infant feeding, selected health conditions, cognitive function, tobacco use, and reproductive history.

## SAMPLE DESIGN AND ANALYSIS GUIDELINES

### Sample Design

The general structure of the NHANES III sample design is the same as that of the previous NHANES. Each of these surveys used a stratified, multi-stage probability design. The major design parameters of the two previous NHANES and the special Hispanic HANES, as well as NHANES III, have been previously summarized (Miller, 1973; McDowell, 1981; NCHS, 1985; NCHS, 1994). The NHANES III sample was designed to be self-weighting within a primary sampling unit (PSU) for subdomains (age, sex, and race-ethnic groups). While the sample was fairly close to self-weighting nationally for each of these subdomain groups, it was not representative of the total population, which includes institutionalized, non-civilian persons that were outside the scope of the survey.

The NHANES III sample represented the total civilian, noninstitutionalized population, two months of age or over, in the 50 states and the District of Columbia of the United States. The first stage of the design consisted of selecting a sample of 81 PSU's that were mostly individual counties. In a few cases, adjacent counties were combined to keep PSU's above a minimum population size. The PSU's were stratified and selected with probability proportional to size (PPS). Thirteen large counties (strata) were chosen with certainty (probability of one). For operational reasons, these 13 certainty PSU's were divided into 21 survey locations. After the 13 certainty strata were designated, the remaining PSU's in the United States were grouped into 34 strata, and two PSU's were selected per stratum (68 survey locations). The selection was done with PPS and without replacement. The NHANES III sample therefore consists of 81 PSU's or 89 locations.

The 89 locations were randomly divided into two groups, one for each phase. The first group consisted of 44 and the other of 45 locations. One set of PSU's was allocated to the first three-year survey period (1988-91) and the other set to the second three-year period (1991-94). Therefore, unbiased estimates (from the point of view of sample selection) of health and nutrition characteristics can be independently produced for both Phase 1 and Phase 2 as well as for both phases combined.

For most of the sample, the second stage of the design consisted of area segments composed of city or suburban blocks, combinations of blocks, or other area segments in places where block statistics were not produced in the 1980 Census. In the first phase of NHANES III, the area segments were used only for a sample of persons who lived in housing units built before 1980. For units built in 1980 and later, the second stage consisted of sets of addresses selected from building permits issued in 1980 or later. These are referred to as "new construction segments." In the second phase, 1990 Census data and maps were used to define the area segments. Because the second phase followed within a few years of the 1990 Census, new construction did not account for a significant part of the sample, and the entire sample came from the area segments.

The third stage of sample selection consisted of households and certain types of group quarters, such as dormitories. All households and eligible

group quarters in the sample segments were listed, and a subsample was designated for screening to identify potential sample persons. The subsampling rates enabled production of a national, approximately equal-probability sample of households in most of the United States with higher rates for the geographic strata with high Mexican-American populations. Within each geographic stratum, there was a nearly equal-probability sample of households across all 89 stands.

Persons within the sample of households or group quarters were the fourth stage of sample selection. All eligible members within a household were listed, and a subsample of individuals was selected based on sex, age, and race or ethnicity. The definitions of the sex, age, race or ethnic classes, subsampling rates, and designation of potential sample persons within screened households were developed to provide approximately self-weighting samples for each subdomain within geographic strata and at the same time to maximize the average number of sample persons per sample household. Previous NHANES indicated that this increased the overall participation rate. Although the exact sample sizes were not known until data collection was completed, estimates were made. Below is a summary of the sample sizes for the full six-year NHANES III at each stage of selection:

Number of PSU's	81
Number of stands (survey locations)	89
Number of segments	2,144
Number of households screened	93,653
Number of households with sample persons	19,528
Number of designated sample persons	39,695
Number of interviewed sample persons	33,994
Number of MEC-examined sample persons	30,818
Number of home-examined sample persons	493

More detailed information on the sample design and weighting and estimation procedures for NHANES III can be found in the Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94 (NCHS, 1994; U.S. DHHS, 1996) and in the Analytic and Reporting Guidelines: Third National Health and Nutrition Examination Survey (NHANES III), 1988-94 (U.S. DHHS, 1996).

#### Analysis Guidelines

Because of the complex survey design used in NHANES III, traditional methods of statistical analysis based on the assumption of a simple random sample are not applicable. Detailed descriptions of this issue and possible analytic methods for analyzing NHANES data have been described earlier (NCHS, 1985; Yetley, 1987; Landis, 1982; Delgado, 1990). Recent analytic and reporting guidelines that should be used for most NHANES III analyses and publications are contained in Analytic and Reporting Guidelines (U.S. DHHS, 1996). These recommendations differ slightly from those used by analysts for previous NHANES surveys. These suggested guidelines provide a framework to users for producing estimates that conform to the analytic design of the survey. All users are strongly urged to review these analytic and reporting guidelines before beginning any analyses of NHANES III data.

