MEPS HC-020:
1997 Full Year
Consolidated Data File

May 2001

Agency for Healthcare Research and Quality
Center for Cost and Financing Studies
2101 East Jefferson Street, Suite 501
Rockville, MD 20852
(301) 594-1406

Table of Contents

A. Data Use Agreement A-1

B. Background B-1
1.0 Household Component B-1
2.0 Medical Provider Component B-2
3.0 Insurance Component B-3
4.0 Nursing Home Component B-3
5.0 Survey Management B-4

C. Technical and Programming Information C-1
1.0 General Information C-1
2.0 Data File Information C-1
2.1 Codebook Structure C-2
2.2 Reserved Codes C-2
2.4 Variable Naming C-3
2.5 File Contents C-3
2.5.1 Survey Administration Variables C-3
2.5.2 Navigating the MEPS Data with Information on Person Disposition Status C-13
2.5.3 Geographic Variables C-18
2.5.4 Demographic Variables C-19
2.5.5 Income and Tax Filing Variables C-24
2.5.5.1 Income Top-Coding C-27
2.5.5.2 Poverty Status C-27
2.5.6 Employment Variables C-28
2.5.7 Health Insurance Variables C-34
2.5.8 Health Status Variables C-37
2.5.8.1 Perceived Health Status and ADL and IADL Limitations C-38
2.5.8.2 Functional and Activity Limitations C-39
2.5.8.3 Vision Problems C-42
2.5.8.4 Hearing Problems C-43
2.5.8.5 Any Limitation Rounds 3, 4, and 5 (Panel 1) / Rounds 1, 2, and 3 (Panel 2) C-44
2.5.8.6 Children's Health Status C-44
2.5.9 Utilization, Expenditures and Source of Payment Variables (TOTTCH97-RXOSR97) C-47
2.5.9.1 Expenditures Definition C-48
2.5.9.2 Utilization and Expenditure Variables by Type of Medical Service C-52
3.0 Survey Sample Information C-58
3.1 Sample Design and Response Rates C-58
3.2 Person Level Estimation using this MEPS PUF C-61
3.3 Family Level Estimation Using this MEPS PUF C-63

D. Variable-Source Crosswalk D-1

Appendix 1: Estimation Procedures in the Household Component of the 1996 Medical Expenditure Panel Survey A1-1

Appendix 2: Sample Design of the 1997 Medical Expenditure Panel Survey: Household Component A2-1

Appendix 3: Summary of Utilization and Expenditure Variables by Health Service Category A3-1

A. Data Use Agreement

Individual identifiers have been removed from the micro-data contained in the files on this CD-ROM. Nevertheless, under sections 308 (d) and 903 (c) of the Public Health Service Act (42 U.S.C. 242m and 42 U.S.C. 299 a-1), data collected by the Agency for Health Care Policy and Research (AHCPR) and /or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which it was supplied; any effort to determine the identity of any reported cases, is prohibited by law.

Therefore in accordance with the above referenced Federal Statute, it is understood that:

1. No one is to use the data in this data set in any way except for statistical reporting and analysis; and

2. If the identity of any person or establishment should be discovered inadvertently, then (a) no use will be made of this knowledge, (b) The Director Office of Management AHCPR will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHCPR, and (d) no one else will be informed of the discovered identity.

3. No one will attempt to link this data set with individually identifiable records from any data sets other than the Medical Expenditure Panel survey or the National Health Interview Survey.

By using this data you signify your agreement to comply with the above stated statutorily based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates 18 U.S.C. 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

The Agency for Health Care Policy and Research requests that users cite AHCPR and the Medical Expenditure Panel Survey as the data source in any publications or research based upon these data.

B. Background

This documentation describes one in a series of public use files from the Medical Expenditure Panel Survey (MEPS). The survey provides a new and extensive data set on the use of health services and health care in the United States.

The Medical Expenditure Panel Survey (MEPS) is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. MEPS also includes a nationally representative survey of nursing homes and their residents. MEPS is cosponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS).

MEPS comprises four component surveys: the Household Component (HC), the Medical Provider Component (MPC), the Insurance Component (IC), and the Nursing Home Component (NHC). The HC is the core survey, and it forms the basis for the MPC sample and part of the IC sample. The separate NHC sample supplements the other MEPS components. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

MEPS is the third in a series of national probability surveys conducted by AHRQ on the financing and use of medical care in the United States. The National Medical Care Expenditure Survey (NMCES, also known as NMES-1) was conducted in 1977, the National Medical Expenditure Survey (NMES-2) in 1987. Beginning in 1996, MEPS continues this series with design enhancements and efficiencies that provide a more current data resource to capture the changing dynamics of the health care delivery and insurance system.

The design efficiencies incorporated into MEPS are in accordance with the Department of Health and Human Services (DHHS) Survey Integration Plan of June 1995, which focused on consolidating DHHS surveys, achieving cost efficiencies, reducing respondent burden, and enhancing analytical capacities. To accommodate these goals, new MEPS design features include linkage with the National Health Interview Survey (NHIS), from which the sampled households for the MEPS HC are drawn, and continuous longitudinal data collection for core survey components. The MEPS HC augments NHIS by selecting a sample of NHIS respondents, collecting additional data on their health care expenditures, and linking these data with additional information collected from the respondents' medical providers, employers, and insurance providers.

1.0 Household Component

The MEPS HC, a nationally representative survey of the U.S. civilian non-institutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment.

The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a 2 1/2 year period. Employing computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched each year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures.

The sample of households selected for the MEPS HC is drawn from among respondents to the NHIS, conducted by NCHS. The NHIS provides a nationally representative sample of the U.S. civilian non-institutionalized population, with oversampling of Hispanics and blacks.

2.0 Medical Provider Component

The MEPS MPC supplements and validates information on medical care events reported in the MEPS HC by obtaining data directly from medical providers and pharmacies identified by household respondents. The MPC sample includes all hospitals, hospital physicians, home health agencies, and pharmacies reported in the HC. Also included in the MPC are office-based physicians:

- Providing care for HC respondents receiving Medicaid
- Identified through a 75 percent sample of HC households receiving care through an HMO (health maintenance organization) or managed care plan
- Identified through a 25 percent sample of the remaining HC households

Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents, including:

- Diagnoses coded according to ICD-9-CM (9th Revision, International Classification of Diseases) and DSM-IV (Fourth Edition, *Diagnostic and Statistical Manual of Mental Disorders*)
- Physician procedure codes classified by CPT-4 (Common Procedure Terminology, Version 4)
- Inpatient stay codes classified by DRGs (diagnosis-related groups)
- Prescriptions coded by national drug code (NDC), medication names, strength, and quantity dispensed
- Charges, payments, and the reasons for any difference between charges and payments

The MPC is conducted through telephone interviews and mailed survey materials.

3.0 Insurance Component

The MEPS IC collects data on health insurance plans obtained through employers, unions, and other sources of private health insurance. Data obtained in the IC include the number and types of private insurance plans offered, benefits associated with these plans, premiums, contributions by employers and employees, eligibility requirements, and employer characteristics.

Establishments participating in the MEPS IC are selected through four sampling frames:

- A list of employers or other insurance providers identified by MEPS HC respondents who report having private health insurance at the Round 1 interview.
- A Bureau of the Census list frame of private sector business establishments.
- The Census of Governments from Bureau of the Census.
- An Internal Revenue Service list of the self-employed.

To provide an integrated picture of health insurance, data collected from the first sampling frame (employers and insurance providers) are linked back to data provided by the MEPS HC respondents. Data from the other three sampling frames are collected to provide annual national and State estimates of the supply of private health insurance available to American workers and to evaluate policy issues pertaining to health insurance.

The MEPS IC is an annual panel survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone followup for nonrespondents.

4.0 Nursing Home Component

The 1996 MEPS NHC was a survey of nursing homes and persons residing in or admitted to nursing homes at any time during calendar year 1996. The NHC gathered information on the demographic characteristics, residence history, health and functional status, use of services, use of prescription medications, and health care expenditures of nursing home residents. Nursing home administrators and designated staff also provided information on facility size, ownership, certification status, services provided, revenues and expenses, and other facility characteristics. Data on the income, assets, family relationships, and care-giving services for sampled nursing home residents were obtained from next-of-kin or other knowledgeable persons in the community.

The 1996 MEPS NHC sample was selected using a two-stage stratified probability design. In the first stage, facilities were selected; in the second stage, facility residents were sampled, selecting both persons in residence on January 1, 1996, and those admitted during the period January 1 through December 31.

The sample frame for facilities was derived from the National Health Provider Inventory, which is updated periodically by NCHS. The MEPS NHC data were collected in person in three rounds of

data collection over a 1 ½-year period using the CAPI system. Community data were collected by telephone using computer-assisted telephone interviewing (CATI) technology. At the end of three rounds of data collection, the sample consists of approximately 815 responding facilities, 3,100 residents in the facility on January 1, and 2,200 eligible residents admitted during 1996.

5.0 Survey Management

MEPS data are collected under the authority of the Public Health Service Act. They are edited and published in accordance with the confidentiality provisions of this act and the Privacy Act. NCHS provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports and microdata files. Summary reports are released as printed documents and electronic files. Microdata files are released on CD-ROM and/or as electronic files. A catalog of all MEPS products released to date is provided on the AHRQ web site (http://www.meps.ahrq.gov/).

Printed documents and CD-ROMs are available through the AHRQ Publications Clearinghouse. Write or call:

AHRQ Publications Clearinghouse
Attn: (publication number)
P.O. Box 8547
Silver Spring, MD 20907
800/358-9295
410/381-3150 (callers outside the United States only)
888/586-6340 (toll-free TDD service; hearing impaired only)

Be sure to specify the AHRQ number of the document or CD-ROM you are requesting. Selected electronic files are available from the Internet on the AHRQ home page: http://www.meps.ahrq.gov/.

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Cost and Financing Studies, Agency for Healthcare Research and Quality, 2101 East Jefferson Street, Suite 500, Rockville, MD 20852 (301/594-1406).

C. Technical and Programming Information

1.0 General Information

This documentation describes the 1997 full-year consolidated data file from the Medical Expenditure Panel Survey Household Component (MEPS HC). Released as an ASCII file (with related SAS programming statements), this public use file provides information collected on a nationally representative sample of the civilian non-institutionalized population of the United States for calendar year 1997. This file consists of MEPS survey data obtained in Rounds 3, 4, and 5 of Panel 1 and Rounds 1, 2, and 3 of Panel 2 (i.e., the rounds for the MEPS panels covering calendar year 1997) and contains variables pertaining to survey administration, demographics, employment, health status, health insurance, income, and health care utilization and expenditures.

The following documentation offers a brief overview of the types and levels of data provided, and the content and structure of the files and the codebook. It contains the following sections:

- Data File Information
- Survey Sample Information
- Variable-Source Crosswalk (Section D)

Detailed information on sample design can be found in Appendixes 1 and 2. Appendix 3 provides an overview of the utilization and expenditure variables included in this file. The codebook is provided in a separate file, H20CB.PDF. A catalog of all MEPS products released to date and a matrix indicating the major MEPS data items on public use files that have been released to date are available on the AHRQ home page: http://www.meps.ahrq.gov/.

2.0 Data File Information

This public use dataset contains variable and frequency distributions for a total of 34,551 persons who participated in the MEPS Household Component of the Medical Panel Expenditure Survey in 1997. This count includes all household survey respondents who resided in eligible responding households. The persons were part of one of the two MEPS panels that collected data about 1997: Rounds 3, 4, and 5 of Panel 1 or Rounds 1, 2, and 3 of Panel 2. Of these persons, 32,636 were assigned a positive person level weight. Both weighted and unweighted frequencies are provided for each variable. In conjunction with the weight variable (WTDPER97) provided on this file, data for these persons can be used to make estimates for the civilian non-institutionalized U. S. population for 1997.

The records on this file can be linked to all other 1997 MEPS-HC public use data sets by the sample person identifier (DUPERSID). Panel 1 cases (PANEL97=1) can be linked back to the 1996 MEPS-HC public use data files. However, the user should be aware that at this time no weight is being provided to facilitate 2 year analysis of Panel 1 data.

2.1 Codebook Structure

The codebook and data file sequence lists variables in the following order:

- Unique person identifiers
- Geographic variables
- Demographic variables
- Income and Tax Filing variables
- Employment variables
- Health Insurance variables
- Health Status variables
- Utilization and Expenditure variables
- Weight and variance estimation variables

2.2 Reserved Codes

The following reserved code values are used:

VALUE		DEFINITION
-1	INAPPLICABLE	Question was not asked due to skip pattern
-2	DETERMINED IN PREVIOUS ROUND	Question was not asked in round because there was no change in employment status or no change in current main job since previous round.
-7	REFUSED	Question was asked and respondent refused to answer question.
-8	DK	Question was asked and respondent did not know answer
-9	NOT ASCERTAINED	Interviewer did not record the data
-10	HOURLY WAGE >= $43.75	Hourly wage was top-coded for confidentiality.

Note: A Reserved Code of "-3" had been used in the 1996 Full Year files to designate "No data in round"; the –3 code is no longer used in 1997. The analyst can used the INSCOPE variable (see Section 2.5.1 Survey Administration Variables) to determine whether or not a person would have data in a specific round.

2.3 Codebook Format

This codebook describes an ASCII data set and provides the following programming identifiers for each variable:

IDENTIFIER	DESCRIPTION
Name	Variable name (maximum of 8 characters)
Description	Variable descriptor (maximum 40 characters)
Format	Number of bytes
Type	Type of data: numeric (indicated by NUM) or character (indicated by CHAR)
Start	Beginning column position of variable in record
End	Ending column position of variable in record

2.4 Variable Naming

In general, variable names reflect the content of the variable, with an eight-character limitation. Edited variables end in an X, and are so noted in the variable label. The last two characters in round-specific variables denote the rounds of data collection, Round 3, 4, or 5 of Panel 1 and Round 1, 2, or 3 of Panel 2. Unless otherwise noted, variables that end in 97 represent status as of December 31, 1997.

Variables contained in this delivery were derived either from the questionnaire itself or from the CAPI. The source of each variable is identified in the section of the documentation entitled "Appendix D. Variable-Source Crosswalk." Sources for each variable are indicated in one of four ways: (1) variables derived from CAPI or assigned in sampling are so indicated; (2) variables derived from complex algorithms associated with re-enumeration are labeled "RE Section"; (3) variables that are collected by one or more specific questions in the instrument have those question numbers listed in the Source column; (4) variables constructed from multiple questions using complex algorithms are labeled "Constructed."

2.5 File Contents

2.5.1 Survey Administration Variables

The survey administration variables contain information related to conducting the interview, household and family composition, and person-level and RU-level status codes. Data for the survey administration variables were derived from the sampling process, the CAPI programs, or were computed based on information provided by the respondent in the re-enumeration section of the questionnaire. Most Survey Administration variables on this file are asked during every round of the MEPS interview. They describe data for Rounds 3/1, 4/2, and 5/3 status as of December 31, 1997. Variable names ending in "xy" represent variables relevant to Round "x" of Panel 1 or Round "y"

of Panel 2. For example, RULETR53 is a variable relevant to Round 5 of Panel 1 or Round 3 of Panel 2.

The December 31, 1997 variables were developed in two ways. Those used in the construction of eligibility, inscope, and the end reference date, were based on an exact date. The remaining variables were constructed using data from specific rounds, if available. If data were missing from the target round, but were available in another round, data from that other round were used in the variable construction. If no valid data were available during any round of data collection, an appropriate reserved code was assigned.

Dwelling Units, Reporting Units, and Families

The definitions of Dwelling Units (DUs) in the MEPS Household Survey are generally consistent with the definitions employed for the National Health Interview Survey. The dwelling unit ID (DUID) is a five-digit random ID number assigned after the case was sampled for MEPS. The person number (PID) uniquely identifies all persons within the dwelling unit. The variable DUPERSID is the combination of the variables DUID and PID.

A Reporting Unit (RU) is a person or group of persons in the sampled dwelling unit who are related by blood, marriage, adoption, foster care or other family association. Each RU was interviewed as a single entity for MEPS. Thus, the RU serves chiefly as a family-based "survey" operations unit rather than an analytic unit. Members of each RU within the DU are identified in the first three rounds by the round-specific variables RULETR31, RULETR42 and RULETR53. End-of-year status (as of December 31, 1997) is indicated by the RULETR97 variable. Regardless of the legal status of their association, two persons living together as a "family" unit were treated as a single reporting unit if they chose to be so identified. Examples of different types of reporting units are:

1. A married daughter and her husband living with her parents in the same dwelling unit constitute a single reporting unit.

2. A husband and wife and their unmarried daughter, age 18, who is living away from home while at college constitute two reporting units.

3. Three unrelated persons living in the same dwelling unit would each constitute a distinct reporting unit (a total of three reporting units)

Unmarried college students (less than 24 years of age) who usually live in the sampled household, (but were living away from home and going to school at the time of the Round 3/1 MEPS interview) were treated as a reporting unit separate from that of their parents for the purpose of data collection. The round-specific variables RUSIZE31, RUSIZE42, RUSIZE53 and the end-of-year status variable RUSIZE97 indicate the number of persons in each RU, treating each student as a single RU separate from their parents. Thus, students are not included in the RUSIZE count of their parents RU. However, for many analytic objectives, the student reporting units would be combined with their parents' reporting unit, treating the combined entity as a single family. Family identifier and size variables are described below and include students with their parent's reporting unit.

PANEL97 is a constructed variable used to specify the panel number for the interview. PANEL97 will indicate either Panel 1 or Panel 2 for each interview.

The round-specific variables FAMID31, FAMID42, FAMID53 and the end-of-year status variable FAMID97 identify a family (i.e., persons related to one another by blood, marriage, adoption, foster care, or self-identified as a single unit) for each round and as of December 31, 1998. The FAMID variables differ from RU only in that student reporting units are combined with their parent's reporting unit.

Two other family identifiers, FAMIDYR and CPSFAMID are provided on this file. The annualized family ID letter, FAMIDYR, identifies eligible members of the eligible annualized families within a DU. The CPSFAMID identifies eligible members of eligible CPS-like families at 12/31/1997. CPSFAMID represents a redefinition of MEPS families into families defined by the Current Population Survey (CPS). Some of the distinctions between CPS and MEPS defined families are that MEPS families include and CPS families do not include: non-married partners, foster children, and in-laws. These persons are considered as members of separate families for CPS-like families. The reason CPS-like families are defined is so that a poverty status classification variable consistent with established definitions of poverty can be assigned to the CPS-like families and used for weight poststratification purposes. In order to identify a person's family affiliation users must create a unique set of FAMID variables by concatenating the DU identifier and the FAMID variable. Instructions to create family estimates are described in section 3.3.

The round-specific variables FAMSZE31, FAMSZE42, FAMSZE53 and the end-of-year status variable FAMSZE97 indicate the number of persons associated with a single family unit after students are linked to their associated parent RUs for analytical purposes. Family-level analyses should use the FAMSZE variables.