It is important to remember that this set of statistical guidelines is not absolute. When conducting analyses, the analyst needs to use his/her subject matter knowledge (including methodological issues) as well as

information about the survey design. The more one deviates from the original analytic categories defined in the sample design, the more important it is to evaluate the results carefully and to interpret the findings cautiously.

In NHANES III, 89 survey locations were randomly divided into two sets or phases, the first consisting of 44 and the other of 45 locations. One set of PSU's was allocated to the first three-year survey period (1988-91) and the other set to the second three-year period (1991-94). Therefore, unbiased national estimates of health and nutrition characteristics can be independently produced for each phase as well as for both phases combined. Computation of national estimates from both phases combined (i.e., total NHANES III) is the preferred option; individual phase estimates may be highly variable. In addition, individual phase estimates are not statistically independent. It is also difficult to evaluate whether differences in individual phase estimates are real or due to methodological differences. That is, differences may be due to changes in sampling methods or data collection methodology over time. At this time, there is no valid statistical test for examining differences between Phase 1 and Phase 2. Therefore, although point estimates can be produced separately for each phase, no test is available to test whether those estimates are significantly different from each other.

NHANES III is based on a complex, multi-stage probability sample design. Several aspects of the NHANES design must be taken into account in data analysis, including the sample weights and the complex survey design. Appropriate sample weights are needed to estimate prevalence, means, medians, and other statistics. Sample weights are used to produce correct population estimates because each sample person does not have the same probability of selection. The sample weights incorporate the differential probabilities of selection and include adjustments for noncoverage and nonresponse. A detailed discussion of nonresponse adjustments and issues related to survey coverage have been published (U.S. DHHS, 1996). With the large oversampling of young children, older persons, black persons, and Mexican-Americans in NHANES III, it is essential that the sample weights be used in all analyses. Otherwise, a misinterpretation of results is highly likely. Other aspects of the design that must be taken into account in data analyses are the strata and PSU pairings from the sample design. These pairings should be used to estimate variances and test for statistical significance. For weighted analyses, analysts can use special computer software packages that use an appropriate method for estimating variances for complex samples such as SUDAAN (Shah, 1995) and WesVarPC (Westat, 1996).

Although initial exploratory analyses may be performed on unweighted data using standard statistical packages and assuming simple random sampling, final analyses should be done on weighted data using appropriate sample weights. A summary of the weighting methodology and the type of sample weights developed for NHANES III is included in Weighting and Estimation Methodology (U.S. DHHS, 1996).

The purpose of weighting the sample data is to permit analysts to produce estimates of statistics that would have been obtained if the entire sampling frame (the United States) had been surveyed. Sample weights can be considered as measures of the number of persons the particular sample

observation represents. Weighting takes into account several features of the survey: the specific probabilities of selection for the individual domains that were oversampled as well as nonresponse and differences between

the sample and the total U.S. population. Differences between the sample and the population may arise due to sampling variability, differential undercoverage in the survey among demographic groups, and possibly other types of response errors, such as differential response rates or misclassification errors. Sample weighting in NHANES III was used to:

1. Compensate for differential probabilities of selection among subgroups (i.e., age-sex-race-ethnicity subdomains where persons living in different geographic strata were sampled at different rates);
2. Reduce biases arising from the fact that nonrespondents may be different from those who participate;
3. Bring sample data up to the dimensions of the target population totals;
4. Compensate, to the extent possible, for inadequacies in the sampling frame (resulting from omissions of some housing units in the listing of area segments, omissions of persons with no fixed address, etc.); and
5. To reduce variances in the estimation procedure by using auxiliary information that is known with a high degree of accuracy.

In NHANES III, the sample weighting was carried out in three stages. The first stage involved the computation of weights to compensate for unequal probabilities of selection (objective 1, above). The second stage adjusted for nonresponse (objective 2). The third stage used poststratification of the sample weights to Census Bureau estimates of the U.S. population to accomplish the third, fourth, and fifth objectives simultaneously. In NHANES III, several types of sample weights (see the sample weights table that follows) were computed for the interviewed and examined sample and are included in the NHANES III data file. Also, sample weights were computed separately for Phase 1 (1988-91), Phase 2 (1991-94), and total NHANES III (1988-94) to facilitate analysis of items collected only in Phase 1, only in Phase 2, and over six years of the survey. Three sets of pseudo strata and PSU pairings are provided to use with SUDAAN in variance estimation. Since NHANES III is based on a complex, multi-stage sample design, appropriate sample weights should be used in analyses to produce national estimates of prevalence and associated variances while accounting for unequal probability of selection of sample persons. For example, the final interview weight, WTPFQX6, should be used for analysis of the items or questions from the family or household questionnaires, and the final MEC examination weight, WTPFEX6, should be used for analysis of the questionnaires and measurements administered in the MEC. Furthermore, for a combined analysis of measurements from the MEC examinations and associated medical history questions from the household interview, the final MEC examination weight, WTPFEX6, should be used. We recommend using SUDAAN (Shah, 1995) to estimate statistics of interest and the associated variance. However, one can also use other published methods for variance estimation. Application of SUDAAN and alternative methods, such as the average design effect approach, balance repeated replication (BRR) methods, or jackknife methods for variance estimation, are discussed in Weighting and Estimation Methodology (U.S. DHHS, 1996).

Appropriate Uses of the NHANES III Sample Weights

Final interview weight, WTPFQX6

Use only in conjunction with the sample interviewed at home and with items collected during the household interview.

Final examination (MEC only) weight, WTPFEX6

Use only in conjunction with the MEC-examined sample and with interview and examination items collected at the MEC.

Final MEC+home examination weight, WTPFHX6

Use only in conjunction with the MEC+home-examined sample and with items collected at both the MEC and home.

Final allergy weight, WTPFALG6

Use only in conjunction with the allergy subsample and with items collected as part of the allergy component of the exam.

Final CNS weight, WTPFCNS6

Use only in conjunction with the CNS subsample and with items collected as part of the CNS component of the exam.

Final morning examination (MEC only) subsample weight, WTPFSD6

Use only in conjunction with the MEC-examined persons assigned to the morning subsample and only with items collected in the MEC exam.

Final afternoon/evening examination (MEC only) subsample weight, WTPFMD6

Use only in conjunction with the MEC-examined persons assigned to the afternoon/evening subsample and only with items collected in the MEC exam.

Final morning examination (MEC+home) subsample weight, WTPFHSD6

Use only in conjunction with the MEC- and home-examined persons assigned to the morning subsample and with items collected during the MEC and home examinations.

Final afternoon/evening examination (MEC+home) weight, WTPFHMD6

Use only in conjunction with the MEC- and home-examined persons assigned to the afternoon/evening subsample and with items collected during the MEC and home examinations.

#### DATA PREPARATION AND PROCESSING PROCEDURES

Automated data collection procedures for the survey were introduced in NHANES III. In the mobile examination centers, data for the interview and examination components were recorded directly onto a computerized data collection form. With the exception of a few independently automated systems, the system was centrally integrated. This operation allowed for ongoing monitoring of much of the data. Before the introduction of the computer-assisted personal interview (CAPI), the household questionnaire data were reviewed manually by field editors and interviewers. CAPI (1992-1994 only) questionnaires featured built-in edits to prevent entering inconsistencies and out-of-range responses. The multi-level data collection and quality control systems are discussed in detail in the Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-1994 (NCHS, 1994; U.S. DHHS, 1996). All interview, laboratory, and examination data were sent to NCHS for final processing.

Guidelines were developed that provided standards for naming variables, filling missing values and coding conventional responses, handling missing records, and standardizing two-part quantity/unit questionnaire variables. NCHS staff, assisted by contract staff, developed data editing specifications that checked data sets for valid codes, ranges, and skip pattern consistencies and examined the consistency of values between interrelated variables. Comments, collected in both interviews and examination components, were reviewed and recoded when possible. Responses to "Other" and "Specify" were recoded either to existing code categories or to new categories. The documentation for each data set includes notes for those variables that have been recoded and standardized and for those variables that differ significantly from what appears in the original data collection instrument. While the data have undergone many quality control and editing procedures, there still may be values that appear extreme or illogical. Values that varied considerably from what was expected were examined by analysts who checked for comments or other responses that might help to clarify unusual values. Generally, values were retained unless they could not possibly be true, in which case they were changed to "Blank but applicable." Therefore, the user must review each data set for extreme or inconsistent values and determine the status of each value for analysis.

Several editing conventions were used in the creation of final analytic data sets:

1. Standardized variables were created to replace all two-part quantity/unit questions using standard conversion factors. Standardized variables have the same name as the variable of the two-part question with an "S" suffix. For instance, MAPF18S (Months received WIC benefits) in the MEC Adult Questionnaire was created from the two-part response option to question F18, "How long did you receive benefits from the WIC program?," using the conversion factor 12 months

per year.

2. Recoded variables were created by combining responses from two or more like variables, or by collapsing responses to create a summary variable for the purpose of confidentiality. Recoded variables have the original variable name with an R suffix. For example, place of birth variable (HFA6X) in the Family Questionnaire was collapsed to a three level response category (U.S., Mexico, Other) and renamed HFA6XR. Generally, only the recoded variable has been included in the data file.
3. Fill values, a series of one or more digits, were used to represent certain specific conditions or responses. Below is a list of the fill values that were employed. Some of the fill values pertain only to questionnaire data, although 8-fill and blank-fill values are found in all data sets. Other fill values, not included in this list, are used to represent component-specific conditions.