Note that the variables RUSIZE31, RUSIZE42, RUSIZE53, RUSIZE97, FAMSZE31, FAMSZE42 FAMSZE53, and FAMSZE97 exclude persons who are ineligible for data collection (i.e., those where ELGRND31^= 1, ELGRND42 ^= 1, ELGRND53 ^= 1 or ELGRND97 ^= 1); analysts should exclude ineligible persons in a given round from all family-level analyses for that round.

The round-specific variables RURSLT31, RURSLT42, and RURSLT53 indicate the RU response status for each round. Users should note that the values for RURSLT31 differ from those for RURSLT42 and RURSLT53. The values for RURSLT31 include the following:

-1	Inapplicable
60	Complete with RU member
61	Complete with proxy--all RU members deceased
62	Complete with proxy--all RU members institutionalized or deceased
63	Complete with proxy, other
80	Entire RU merged with other RU
81	Entire RU deceased before 1/1/97
82	Entire RU is in military BEFORE 1/1/97
83	RU institutionalized before 1/1/97
84	Entire RU left U.S. before 1/1/97
85	RU ineligible before 1/1/97, multi-reason
86	RU ineligible, non-key NHIS study
87	Re-enumeration complete, no eligible RU member, Ineligible RU
88	Unavailable during field period
89	Too ill, No proxy
90	Physical/Mental incompetent, No proxy
91	Final Refusal
92	Final Breakoff
93	Unable to locate
94	Entire RU is military or left U.S. AFTER 1/1/97
95	RU member instituitionalized after 1/1/97, No proxy
96	RU member deceased after 1/1/97, No proxy
97	Re-enumeration complete, no RU member, Non-Response
98	RU moved too far away to interview
99	Final other Non-Response

The values for RURSLT42 and RURSLT53 include the following:

-1	Inapplicable
60	Complete with RU member
61	Complete with proxy--all RU members deceased
62	Complete with proxy--all RU members institutionalized or deceased
63	Complete with proxy, other
70	Entire RU merged with other RU
71	Re-enumeration complete, no eligible RU member, Ineligible RU
72	RU institutionalized in prior round; still institutionalized
81	Entire RU deceased before 1/1/97
82	Entire RU is in military BEFORE 1/1/97
83	RU institutionalized before 1/1/97
84	Entire RU left U.S. before 1/1/97
85	RU ineligible before 1/1/97, multi-reason
86	RU ineligible, non-key NHIS study
87	Language Barrier
88	Unavailable during field period
89	Too ill, No proxy
90	Physical/Mental incompetent, No proxy
91	Final Refusal
92	Final Breakoff
93	Unable to locate
94	Entire RU is military or left U.S. AFTER 1/1/97
95	RU member instituitionalized after 1/1/97, No proxy
96	RU member deceased after 1/1/97, No proxy
97	Re-enumeration complete, no RU member, Non-Response
98	RU moved too far away to interview
99	Final other Non-Response

Standard, or primary RUs are the original RUs from NHIS. All primary RUs are classified as standard RUs. A new RU is one created when members of the household leave the primary RU and are followed according to the rules of the survey. A student RU is an unmarried college student (under 24 years of age) that is considered a usual member of the household (but was living away from home while going to school) and was treated as a Reporting Unit (RU) separate from their parents RU for the purpose of data collection. RUCLAS97 was set based on the RUCLASS values from Rounds 3/1, 4/2, and 5/3. If the person was present in the responding RU in Round 5/3, then RUCLAS97 was set to RUCLAS53. If the person was not present in a responding RU in Round 5/3, but was present in Round 4/2, then RUCLAS97 was set to RUCLAS42. If the person was not present in either Rounds 4/2 or 5/3, but was present in Round 3/1, then RUCLAS97 was set to RUCLAS31. If the person was not linked to a responding RU during any round then RUCLAS97 was set to -9.

Reference Period Dates

The reference period is the period of time for which data were collected in each round for each person. The reference period dates were determined during the interview for each person by the CAPI program. The round-specific beginning reference period dates are included for each person. These variables include BEGRFM31, BEGRFD31, BEGRFY31, BEGRFM42, BEGRFD42, BEGRFY42, BEGRFM53, BEGRFD53, and BEGRFY53. The reference period for Round 1 for most persons began on January 1, 1997 and ended on the date of the Round 1 interview. For RU members who joined later in Round 1, the beginning Round 1 reference date was the date the person entered the RU. For all subsequent rounds, the reference period for most persons began on the date of the previous round's interview and ended on the date of the current round's interview. Persons who joined after the previous round's interview had their beginning reference date for the round set as the day they joined the RU.

The round specific ending reference period dates for Rounds 3/1, 4/2, and 5/3 as well as the end-of-year reference period end date variables are also included for each person. These variables include ENDRFM31, ENDRFD31, ENDRFY31, ENDRFM42, ENDRFD42, ENDRFY42, ENDRFM53, ENDRFD53, ENDRFY53, ENDRFM97, ENDRFD97, and ENDRFY97. For most persons in the sample, the date of the interview is the reference period end date. Note that the end date of the reference period is prior to the date of the interview if the person was deceased during the round, left the RU, or was institutionalized prior to that round's interview, or left the RU to join the military.

Reference Person Identifiers

The round specific variables REFPRS31, REFPRS42 and REFPRS53 and the end-of-year status variable REFPRS97 identify the reference person for Rounds 3/1, 4/2 and 5/3, and as of December 31, 1997. In general, the reference person is defined as the household member 16 years of age or older who owns or rents the home. If more than one person meets this description, the household respondent identifies one from among them. If the respondent was unable to identify a person fitting this definition, the questionnaire asked for the head of household and this person was then considered the reference person for that RU. This information was collected in the Re-enumeration section of the CAPI questionnaire.

Respondent Identifiers

The respondent is the person who answered the interview questions for the reporting unit (RU). The round specific variables RESP31, RESP42, and RESP53 and the end-of-year status variable RESP97 identify the respondent for Rounds 3/1, 4/2, and 5/3 and as of December 31, 1997. Only one respondent is identified for each RU. In instances where the interview was completed in more than one session, only the first respondent is indicated.

There are two types of respondents. The respondent can be either a RU member or a non-RU member proxy. The round specific variables PROXY31, PROXY42, PROXY53 and the end-of-year status variable PROXY97 identify the type of respondent for Rounds 3/1, 4/2, 5/3 and as of December 31, 1997.

Person Status

A number of variables describe the various components reflecting each person's status for each round of data collection. These variables provide information about a person's inscope status, Keyness status, eligibility status, and disposition status. These variables include: INSCOPE, INSCOP31, INSCOP42, INSCOP53, INSCOP97, KEYNESS, ELIGIBLE, ELGRND31, ELGRND42, ELGRND53, ELGRND97, PSTATS31, PSTATS42, and PSTATS53. These variables are set based on sampling information and responses provided in the re-enumeration section of the CAPI questionnaire.

Through the re-enumeration section of the CAPI questionnaire, each member of a reporting unit was classified as "key" or "non-key", "inscope" or "out-of-scope", and "eligible" or "ineligible" for MEPS data collection. To be included in the set of persons used in the derivation of MEPS person level estimates; a person had to be a member of the civilian non-institutionalized population for at least one day during 1997. Because a person's eligibility for the survey might have changed since the NHIS interview, a sampling re-enumeration of household membership was conducted at the start of each round's interview. Only persons who were "inscope" sometime during the year, "key", and responded for the full period in which they were inscope were assigned positive person level weights and thus are to be used in the derivation of person level national estimates from the MEPS.
Note: if analysts want to subset to infants born during 1997, then newborns should be identified using AGE97X = 0 rather than PSTATS = 51.

Inscope

A person was considered as inscope during a round if he or she was a member of the U.S. civilian, non-institutionalized population at some time during that round. The round specific variables INSCOP31, INSCOP42, and INSCOP53 indicate a person's inscope status for rounds 3/1, 4/2, and 5/3. INSCOP97 indicates a person's inscope status for the portion of round 5/3 that covers 1997. The values of these three variables taken in conjunction allow one to determine inscope status over time (for example, becoming inscope in the middle of a round, as would be the case for newborns). The

INSCOPE variable indicates whether a person was ever inscope during the calendar year 1997. INSCOP31, INSCOP42, INSCOP53, and INSCOP97 will contain the following values and corresponding label (for INSCOP97, "reference period" in the description below is the portion of Round 5/3 in 1997):

0 Incorrectly listed, or on NHIS roster but out-of-scope prior to January 1, 1997

1 Person is inscope for the whole reference period

2 Person is inscope at the start of the RU reference period, but not at the end of the RU reference period.

3 Person is not inscope at the start of RU reference period, but is inscope at the end of the RU reference period. (E.g., the person is inscope from the date the person joined the RU or the person was in the military in the previous round, but is no longer in the military in the current round.)

4 Person is inscope during the reference period, but neither at the reference start date nor on the reference end date. (E.g., Person leaves an institution, goes into community, and then dies.)

5 Person is out-of-scope for all of the reference period during which they are in an RU member (i.e. The person is in the military.)

6 Person is out-of-scope for the entire reference period and is not a member of the RU during this time period and was inscope and an RU member in an earlier round.

7 Person is not in an RU, joined in a later round (or joined RU after December 31, 1997 for INSCOP97)

8 RU Non-response and key persons who left an RU with no tracing info and so a new RU was not formed

9 Person is non-key or full time in the military, not a member of an RU during this time period, and was an RU member in an earlier round

Keyness

The term "keyness" is related to an individual's chance of being included in MEPS. A person is key if that person is linked for sampling purposes to the set of NHIS sampled households designated for inclusion in MEPS. Specifically, a key person was a member of an NHIS household at the time of the NHIS interview, or became a member of such a household after being out-of-scope at the time of the NHIS (examples of the latter situation include newborns and persons returning from military service, an institution, or living outside the United States).

A non-key person is one whose chance of selection for the NHIS (and MEPS) was associated with a household eligible but not sampled for the NHIS, and who later became a member of a MEPS reporting unit. MEPS data, (e.g., utilization and income) was collected for the period of time a non-

key person was part of the sampled unit to provide information for family level analyses. However, non-key persons who leave a sample household unaccompanied by a key, inscope member were not followed for subsequent interviews. Non-key individuals do not receive sample person-level weights and thus do not contribute to person level national estimates.

The variable KEYNESS indicates a person's keyness status. This variable is not round specific. Instead, it is set at the time the person enters MEPS, and the person's keyness status never changes. Once a person is determined to be key, that person will always be key.

It should be pointed out that a person might be key even though not part of the civilian, non-institutionalized portion of the U.S. population. For example, a person in the military may have been living with his or her civilian spouse and children in a household sampled for NHIS. The person in the military would be considered a key person for MEPS; however, such a person would not be eligible to receive a person-level sample weight if he or she was never inscope during 1997.

Eligibility

The eligibility of a person for MEPS pertains to whether or not data were to be collected for that person. All of the key, inscope, persons of a sampled RU were eligible for data collection. The only non-key persons eligible for data collection were those who happened to be living in a RU with at least one key, inscope person. Their eligibility continued only for the time that they were living with at least one such person. The only out-of-scope persons eligible for data collection were those who were living with key inscope persons, again only for the time they were living with such a person. Only military persons can meet this description (for example, a person on full time active duty military, living with a spouse who is key).

A person may be classified as eligible for an entire round or for some part of a round. For persons who are eligible for only part of a round (for example, persons may have been institutionalized during a round) data were collected for that person only for the period of time for which that person was classified as eligible. The round specific variables ELGRND31, ELGRND42, ELGRND53 and the end-of-year status variable ELGRND97 indicate a person's eligibility status for Rounds 3/1, 4/2 and 5/3 and as of December 31, 1997. The ELIGIBLE variable indicates if a person was ever eligible during the calendar year 1997.

Person Disposition Status

The round-specific variables PSTATS31, PSTATS42, and PSTATS53 indicate a person's response and eligibility status for each round of interviewing. The PSTATS variables indicate the reasons for either continuing data collection for a person or terminating data collection for each person in the MEPS. Using this variable, one could identify persons who moved during the reference period, died, were born, institutionalized or who were in the military. Analysts should note that PSTATS53 provides a summary for all of Round 5/3, including transitions that occurred after 1997. However, PSTATS53 is still a useful guide to following transitions that occur over time in the sample for 1997.

The following codes specify the value labels for the PSTATS variables.

-1	The person was not fielded during the round or the RU was non-response
0	Incorrectly listed in RU at NHIS -applies to MEPS Round 1 only
11	Person in original RU , not full time active military duty
12	Person in original RU, full time active military duty, out-of-scope for whole reference period.
13	Full time student living away from home, but associated with sampled RU
14	The person is full time active military duty during round and is inscope for part of the reference period and is in the RU at the end of the reference period
21	The person remains in a health care institution for the whole round - rounds 4/2 and 5/3 only
22	The person leaves a health care instnitution and rejoins the community - rounds 4/2 and 5/3 only
23	The person leaves a health care institution, goes into community and then dies - rounds 4/2 and 5/3 only
24	The person dies in a health care institution during the round (former RU member) - rounds 4/2 and 5/3 only
31	Person from original RU, dies during reference period
32	Went to health care institution during reference period
33	Went to non-healthcare institution during reference period
34	Moved from original RU, outside U.S. (not as student)
35	Moved from original RU, to a military facility while on full time active military duty
36	Went to institution (type unknown) during reference period
41	Moved from the original RU, to new RU within U.S. (new RUs include RUs originally classified as “Student RU” but which converted to “New RU”)
42	The person joins RU and is not full time military during round
43	The person's disposition as to why the person is not in the RU is unknown or the person moves and it is unknown whether the person moved inside or outside the U.S.
44	The person leaves an RU and joins an existing RU and is not both in the military and coded as inscope during the round
51	Newborn in reference period
61	Died prior to reference period (not eligible)-Round 1 only

62 Institutionalized prior to reference period (not eligible)-Round 1 only

63 Moved outside U.S., prior to reference period (not eligible)-Round 1 only

64 Full time military, living on a military facility, moved prior to reference period (not eligible)-Round 1 only

71 Student under 24 living away at school in grades 1-12 (non-key)

72 Person is dropped from the RU roster as ineligible: the person is a non-key student living away or the person is not related to reference person or the RU is the person's residence only during the school year

73 Not key and not full-time military, moved w/o someone key and inscope (not eligible)

74 Moved as full-time military but not to a military facility and w/o someone key and inscope (not eligible this round)

81 Person moved from original RU, full time student living away from home, did not respond

In addition, the variable INRU1231 indicates if a person was present in the RU on December 31, 1997. Persons living in the RU as well as any person coded as "living away in grades 1-12" will have a value of "1" indicating "Yes, the person was present on December 31, 1997."

2.5.2 Navigating the MEPS Data with Information on Person Disposition Status

Since the variables PSTATS31, PSTATS42, and PSTATS53 indicate the reasons for either continuing or terminating data collection for each person in MEPS, these variables can be used to explain the beginning and ending dates for each individual's reference period of data collection, as well as which sections in the instrument each individual received. By using the information included in the table below, analysts will be able to determine for each individual which sections of the MEPS questionnaire collected data elements for that person.

Some individuals have a reference period that spans an entire round, while other individuals may have data collected only for a portion of the round. When an individual's reference period does not coincide with the RU reference period, the individual's start date may be a later date, or their end date may be an earlier date, or both. In addition, some individuals have reference period information coded as inapplicable (e.g., for individuals who were not actually in the household). The information in this table indicates the beginning and ending dates of reference periods for persons with various values of PSTATS31, PSTATS42, and PSTATS53. The actual dates for each individual can be found in the following variables included on this file: BEGRFM31, BEGRFM42, BEGRFM53, BEGRFD31, BEGRFD42, BEGRFD53, BEGRFY31, BEGRFY42, BEGRFY53, ENDRFM31, ENDRFM42, ENDRFM53, ENDRFD31, ENDRFD42, ENDRFD53, ENDRFY31, ENDRFY42, ENDRFY53, ENDRFM97, ENDRFD97, and ENDRFY97.

The table below also describes the section or sections of the questionnaire, which were **NOT** asked for each value of PSTATS31, PSTATS42, and PSTATS53**.** For example, the condition enumeration (CE) and alternative/preventive care (AP) sections have questions, which are not asked for deceased persons. The closing section (CL) also contains some questions or question rosters (see CL06A, CL35 through CL37, CL48 through CL50, CL54, CL58, and CL64) that exclude certain persons depending on whether the person died, became institutionalized, or otherwise left the reporting unit; however, no one is considered to have skipped the entire section. Some questions or sections (e.g., health status (HE), employment (RJ, EM, EW)) are skipped if individuals are not within a certain age range. Since the PSTATS variables do not address skip patterns based on age, analysts will need to use the appropriate age variables.

Please note that the end reference date shown below for PSTATS53 reflects the Round 5/3 reference period rather than the portion of Round 5/3 that occurred during 1997.