6-fills = Varies/varied. (Questionnaires only)

7-fills = Fewer than the smallest number that could be reported within the question structure (e.g., fewer than one cigarette per day). (Questionnaires only)

8-fills = Blank but applicable/cannot be determined. This means that a respondent was eligible to receive the question, test, or component but did not because of refusal, lack of time, lack of staff, loss of data, broken vial, language barrier, unreliability, or other similar reasons.

9-fills = Don't know. This fill was used only when a respondent did not know the response to a question and said, "I don't know." (Questionnaires only)

Blank fills = Inapplicable. If a respondent was not eligible for a questionnaire, test, or component because of age, gender, or specific reason, the variable was blank-filled. In the questionnaire, if a respondent was not asked a question because of a skip-pattern, variables corresponding to the question were blank-filled. For examination or laboratory components, if a person was excluded by a defined protocol (e.g., screening exclusion questions) and these criteria are included in the data set, then the corresponding variables were blank-filled for that person. For home examinees, variables for examination components and blood tests not performed as part of the home examination protocol were blank-filled.

4. For variables describing discrete data, codes of zero (0) were used to mean "none," "never," or the equivalent. Value labels for which "0" is used include: "has not had," "never regularly," "still taking," or "never stopped using." Unless otherwise labeled, for variables containing continuous data, "zero" means "zero."
5. Where there are logical skip patterns in the flow of the questionnaire or examination component, the skip was indicated by placing the variable label of the skip destination in parentheses as part of the value label of the response generating the skip. For example, in the Physical Function Evaluation, the variable PFPWC (in wheelchair) has a

value label, "2 No (PFPSCOOT)" that means that the next item for persons not in a wheelchair would be represented by the variable, PFPSCOOT.

## Variable Nomenclature

A unique name was assigned to every NHANES III variable using a standard convention. By following this naming convention, the origin of each variable is clear, and there is no chance of overlaying similar variables across multiple components. Variables range in length from three to eight characters. The first two variable characters represent the topic (e.g., analyte, questionnaire instrument, examination component) and are listed below alphabetically by topic. For questionnaires administered in the household, the remainder of the variable name following the first two characters indicates the question section and number. For example, data for the response to the Household Adult Questionnaire question B1 are contained in the variable HAB1. For most laboratory and examination variables, as well as some other variables, a "P" in the third position refers to "primary" and the remainder of the variable name is a brief description of the item. For instance, in the Laboratory Data File, information on the length of time the person fasted before the first blood draw is contained in the variable PHPFAST. The variable PHPFAST was derived as follows: characters 1-2 (PH) refer to "phlebotomy," character 3 (P) refers to "primary," characters 4-8 (FAST) refer to an abbreviation for "fasting."

CODE	TOPIC
AT	Alanine aminotransferase (from biochemistry profile)
AM	Albumin (from biochemistry profile)
AP	Alkaline phosphatase (from biochemistry profile)
AL	Allergy skin test
AC	Alpha carotene
AN	Anisocytosis
TM	Antimicrosomal antibodies
TA	Antithyroglobulin antibodies
AA	Apolipoprotein (AI)
AB	Apolipoprotein (B)
AS	Aspartate aminotransferase (from biochemistry profile)
LA	Atypical lymphocyte
AU	Audiometry
BA	Band
BO	Basophil
BS	Basophilic stippling
BC	Beta carotene
BX	Beta cryptoxanthin
BL	Blast
BU	Blood urea nitrogen (BUN) (from biochemistry profile)
BM	Body measurements
BD	Bone densitometry
C1	C-peptide (first venipuncture)
C2	C-peptide (second venipuncture)
CR	C-reactive protein
UD	Cadmium
CN	Central nervous system function evaluation
CL	Chloride (from biochemistry profile)
CO	Cotinine
CE	Creatinine (serum)(from biochemistry profile)
UR	Creatinine (urine)

CODE	TOPIC
DM	Demographic
DE	Dental examination
MQ	Diagnostic interview schedule
DR	Dietary recall (total nutrient intakes)
EO	Eosinophil
EP	Erythrocyte protoporphyrin
FR	Ferritin
FB	Fibrinogen
RB	Folate (RBC)
FO	Folate (serum)
FH	Follicle stimulating hormone (FSH)
FP	Fundus photography
GG	Gamma glutamyl transferase (GGT) (from biochemistry profile)
GU	Gallbladder ultrasonography
GB	Globulin (from biochemistry profile)
G1	Glucose (first venipuncture)
G2	Glucose (second venipuncture)
SG	Glucose (from biochemistry profile)
GH	Glycated hemoglobin
GR	Granulocyte
C3	HCO <sub>3</sub> (Bicarbonate)(from biochemistry profile)
HD	HDL cholesterol
HP	Helicobacter pylori antibody
HT	Hematocrit
HG	Hemoglobin
AH	Hepatitis A antibody (HAV)
HB	Hepatitis B core antibody (anti-HBc)
SS	Hepatitis B surface antibody (anti-HBs)
SA	Hepatitis B surface antigen (HBsAg)
HC	Hepatitis C antibody (HCV)
DH	Hepatitis D antibody (HDV)
H1	Herpes 1 antibody
H2	Herpes 2 antibody
HX	Home examination (general)
HO	Homocysteine
HF	Household family questionnaire
HA	Household adult questionnaire
HQ	Household questionnaire variables (composite)
HS	Household screener questionnaire
HY	Household youth questionnaire
HZ	Hypochromia
I1	Insulin (first venipuncture)
I2	Insulin (second venipuncture)
UI	Iodine (urine)
FE	Iron
SF	Iron (from biochemistry profile)
LD	Lactate dehydrogenase (from biochemistry profile)
L1	Latex antibody
LC	LDL cholesterol (calculated)
PB	Lead
LP	Lipoprotein (a)
LH	Luteinizing hormone

CODE	TOPIC
LU	Lutein/zeaxanthin
LY	Lycopene
LM	Lymphocyte
MR	Macrocyte
MC	Mean cell hemoglobin (MCH)
MH	Mean cell hemoglobin concentration (MCHC)
MV	Mean cell volume (MCV)
PV	Mean platelet volume
MA	MEC adult questionnaire
MX	MEC examination (general)
FF	Dietary food frequency (ages 12-16 years)
MP	MEC proxy questionnaire
MY	MEC youth questionnaire
ME	Metamyelocyte
MI	Microcyte
MO	Monocyte
MN	Mononuclear cell
ML	Myelocyte
IC	Normalized calcium (derived from ionized calcium)
OS	Osmolality (from biochemistry profile)
PH	Phlebotomy data collected in MEC (e.g., questions)
PS	Phosphorus (from biochemistry profile)
PF	Physical function evaluation
PE	Physician's examination
PL	Platelet
DW	Platelet distribution width
PK	Poikilocytosis
PO	Polychromatophilia
SK	Potassium (from biochemistry profile)
PR	Promyelocyte
RC	Red blood cell count (RBC)
RW	Red cell distribution width (RDW)
RE	Retinyl esters
RF	Rheumatoid factor antibody
RU	Rubella antibody
WT	Sample weights
SE	Selenium
SI	Sickle cell
NA	Sodium (from biochemistry profile)
SH	Spherocyte
SP	Spirometry
SD	Survey design
TT	Target cell
TE	Tetanus
TH	Thyroid Stimulating Hormone (TSH)
T4	Thyroxine
TB	Total bilirubin (from biochemistry profile)
CA	Total calcium
SC	Total calcium (from biochemistry profile)
TC	Total cholesterol
CH	Total cholesterol (from biochemistry profile)
TI	Total iron binding capacity (TIBC)
TP	Total protein (from biochemistry profile)
TX	Toxic granulation

CODE	TOPIC
TO	Toxoplasmosis antibody
PX	Transferrin saturation
TG	Triglycerides
TR	Triglycerides (from biochemistry profile)
TY	Tympanometry
UA	Uric acid (from biochemistry profile)
UB	Urinary albumin
VU	Vacuolated cells
VR	Varicella antibody
VA	Vitamin A
VB	Vitamin B12
VC	Vitamin C
VD	Vitamin D
VE	Vitamin E
WC	White blood cell count (WBC)
WW	WISC/WRAT cognitive test