PSTATS Value	**PSTATS Description**	**Sections in the instrument which persons with this PSTATS value do NOT receive**	**Begin Reference Date**	**End Reference Date**
-1	The person was not fielded during the round or the RU was non-response	ALL sections	Inapplicable	Inapplicable
0	Incorrectly listed in RU at NHIS - Round 3/1 only	ALL sections after RE	Inapplicable	Inapplicable
11	Person in original household, not FT active military duty (Person is in the same RU as the previous round)	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Interview date
12	Person in original household, FT active military duty, out-of-scope for whole reference period.	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Interview date
13	FT student living away from home, but associated with sampled household	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Interview date
14	The person is FT active military duty during round and is inscope for part of the reference	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53:	PSTATS31: Interview date PSTATS42 and PSTATS53: If the

PSTATS Value	PSTATS Description	Sections in the instrument which persons with this PSTATS value do NOT receive	Begin Reference Date	End Reference Date
	period and is in the RU at the end of the reference period		Prior round interview date	person is living w/ someone key and inscope, then the interview date. If not living w/ someone who is key and inscope, then the date the person joined the military
21	The person remains in a health care institution for the whole round - rounds 4/2 and 5/3 only	All sections after RE	Inapplicable	Inapplicable
22	The person leaves a health care institution and rejoins the community - rounds 4/2 and 5/3 only	--	Date rejoined the community	Interview date
23	The person leaves a health care institution, goes into community and then dies - rounds 4/2 and 5/3 only	Part of CE - Condition enumeration: Skip CE1 to CE5 HE - Health status AC - Access to care Part of AP - Alternative/Preventive care: Skip AP12 to AP22	Date rejoined the community	Date of Death
24	The person dies in a health care institution during the round (former household member) - rounds 4/2 and 5/3 only	All sections after RE	Inapplicable	Inapplicable
31	Person from original household, dies during reference period	Part of CE - Condition enumeration: Skip CE1 to-CE5 HE - Health status AC - Access to care Part of AP - Alternative/Preventive care: Skip AP12 to AP22	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Date of Death
32	Went to healthcare institution during reference period	Access to care (AC)	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round	Date institutionalized

PSTATS Value	PSTATS Description	Sections in the instrument which persons with this PSTATS value do NOT receive	Begin Reference Date	End Reference Date
			interview date	
33	Went to non-healthcare institution during reference period	Access to care (AC)	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Date institutionalized
34	Moved from original household, outside US	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Date left the RU
35	Moved from original household, to a military facility while on FT active military duty	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Date left the RU
36	Went to institution (type unknown) during reference period	Access to care (AC)	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Date institutionalized
41	Moved from the original household, to new household within US (new households include RUs originally classified as a student RU but which converted to a new RU. These are individuals in an RU that has split from an RU since the previous round	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Interview date
42	The person joins household and is not full time military during round	--	The later date of January 1, 1997 and the date the person joined the RU	Interview date
43	The person's disposition as to why the person is	All sections after RE	Inapplicable	Inapplicable

PSTATS Value	PSTATS Description	Sections in the instrument which persons with this PSTATS value do NOT receive	Begin Reference Date	End Reference Date
	not in the RU is unknown or the person moves and it is unknown whether the person moved inside or outside the U.S.			
44	The person leaves an RU and joins an existing RU and is not both in the military and coded as inscope during the round	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date of the RU the person has joined. This may not be the interview date of the RU that the person came from	Interview date
51	Newborn in reference period	Questions where age must be > 1 (see Health status (HE), Disability days (DD) Employment (RJ/EM/EW) will be skipped	PSTATS31: January 1, 1997 if born prior to 1997. The date of birth if born in 1997. PSTATS42 and PSTATS53: The later of the Prior round interview date and date of birth	Interview date
61	Died prior to reference period (not eligible)--Round 3/1 only	All sections after RE	Inapplicable	Inapplicable
62	Institutionalized prior to reference period (not eligible)--Round 3/1 only	All sections after RE	Inapplicable	Inapplicable
63	Moved outside U.S., prior to reference period (not eligible)--Round 3/1 only	All sections after RE	Inapplicable	Inapplicable
64	FT military, moved prior to reference period (not eligible)--Round 3/1 only	All sections after RE	Inapplicable	Inapplicable

PSTATS Value	PSTATS Description	Sections in the instrument which persons with this PSTATS value do NOT receive	Begin Reference Date	End Reference Date
71	Student under 24 living away at school in grades 1 thru 12 (non-key)	--	PSTATS31: January 1, 1997 PSTATS42 and PSTATS53: Prior round interview date	Interview date
72	Person is dropped from the RU roster as ineligible: the person is a non-key student living away or the person is not related to reference person or the RU is the person's residence only during the school year	All sections after RE	Inapplicable	Inapplicable
73	Not key and not full-time military, moved w/o someone key and inscope (not eligible)	All sections after RE	Inapplicable	Inapplicable
74	Moved as full-time military but not to a military facility and w/o someone key and inscope (not eligible)	All sections after RE	Inapplicable	Inapplicable
81	Person moved from original household, FT student living away from home, did not respond	No data was collected	Inapplicable	Inapplicable

2.5.3 Geographic Variables

The round-specific variables REGION31, REGION42, REGION53, and the end-of-year status variable REGION97 indicate the Census region for the RU. REGION97 indicates the region for the 1997 portion of Round 5/3. For most analyses, REGION97 should be used. The round-specific variable MSA53 and the end-of-year status variable MSA97 indicate whether or not the RU is found in a metropolitan statistical area. MSA53 indicates the MSA status at the time of the Round 5/3 interview. MSA97 indicates the MSA status for the 1997 portion of Round 5/3. For most analyses, analysts should use MSA97 rather than MSA53.

2.5.4 Demographic Variables

General Information

Demographic variables provide information about the demographic characteristics of each person from the MEPS-HC. The characteristics include age, sex, race, ethnicity, educational attainment, marital status, and military service. As noted below, some variables have edited and imputed values. Most demographic variables on this file are asked during every round of the MEPS interview. These variables describe data for Rounds 3, 4, and 5 of Panel 1 (1996 Panel); Rounds 1, 2 and 3 of Panel 2 (1997 Panel); and status as of December 31, 1997. Demographic variables that are round specific are identified by names including numbers "*xy*," where *x* and *y* refer to Round numbers of Panels 1 and 2 respectively. Thus, for example, AGE31X represents the age data relevant to Round 3 of Panel 1 or Round 1 of Panel 2. As mentioned in Section 2.5.1 Survey Administrative Variables, the variable PANEL97 indicates from which Panel the data were derived. A value of 1 indicates Panel 1 data and a value of 2 indicates Panel 2 data. The remaining demographic variables on this file are not round specific.

The variables describing demographic status of the person as of December 31, 1997 were developed in two ways. First, the age variable (AGE97X) represents the exact age as of 12/31/97, calculated from date of birth. For the remaining December 31st variables [i.e., related to marital status (MARRY97X, SPOUID97, SPOUIN97), educational attainment (EDUCYR97, HIDEG97), student status (FTSTUD97X) and the relationship to reference persons (RFREL97X)], the following algorithm was used: data were taken from Round 5/3 counterpart if non-missing; else, if missing, data were taken from the Round 4/2 counterpart; else from the Round 3/1 counterpart. If no valid data was available during any of these Rounds of data collection, the same algorithm was followed to assign a missing value other than -1 (Inapplicable).

Age

Date of birth and age for each RU member were asked or verified during each MEPS interview (DOBMM, DOBYY, AGE31X, AGE42X, AGE53X). If date of birth was available, age was calculated based on the difference between date of birth and date of interview (or the date of death, if the person died prior to the interview date). Inconsistencies between the calculated age and the age reported during the CAPI interview were reviewed and resolved. For purposes of confidentiality, the variables AGE31X, AGE42X, AGE53X and AGE97X were top coded at 90 years.

When date of birth was not provided but age was provided (either from the MEPS interviews or the 1995-1996 NHIS data), the month and year of birth were assigned randomly from among the possible valid options. For any cases still not accounted for, age was imputed using

(1) the mean age difference between MEPS participants with certain family relationships (where available) or
(2) the mean age value for MEPS participants.

For example, a mother's age is imputed as the average age of her children plus 26, where 26 is the mean age difference between MEPS mothers and their children. Or a wife's age is imputed as the husband's age minus 3, where 3 is the mean age difference between MEPS wives and husbands.

Age was imputed in this way for 20 persons on this file.

Sex

Data on the sex of each RU member (SEX) were initially determined from the 1995 NHIS for Panel 1 and from the 1996 NHIS for Panel 2. The SEX variable was verified and, if necessary, corrected during each MEPS interview. The data for new RU members (persons who were not members of the RU at the time of the NHIS interviews) was also obtained during each MEPS Round. When sex of the RU member was not available from the NHIS interviews and was not ascertained during one of the subsequent MEPS interviews, it was assigned in the following way. The person's first name was used to assign sex, if obvious (25 cases were resolved in this way). If the person's first name provided no indication of gender, then family relationships were reviewed (11 cases). If neither of these approaches made it possible to determine the individual's sex, sex was randomly assigned (3 cases).

Race, Race/Ethnicity, Hispanic Ethnicity, and Hispanic Ethnicity Group

Race (RACEX) and Hispanic ethnicity (HISPANX) questions were initially asked for each RU member during the Round 1 MEPS interview. If this information was not obtained in Round 1, the questions were asked in subsequent Rounds. When race and/or ethnicity was not reported in the Rounds, values for these variables were obtained based on the following priority order. When available, they were obtained from the originally collected NHIS data (1995 or 1996, depending on the Panel). If not ascertained, the race, and/or ethnicity were assigned based on relationship to other members of the RU using a priority ordering that gave precedence to blood relatives in the immediate family. This approach was used in the resolution of a residual group of 24 cases, 17 of which were missing both race and ethnicity and 7 of which were missing only ethnicity. The variable RACETHNX indicating both race and ethnicity (e.g., with categories such as "Hispanic" and "black but not Hispanic") reflects the imputations done for RACEX and HISPANX. The specific Hispanic ethnicity group is given in the unedited variable HISPCAT.

Marital Status and Spouse ID

Current marital status was collected and/or updated during every Round of the MEPS interview. This information was obtained in RE13 and RE97 and is reported as MARRY31X, MARRY42X, MARRY53X and MARRY97X. Persons under the age of 16 were coded as 6 (under 16 – inapplicable). If marital status of a specified round differed from that of the previous Round, then the marital status of the specified Round was edited to reflect a change during the Round (e.g., married in Round, divorced in Round, separated in Round, or widowed in Round).
In instances where there were discrepancies between the marital status of two individuals within a family, other person-level variables were reviewed to determine the edited marital status for each

individual. Thus, when one spouse was reported as married and the other spouse reported as widowed, the data were reviewed to determine if one partner should be coded as 8 (widowed in Round).

Four edits were performed to ensure minimal consistency across rounds. First, a person could not be coded as "Never Married" after previously being coded as any other marital status (e.g. "Widowed"). Second, a person could not be coded as "Under 16 – Inapplicable" after being previously coded as any other marital status. Third, a person could not be coded as "Married in Round" after being coded as "Married" in the Round immediately preceding. Fourth, a person could not be coded as an "in Round" code (e.g., "widowed in Round") in two subsequent Rounds. Because no other edits were performed, and since marital status can change across Rounds, unlikely sequences for marital status across the Round-specific variables do exist.

The person identifier for each individual's spouse is reported in SPOUID31, SPOUID42, SPOUID53 and SPOUID97. These are the PIDs (within each family) of the person identified as the spouse during Round 3/1, Round 4/2, Round 5/3 and as of December 31, 1997, respectively. If no spouse was identified in the household, the variable was coded as 995 (No spouse in household). Those with unknown marital status are coded as 996 (Marital status unknown). Persons under the age of 16 are coded as 997 (Less than 16 years old).

The SPOUIN31, SPOUIN42, SPOUIN53 and SPOUIN97 variables indicate whether a person's spouse was present in the RU during Round 3/1, Round 4/2, Round 5/3 and as of December 31, 1997 respectively. If the person had no spouse in the household, the value was coded as 2 (Not married/No spouse). For persons under the age of 16 the value was coded as 3 (Under 16 – Inapplicable).

The SPOUID and SPOUIN variables were obtained from RE76 and RE77, where the respondent was asked to identify how each pair of persons in the household were related. Analysts should note that this information was collected in a set of questions separate from the questions that asked about marital status. While editing was performed to ensure that SPOUID and SPOUIN are consistent within each Round, there was no consistency check between these variables and marital status in a given Round. Apparent discrepancies between marital status and spouse information may be due to any of the following causes:

1. Ambiguity as to when during a Round a change in marital status occurred. This is a result of relationship information being asked for all persons living in the household at any time during the Round, while marital status is asked as of the interview date (e.g., If one spouse died during the reference period, the surviving spouse's marital status would be "widowed in Round", but SPOUIN and SPOUID for the same round would indicate that a spouse was present).

2. Valid discrepancies in the case of persons who are married but not living with their spouse, or separating but still living together.

3. Discrepancies that cannot be explained for either of the previous reasons.

Student Status and Educational Attainment

The variables FTSTU31X, FTSTU42X, FTSTU53X and FTSTU97X indicate whether the person was a full-time student at the interview date (or 12/31/97 for FTSTU97X). These variables have valid values for all persons between the ages of 17 - 23 inclusive. When this question was asked during Round 1 of Panel 2, it was based on age as of the 1996 NHIS interview date; for persons who were 17 years old at the Round 1 but 16 years old at the time of the NHIS, FTSTU31X was set to -9 (Not ascertained) resulting in a large number of -9 values.

Completed years of education are indicated in the variables EDUCYR31, EDUCYR42, EDUCYR53 and EDUCYR97. Information was obtained from questions RE 103-105. Children who are 5 years of age or older and who never attended school were coded as 0; children under the age of 5 years were coded as -1 (Inapplicable) regardless of whether or not they attended school. However, among the cases coded as inapplicable, there is no distinction between those who were under the age of five and others who were inapplicable, such as persons who may be institutionalized for an entire round.

The variables indicating highest degree (HIDEG31, HIDEG42, HIDEG53 and HIDEG97) were obtained from three questions: highest grade completed (RE103), high school diploma (RE 104) and highest degree (RE 105). Persons under 16 years of age were coded as 8 (Under 16- Inapplicable). In cases where the response to the highest degree question was “No degree” and the response to the highest grade question was 13 through 17 (1 or more years of college), the variable HIDEG was coded as 3 (high school diploma). If highest grade completed was “refused“ or “don’t know” for those with a “No degree” response for the highest degree question, the variable HIDEG was coded as 1 (no degree).

The user should note that the EDUCYR and HIDEG variables are unedited variables and minimal data cleaning was performed on these variables. Therefore, discrepancies across rounds of data remain for these two sets of variables. Decisions as to how to handle these discrepancies are left to the analyst.

Military Service and Service Era

Information on active duty military status was collected during each Round of the MEPS interview. Persons currently on full-time active duty status are identified in the variables ACTDTY31, ACTDTY42, and ACTDTY53. Those under 16 years of age were coded as 3 (Under 16 – Inapplicable) and those over the age of 59 were coded as 4 (Over 59 – Inapplicable).

The variable DIDSERVE is only collected during Round 1 of the MEPS interview. It indicates if the person ever served in the Armed Forces. Persons under the age of 16 were coded as 3 (Under 16 – Inapplicable). Individuals currently on active duty military service were coded as 4 (Now active duty). Those individuals entering a MEPS household after Round 1 have DIDSERVE set to –1 (Inapplicable). Like DIDSERVE, data on service in specific eras was only collected during Round 1 of the MEPS interview. Individuals who were ever in the military based on the DIDSERVE and ACTDTY question(s) of Round 1 were also asked if they served in either World War I or World War

II (VETWW), the Korean War era (VETKOR), the Vietnam War era (VETVIET), the Post-Vietnam War era (VETPVIET), or another service era (VETOTH). Those under the age of 16 were coded as 3 (Under 16 –Inapplicable) and those who never served in the military were coded as 4 (Never in military). Persons entering a MEPS household after Round 1 have these variables set to –1 (Inapplicable).

Because DIDSERVE and veteran status variables are only asked during Round 1, and because the 1997 FY file only contains data from Rounds 3, 4, and 5 of Panel 1, these variables would have been missing for persons in Panel 1. Consequently, an analyst would have had to go back to the 1996 full year file (MEPS HC-008) in order to determine the military service and veteran status values for those Panel 1 persons. Therefore, to provide a better estimation of military service and veteran status for this 1997 full year file, DIDSERVE, VETWW, VETKOR, VETVIET, VETPVIET, and VETOTH from Panel 1, Round 1 (on the 1996 file) were brought forward onto the 1997 Full Year file.

The user should note that the DIDSERVE and veteran status variables were reviewed for consistency. The veteran status variables were minimally edited to ensure that all individuals under 16 years of age were coded as 3 (Under 16 – Inapplicable) for the specific veteran-era variables. However, no other age editing was performed, and, thus, it is possible for age/era inconsistencies to exist (e.g., AGE31X=17 and VIETVET=Yes).

Relationship to the Reference Person within Reporting Units

For each reporting unit (RU), the person who owns or rents the dwelling unit is usually defined as the reference person. For student RUs, the student is defined as the reference person. (For additional information on reference persons, see the documentation on survey administration variables.) The variables RFREL31X, RFREL42X, RFREL53X and RFREL97X indicate the relationship of each individual to the reference person of the reporting unit (RU) in a given round. For the reference person, this variable has the value “self;” for all other persons in the RU, relationship to the reference person is indicated by codes representing “husband/spouse,” “wife/spouse,” “son,” “daughter,” “female partner,” “male partner,” etc. A code of 91, meaning “other related, specify,” was used to indicate rarely observed relationship descriptions such as “mother of partner.” If the relationship of an individual to the reference person was not ascertained during the Round-specific interview, relationships between other RU members were used, where possible, to assign a relationship to the reference person. If MEPS data from calendar year 1997 were not sufficient to identify the relationship of an individual to the reference person, relationship variables from the 1996 MEPS or NHIS data were used to assign a relationship. In the event that a meaningful value could not be determined or data were missing, the relationship variable was assigned a missing value code.

For 339 cases, where two individual’s relationship indicated they were spouses, but both had marital status indicating they were not married, their relationship was changed to non-marital partners. In addition, the relationship variables were edited to insure that they did not change across rounds for RUs in which the reference person did not change, with the exception of relationships identified as partner, spouse, or foster relationships.

Parent Identifiers

The variables MOPID31X, MOPID42X, MOPID53X and DAPID31X, DAPID42X DAPID53X are round specific and are used to identify the parents (biological, adopted, or step) of the person represented on that record. MOPID##X contains the person identifier (PID) for each individual's mother if she lived in the dwelling unit in that panel/round of the survey, or a value of –1 (Inapplicable) if she did not. Similarly, DAPID##X contains the person identifier (PID) for each individual's father if he lived in the dwelling unit in that panel/round of the survey, or a value of –1 (Inapplicable) if he did not. MOPID##X and DAPID##X were constructed based on information collected in the relationship grid of the instrument each round at questions RE76 and RE77 and include biological, adopted, and step parents. Foster parents were not included.

Edits were performed to ensure that MOPID##X and DAPID##X were consistent with each individual's age, sex, and other relationships within the family. For instance, the gender of the parent must be consistent with the indicated relationship; mothers are at least 12 years older than the person and no more than 55 years older than the person; fathers are at least 12 years older than the person; each person has no more than one mother and no more than one father; any values set for MOPID##X and DAPID##X were removed from any person identified as a foster child; and the PID for the person's mother and father are valid PIDS for that person's DU in that round. For persons who were not present in the household during a round, MOPID##X and DAPID##X have values of –1 (Inapplicable).

2.5.5 Income and Tax Filing Variables

The file provides income and tax-related variables that were constructed primarily from data collected in the Round 3 Income Section. Person-level income amounts have been edited and imputed for every record on the full-year file, with detailed imputation flags provided as a guide to the method of editing. The tax-filing variables and some program participation variables are unedited as discussed below.

Logical editing or weighted, sequential hot-deck imputation was used to impute income amounts for missing values (both for item non-response and for persons in the full-year file who were not in Round 3). Reported income components were generally left unedited (with the few exceptions noted below). Thus, analysts using these data may wish to apply additional checks for outlier values that would appear to stem from mis-reporting.

The editing process began with wage and salary income, WAGEP97X. Complete responses were left unedited, and this group of people was assigned WAGEIMP97 = 1, where WAGEIMP97 is the imputation flag for wage and salary data. The only exception was for a small number of persons who reported zero wage and salary income despite having been employed for pay during the year according to round level data (see below). Since data on tax filing and on taxable income sources were collected using an approach that encouraged respondents to provide information from their

federal tax returns, logical edits were used to assign separate income amounts to married persons whose responses were based on combined income amounts on their joint tax returns.

Persons assigned WAGIMP97=2 were those providing broad income ranges rather than giving specific dollar amounts. Weighted sequential hot-decking was used to provide these individuals with specific dollar amounts. For this imputation, donors were persons who reported specific dollar amounts within the corresponding broad income ranges. All WAGEP97X hot-deck imputations used cells defined on the basis of a conventional list of person-level characteristics including age, education, employment status, race, sex, and region.

Persons assigned WAGIMP97=3 were those who did not report wage and salary income and who were assigned WAGEP97X=0 based on either being under 16 or not having been employed during the year.

Persons assigned WAGIMP97=4 were those who did not provide valid dollar amounts or dollar ranges, but for whom we had information from the employment sections of Rounds 1, 2, and 3 concerning wages, hours, and weeks worked (in all jobs). These data were used to construct annualized wage amounts to be used in place of missing annual wage and salary data. Comparisons of reported and constructed wages and salaries using persons who provided both sorts of information made us highly confident that employment data could be reliably used to derive values to serve in place of missing wage and salary information (the two measures were highly correlated and the means differed by less than $20). To implement this approach, part-year responders were assumed to be fully-employed during the remainder of the year if they were employed during the period in which they provided data. An exception was made for those who either died or were institutionalized. These persons were assigned zero wages and salaries for the time they were not in MEPS.

Hot deck imputation was used for the remaining persons with missing WAGEP97X. Donor pools included persons whose WAGEP97X amounts were edited in the steps described above. Whenever possible, the hot-deck imputations used data on whether or not the person had been employed at any point during the year (and, if available, the number of weeks worked). Imputations for persons deemed to have been employed were conditional in nature, using only donors with positive WAGEP97X amounts (WAGIMP97=5). Imputations for WAGEP97X for the remaining persons were unconditional, using both workers and non-workers as donors (WAGIMP97=6).

Having edited WAGEP97X for all persons in the full-year file, we then edited the remaining income sources in the following sequence: INTRP97X, BUSNP97X, FARMP97X, DIVDP97X, REFDP97X, ALIMP97X, SALEP97X, TRSTP97X, PENSP97X, IRASP97X, SSECP97X, UNEMP97X, WCMPP97X, VETSP97X, CASHP97X, OTHRP97X, CHLDP97X, SSIP97X, and PUBP97X. Income components were edited sequentially, in each case using information regarding income amounts that had already been edited (so as to maintain patterns of correlation across income sources whenever possible). In all cases, bracketed responses were edited first (using hot-deck imputations from donors in corresponding brackets who gave specific dollar amounts), followed by imputations for remaining missing values. The hot-deck imputations used cells defined on the basis of income

amounts already edited and a conventional list of person-level characteristics such as age, education, employment status, race, sex, and region. In addition, hot-deck imputations for CHLDP97X used family-level information concerning marital status and the number of children. Hot-deck imputations for SSIP97X and PUBP97X were also assigned using, in part, simulated program eligibility indicators that integrated state-level program eligibility criteria with data on family composition and income.

Reported income amounts of less than one dollar were treated as missing amounts (to be hot-decked from donors with positive amounts of the corresponding income source). Also, a very few cases of outlier responses were edited (primarily public sources of income that exceeded possible amounts). With only one other exception, reported income amounts were left unchanged. The exception was Social Security Income, SSECP97X, which was under-reported in the MEPS relative to the March 1998 Current Population Survey (CPS). Comparison with the CPS identified the source of the MEPS under-reporting to be persons aged 65 and older who failed to report any SSECP97X despite having also reported no earned income. Persons over 65 with neither earnings nor Social Security income are quite rare in the CPS, giving us confidence in editing these responses. Using the CPS, a probabilistic model was developed to select persons/couples whose values of SSECP97X were changed from zero to a positive (imputed) amount.

For all of the income components, *xx*IMP97 variables contain indicators concerning the method for editing/imputation. All the flag variables have the following formatted values:

- 1=Original response used;
- 2=Bracket converted;
- 3=Missing value set to 0;
- 4=Weeks worked/earnings used (WAGIMP97 only);
- 5=Conditional hot-deck;
- 6=Unconditional hot-deck.

Missing values were set to zero when there were too few recipients to warrant hot-deck imputations of positive values (as in the case of ALIMP97X received by males or WAGEP97X received by persons under age 16). “Conditional hot-decks” indicate instances where the respondent indicated receipt but not a specific dollar amount. In these cases, the donor pool was restricted to persons with nonzero amounts of the income source in question. “Unconditional hot-decks” indicate instances where the donor pool included persons receiving both zero and nonzero amounts (implemented in cases where we had little or no information about a person’s income source).

Total person-level income (TTLP97X) is the sum of all income components with the exception of REFDP97X and SALEP97X (so that we are following as closely as possible the CPS definition of income). Some researchers may wish to define their own income measure by adding in one or both of these excluded components.

The tax variables, food stamp variables, SSI disability flag, and AFDC participation flag are all completely unedited. In particular, while the tax variables are provided to assist researchers building

tax simulation programs, there is substantial item non-response. No effort was made to eliminate inconsistencies with other MEPS data. All of these unedited variables should be used with great care.

2.5.5.1 Income Top-Coding

All income amounts on the file, including both total income and the separate sources of income, were top coded to preserve confidentiality. For each income source, top codes were applied to the top percentile of all cases (including negative amounts that exceeded income thresholds in absolute value). In cases where fewer than one percent of all persons received a particular income source, we top-coded all recipients. Top-coded income amounts were masked using a regression-based approach. The regressions relied on many of the same variables used in the hot-deck imputations, with the dependent variable in each case being the natural logarithm of the amount that the income component was in excess of its top-code threshold. Predicted values from this regression were reconverted from logarithms to levels using a smearing correction, and these predicted amounts were then added back to the top-code thresholds. This approach preserves the component-by-component weighted means (both overall and among top-coded cases), while also preserving much of the income distribution conditional on the variables contained in our regressions. At the same time, this approach ensures that every reported amount in excess of its respective threshold is altered on the public use file. The process of top-coding income amounts in this way inevitably introduces measurement error in cases where income amounts were reported correctly by respondents. Note, however, that top-coding can also help to reduce the impact of outliers that occur due to mis-reporting.

Total income is constructed as the sum of the adjusted income components. Having constructed total income in this manner, we then top-coded this total using the same regression-based procedure described above (again masking the top percentile of cases). Finally, we scaled the components of income up or down in order to make the sources of income consistent with the newly-adjusted totals.

2.5.5.2 Poverty Status

The file includes a categorical variable for 1997 family income as a percentage of poverty (POVCAT97). This variable was constructed primarily from data collected in the Round 3 Income Section. Logical editing or weighted, sequential hot-deck imputation was used to impute income amounts for missing values (both for item non-response and or persons in the full-year file who were not in Round 3). Round-level data on employment status, hours worked, and wages were used to supplement earnings data collected in the Income Section. Family income was derived by constructing person-level total income comprising annual earnings from wages, salaries, bonuses, tips, commissions; business and farm gains and losses; unemployment and workman's compensation; interest and dividends; alimony, child support, and other private cash transfers; private pensions, IRA withdrawals, social security, and veterans payments; supplemental security income and cash welfare payments from public assistance, Aid to Families with Dependent Children, and Aid to Dependent Children; gains or losses from estates, trusts, partnerships, S corporations, rent, and royalties; and a small amount of "other"income. Family income excluded tax refunds and capital gains. Person-level income totals were then summed over family members as defined by

CPSFAMID to yield the family-level total. POVCAT97 is constructed by dividing family income by the applicable poverty line (based on family size and composition), with the resulting percentages grouped into 5 categories; negative or poor (less than 100%), near poor (100 to less than 125%), low income (125 to less than 200%), middle income (200 to less than 400%), and high income (400%+). Persons missing CPSFAMID were treated as one-person families in constructing POVCAT97. Family income as well as the components of person level income have been subjected to internal editing patterns and derivation methods that are in accordance to specific definitions, and are not being released at this time. Researchers working with a family definition other than CPSFAMID may wish to create their own versions of total family income (and perhaps POVCAT97).

2.5.6 Employment Variables

Employment questions were asked of all persons 16 years and older at the time of the interview. Employment variables consist of person-level indicators such as employment status and job-related variables such as hourly wage. All job-specific variables refer to a person's current main job. The current main job, defined by the respondent, indicates the main source of employment.

Most employment variables pertain to the round interview date. The round dates are indicated by two numbers following the variable name; the first number representing the round for Panel 1 persons, the second number representing the round for Panel 2 persons. For example, EMPST31 refers to employment status on the Round 3 interview date for Panel 1 persons and employment status on the Round 1 interview date for Panel 2 persons.

With the exception of health insurance held or offered from a current main job, no attempt has been made to logically edit any employment variables. When missing, values were imputed for certain persons' hourly wages; however, there was no editing performed on any values reported by the respondent. Due to confidentiality concerns, hourly wages greater than or equal to $43.75 were top-coded to –10 and the number of employees variable was top-coded at 500. With the exception of a variable indicating whether the employer has more than one location (MORE), all employer-specific variables refer to the establishment that is the location of a person's current main job.

The MEPS employment section used dependent interviewing in Rounds 2 through 5. If employment status and certain job characteristics did not change from the previous round, the respondent was skipped through the employment section. A code of "–2" is used to indicate that the information in question was obtained in a previous round. For example, if the HRWG42X (Round 4 interview date hourly wage for Panel 1 persons/Round 2 interview date hourly wage for Panel 2 persons) is coded as -2, refer to HRWG31X (Round 3 interview date hourly wage for Panel 1 persons/Round 1 interview date hourly wage for Panel 2 persons) for the value for HRWG42X. Note that there may be a value for the Round 3/1 hourly wage or there may be an inapplicable code. The –2 value for HRWG42X simply indicates that the person was skipped past the question at the time of the interview. Obviously, to determine who should be skipped through various employment questions, certain information, such as employment status, had to be asked in every round and, thus, -2 codes do not apply to employment status. Additionally, information on whether the person currently worked at more than one job or whether the person

held health insurance from a current main employer was asked in every round, and, therefore, those variables also have no –2 codes.

For Panel 1 persons who have a current main job in Round 3 that continues from Round 1 or 2, the –2 code is not sufficient for those variables that the person was skipped past at the time of the interview. This is because the Panel 1 Round 1 and 2 data will not be included on this release and therefore there are no data to which to refer. For such persons, the values for the variables for these skipped questions are copied from the Round 1 or 2 constructed variable, depending on the round in which the job first became the current main job. The accompanying variable RNDFLG31 indicates the round in which these data were collected. For example, if the person has a Round 3 current main job that continues from Round 2 and was first reported as the current main job in Round 2, HRWG31X will be a copy of the HRWAG2X from the 1996 Full Year Public Use Release and RNDFLG31 will be '2', indicating the round in which the job was first reported as the current main job.

Employment Status (EMPST31, EMPST42, and EMPST53)

Employment status was asked for all persons aged 16 or older. Allowable responses to the employment status question were as follows:

- "currently employed" if the person had a job at the interview date;
- "has a job to return to" if the person did not work during the reference period but had a job to return to as of the interview date;
- "employed during the reference period" if the person had no job at the interview date but did work during the round;
- "not employed with no job to return to" if the person did not have a job at the interview date, did not work during the reference period, and did not have a job to return to.

These responses were mutually exclusive. A current main job was defined for persons reporting that they were currently employed and identified a current main job, and for persons who reported and identified a job to return to. Therefore, job-specific information such as hourly wage exists for persons not presently working at the interview date but who have a job to return to as of the interview date.

Data Collection Round for Round 3/1 CMJ (RNDFLG31)

For Panel 1, if the Round 3 current main job (CMJ) is a continuation CMJ from Round 2 or Round 1, the value of most "31" variables will be copied forward from the variable representing the round in which the job was first reported as the CMJ. For persons in Panel 1, RNDFLG31 indicates the round in which the Round 3 CMJ was first reported as the CMJ and provides a timeframe for the reported wage information and other job details. RNDFLG31 is used with many "31" variables to indicate the round on which the reported information is based.

For persons in either panel, RNDFLG31 is set to inapplicable (–1) for persons who are under age 16 or who do not have a CMJ in Panel 1 Round 3 or Panel 2 Round 1. For persons who are part of Panel 1, RNDFLG31 is also set to inapplicable (–1) if person is out-of-scope in the 1997 portion of Round 3. For persons who are part of Panel 2, RNDFLG31 is also set to inapplicable (–1) if person is out-of-scope in Round 1. For persons who are part of Panel 1, other values for RNDFLG31 are set as follows:

- 1 - for continuing Round 3 CMJs reported first in Round 1;
- 2 - for continuing Round 3 CMJs reported first in Round 2;
- 3 - for jobs newly reported as current main in Round 3;
- -9 - Round 3 CMJ is a continuation CMJ (wage information and other details were not collected in Round 3) but the Round 2 CMJ record either does not exist or is not the same job. This can occur in rare instances because corrections made to a person's record in a current file cannot be made to that record in an earlier file due to data base processing constraints.

For persons who are part of Panel 2 and reported a Round 1 CMJ, RNDFLG31 equals "1" indicates that the job information represented in the "31" variables was collected in Round 1.

Self-employed (SELFCM31, SELFCM42, and SELFCM53)

Information on whether an individual was self-employed at the current main job was obtained for all persons who reported a current main job. Certain questions, namely those regarding benefits and hourly wage, were not asked of the self-employed. These variables indicate whether the establishment reported by wage earners as the main source of employment offered the following benefits:

- Paid leave to visit a doctor (PAYDR31, PAYDR42, and PAYDR53);
- Paid sick leave (SICPAY31, SICPAY42, and SICPAY53);
- Paid vacation (PAYVAC31, PAYVAC42, and PAYVAC53);
- Pension plan (RETPLN31, RETPLN42, and RETPLN53).

Those who were self-employed at their current main job are coded as inapplicable (-1) for all these variables. Additionally, information on whether the firm has more than one establishment (MORE31, MORE42, and MORE53) and whether the establishment is a private for-profit, nonprofit, or a government entity (JOBORG31, JOBORG42, and JOBORG53) is not applicable for self-employed persons. Conversely, the variables that measure whether a business is incorporated, a proprietorship, or a partnership (BSNTY31, BSNTY42, and BSBTY53) apply only to those who are self-employed at their current main job.

Hourly wage (HRWG31X, HRWG42X, HRWG53X)

Hourly wage was asked of all persons who reported a current main job that was not self-employment (SELFCM). An hourly wage was imputed using a weighted sequential hot-deck procedure for those identified as having a current main job who were not self-employed and who did not know their

wage, or refused to report a wage. Hourly wage for persons for whom employment status was not known was coded as not ascertained (-9). Additionally, wages were imputed for wage earners reporting a wage range and not a specific value. For these persons, values were imputed from donors within the reported range. All imputed wages can be identified as such by three wage imputation flags (HRWGIM31, HRWGIM42, HRWGIM53). Note that wages were imputed only for persons with a positive person weight.

For reasons of confidentiality, the hourly wage variable was top-coded. A value of –10 indicates that the hourly wage was greater than or equal to $43.75. The hourly wage variables on this file (HRWG31X, HRWG42X, HRWG53X) should be considered along with their accompanying variables--HRHOW31, HRHOW42, and HRHOW53—which indicate how the respective round hourly wage was constructed. Hourly wage could be derived, as applicable, from a large number of source variables. In the simplest case, hourly wage was reported directly by the respondent. For other persons, construction of the hourly wage was based upon salary, the time period on which the salary was based, and the number of hours worked per time period. If the number of hours worked per time period was not available, a value of 40 hours per week was assumed, as identified in the HRHOW variable. It should be noted that HRHOW and HRWGIM may differ. As mentioned above, wage imputations were performed on persons with positive weights only, while HRHOW will apply to persons with a zero person-level weight.

Health Insurance (HELD31X, HELD42X, HELD53X, OFFER31X, OFFER42X, OFFER53X, CHOIC31, CHOIC42, CHOIC53, DISVW31X, DISVW31X, DISVW42X, DISVW53X)

There are several employment-related health insurance measures included in this release: health insurance held from a current main job (HELD31X, HELD42X, HELD53X), health insurance offered from a current main job (OFFER31X, OFFER42X, OFFER53X), and whether the individual had a choice of health plans to choose from at the current main job (CHOIC31, CHOIC42, CHOIC53). The HELD and OFFER variables were logically edited using health insurance information.

Several persons indicated that they held health insurance through a current main job in the employment section and then denied this coverage later in the interview in the health insurance section. Employment section health insurance HELD variables were edited for consistency to match the health insurance measures obtained in the health insurance section. To allow for easy identification of these individuals, round-specific flag variables were constructed (DISVW31X, DISVW42X, DISVW53X).

Responses in the employment section for health insurance held were recoded to be consistent with the variables in the health insurance section of the survey. Due to questionnaire skip patterns, the responses to health insurance offered were affected by editing the HELD variable. For example, if a person responded that health insurance was held from a current main job, the question relating to whether health insurance was offered was skipped. For persons who responded in the employment section that they held health insurance coverage and then disavowed the coverage in the health

insurance section, we could not ascertain whether they were offered a policy. These individuals are coded as –9 for the OFFER variables.

Finally, persons under age 16 as well as persons aged 16 and older who did not hold a current main job or who were self-employed with no employees were coded as inapplicable for the health insurance-related employment variables.

Hours (HOUR31, HOUR42, HOUR53)

Hours worked per week is a combination of two MEPS employment section measures. For salaried persons in Panel 1, the hours measure refers to the hours per week on which the salary is based. For all others, the hours measure refers to usual hours worked per week.

Number of Employees (NUMEMP31, NUMEMP42, NUMEMP53)

Due to confidentiality concerns, the variable indicating the number of employees at the establishment has been top coded at 500 or more employees. NUMEMP indicates the number of employees at the location of the person's current main job. For persons who reported a categorical size, we report a median estimated size from donors within the reported range.

Other Employment Variables

Information about industry and occupation types for a person's current main job at the interview date is also contained in this release. Based on verbatim text fields collected during the interview, industry and occupation types were first coded by trained coders into the three-digit codes defined by the Bureau of the Census for the 1990 Census. For confidentiality reasons, these codes were then condensed. CIND31, CIND42, and CIND53 represent the condensed industry codes for a person's current main job at the interview date. COCCP31, COCCP42 and COCCP53 represent the condensed occupation codes for a person's current main job at the interview date.

Information indicating whether a person belonged to a labor union (UNION31, UNION42, and UNION53) and whether a person worked an irregular work shift (SHFTWK31, SHFTWK42, and SHFTWK53) is also contained in this release. In addition, there are three round specific variables, which show the usual daily start time of the current main job (BGNWK31, BGNWK42, and BGNWK53). There are also three measures of the usual daily ending time of the current main job (ENDWK31, ENDWK42, and ENDWK53). The values for these variables are coded in 24-hour military time and reflect the hours that the respondent reported as the usual starting and ending times. There is an additional allowable value of '95' indicating respondents who reported that their usual start and end times varied.

The day, month, and year that the current main job started for Rounds 3, 4, and 5 of Panel 1 and Rounds 1, 2, and 3 of Panel 2 are provided on this release (STJBDD31, STJBMM31, STJBYY31, STJBDD42, STJBMM42, STJBYY42, STJBDD53, STJBMM53, and STJBYY53).