## GENERAL REFERENCES

- Delgado JL, Johnson CL, Roy I, Trevino FM. Hispanic Health and Nutrition Examination Survey: methodological considerations. *Amer J Pub Health* 80(suppl.):6-10. 1990.
- Engel A, Murphy RS, Maurer K, Collins E. Plan and operation of the HANES I Augmentation Survey of Adults 25-74 Years, United States, 1974-75. National Center for Health Statistics. *Vital Health Stat* 1(14). 1978.
- Freeman DH, Freeman JL, Brock DB, Koch GG. Strategies in the multivariate analysis of data from complex surveys II: an application to the United States National Health Interview Survey. *Int Stat Rev* 40(3):317-30. 1976.
- Khare M, Mohadjer LK, Ezzati-Rice TM, Waksberg J. An evaluation of nonresponse bias in NHANES III (1988-91). 1994 Proceedings of the Survey Research Methods section of the American Statistical Association. 1994.
- Landis JR, Lepkowski JM, Eklund SA, Stehouwer SA. A statistical methodology for analyzing data from a complex survey, the first National Health and Nutrition Examination Survey. National Center for Health Statistics. *Vital Health Stat* 2(92). 1982.
- McDowell A, Engel A, Massey JT, Maurer K. Plan and operation of the second National Health and Nutrition Examination Survey, 1976-80. National Center for Health Statistics. *Vital Health Stat* 1(15). 1981.
- Miller HW. Plan and operation of the Health and Nutrition Examination Survey, United States, 1971-1973. National Center for Health Statistics. *Vital Health Stat* 1(10a) and (10b). 1973.
- National Center for Health Statistics. Plan and initial program of the Health Examination Survey. *Vital Health Stat* 1(4). 1965.
- National Center for Health Statistics. Plan and operation of a health examination survey of U.S. youths 12-17 years of age. *Vital Health Stat* 1(8). 1969.
- National Center for Health Statistics. Plan and operation of the Hispanic Health and Nutrition Examination Survey, 1982-84. *Vital Health Stat* 1(19). 1985.
- National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. *Vital Health Stat* 1(32). 1994.
- National Center for Health Statistics. Plan, operation, and response results of a program of children's examinations. *Vital Health Stat* 1(5). 1967.
- Shah BV, Barnwell BG, Bieler GS. SUDAAN User's Manual: Software for Analysis of Correlated Data. Research Triangle Park, NC: Research Triangle Institute. Release 6.04. 1995.

Skinner CJ. Aggregated analysis: standard errors and significance tests.  
In: Skinner CJ, Holt D, Smith TMF, eds. Analysis of complex surveys. New  
York: John Wiley and Sons, Inc. 1989.

U.S. Department of Health and Human Services (DHHS). National Center for  
Health Statistics. NHANES III reference manuals and reports (CD-ROM).  
Hyattsville, MD: Centers for Disease Control and Prevention, 1996.  
Available from National Technical Information Service (NTIS), Springfield,  
VA. Acrobat .PDF format; includes access software: Adobe Systems, Inc.  
Acrobat Reader 2.1.

Westat, Inc. A User's Guide to WesVarPC. Rockville, MD. Westat, Inc.  
1996.

Yetley E, Johnson C. Nutritional applications of the Health and Nutrition  
Examination Surveys (HANES). Annu Rev Nutr 7:441-63. 1987.  
VARIABLE INGREDIENTS FILE

## Ingredient Information

The approach used to classify elemental foods and recipe foods was described in the NHANES III Second Exam Individuals Foods File (IFFSE) documentation. The Individual Foods File provides information about the component foods that were reported during the dietary interview. Many component foods were recipe foods. Recipe foods have ingredient records, some of which were variable ingredients. Variable ingredients were ingredients that the respondent could specify during the interview. Many types of ingredients were variable; the DDC System targeted sources of fat and sodium in food. Information about variable ingredients is reported in the NHANES III Second Exam Variable Ingredients File (VIFSE).

An important concept to understand when using the NHANES III foods files is that many foods can be component foods as well as ingredients of component foods, depending on their use. For example, margarine is included in the Individual Foods File as a component food. Two examples of margarine as a component food were margarine spread on bread and margarine added to mashed potatoes at the table; in both examples, margarine is a component food and has a food gram weight, USDA food code, possibly a brand name, nutrients, and so forth. The SAME margarine product may also be used as an ingredient in a recipe food such as homemade cookies. If a respondent reported eating homemade cookies, a probe as to the type of fat used in the recipe was asked during the interview. The Individual Foods File record for this food would report the type of cookie, the amount eaten, and nutrients for the cookie. The Variable Ingredients File (VIFSE) reports information about the margarine ingredient that was used to prepare the cookies; the VIFSE record includes a USDA food code for the margarine and possibly a brand name.

The VIFSE reports information pertaining to the variable ingredients for many recipe foods in the Individual Foods File. Only ingredients that the survey respondents were asked to specify are included in the VIFSE; other recipe ingredients that were not presented to respondents are excluded from the VIFSE. For example, if a respondent reported eating tuna salad, the variable ingredient probes included a probe as to the type of tuna fish, a probe as to whether the tuna was rinsed or drained, and a probe as to the type of mayonnaise or salad dressing used. The VIFSE provides information about these variable ingredients. On the other hand, the tuna salad may also have celery, pickles, and onion ingredients, but these ingredients were not variable ingredients. Again, the DDC System variable ingredient probes targeted sources of fat and sodium in the diet.

## Notes to Data Users:

### 1. Atypical Recipes and Modified Recipes

If a respondent reported that a food was prepared using what might be considered to be atypical or unusual ingredients, the dietary interviewers noted this. For example, some respondents used yogurt instead of mayonnaise to prepare salads. Additionally, respondents added unusual components to foods. The interviewers were instructed to note information about the ingredients that were used to prepare foods. NCHS evaluated the interviewers' notes and finalized the entries for all foods reported during the survey.

Respondents also modified recipes by omitting fat, substituting lower fat ingredients, using egg substitute products instead of whole eggs, etc. The recipe information that was recorded during the dietary interview was used to modify a standard recipe or locate another suitable recipe so that the food could be entered into the DDC System as a multi-component/combination food. Modified recipe food components included ingredient-type items such as flour and salt. The Individual Foods File reports the component level information. The multi-component foods have descriptive data in the Combination Foods File.

### 2. Default Ingredients in the VIFSE

Respondent-specified variable ingredients are reported in the VIFSE. If the respondent did not know anything about the ingredients that were used to prepare a food, the DDC System assigned a default ingredient automatically. There are default ingredients for home-prepared and commercially-prepared foods. For example, if a respondent reported eating brownies that were purchased at a bakery, and the fat ingredient information was unknown, the DDC System default for commercial brownies, purchased at a bakery would be assigned to the food. Similarly, ingredients that were used to prepare home-prepared foods also had default ingredient options. If a respondent ate a homemade meatloaf and could not specify the type of meat used, the DDC System assigned a default ingredient to the ingredient probes for the food. The variable ingredient default code (DRRVIDC) denotes when default variable ingredients were assigned to foods; the DRRVIDC codes have a descriptive text with them. For example, the type of fat used to fry a commercial food might be unknown so the DRRVIDC description might read: "fat used in frying unknown-commercially prepared".

### 3. No Ingredients Added

If the respondent stated that variable ingredients were not used to prepare a particular food, the VIFSE will include an ingredient record for the omitted ingredient(s), but the ingredient food code (DRRICODE) field is blank for each omitted ingredient. For these omitted ingredients, the ingredient identification code variable, DRRINGID, matches the primary file variable DRPINGID, which is linked to text descriptions in a look-up table called IDCODE. For example, cooked vegetables have fat and salt ingredient probes. If the respondent stated that no fat or salt was added in preparation, the VIFSE ingredient identification code (DRRINGID) descriptions for each variable ingredient will link to text information stating that no fat or salt was added in preparation.

Note: The same look-up table called "IDCODE" is used for Food Identification Code (DRPFID) and Ingredient Identification Code (DRPINGID) text descriptions.

As noted in other sections of this documentation, these codes match those in the Second Exam file, DRRFID and DRRINGID. DRPFID provides component level information as described in the Individual Foods File documentation; DRPINGID provides ingredient level information found in the Variable Ingredients File.

#### Summary

The VIFSE provides information about respondent-specified variable ingredients, including default variable ingredients. The VIFSE variables are sorted by sequence number (SEQN), meal number (DRRMN), food number (DRRFN), component number (DRRCN) and ingredient number (DRRIN). All variable ingredients that were added to a recipe food have a food code (DRRICODE); ingredient food codes in DRRICODE match those in the primary file variable DRPICODE, which is linked to food code descriptions in the look-up table "CODEBOOK". If a variable ingredient was reported by brand name, this information was included in the VIFSE; the brand name code, DRRCOMM, uses the same codes as the primary file variable, DRPCOMM, which is linked to a brand name description in a look-up table called "BRANDS". The ingredient identification code (DRRINGID) for all variable ingredients provides descriptive information about variable ingredients; this variable matches the primary file variable, DRPINGID, which is linked to the look-up table "IDCODE".

NHANES III Variable Ingredient Foods Data File Index  
from the Dietary Recall

Description	Variable Name	Positions
Respondent identification number .....	SEQN	1-5
Meal Number .....	DRRMN	6-7
Food Number .....	DRRFN	8-9
Component Number .....	DRRCN	10-11
Ingredient Number .....	DRRIN	12-13
Parent. USDA Food Code .....	DRRFCD	14-20
USDA Ingredient Food Code .....	DRRICD	21-27
Ingredient ID. Table look-up description ....	DRRINGID	28-31
Brand Id/Commercial Code .....	DRRCOMM	32-35
Variable ingredient Default Code .....	DRRVIDC	36-38

NHANES III Variable Ingredient Foods Data File  
from the Dietary Recall

Positions		Item description		Notes
SAS name	Counts	and code		
DATASET=VIFSE				
DOCUMENTATION DATE=06/22/99				
N=7653				
1-5		Sample person identification number		
SEQN	7653	00009-53594		
6-7		Meal number		See note
DRRMN	7653	01-11		
8-9		Food number		See note
DRRFN	7653	01-14		
10-11		Component number		See note
DRRCN	7653	01-13		
12-13		Ingredient number		See note
DRRIN	7653	01-36		
14-20		Component food code		See note
DRRFCODE	7653	1151200-9261301		
21-27		Ingredient food code		See note
DRRICODE	5696	0088270-9241071		
	1957	Blank		
28-31		Ingredient ID code		See note
DRRINGID	6835	0282-7278		
	818	Blank		
32-35		Brand or commercial ID code		See note
DRRCOMM	818	0746-6758		
	6835	Blank		