There are two measures included in this release that relate to a person's work history over a lifetime. One indicates whether a person ever retired from a job as of the Round 5 interview date for Panel 1 persons or the Round 3 interview date for Panel 2 persons (EVRETIRE). The other indicates whether a person ever worked for pay as of the Round 5 interview date for Panel 1 persons or the Round 3 interview date for Panel 2 persons (EVRWRK). The latter was asked of everyone who indicated that they were not working as of the round interview date. Therefore, anyone who indicated current employment or who had a job during any of the previous or current rounds was skipped past the question identifying whether the person every worked for pay. These individuals were coded as inapplicable (-1). The ever retired question was asked of all persons who ever reported a job and were 55 years or older as of the round interview date. Since both of these variables are not round specific, there are no –2 codes.

This release contains variables indicating the main reason a person did not work since the start of the reference period (NWK31, NWK42, and NWK53). If a person was not employed at all the during the reference period (at the interview date or at any time during the reference period) but was employed some time prior to the reference period, the person was asked to choose from a list the main reason he or she did not work during the reference period. The inapplicable (-1) category for the NWK variables includes persons:

- who were employed during the reference period;
- who were not employed during the reference period and who were never employed;
- who were out-of-scope the entire reference period;
- who were less than 16 years old.

A measure of whether an individual had more than one job on the round interview date (MORJOB31, MORJOB42, and MORJOB53) is provided on this release. In addition to those under 16 and those individuals who were out of scope, the inapplicable category includes those who did not report having a current main job. Because this is not a job-specific variable, there are no –2 codes.

This release contains variables indicating if a current main job changed between the third and fourth rounds for Panel 1 persons or between the first and second rounds for Panel 2 persons (CHGJ3142) and between the fourth and fifth rounds for Panel 1 persons or between the second and third rounds for Panel 2 persons (CHGJ4253). In addition to the inapplicable, refused, don't know, and not ascertained categories, the change job variables were coded to represent the following:

- 1 – person left previous round current main job and now has a new current main job;
- 2 – person still working at the previous round's current main job but, as of the new round, no longer considers this job to be the current main job and defines a new main job (previous round's current main job is now a current miscellaneous job);
- 3 – person left previous round's current main job and does not have a new job;
- 4 – person did not change current main job.

Finally, this release contains the reason given by the respondent for the job change (YCHJ3142 and YCHJ4253). The reasons for a job change were listed in the CAPI questionnaire and a respondent was asked to choose the main reason from this list. In addition to those out of scope, those under 16, and those not having a current main job, the inapplicable category for YCHJ3142 and YCHJ4253 includes workers who did not change jobs.

2.5.7 Health Insurance Variables

Constructed and edited variables are provided that indicate any coverage in each month of 1997 for the sources of health insurance coverage collected during the MEPS interviews (Panel 1, Rounds 3 through 5 and Panel 2, Rounds 1 through 3). In Rounds 2, 3, 4, and 5, insurance that was in effect at the previous round's interview date was reviewed with the respondent. Most of the insurance variables have been logically edited to address issues that arose during such reviews in Rounds 2, 3, 4, and 5. One edit to the private insurance variables corrects for a problem concerning covered benefits which occurred when respondents reported a change in any of their private health insurance plan name. Additional edits address issues of missing data on the time period of coverage for both public and private coverage that was either reviewed or initially reported in a given round. For CHAMPUS/CHAMPVA coverage, respondents who were classified as active duty military or who were over age 65 had their reported CHAMPUS/CHAMPVA coverage overturned. Additional edits, described below, were performed on the Medicare and Medicaid variables to assign persons to coverage from these sources. Observations that contain edits assigning persons to Medicare or Medicaid coverage can be identified by comparing the edited and unedited versions of the Medicare and Medicaid variables.

Public sources include Medicare, CHAMPUS/CHAMPVA, Medicaid and other public hospital/physician coverage. State-specific program participation in non-comprehensive coverage (STAJA97-STADE97) was also identified but is not considered health insurance for the purpose of this survey.

Medicare

Medicare (MCRJA97-MCRDE97) coverage was edited (MCRJA97X-MCRDE97X) for persons age 65 or over. Within this age group, individuals were assigned Medicare coverage if:

> They answered yes to a follow-up question on whether or not they received Social Security benefits; or
>
> They were covered by Medicaid, other public hospital/physician coverage or Medigap coverage; or
>
> Their spouse was age 65 or over and covered by Medicare; or
>
> They reported CHAMPUS/CHAMPVA coverage.

Medicaid and Other Public Hospital/Physician Coverage

Questions about other public hospital/physician coverage were asked in an attempt to identify Medicaid recipients who may not have recognized their coverage as Medicaid. These questions were asked only if a respondent did not report Medicaid directly. Respondents reporting other public hospital/physician coverage were asked follow-up questions to determine if their coverage was through a specific Medicaid HMO or if it included some other managed care characteristics. Respondents who identified managed care from either path were asked if they paid anything for the coverage and/or if a government source paid for the coverage.

The Medicaid variables (MCDJA97-MCDDE97) have been edited (MCDJA97X-MCDDE97X) to include persons who paid nothing for their other public hospital/physician insurance when such coverage was through a Medicaid HMO or reported to include some other managed care characteristics. In addition, a small number of persons reporting AFDC or SSI coverage (questions included in the MEPS health insurance sections for this purpose) were assigned Medicaid coverage.

To assist users in further editing sources of insurance, this file contains variables constructed from the other public hospital/physician series that measure whether:

> the respondent reported some type of managed care and paid something for the coverage, Other Public A Insurance (OPAJA97-OPADE97); and
>
> the respondent did not report any managed care, Other Public B Insurance (OPBJA97-OPBDE97).

The variables OPAJA97-OPADE97 and OPBJA97-OPBDE97 are provided only to assist in editing and should not be used to make separate insurance estimates for these types of insurance categories.

Any Public Insurance in Month

The file also includes summary measures that indicate whether or not a sample person has any public insurance in a month (PUBJA97X-PUBDE97X). Persons identified as covered by public insurance are those reporting coverage under CHAMPUS/CHAMPVA, Medicare, Medicaid or other public hospital/physician programs. Persons covered only by state-specific programs that did not provide comprehensive coverage (STAJA97-STADE97), for example, Maryland Kidney Disease Program, were not considered to have public coverage when constructing the variables PUBJA97X-PUBDE97X.

Private Insurance

Variables identifying private insurance in general (PRIJA97-PRIDE97) and specific private insurance sources [such as employer/union group insurance (PEGJA97-PEGDE97); non-group (PNGJA97-PNGDE97); and other group (POGJA97-POGDE97)] were constructed. Private insurance sources identify coverage in effect at any time during each month of 1997. Separate

variables identify covered persons and policyholders (policyholder variables begin with the letter "H"). These variables indicate coverage or policyholder status within a source and do not distinguish between persons who are covered or are policyholders on one or more than one policy within a given source. In some cases, the policyholder was unable to characterize the source of insurance (PDKJA97-PDKDE97). Covered persons (but not policyholders) are identified when the policyholder is living outside the RU (POUJA97-POUDE97). An individual was considered to have private health insurance coverage if, at a minimum, that coverage provided benefits for hospital and physician services (including Medigap coverage). Sources of insurance with missing information regarding the type of coverage were assumed to contain hospital/physician coverage. Persons without private hospital/physician insurance were not counted as privately insured.

Health insurance through a job or union (PEGJA97-PEGDE97, PRSJA97-PRSDE97) was initially asked about in the Employment Section of the interview and later confirmed in the Health Insurance Section. Respondents also had an opportunity to report employer and union group insurance (PEGJA97-PEGDE97) for the first time in the Health Insurance Section, but this insurance was not linked to a specific job.

All insurance reported to be through a job classified as self-employed with firm size of 1 (PRSJA97-PRSDE97) was initially reported in the Employment Section and verified in the Health Insurance Section. Unlike the other employment-related variables (PEGJA97-PEGDE97), self-employed-firm size 1 (PRSJA97-PRSDE97) health insurance could not be reported in the Health Insurance section for the first time. The variables PRSJA97-PRSDE97 have been constructed to allow users to determine if the insurance should be considered employment-related.

Private insurance that was not employment-related (POGJA97-POGDE97, PNGJA97-PNGDE97, PDKJA97-PDKDE97 and POUJA97-POUDE97) was reported in the Health Insurance Section only.

Any Insurance in Month

The file also includes summary measures that indicate whether or not a sample person has any insurance in a month (INSJA97X-INSDE97X). Persons identified as insured are those reporting coverage under CHAMPUS/CHAMPVA, Medicare, Medicaid or other public hospital/physician or private hospital/physician insurance (including Medigap plans). A person is considered uninsured if not covered by one of these insurance sources.

Persons covered only by state-specific programs that provide non-comprehensive coverage (STAJA97-STADE97), for example, Maryland Kidney Disease Program, and those without hospital/physician benefits (for example, private insurance for dental or vision care only, accidents or specific diseases) were not considered to be insured when constructing the variables INSJA97X-INSDE97X.

1997 Summary Insurance Coverage Indicators (PRVEV97 - INSCOV97)

The variables PRVEV97-UNINS97 summarize health insurance coverage for the person in 1997

for the following types of insurance: private (PRVEV97); CHAMPUS/CHAMPVA (CHPEV97); Medicaid (MCDEV97); Medicare (MCREV97); other public A (OPAEV97); other public B (OPBEV97). Each variable was constructed based on the values of the corresponding 12 month to month health insurance variables described above. A value of 1 indicates that the person was covered for at least one day of at least one month during 1997. A value of 2 indicates that the person was not covered for a given type of insurance for all of 1997. The variable UNINS97 summarizes PRVEV97-OPBEV97. Where PRVEV97-OPBEV97 are all equal to 2, then UNINS97 equals 1; person was uninsured for all of 1997. Otherwise UNINS97 is set to 2, not uninsured for some portion of 1997.

For user convenience this file contains a constructed variable INSCOV97 that summarizes health insurance coverage for the person in 1997, with the following 3 values:

1 = ANY PRIVATE (Person had any private insurance coverage **(including Champus/VA)** any time during 1997)
2 = PUBLIC ONLY (Person had only public insurance coverage during 1997)
3 = UNINSURED (Person was uninsured during all of 1997)

Please note this variable categorizes Champus as private coverage. If an analyst wishes to consider Champus public coverage, the variable can easily be reconstructed using the PRVEV97 and CHMPEV97 variables.

2.5.8 Health Status Variables

This data release incorporates information from calendar year 1997. However, health status data obtained in Round 3 of Panel 2 are also included in variables that have names ending in “53”. Panel 2 Round 3 extended into 1998. Therefore, for variables that have names ending in “53”, some information from early 1998 is included.

Health status variables in this data release can be classified into several conceptually distinct sets:

1) Perceived health status and ADL and IADL limitations
2) Functional limitations and activity limitations
3) Vision problems
4) Hearing problems
5) Children’s health status

Variables in the first set were measured in all Rounds. Variables in set 2 were measured in Rounds 3 and 5 for Panel 1 and Rounds 1 and 3 for Panel 2. Variables in sets 3, 4, and 5 were measured only in Round 4 for Panel 1 and Round 2 for Panel 2.

In general, Health Status variables involved the construction of person-level variables based on information collected in the Condition Enumeration and Health Status sections of the questionnaire. Many Health Status questions were initially asked at the family level to ascertain if

anyone in the household had a particular problem or limitation. These were followed up with questions to determine which household member had each problem or limitation. All information ascertained at the family level has been brought to the person level for this file. Logical edits were performed in constructing the person-level variables to assure that family-level and person-level values were consistent. Particular attention was given to cases where missing values were reported at the family level, to ensure that appropriate information was carried to the person level.

Inapplicable cases occurred when a question was never asked because of a skip pattern in the survey (e.g., individuals who were 13 years of age or older were not asked some follow-up verification questions; individuals older than 17 were not asked questions pertaining to children's health status). Inapplicable cases are coded as -1. In addition, deceased persons were coded as inapplicable and received a code of -1.

Each of the sets of variables listed above will be described in turn.

2.5.8.1 Perceived Health Status and ADL and IADL Limitations

Perceived Health Status. Perceived health status (RTHLTH31, RTHLTH42, and RTHLTH53) and perceived mental health status (MNHLTH31, MNHLTH42**,** and MNHLTH53) were collected in the Condition Enumeration section. These questions (CE01 and CE02) asked the respondent to rate each person in the family according to the following categories: excellent, very good, good, fair, and poor. The corresponding dichotomous variables RTPROX31, RTPROX42, RTPROX53, MNPROX31, MNPROX42**,** and MNPROX53 each indicate whether the ratings of physical and mental health were provided by oneself or by someone else.

IADL Help. The Instrumental Activities of Daily Living (IADL) Help or Supervision variables (IADLHP31, IADLHP42**,** and IADLHP53) were each constructed from a series of three questions administered in the Health Status section of the interview. The initial question (HE01) determined if anyone in the family received help or supervision with IADLs such as using the telephone, paying bills, taking medications, preparing light meals, doing laundry, or going shopping. If the response was "yes," a follow-up question (HE02) was asked to determine which household member received this help or supervision. For persons under age 13, a final verification question (HE03) was asked to confirm that the IADL help or supervision was the result of an impairment or physical or mental health problem. If the response to the final verification question was "no," IADLHP31**,** IADLHP42, and IADLHP53 were coded "no" for persons under the age of 13.

If no one in the family was identified as receiving help or supervision with IADLs, all members of the family were coded as receiving no IADL help or supervision. In cases where the response to the family-level question was "refused" (-7), "don't know" (-8), or not ascertained (-9), all persons were coded according to the family-level response. In cases where the response to the family-level question (HE01) was "yes" but no specific individuals were identified in the follow-up question as having IADL difficulties, all persons were coded as "don't know" (-8).

ADL Help. The Activities of Daily Living (ADL) Help or Supervision variable (ADLHLP31, ADLHLP42**,** and ADLHLP53) were each constructed in the same manner as the IADL help variables, but using questions HE04-HE06. Coding conventions for missing data were the same as for the IADL variables.

2.5.8.2 Functional and Activity Limitations

Functional Limitations. A series of questions pertained to functional limitations, defined as difficulty in performing certain specific physical actions. WLKLIM31 and WLKLIM53 were the filter questions, depending on the Round. These variables were derived from a question (HE09) that was asked at the family level: "Does anyone in the family have difficulties walking, climbing stairs, grasping objects, reaching overhead, lifting, bending or stooping, or standing for long periods of time?" If the answer was "no" then all family members were coded as "no" (2) on WLKLIM31 or WLKLIM53. If the answer was "yes," then the specific persons who had any of these difficulties were identified and coded as "yes" (1), and remaining family members were coded as "no". If the response to the family-level question was "don't know" (-8), "refused" (-7), "missing" (-9), or "inapplicable" (-1), then the corresponding missing value code was applied to each family member's value for WLKLIM31 or WLKLIM53. If the answer to HE09 was "yes," but no specific individual was named as experiencing such difficulties, then each family member was assigned -8. Deceased respondents were assigned a -1 code ("inapplicable") for WLKLIM31 or WLKLIM53.

For Round 3 (Panel 1) and Round 1 (Panel 2), if any family member was coded "yes" to WLKLIM31 a subsequent series of questions was administered. The series of questions for which WLKLIM31 served as a filter was as follows:

LFTDIF31 - difficulty lifting 10 pounds
STPDIF31 - difficulty walking up 10 steps
WLKDIF31 - difficulty walking 3 blocks
MILDIF31 - difficulty walking a mile
STNDIF31 - difficulty standing 20 minutes
BENDIF31 - difficulty bending or stooping
RCHDIF31 - difficulty reaching over head
FNGRDF31 - difficulty using fingers to grasp

The series of questions was asked separately for each person who was coded "yes" to WLKLIM31. The series of questions was not asked for other individual family members for whom WLKLIM31 was "no." In addition, this series was not asked about family members who were less than 13 years of age, regardless of their status on WLKLIM31. Finally, these questions were not asked about deceased family members. In such cases (i.e., WLKLIM31 = 2, or age < 13, or PSTATS31 = 31), each question in the series was coded as "inapplicable" (-1). Finally, if responses to WLKLIM31 were "refused" (-7), "don't know" (-8), "not ascertained" (-9), or otherwise inapplicable (-1), then each question in this series was coded as "inapplicable" (-1).

Analysts should note that, for WLKLIM31, there was no minimum age criterion that was used to determine a skip pattern, whereas, for the subsequent series of questions, persons less than 13 years old were skipped and coded as "inapplicable". Therefore, it is possible for someone aged 12 or less to have a code of 1 ("yes") on WLKLIM31, and also to have codes of "inapplicable" on the subsequent series of questions.

For Round 5 (Panel 1) and Round 3 (Panel 2), the corresponding filter question was WLKLIM53. The series of questions for which WLKLIM53 served as a filter was as follows:

LFTDIF53 - difficulty lifting 10 pounds
STPDIF53 - difficulty walking up 10 steps
WLKDIF53 - difficulty walking 3 blocks
MILDIF53 - difficulty walking a mile
STNDIF53 - difficulty standing 20 minutes
BENDIF53 - difficulty bending or stooping
RCHDIF53 - difficulty reaching over head
FNGRDF53 - difficulty using fingers to grasp

Editing conventions were the same for this set of variables as they were for the corresponding set described above.

Use of Assistive Technology and Social/Recreational Limitations . The variables indicating use of assistive technology (AIDHLP31 and AIDHLP53, from question HE07) and social/recreational limitations (SOCLIM31 and SOCLIM53, from question HE22) were collected initially at the family level. If there was a "yes" response to the family-level question, a second question identified which specific individual(s) the "yes" response pertained to. Each individual identified as having the difficulty was coded "yes" on the appropriate variable; all remaining family members were coded "no." If the family-level response was "don't know" (-7), "refused" (-8), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was "yes" but no specific individual was identified as having difficulty, all family members were coded as "don't know" (-8).

Work, Housework, and School Limitations . The variables indicating any limitation in work, housework, or school (ACTLIM31 and ACTLIM53) were constructed using questions HE19-HE20. Specifically, information was collected initially at the family level. If there was a "yes" response to the family-level question (HE19), a second question (HE20) identified which specific individual(s) the "yes" response pertained to. Each individual identified as having a limitation was coded "yes"; all remaining family members were coded "no." If the family-level response was "don't know"(-7), "refused" (-8), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was "yes" but no specific individual was identified as having difficulty, all family members were coded as "don't know" (-8). Persons less than five years old were coded as inapplicable (-1) on ACTLIM31 and ACTLIM53.

For Round 3 (Panel 1) or Round 1 (Panel 2) if ACTLIM31 was “yes” and the person was 5 years of age or older, a follow-up question (HE20A) was asked to identify the specific limitation or limitations for each person. These included working at a job (WRKLIM31), doing housework (HSELIM31), or going to school (SCHLIM31). Respondents could answer “yes” to each activity; one person could thus report limitation in multiple activities. WRKLIM31, HSELIM31, and SCHLIM31 have values of “yes” or “no” only if ACTLIM31 was “yes;” each variable was coded as inapplicable (-1) if ACTLIM31 was “no,” “refused” (-7), or not ascertained (-9). When ACTLIM31 was “don’t know” (-8), these variables were all coded as “don’t know” (-8). If a person was under 5 years old or was deceased, WRKLIM31, HSELIM31, and SCHLIM31 were each coded as “inapplicable” (-1).

A second question (UNABLE31) asked if the person was completely unable to work at a job, do housework, or go to school. This question was asked only of the same set of respondents who provided data on WRKLIM31, HSELIM31, and SCHLIM31. Therefore, those respondents who were coded “no” on ACTLIM31, or were under 5 years of age, or were deceased, were coded as inapplicable (-1) on UNABLE31. UNABLE31 was asked once for whichever set of WRKLIM31, HSELIM31, and SCHLIM31 the respondent had limitations; if a respondent was limited in more than one of these three activities, UNABLE31 did <u>not</u> specify if the respondent was completely unable to perform all of them, or only some of them.

For Rounds 5 (Panel 1) and 3 (Panel 2) corresponding variables were ACTLIM53, WRKLIM53, HSELIM53, SCHLIM53, and UNABLE53. Editing conventions were the same as those described above.

<u>Cognitive Limitations</u> . The variables indicating any cognitive limitation (COGLIM31 or COGLIM53, depending on the round) were collected at the family level as a three-part question (HE24-01 to HE24-03) indicating if any of the adults in the family (1) experience confusion or memory loss, (2) have problems making decisions, or (3) require supervision for their own safety. If a “yes” response was obtained to any item, the persons affected were identified in HE25 and COGLIM31 was coded as “yes.” Remaining family members not identified were coded as “no” for COGLIM31.

If responses to HE24-01 though HE24-03 were all “no,” or if two of three were “no” and the remaining was “don’t know,” “refused,” or not ascertained, all family members were coded as “no.” If responses to the three questions were combinations of “don’t know,” “refused,” and missing, all persons were coded as “don’t know” (-8). If the response to any of the three questions was “yes” but no individual was identified in HE25, all persons were coded as “don’t know” (-8).

Analogous editing specifications were implemented for COGLIM53.

The cognitive limitations variables (COGLIM31 or COGLIM53) reflect whether <u>any</u> of the three component questions is “yes.” Respondents with one, two, or three specific cognitive limitations

cannot be distinguished. In addition, because the question asked specifically about adult family members, all persons less than 18 years of age are coded as inapplicable (-1) on this question.

2.5.8.3 Vision Problems

A series of questions (HE26 to HE32) provides information on visual impairment. These questions were asked of all household members, regardless of age. Deceased respondents were coded as not applicable (-1).

WRGLAS42 indicates whether a person wears eyeglasses or contact lenses. This variable was based on two questions, HE26 and HE27. The initial question (HE26) determined if anyone in the family wore eyeglasses or contact lenses. If the response was “yes,” a follow-up question (HE27) was asked to determine which household member(s) wore eyeglasses or contact lenses. If the family-level response was “don’t know”(-8), “refused” (-7), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was “yes” but no specific individual was identified as wearing glasses or contact lenses, all family members were coded as “don’t know” (-8).

SEEDIF42 indicates whether anyone in the family had difficulty seeing (with glasses or contacts, if used). This variable was based on two questions, HE28 and HE29. The initial question (HE28) determined if anyone in the family had difficulty seeing. If the response was "yes," a follow-up question (HE29) was asked to determine which household member(s) had a visual impairment. If the family-level response was “don’t know”(-8), “refused” (-7), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was “yes” but no specific individual was identified as having difficulty seeing, all family members were coded as “don’t know” (-8).

Three subsequent questions were asked only for individuals who had difficulty seeing (i.e., SEEDIF42 = 1). Persons with no visual impairment were coded as not applicable (-1) for these questions, as were persons with don’t know (-8), refused (-7), or not ascertained (-9) responses to SEEDIF42. BLIND42 determined if a person with difficulty seeing was blind. For persons who were not blind (BLIND42 = 2), READNW42 asked whether the person could see well enough to read ordinary newspaper print (with glasses or contacts, if used); persons who were blind were not asked this question and were coded as not applicable (-1). For persons who could not read ordinary newspaper print (READNW42 = 2), RECPEP42 asked if the person could see well enough to recognize familiar people standing two or three feet away. Persons who were blind or who could read newsprint were not asked this question and were coded as not applicable (-1).

VISION42 summarizes the pattern of responses to the set of visual impairment questions. Codes for VISION42 are as follows:

1 - No difficulty seeing (SEEDIF42 = 2)
2 - Some difficulty seeing, can read newsprint (SEEDIF42 = 1 and READNW42 = 1)
3 - Some difficulty seeing, can not read newsprint, can recognize familiar people

(SEEDIF42 = 1 and READNW42 = 2 and RECPEP42 = 1)
4 - Some difficulty seeing, can not read newsprint, can not recognize familiar people but is not blind (SEEDIF42 =1 and READNW42 = 2 and RECPEP42 = 2)
5 - Blind (SEEDIF42 = 1 and BLIND42 = 1).

2.5.8.4 Hearing Problems

A series of questions (HE33 to HE39) provides information on hearing impairment. These questions were asked of all household members, regardless of age. Deceased respondents were coded as not applicable (-1).

HEARAD42 indicates whether a person wears a hearing aid. This variable was based on two questions, HE33 and HE34. The initial question (HE33) determined if anyone in the family wore a hearing aid. If the response was “yes,” a follow-up question (HE34) was asked to determine which household member(s) wore a hearing aid. If the family-level response was “don’t know”(-8), “refused” (-7), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was “yes” but no specific individual was identified as wearing a hearing aid, all family members were coded as “don’t know” (-8).

HEARDI42 indicates whether anyone in the family had difficulty hearing (with a hearing aid, if used). This variable is based on two questions, HE35 and HE36. The initial question (HE35) determined if anyone in the family had difficulty hearing. If the response was “yes,” a follow-up question (HE36) was asked to determine which household member had an aural impairment. If the family-level response was “don’t know”(-8), “refused” (-7), or not ascertained (-9), all persons were coded with the family-level response. In cases where the family-level response was “yes” but no specific individual was identified as using a hearing aid, all family members were coded as “don’t know” (-8).

Three subsequent questions were asked only for individuals who had difficulty hearing (i.e., HEARDI42 = 1). Persons with no hearing impairment were coded as inapplicable (-1) for these questions, as were persons with don’t know (-8), refused (-7), or not ascertained (-9) responses to HEARDI42. DEAF42 determined if a person with difficulty hearing was deaf. For persons who were not deaf (DEAF42 = 2), HEARMO42 asked whether the person could hear well enough to hear most of the things people say (with a hearing aid, if used); persons who were deaf were not asked this question and were coded as not applicable (-1). For persons who could not hear most things people say (HEARMO42 = 2), HEARSM42 asked if the person could hear well enough to hear some of the things that people say. Persons who were deaf or who could hear most conversation were not asked this question and were coded as inapplicable (-1).

HEARNG42 summarizes the pattern of responses to the set of hearing impairment questions. Codes for HEARNG42 are as follows:

1 - No difficulty hearing (HEARDI42 = 2)
2 - Some difficulty hearing, can hear most things people say (HEARDI42 = 1 and

HEARMO42 = 1)

3 - Some difficulty hearing, can not hear most things people say, can hear some things people say (HEARDI42 = 1 and HEARMO42 = 2 and HEARSM42 = 1)

4 - Some difficulty hearing, can not hear most things people say, can not hear some things people say, but not deaf (HEARDI42 =1 and HEARMO42 = 2 and HEARSM42 = 2)

5 - Deaf (HEARDI42 = 1 and DEAF42 = 1).

2.5.8.5 Any Limitation Rounds 3, 4, and 5 (Panel 1) / Rounds 1, 2, and 3 (Panel 2)

ANYLIM97 summarizes whether the respondent has any ADL, IADL, activity, functional, or sensory limitations. This variable was derived based on data from Rounds 3, 4, and 5 (Panel 1) or Rounds 1, 2, and 3 (Panel 2). ANYLIM97 was built upon component variables IADLHP31, IADLHP42, IADLHP53, ADLHLP31, ADLHLP42, ADLHLP53, WLKLIM31, WLKLIM53, ACTLIM31, ACTLIM53, SEEDIF42, and HEARDIF42. (The latter two variables, discussed above, indicate any visual or hearing impairment, respectively.) If any of these components was coded “yes”, then ANYLIM97 was coded “yes” (1). If all components equaled “no”, then ANYLIM97 equaled “no” (2). If all the components had missing value codes (i.e., -7, -8, -9, or –1), then ANYLIM97 was coded as not ascertained (-9). If some components were “no” and others had missing value codes, ANYLIM97 was coded as not ascertained (-9). The exception to this latter rule was for children less than five years old, who did not receive the ACTLIM31 or ACTLIM53 questions; for these respondents, if all other components were “no”, then ANYLIM97 was coded as “no” (2).

The variable label for ANYLIM97 departs slightly from conventions. Typically, variables that end in "97" refer only to 1997. However, because some of the variables utilized to construct ANYLIM97 spanned 1997 and 1998, some information from early 1998 is included in this variable.

2.5.8.6 Children’s Health Status

Play Limitations (Children age 4 and under). The variable LIMACT42, indicating limitation in activities for children ages 0 through 4, was constructed using questions HE40 and HE41. The initial question (HE40) determined if any child aged 4 or under in the family was limited in any way, including play activity, because of an impairment or physical or mental health problem. If the response was “yes,” the follow-up question determined which child should be coded “yes.” If there were other children aged 4 or under in the family who were not identified as having limitations, they were coded “no.” If the answer to LIMACT42 was “no,” all children aged four or under in the family were coded “no.” If there was an indication that a child had a limitation, but no child was identified, all children within the age category were coded “don’t know” (-8). In cases where the response to the family-level question was “don’t know” (-8), refused (-7), or not ascertained (-9), all children ages 4 and under were coded according to the family-level response. If a person's age (as measured by the Panel 1 Round 4/Panel 2 Round 2 age variable) was greater than 4, LIMACT42 was coded -1.

Other variables indicate if children aged 0 to 4 were limited in the kind or amount of play activities (PLYLIM42), were unable to play (CANTPL42), or participated in special programs or early interventions (SPCPRO42). If a person aged 4 or under had no activity limitations (i.e., LIMACT42 = 2), then PLYLIM42, CANTPL42, and SPCPRO42 were each coded -1. If a person's age (as measured by the Panel 1 Round 4/Panel 2 Round 2 age variable) was greater than 4, PLYLIM42, CANTPL42, and SPCPRO42 were coded -1.

Immunization Variables (Children ages 0 through 6). Immunization information was collected at the person level for children ages 0 through 6 by questions HE45 to HE49A. If age of child, as measured by the Panel 1 Round 4/Panel 2 Round 2 age variable, was greater than 6, all immunization variables were coded -1. For questions about diphtheria, whooping cough and tetanus (DPT) or polio immunization (DPTSHT42, POLSHT42), there were follow up questions that asked about the frequency of the immunization shots or drops (NUMDPT42, NUMPOL42). If the answer to DPTSHT42 or POLSHT42 was "no," "don't know," or "refused," the respective follow-up variables NUMDPT42 and NUMPOL42 were coded -1. For questions about immunization for measles/mumps/rubella (MMRSHT42) and for hepatitis (HEPSHT42), there were no follow-up questions.

Behavioral Problem Variables (Children ages 5 to 17) The series of questions HE50_01 to HE50_13 inquired about possible child behavioral problems. Variables in this set include:

MOMPRO42: problem getting along with mother

DADPRO42: problem getting along with father

UNHAP42: feeling unhappy or sad

SCHLBH42: problem with behavior at school

HAVFUN42: problem having fun

ADUPRO42: problem getting along with adults

NERVAF42: problem with child feeling nervous or afraid

SIBPRO42: problem getting along with siblings

KIDPRO42: problem getting along with other kids

SPRPRO42: problem engaging in sports or hobbies

SCHPRO42: problem doing schoolwork

HOMEBH42: problem with behavior at home

TRBLE42: problem staying out of trouble.

If the age of the child (as measured by the Panel 1 Round 4/Panel 2 Round 2 age variable) was less than 5 or greater than 17, the variables MOMPRO42 to TRBLE42 were coded -1.

Certain questions in this series could be inapplicable for a specific child. For example, if a child's mother was deceased, a question about how a child gets along with his/her mother is inapplicable. Similarly, the question about problems getting along with siblings would be inapplicable for only children. In such instances, the relevant variable was coded 99 to indicate that it was inapplicable.

Special Education and Special Services (Children ages 5-17). A series of questions asked about participation in special education programs or receipt of therapy or special services. If the respondent was not in the age range of 5-17 years of age (as measured by the Panel 1 Round 4/Panel 2 Round 2 age variable), or if the respondent was deceased, these questions were coded as inapplicable (-1).

SPCSCH42 is based on question HE51, which asked whether the child had an impairment or a physical or mental health problem that limited school attendance or required a special school program. This question served as a filter for subsequent questions. If the response "no"(2), "refused" (-7), or "don't know" (-8), then SPECED42 through CANTSC42 were coded as inapplicable (-1).

If the response to SPCSCH42 was "yes" (1) then question HE52 (SPECED42) was asked. SPECED42 asked whether the child was enrolled in any type of special education or received related services. Possible responses to this question were "yes, enrolled in special education" (1), "yes, enrolled in related services," (2), "yes, both special education and special services," (3), "no" (4), and "other"(91).

If responses to SPECED42 were coded as 2 or 3, then respondents were presented with a list of other related services and asked to indicate which one(s) the child had received. Respondents could indicate more than one type of service. These questions constitute variables SPCHTH42 to OTHSVC42.

SPCHTH42: Received speech therapy
OCUPTH42: Received occupational therapy
VOCSVC42: Received vocational services
TUTOR42: Received tutoring
READIN42: Uses a reader or interpreter
PHYTHR42: Received physical therapy
LIFSKL42: Received life skills training
PSYCNS42: Received psychological counseling
FAMCNS42: Received family counseling
RECTH4R2: Received recreational therapy
OTHSVC42: Received other school services

Responses to these questions were coded as inapplicable (-1) if the response to SPECED42 was 1 (enrolled in special education only), or -7 (refused), or -8 (don't know).

If the response to SPCSCH42 was “yes”(1) then question HE53 (CANTSC42) was asked. This question asked whether the child was limited in attendance or unable to attend school due to an impairment or a physical or mental health problem. Responses of “limited in attendance” were coded 1, “unable to attend” as 2, and “neither” as 3.

Question HE54 (LMOACT42) was asked of all children ages 5-17. This question ascertained whether the child was limited in any way in activities other than school because of an impairment or a physical or mental health problem.

Children’s Health Status: General Questions (ages 0 - 17)

Several questions were asked about all children ages 0 through 17. Respondents who were older than 17 or who were deceased were coded as not applicable (-1) for these variables. Three questions asked for ratings of the child’s health on a 4-point Likert scale, ranging from “definitely false”(1) to “definitely true”(4). These questions were:

HLTHY42: Child resists illness.
NTHLTH42: Child seems to be less healthy than other children.
GETSIC42: Child seems to catch diseases that are going around.

In addition, information was provided on each child’s height in feet (HGTFT42) and inches (HGTIN42), as well as each child’s weight in pounds (WGTLB42) and in ounces (WGTOZ42).

Finally, CHLIM42 was constructed to reflect each child’s inability to perform age-appropriate social roles. For children aged 0 to 4, this variable was based on responses to LIMACT42, PLYLIM42 and CANTPL42; for children aged 5-17, it was based on responses to SPCSCH42, CANTSC42 and LMOACT42. If any one of these variables had a “yes” response (i.e., codes of 1 for LIMACT42, PLYLIM42, CANTPL42, SPCSCH42, or LMOACT42, or codes of 1 or 2 for CANTSC42), then CHLIM42 was coded as “yes”(1). If the relevant variables were all “no”, then CHLIM42 was coded as “no”(2). CHLIM42 was coded as “not ascertained”(-9) if the relevant variables were combinations of “refused”(-7), “don’t know”(-8), or not ascertained (-9).

2.5.9 Utilization, Expenditures and Source of Payment Variables (TOTTCH97-RXOSR97)

The MEPS Household Component (HC) collects data in each round on use and expenditures for office and hospital-based care, home health care, dental services, vision aids, and prescribed medicines. Data were collected for each sample person at the event level (e.g. doctor visit, hospital stay) and summed across rounds 1-3 (excluding 1997 events covered in round 3) to produce the annual utilization and expenditure data for 1997 in this file. In addition, the MEPS Medical Provider Component (MPC) is a follow-back survey that collected data from a sample of medical providers and pharmacies that were used by sample persons in 1997. Expenditure data collected in the MPC are generally regarded as more accurate than information collected in

the HC and were used to improve the overall quality of MEPS expenditure data in this file (see below for description of methodology used to develop expenditure data).

This file contains utilization and expenditure variables for several categories of health care services. In general, there is one utilization variable (based on HC responses only), 13 expenditure variables (derived from both HC and MPC responses), and 1 charge variable for each category of health care service. The utilization variable is typically a count of the number of medical events reported for the category. The 13 expenditure variables consist of an aggregate total payments variable, 10 main component source of payment category variables, and 2 additional source of payment category variables (see below for description of source of payment categories). Expenditure variables for all categories of health care combined are also provided.

The table in Appendix 3 provides an overview of the utilization and expenditure variables included in this file. For each health service category, the table lists the corresponding utilization variable(s) and provides a general key to the expenditure variable names (13 per service category). The first 3 characters of the expenditure variable names reflect the service category (except only 2 characters for prescription medicines) while the subsequent 3 characters (*** in table) reflect the naming convention for the source of payment categories described below (except only 2 characters for Veterans Administration). The last 2 positions of all utilization and expenditure variable names reflect the survey year (i.e. 97). More details are provided on the utilization and expenditure variables in sections 2.5.9.1 and 2.5.9.2 below.

2.5.9.1 Expenditures Definition

Expenditures on this file refer to what is paid for health care services. More specifically, expenditures in MEPS are defined as the sum of direct payments for care provided during the year, including out-of-pocket payments and payments by private insurance, Medicaid, Medicare, and other sources. Payments for over the counter drugs and for alternative care services are not included in MEPS total expenditures. Indirect payments not related to specific medical events, such as Medicaid Disproportionate Share and Medicare Direct Medical Education subsidies, are also not included.

The definition of expenditures used in MEPS is somewhat different from the 1987 NMES and 1977 NMCES surveys where “charges” rather than “sum of payments” were used to measure expenditures. This change was adopted because charges became a less appropriate proxy for medical expenditures during the 1990’s due to the increasingly common practice of discounting charges. Another change from the two prior surveys is that charges associated with uncollected liability, bad debt, and charitable care (unless provided by a public clinic or hospital) are not counted as expenditures because there are no payments associated with those classifications.

While the concept of expenditures in MEPS has been operationalized as payments for health care services, variables reflecting charges for services received are also provided on the file (see below). Analysts should use caution when working with the charge variables because they do not typically represent actual dollars exchanged for services or the resource costs of those services.