NHANES III Variable Ingredient Foods Data File  
from the Dietary Recall

SECOND EXAM DATA

Positions SAS name	Counts	Item description and code	Notes
36-38		Variable ingredient default code	See note
DRRVIDC	1	BKY Unknown-prepared in bakery	
	492	C Unknown-commercially prepared	
	111	H Unknown-prepared at home	
	41	MIX Unknown-prepared from commercial mix	
	2	RCP Unknown-prepared from recipe	
	152	UF Unknown if fat used	
	424	UK Unknown	
	1	UMH Unknown if prepared from mix or at home	
	328	US Unknown if salt added in preparation	
	314	UTF Unknown type of fat used	
	5787	Blank	

## NOTES

### DRRMN Meal number

Meal numbers in recalls that were complete and reliable (DRRSTAT=1) always begin with meal number=1; meal numbers increase by one for each consecutive meal or snack reported during the dietary interview. If a recall was coded reliable, but incomplete (DRRSTAT=2), the meal numbers may not be consecutive; information is reported for meals that were reported during the dietary interview. Meal numbers are not sorted by the time of day.

### DRRFN Food number

Every food has a food number. Foods are numbered within meals. If the recall was coded complete and reliable (DRRSTAT=1), the first food in each meal has a food number=1, and the other foods reported in the same meal are numbered consecutively. If the recall was coded reliable, but incomplete (DRRSTAT=2), the food numbers may not be consecutive; information is reported for all foods that were reported by the respondent.

### DRRCN Component number

Foods are comprised of one or more components. An example of a single component food is a slice of bread. A sandwich is an example of a multiple component food or combination food; in this example, the component foods consist of bread and sandwich filling components. If a recall was coded reliable and complete (DRRSTAT=1), all components are numbered consecutively within a given food; the component numbering sequence for the first food begins with component number=1 and increases by one for each additional component in the food. The numbering sequence is repeated for each additional food reported. If the recall was coded reliable, but incomplete (DRRSTAT=2), the component numbers may not be consecutive; information is reported for the components that were reported by the respondent.

### DRRIN Ingredient number

Component foods may have ingredient records associated with them. The first ingredient of a component has ingredient number=1; ingredient numbers increase by one for each additional ingredient of the component food. Variable ingredients are reported in the Variable Ingredients File.

### DRRFCODE: Component food code

A food code was assigned to all component foods reported in the survey. Most of the food codes that were assigned to the component foods were USDA Survey Nutrient Data Base food codes. A small number of NCC food codes were used when USDA food codes were not available; NCC nutrient values were used for components that were coded using NCC food codes. DRRFCODE uses codes

that match those in the primary file variable, DRPFCD, which is linked to text descriptions in the look-up table called "CODEBOOK".

DRRICODE Ingredient food code

An ingredient food code was assigned to all ingredients that were used to prepare recipe foods. Most of the ingredient food codes were USDA Survey Nutrient Data Base food codes; NCC food codes were used to code some ingredients. NCC nutrient values were used for the ingredients that were assigned NCC food codes. The codes in DRRICODE match those in the primary file variables, DRPICD, which is linked to text descriptions in a look-up table called "CODEBOOK".

Blank Blank values in the DRRICODE field denote that one or more variable ingredients were omitted during food preparation. The ingredient food identification code (DRRINGID) provides a unique food identification code for variable ingredients that were omitted during food preparation; DRRINGID matches the primary file variable DRPINGID, which is linked to text descriptions in the look-up table called "IDCODE".

DRRINGID Ingredient ID code

A unique 4-digit code that provides additional descriptive information about the ingredients that were used to prepare recipe foods. All ingredient ID codes in DRRINGID match those in the primary file variable, DRPINGID, which has a corresponding text description found in the look-up table called "IDCODE".

DRRCOMM Brand ID or fast food code

All brand name and fast food restaurant items reported during NHANES III were assigned a 5-digit DRRCOMM. DRRCOMM codes match those in the primary file variable, DRPCOMM, which is linked to a table of commercial foods text descriptions in the look-up table called "BRANDS".

DRRVIDC Variable ingredient default code

The DDC System assigned variable ingredients to foods automatically when the respondent was unable to specify the ingredients that were used to prepare recipe foods. Each DRRVIDC has a text description associated with it in the data file documentation.