Data Sources on Expenditures

The expenditure data included on this file were derived from the MEPS Household and Medical Provider Components. Only HC data were collected for nonphysician visits, dental and vision services, other medical equipment and services, and home health care not provided by an agency while data on expenditures for care provided by home health agencies were only collected in the MPC. In addition to HC data, MPC data were collected for some office-based visits to physicians (or medical providers supervised by physicians), hospital-based events (e.g. inpatient stays, emergency room visits, and outpatient department visits), and prescribed medicines. For these types of events, MPC data were used if complete; otherwise HC data were used if complete. Missing data for events where HC data were not complete and MPC data were not collected or complete were derived through an imputation process (see below).

A series of logical edits were applied to both the HC and MPC data to correct for several problems including outliers, copayments or charges reported as total payments, and reimbursed amounts that were reported as out of pocket payments. In addition, edits were implemented to correct for misclassifications between Medicare and Medicaid and between Medicare HMO's and private HMO's as payment sources. Data were not edited to insure complete consistency between the health insurance and source of payment variables on the file.

Imputation for Missing Expenditures and Data Adjustments

Expenditure data were imputed to 1) replace missing data, 2) provide estimates for care delivered under capitated reimbursement arrangements, and 3) to adjust household reported insurance payments because respondents were often unaware that their insurer paid a discounted amount to the provider. This section contains a general description of the approaches used for these three situations. A more detailed description of the editing and imputation procedures will be provided in the documentation for the forthcoming MEPS event level files.

Missing data on expenditures were imputed using a weighted sequential hot-deck procedure for most medical visits and services. In general, this procedure imputes data from events with complete information to events with missing information but similar characteristics. For each event type, selected predictor variables with known values (e.g., total charge, demographic characteristics, region, provider type, and characteristics of the event of care, such as whether it involved surgery) were used to form groups of donor events with known data on expenditures, as well as identical groups of recipient events with missing data. Within such groups, data were assigned from donors to recipients, taking into account the weights associated with the MEPS complex survey design. Only MPC data were used as donors for hospital-based events while data from both the HC and MPC were used as donors for office-based physician visits. The general approach that was used to impute missing expenditure data on prescribed medicines is described in section 2.5.9.2 below.

Because payments for medical care provided under capitated reimbursement arrangements and through public clinics and Veterans' Hospitals are not tied to particular medical events,

expenditures for events covered under those types of arrangements and settings were also imputed. Events covered under capitated arrangements were imputed from events covered under managed care arrangements that were paid based on a discounted fee-for-service method, while imputations for visits to public clinics and Veterans' Hospitals were based on similar events that were paid on a fee-for-service basis. As for other events, selected predictor variables were used to form groups of donor and recipient events for the imputations.

An adjustment was also applied to some HC reported expenditure data because an evaluation of matched HC/MPC data showed that respondents who reported that charges and payments were equal were often unaware that insurance payments for the care had been based on a discounted charge. To compensate for this systematic reporting error, a weighted sequential hot-deck imputation procedure was implemented to determine an adjustment factor for HC reported insurance payments when charges and payments were reported to be equal. As for the other imputations, selected predictor variables were used to form groups of donor and recipient events for the imputation process.

Methodology for Flat Fee Expenditures

Most of the expenditures for medical care reported by MEPS participants are associated with single medical events. However, in some situations there is one charge that covers multiple contacts between a medical provider and patient (e.g. obstetrician services, orthodontia). In these situations (generally called flat or global fees), total payments for the flat or global fee were included if the initial service was provided in 1997. For example, all payments for an orthodontist's fee that covered multiple visits over three years were included if the initial visit occurred in 1997. However, if a visit in 1997 to an orthodontist was part of a flat fee in which the initial visit occurred in 1995, then none of the payments for the flat fee were included.

The approach used to count expenditures for flat fees may create what appear to be inconsistencies between utilization and expenditure variables. For example, if several visits under a flat fee arrangement occurred in 1997 but the first visit occurred in 1995, then none of the expenditures were included, resulting in low expenditures relative to utilization for that person. Conversely, the flat fee methodology may result in high expenditures for some persons relative to their utilization. For example, all of the expenditures for an expensive flat fee were included even if only the first visit covered by the fee had occurred in 1997. On average, the methodology used for flat fees should result in a balance between overestimation and underestimation of expenditures in a particular year.

Zero Expenditures

There are some medical events reported by respondents where the payments were zero. This could occur for several reasons including (1) free care was provided, (2) bad debt was incurred, (3) care was covered under a flat fee arrangement beginning in an earlier year, or (4) follow-up visits were provided without a separate charge (e.g. after a surgical procedure). In summary, these types of events have no impact on the person level expenditure variables contained in this file.

Source of Payment Categories

In addition to total expenditures, variables are provided which itemize expenditures according to the major source of payment categories. These categories are:

1. Out of pocket by user or family (SLF);
2. Medicare (MCR);
3. Medicaid (MCD);
4. Private Insurance (PRV);
5. Veterans' Administration, excluding CHAMPVA (VA);
6. CHAMPUS (i.e. TRICARE) or CHAMPVA (CHM);
7. Other Federal Sources--includes Indian Health Service, Military Treatment Facilities, and other care provided by the Federal government (OFD);
8. Other State and Local Source--includes community and neighborhood clinics, State and local health departments, and State programs other than Medicaid (STL);
9. Worker's Compensation (WCP);
10. Other Unclassified Sources--includes sources such as automobile, homeowner's, liability, and other miscellaneous or unknown sources (OSR).

Two additional source of payment variables were created to classify payments for particular persons that appear inconsistent due to differences between the survey questions on health insurance coverage and sources of payment for medical events. These variables include:

11. Other Private (OPR)—any type of private insurance payments reported for persons not reported to have any private health insurance coverage during the year as defined in MEPS (i.e. for hospital and physician services); and

12. Other Public (OPU)—Medicaid payments reported for persons who were not reported to be enrolled in the Medicaid program at any time during the year.

Though relatively small in magnitude, users should exercise caution when interpreting the expenditures associated with the OPR and OPU categories. While these payments stem from apparent inconsistent responses to the health insurance and source of payment questions in the survey, some of these inconsistencies may have logical explanations. For example, private insurance coverage in MEPS is defined as having a major medical plan covering hospital and physician services. If a MEPS sample person did not have such coverage but had a single service type insurance plan (e.g. dental insurance) that paid for a particular episode of care, those payments may be classified as "other private". Some of the "other public" payments may stem from confusion between Medicaid and other state and local programs or may be for persons who were not enrolled in Medicaid, but were presumed eligible by a provider who ultimately received payments from the program.

The naming conventions used for the source of payment expenditure variables are shown in

parentheses in the list of categories above and in the key to the attached table in Appendix 3. In addition, total expenditure variables (EXP in key) based on the sum of the 12 source of payment variables above are provided.

Charge Variables

In addition to the expenditure variables described above, a variable reflecting total charges is provided for each type of service category (**except prescribed medicines**). This variable represents the sum of all fully established charges for care received and usually does not reflect actual payments made for services, which can be substantially lower due to factors such as negotiated discounts, bad debt, and free care (see above). The naming convention used for the charge variables (TCH) is also included in the key to the attached table in Appendix 3. The total charge variable across services (TOTTCH97) excludes prescribed medicines.

2.5.9.2 Utilization and Expenditure Variables by Type of Medical Service

The following sections summarize definitional, conceptual and analytic considerations when using the utilization and expenditure variables in this file. Separate discussions are provided for each MEPS medical service category.

Medical Provider Visits (i.e., Office-Based Visits)

Medical provider visits consist of encounters that took place primarily in office-based settings and clinics. Care provided in other settings such as a hospital, nursing home, or a person's home are not included in this category.

The total number of office based visits reported for 1997 (OBTOTV97) as well as the number of such visits to physicians (OBDRV97) and nonphysician providers (OBOTHV97) are contained in this file. For a small proportion of sample persons, the sum of the physician and nonphysician visit variables (OBDRV97+OBOTHV97) is less than the total number of office-based visits variable (OBTOTV97) because OBTOTV97 contains reported visits where the respondent did not know the type of provider.

Non-physician visits (OBOTHV97) include visits to the following types of providers: chiropractors, midwives, nurses and nurse practitioners, optometrists, podiatrists, physician's assistants, physical therapists, occupational therapists, psychologists, social workers, technicians, receptionists/clerks/secretaries, or other medical providers. Separate utilization variables are included for selected types of more commonly seen non-physician providers including chiropractors (OBCHIR97), nurses/nurse practitioners (OBNURS97), optometrists (OBOPTO97), physician assistants (OBASST97), and physical or occupational therapists (OBTHER97).

Expenditure variables associated with all medical provider visits, physician visits, and non-physician visits in office-based settings can be identified using the attached table in Appendix 3. As for the corresponding utilization variables, the sum of the physician and non-physician visit

expenditure variables (e.g. OBDEXP97+OBOEXP97) is less than the total office-based expenditure variable (OBVEXP97) for a small proportion of sample persons. This can occur because OBVEXP97 includes visits where the respondent did not know the type of provider seen.

Hospital Events

Separate utilization variables for hospital care are provided for each type of setting (inpatient, outpatient department, and emergency room) along with two expense variables per setting; one for basic hospital facility expenses and another for payments to physicians who billed separately for services provided at the hospital. These payments are referred to as "separately billing doctor" or SBD expenses.

Hospital facility expenses include all expenses for direct hospital care, including room and board, diagnostic and laboratory work, x-rays, and similar charges, as well as any physician services included in the hospital charge. Separately billing doctor (SBD) expenses typically cover services provided to patients in hospital settings by providers like radiologists, anesthesiologists, and pathologists, whose charges are often not included in hospital bills.

Hospital Outpatient Visits

Variables for the total number of reported visits to hospital outpatient departments in 1997 (OPTOTV97) as well as the number of outpatient department visits to physicians (OPDRV97) and non-physician providers (OPOTHV97) are contained in this file. For a small proportion of sample persons, the sum of the physician and non-physician visit variables (OPDRV97+OPOTHV97) is less than the total number of outpatient visits variable (OPTOTV97) because OPTOTV97 contains reported visits where the respondent did not provide information on the type of provider seen.

Expenditure variables (both facility and SBD) associated with all medical provider visits, physician visits, and non-physician visits in outpatient departments can be identified using the attached table in Appendix 3. As for the corresponding utilization variables, the sum of the physician and non-physician expenditure variables (e.g. OPVEXP97+OPOEXP97 for facility expenses) is less than the variable for total outpatient department expenditures (OPFEXP97) for a small proportion of sample persons. This can occur because OBFEXP97 includes visits where the respondent did not know the type of provider seen. No expenditure variables are provided for health care consultations that occurred over the telephone.

Hospital Emergency Room Visits

The variable ERTOT97 represents a count of all emergency room visits reported for the survey year. Expenditure variables associated with ERTOT97 are identified in the attached table in Appendix 3. It should be noted that hospitals usually include expenses associated with emergency room visits that immediately result in an inpatient stay with the charges and payments for the inpatient stay. Therefore, to avoid the potential for double counting when imputing

missing expenses, separately reported expenditures for emergency room visits that were identified in the MPC as directly linked to an inpatient stay were included as part of the inpatient stay only (see below). This strategy to avoid double counting resulted in $0 expenditures for these emergency room visits. However, these $0 emergency room visits are still counted as separate visits in the utilization variable ERTOT97.

Hospital Inpatient Stays

Two measures of total inpatient utilization are provided on the file: (1) total number of hospital discharges (IPDIS97) and (2) the total number of nights associated with these discharges (IPNGTD97). IPDIS97 includes hospital stays where the dates of admission and discharge were reported as identical. These "zero night stays" can be included or excluded from inpatient analyses at the user's discretion (see last paragraph of this section). If the number of nights in the hospital could not be computed for any reported stay for a person, then IPNGTD97 was assigned a missing value.

Expenditure variables associated with hospital inpatient stays are identified in the attached table in Appendix 3. To the extent possible, payments associated with emergency room visits that immediately preceded an inpatient stay are included with the inpatient expenditures (see above) and payments associated with healthy newborns are included with expenditures for the mother (see next paragraph for more detail).

Data used to construct the inpatient utilization and expenditure variables for newborns were edited to exclude stays where the newborn left the hospital on the same day as the mother. This edit was applied because discharges for infants without complications after birth were not consistently reported in the survey and charges for newborns without complications are typically included in the mother's hospital bill. However, if the newborn was discharged at a later date than the mother was discharged, then the discharge was considered a separate stay for the newborn when constructing the utilization and expenditure variables.

Some analysts may prefer to exclude zero night stays from inpatient analyses and/or count these stays as ambulatory visits. Therefore, a separate use variable is provided which contains a count of the number of inpatient events where the reported dates of admission and discharge were the same (IPZERO97). This variable can be subtracted from IPDIS97 to exclude "zero night" stays from inpatient utilization estimates. In addition, separate expenditure variables are provided for "zero night" facility expenses (ZIFEXP97) and for separately billing doctor expenses (ZIDEXP97). Analysts who choose to exclude zero-night stays from inpatient expenditure analyses need to subtract the zero-night expenditure variable from the corresponding expenditure variable for total inpatient stays (e.g. IPFEXP97-ZIFEXP97 for facility expenses, IPDEXP97-ZIDEXP97 for separately billing doctor expenses).

Dental Visits

The total number of dental visits variable (DVTOT97) includes those to any person(s) for dental

care including general dentists, dental hygienists, dental technicians, dental surgeons, orthodontists, endodontists, and periodontists. Additional variables are provided for the numbers of dental visits to general dentists (DVGEN97) and to orthodontists (DVORTH97). For a small proportion of sample persons, the sum of the general dentist and orthodontist visit variables (DVGEN97+DVORTH97) is greater than the total number of dental visits (DVTOT97). This result can only occur for persons who were reported to have seen both a general dentist and orthodontist in the same visit(s). When this occurred, expenditures for the visit were included as orthodontist expenses but not as general dentist expenses. Expenditure variables for all three categories of dental providers can be identified using the attached table in Appendix 3.

Home Health Care

In contrast to other types of medical events where data were collected on a per visit basis, information on home health care utilization is collected in MEPS on a per month basis. Variables are provided which indicate the total number of days in 1997 where home health care was received from any type of paid or unpaid caregiver (HHTOTD97), agencies, hospitals, or nursing homes (HHAGD97), self-employed persons (HHINDD97), and unpaid informal caregivers not living with the sample person (HHINFD97). The number of provider days represents the sum across months of the number of days on which home health care was received, with days summed across all providers seen. For example, if a person received care in one month from one provider on 2 different days, then the number of provider days would equal 2. The number of provider days would also equal 2 if a person received care from 2 different providers on the same day. However, if a person received care from 1 provider 2 times in the same day, then the provider days would equal 1. These variables were assigned missing values if the number of provider days could not be computed for any month in which the specific type of home health care was received.

Separate expenditure variables are provided for agency-sponsored home health care (includes care provided by home health agencies, hospitals, and nursing homes) and care provided by self-employed persons. The attached table in Appendix 3 identifies the home health care utilization and expenditure variables contained in the file.

Vision Aids

Expenditure variables for the purchase of glasses and/or contact lenses are identified in the attached table in Appendix 3. Due to the data collection methodology, it was not possible to determine whether vision items that were reported in round 3 had been purchased in 1997 or 1998. Therefore, expenses reported in round 3 were only included if more than half of the person's reference period for the round was in 1997.

Other Medical Equipment and Services

This category includes expenditures for ambulance services, orthopedic items, hearing devices, prostheses, bathroom aids, medical equipment, disposable supplies, alterations/modifications, and other miscellaneous items or services that were obtained, purchased or rented during the year. Respondents were only asked once (in round 3) about their total annual expenditures and were not asked about their frequency of use of these services. Expenditure variables representing the combined expenses for these supplies and services are identified in the Appendix 3 table.

Prescribed Medicines

There is one total utilization variable (RXTOT97) and 13 expenditure variables included on the 1997 full-year file relating to prescribed medicines. These 13 expenditure variables include an annual total expenditure variable (RXEXP97) and 12 corresponding annual source of payment variables (RXSLF97, RXMCR97, RXMCD97, RXPRV97, RXVA97, RXCHM97, RXOFD97, RXSTL97, RXWCP97, RXOSR97, RXOPR97, and RXOPU97). Unlike the other event types, the prescribed medicine events have some remaining inconsistencies in the data when comparing information from the insurance section of the Household Component and source of payment information from the Pharmacy Component (more specifically, discrepancies between Medicare only household insurance responses and Medicaid source of payment provided by pharmacy providers). These inconsistencies remain unedited because there was strong evidence from the Pharmacy Component that these were indeed Medicaid payments. All of these types of Household Component events were either exact matches to events in the Pharmacy Component or refills of exact matches, and in addition, all of these types of events were purchases by persons with positive weights. The total utilization variable is a count of all prescribed medications initially purchased or otherwise obtained during 1997, as well as any additional acquisitions of the medication. The total expenditure variable sums all amounts paid out-of-pocket and by third party payers for each prescription purchased in 1997. No variables reflecting charges for prescription medicines are included because a large proportion of respondents to the pharmacy component survey did not provide charge data (see below).

Prescribed Medicines Data Collected

Data regarding prescription drugs were obtained through the household questionnaire and a pharmacy component survey. During each round of the MEPS HC, all respondents were asked to supply the name of any prescribed medication they or their family members purchased or otherwise obtained during that round. For each medication and in each round, the following information was collected: whether any free samples of the medication were received; the name(s) of any health problems the medication was prescribed for; the number of times the prescription drug was obtained or purchased; the year, month, and day on which the person first used the medication; and a list of the names, addresses, and types of pharmacies that filled the household's prescriptions. Also, during the Household Component, respondents were asked if they send in claim forms for their prescriptions (self-filers) or if their pharmacy providers do this automatically for them at the point of purchase (non-self-filers). For non-self-filers, charge and payment information was collected in the pharmacy component survey. However, charge and

payment information was collected for self-filers in the household questionnaire, because payments by private third party payers for self-filers' purchases would not be available from a pharmacy follow-back survey.

Pharmacy providers identified by the household were contacted by mail for the pharmacy component survey if permission was obtained in writing from the person with the prescription to release their pharmacy records. The signed permission forms were provided to the various establishments prior to making any requests for information. Each establishment was informed of all persons participating in the survey that had prescriptions filled there in 1997 and a computerized printout containing information about these prescriptions was sought. For each medication listed, the following information was requested: date filled; national drug code (NDC); medication name; strength of medicine (amount and unit); quantity (package size and amount dispensed); total charge; and payments by source.

When diabetic supplies, such as syringes and insulin, were reported in the other medical supply section of the MEPS HC questionnaire as having been obtained during the round, the interviewer was directed to collect information on these items in the prescription drug section of MEPS. Data on expenses for these items were collected in and imputed from the pharmacy component survey.

Prescribed Medicines Data Editing and Imputation

The general approach to preparing the household prescription data for this file was to utilize the pharmacy component prescription data to assign expenditure values to the household drug mentions. For self-filers, information on payment sources was retained to the extent that these data were reported by the household in the charge and payment section of the household questionnaire. A matching program was adopted to link pharmacy survey drugs and the corresponding drug information to household drug mentions. To improve the quality of these matches, all drugs on the household and pharmacy files were coded based on the medication names provided by the household and pharmacy, and when available, the national drug code (NDC) provided in the pharmacy survey. Considerable editing was done prior to the matching to correct data inconsistencies in both data sets and fill in missing data and correct outliers on the pharmacy file.

Drug price per unit outliers were analyzed on the pharmacy file by first identifying the average wholesale unit price (AWUP) of the drug by linkage through the NDC to a proprietary data base. In general, prescription drug unit prices were deemed to be outliers by comparing unit prices reported in the pharmacy data base to the AWUP and were edited, as necessary.

Round 3 household drug mentions in MEPS were not identified in the HC as 1997 or 1998 drug events for persons in households in which their Round 3 began in 1997 and ended in 1998. All exact matches to pharmacy survey drug events for persons whose pharmacies participated were classified as 1997 drug purchases. Any remaining Round 3 household drug mentions for persons

with Round 3 spanning both years were randomly allocated to 1997 or 1998 based on the proportion of the household's Round 3 period in each year.

3.0 Survey Sample Information

3.1 Sample Design and Response Rates

The MEPS is designed to produce estimates at the national and regional level over time for the civilian, noninstitutionalized population of the United States and some subpopulations of interest. The health care utilization data in this public use set pertain to calendar year 1997. The data were collected in Rounds 1, 2, and 3 for MEPS Panel 2 and Rounds 3, 4, and 5 for MEPS Panel 1. Note that Round 3 for a MEPS panel overlaps two calendar years. The reference period for Round 3 of Panel 1 covers the end of 1997 and the beginning of 1997 while the reference period for Round 3 of MEPS Panel 2 covers the end of 1997 and the beginning of 1998. The only utilization data that appear on the file are those associated with health care events occurring in calendar year 1997.

The households in this 1997 MEPS database are related to households participating in the National Health Interview Survey in 1995 and 1997. The households (occupied dwelling units) selected for MEPS Panel 1 were a subsample of 1995 NHIS respondents while those in MEPS Panel 2 were a subsample of 1997 NHIS respondents. A household may contain one or more family units, each consisting of one or more individuals. Analysis can be undertaken using either the individual or the family as the unit of analysis.

For MEPS Panel 2 several domains of interest were oversampled to provide increased precision for analytic purposes. These domains included households containing persons with one of the following characteristics based on NHIS data: adults with functional impairments, children with limitations in activity, individuals aged 18-64 with expected high medical expenditures, individuals with family incomes expected to be below 200% of the poverty level in 1997, and adults with other impairments. Because some households could be associated with more than one domain, a hierarchical sample selection procedure was employed. If a household could be associated with multiple domains, it was assigned to the domain given the highest priority in the hierarchy.

For detailed information on the MEPS sample design for Panel 1, see Cohen, S. Sample Design of the 1996 Medical Expenditure Panel Survey Household Component. Rockville (MD): Agency for Health Care Policy and Research; 1996. MEPS Methodology Report, No. 2. AHCPR Pub. No. 97-0027. For detailed information on the MEPS sample design for Panel 2, see Appendix 2: Cohen, S. Sample Design of the 1997 Medical Expenditure Panel Survey Household Component.

MEPS-Linked to the National Health Interview Survey

The sample of 10,639 households (occupied dwelling units) for the MEPS Panel 1 consisted of a nationally representative subsample of the households responding to the 1995 National Health Interview Survey (NHIS). A subsample of 6,300 households was selected for MEPS Panel 2 from among households responding to the 1996 NHIS.

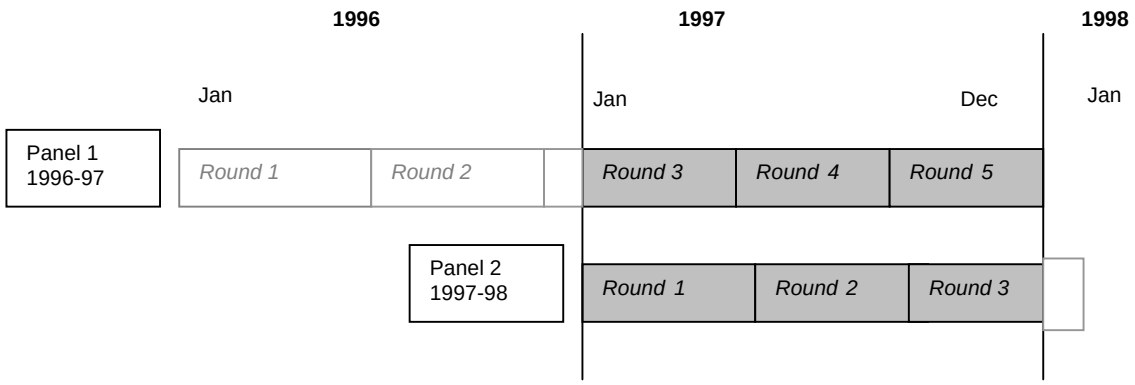
The NHIS sample design has three stages of sample selection: an area sample of PSUs; a sample of segments (single or groups of blocks or block equivalents) within sampled PSUs; and a sample of housing units within segments. Among initially sampled households, those containing Hispanics and blacks were oversampled at rates of approximately 2 and 1.5 times the rate of remaining households. These same rates of oversampling are reflected in the MEPS sample of households. The only major difference in the definition of a household between NHIS and MEPS is that college aged students living away from home during the school year were interviewed at their place of residence for the NHIS but were identified by and linked to their parents' household for MEPS.

Sample Weights and Variance Estimation

In the database MEPS HC-020: 1997 Full Year Population Characteristics weight variables are provided for estimation purposes. Procedures and considerations associated with the construction and interpretation of person and family level estimates using these and other variables are discussed below.

Response Rates

In order to produce annual health care estimates for calendar year 1997 based on the full MEPS sample, data will also need to be pooled across the first two MEPS national samples. More specifically, full calendar year 1997 data collected in Rounds 3 through 5 for the MEPS Panel 1 sample are pooled with data from the first three rounds of data collection for the MEPS Panel 2 sample (illustrated below). Overall, the full 1997 MEPS household sample will consist of approximately 13,000 reporting units which include 32,636 individuals that completed the full series of MEPS interviews for their entire period of eligibility, providing the necessary information to produce national use and expenditure estimates for calendar year 1997.



Panel 1

Conditioned on response to Rounds 1-3 of the Panel 1 MEPS, of 21,696 key and inscope individuals eligible for data collection in 1997, 19,622 (90.44 percent) provided data for their entire period of eligibility. Consequently, after factoring in the impact of survey attrition, the overall Panel 1 MEPS person level response rate for deriving annual estimates was 63.5 percent (.702 x .9044). Of these full year respondents for calendar year 1997, 19,407 were in scope on December 31, 1997.

Panel 2

Conditioned on response to Round 1 of the Panel 2 MEPS, of 14, 644 key and inscope individuals eligible for data collection in 1997, 13,014 (88.87 percent) provided data for their entire period of eligibility. Consequently, after factoring in the impact of survey attrition, the overall Panel 2 MEPS person level response rate for deriving annual estimates was 69.2 percent (.779 x .8887). Of these full year respondents for calendar year 1997, 12,819 were in scope on December 31, 1997.

Combined MEPS Panels: Response Rate for Annual 1997 Estimates

For each independent MEPS sample, the estimation weights were further adjusted for survey attrition over time. Each panel was then given equal weight in the development of sampling weights to produce annual national estimates. Therefore, a pooled response rate for the survey respondents in this data set can be obtained by taking an average of the panel specific response rates. This pooled response rate for the combined panels is 66.4 percent, consisting of a total of 32,636 survey participants. The weighted MEPS population estimate for the civilian non-institutionalized population as of December 31, 1997 was 267,704,802, based on poststratification to population estimates produced from the December 1997 Current Population Survey. Future analyses will examine the impact of survey attrition on health care utilization and expenditure estimates covering calendar year 1997.

3.2 Person Level Estimation using this MEPS PUF

Overview

There is a single person level weight variable called WTDPER97. However, care should be taken in its application as it permits both "point-in-time" and "range of time" estimates, depending on the variables used to define the set of persons of interest for analysis. A person level weight was assigned to each key, inscope person who responded to MEPS for the full period of time that he or she was inscope during the MEPS survey. For Panel 2 this requirement pertained only to 1997, but for Panel 1 it pertained to both 1996 and 1997. (Recall that a person is inscope whenever he or she is a member of the civilian, noninstitutionalized portion of the U.S. population.)

Developing Person Level MEPS Estimates

The data in this file can be used to develop estimates on persons in the civilian, noninstitutionalized population on December 31, 1997 and for the slightly larger population of persons in the civilian, noninstitutionalized population at any time during 1997. To obtain a cross-sectional (point-in-time) estimate for all inscope persons living in the country on December 31, 1997, include cases with both WTDPER97>0 (a positive person level weight) and INSC1231=1 (the person is inscope on December 31, 1997). To obtain an estimate for all persons who were inscope at some time in 1997, include all cases with WTDPER97>0. After selecting the appropriate cases, apply the weight variable WTDPER97 to the analytic variable(s) of interest to obtain national estimates. The following table contains a summary of cases to include and sample sizes for these two populations (for shorthand purposes the term "general" is used to indicate the "civilian, noninstitutionalized" component of the U.S.population).

Population of Interest	**Cases to Include**	**Sample Size**
General Population on December 31, 1997	WTDPER97>0 and INSC1231=1	32,226
General Population over the course of 1997	WTDPER97>0	32,636

Details on Person Weights Construction

Overview

The person level weight WTDPER97 was developed in three stages. A person level weight for Panel 2 was created, including both an adjustment for nonresponse over time and poststratification, controlling to Current Population Survey (CPS) population estimates based on five different variables. Poverty status was not included since income data for assigning persons to a poverty status was yet to be established. Then a person level weight for Panel 1 was created, again including an adjustment for nonresponse over time and poststratification, controlling to CPS population estimates based on the same five variables. In the meantime work proceeded on the MEPS income data and

the assignment of MEPS families to CPS-like family units, since the assignment of poverty status is based on CPS family structures. When poverty status information derived from income variables became available, a 1997 average annual weight was formed from the Panel 1 and Panel 2 weights by multiplying the Panel weights by .5. Then a final poststratification was done on this composite weight variable, including poverty status as well as the original five poststratification variables in the establishment of the final 1997 person level weight.

MEPS Panel 1

The person level weight for MEPS Panel 1 was developed using the 1996 full year weight for an individual as a "base" weight for survey participants present in 1996. For key, inscope respondents who joined an RU some time in 1997 after being out-of-scope in 1996, the 1996 family weight associated with the family the person joined served as a "base" weight. The weighting process included an adjustment for nonresponse over Rounds 4 and 5 as well as poststratification to population control totals from the CPS for December, 1997. These control totals were derived by scaling back the population totals obtained from the March 1998 CPS to reflect the December, 1997 CPS estimated population distribution across age and sex categories as of December, 1997.

Variables used in the establishment of person level poststratification control figures included: census region (Northeast, Midwest, South, West); MSA status (MSA, non-MSA); race/ethnicity (Hispanic, black but non-Hispanic, and other); sex, and age.

Overall, the weighted population estimate for the civilian, noninstitutionalized population on December 31, 1997 is 267,704,802. Key, responding persons not inscope on December 31, 1997 but inscope earlier in the year retained, as their final Panel 1 weight, the weight after the nonresponse adjustment.

MEPS Panel 2

The person level weight for MEPS Panel 2 was developed using the MEPS Round 1 person-level weight as a "base" weight. For key, inscope respondents who joined an RU after Round 1, the Round 1 family weight served as a "base" weight. The weighting process included an adjustment for nonresponse over Round 2 and the 1997 portion of Round 3 as well as poststratification to the same population control figures for December 1997 used for the MEPS Panel 1 weights. The same five variables employed for Panel 1 poststratification (census region, MSA status, race/ethnicity, sex, and age) were used for Panel 2 poststratification. As with Panel 1, Panel 2 key, responding persons not inscope on December 31, 1997 but inscope earlier in the year retained the weight after the nonresponse adjustment as their final Panel 2 weight.

Note that the MEPS round 1 weights (for both panels with one exception as noted below) incorporated the following components: the original household probability of selection for the NHIS; ratio-adjustment to NHIS-based national population estimates at the household (occupied dwelling unit) level; the probability of selection of dwelling units associated with the oversampling of five

population domains of analytic interest (for Panel 2 only); adjustment for nonresponse at the dwelling unit level for Round 1; and poststratification to figures at the family and person level obtained from the March 1997 CPS data base. The five oversampled domains for Panel 2 were households with: persons with functional impairments; children with limitations in activity; individuals 18-64 expected to incur high medical expenditures based on a statistical model; persons with family incomes expected to be below 200 percent of poverty, based on a statistical model; and adults with other impairments.

The Final Weight for 1997

Variables used in the establishment of person level poststratification control totals included: poverty status (below poverty, from 100 to 125 percent of poverty, from 125 to 200 percent of poverty, from 200 to 400 percent of poverty, at least 400 percent of poverty); census region (Northeast, Midwest, South, West); MSA status (MSA, non-MSA); race/ethnicity (Hispanic, black but non-Hispanic, and other); sex, and age. Overall, the weighted population estimate for the civilian, noninstitutionalized population for December 31, 1997 is 267,704,802 (WTDPER97>0 and INSC1231=1). The inclusion of key, inscope persons who were not inscope on December 31, 1997 brings the estimated total number of persons represented by the MEPS respondents over the course of the year, to 270,965,010 (WTDPER97>0).

The weights for persons who died while members of the civilian, noninstitutionalized population in 1997 were adjusted separately for persons under age 65 and those age 65 and older. Control figures were derived from Vital Statistics death registries, the Nursing Home Component of the 1996 MEPS, and the annual Medicare Beneficiary Survey (MCBS).

The weights for persons ascertained to be living in nursing homes for at least one day in 1997 and who were not inscope on December 31, 1997 were adjusted to a control figure derived from the Nursing Home Component of the 1996 MEPS.

Coverage

The target population for MEPS is the 1997 U.S. civilian, noninstitutionalized population. However, the MEPS sampled households are a subsample of the NHIS households interviewed in 1995 (Panel 1) and 1996 (Panel 2). New households created after the NHIS interviews for the respective Panels and consisting exclusively of persons who entered the target population after 1995 (Panel 1) or after 1996 (Panel 2) are not covered by MEPS. These would include families consisting solely of: immigrants; persons leaving the military; U.S. citizens returning from residence in another country; and persons leaving institutions. It should be noted that this set of uncovered persons constitutes only a tiny proportion of the MEPS target population.

3.3 Family Level Estimation Using this MEPS PUF

There are two family weight variables called WTFAMF97 and WTCFAM97 that are provided in this release. In general, WTFAMF97 can be used to make estimates for the cross-section of families in

the U.S. civilian, noninstitutionalized population on December 31, 1997 where families are identified based on the MEPS definition of a family unit. WTCFAM97 is provided to permit the capability of producing estimates for families defined according to the guidelines used by the Current Population Survey (CPS). Such family units are characterized here as "CPS-like". In addition, estimates can be constructed using WTFAMF97. These estimates permit the inclusion of: 1) MEPS families that existed at some time during 1997 but whose members became out-of-scope prior to the end of the year (e.g., all family members moved out of the country, died, etc.) and 2) MEPS families in existence on December 31, 1997.

Definition of Family Estimates

A family is defined in MEPS as two or more persons living together in the same household who are related by blood, marriage, or adoption, as well as foster children. Other MEPS families include unmarried persons living together who consider themselves a family unit. Single persons living with neither a relative nor a person identified as a "significant other" have also been assigned a family ID value as and a family level weight, and thus can be included or excluded from estimates, as desired. Relatives identified as usual residents of the household who were not there at the time of the interview, such as college students living away from their parents' home during the school year, were considered as members of the family that identified them.

To make estimates at the family level, it is necessary to prepare a family level file containing one record per family (see instructions below), family level summary characteristics, and the family-level weight variable (WTFAMF97). Each MEPS family unit is uniquely identified by the combination of the variables DUID and FAMIDYR. The number of persons in a MEPS sample family ranges from 1 to 14 (the positive values for FAMSZEYR). Only persons with positive nonzero family weight values (WTFAMF97>0) are candidates for inclusion in family estimates.

Three sets of families for whom estimates can be obtained are defined in the table below (along with respective sample sizes). Persons with FMRS1231=1 were inscope for the survey on 12/31/97 and therefore part of a MEPS family on 12/31/97. The more expansive definition of families (second row in table) includes families and members of families who were not inscope at the end of the year. The third row is for CPS-like families, and thus excludes foster children. While MEPS includes individual persons as family units (about one-third of all units) to cover the entire civilian, noninstitutionalized population, analysts may restrict their analyses to families with 2 or more members using the family size variables shown in the table.

Population of Interest	Cases to Include	Sample Size	Family Size Variable
Cross-section of Families in the Civilian Noninstitutionalized Population on 12/31/97	WTFAMF97>0 & FMRS1231=1	12,970	FAMS1231
Families in the Civilian Noninstitutionalized Population on 12/31/97 plus families and members of families in existence earlier in 1997 who were not part of the civilian	WTFAMF97>0	13,087	FAMSZEYR

Population of Interest	Cases to Include	Sample Size	Family Size Variable
noninstitutionalized population on 12/31/97			
CPS-like families excluding foster children	WTCFAM97>0	13,246	FCSZ1231

Instructions to Create Family Estimates

The following is a summary of the steps and the variables to be used for family level estimation based on the MEPS type definition of families.

1. Concatenate the variables DUID and FAMIDYR into a new variable (e.g. DUIDFAMY).

2. To create a family level file, sort by DUIDFAMY and then subset to one record per DUIDFAMY value by retaining only the reference person record (FAMRFPYR=1) for each value of DUIDFAMY. If aggregate measures for families are needed for analytic purposes (e.g. means or totals), then those measures need to be computed using person-level information within families and attached to the family record. For other types of variables, analysts frequently use the characteristics of the reference person to represent family characteristics.

3. Apply the weight WTFAMF97 to the analytic variable(s) of interest to obtain national family estimates.

4. Use CPSFAMID, FCRP1231, and WTCFAM97 in place of FAMIDYR, FAMRFPYR, and WTFAMF97, respectively, to make estimates as of 12/31/97 for CPS-like families (thus excluding foster children).

Details on Family Weight Construction and Estimated Number of Families

To develop the family level weight (WTFAMF97), the person level weight (WTDPER97) of the family reference person (FAMRFPYR=1) was used as the “base” weight for all responding full year families. Then, for responding families eligible for weighting and in existence at the end of 1997, these “base” weights were poststratified to population control figures from the Current Population Survey (CPS) for December 1997 (these figures were derived by scaling the population totals obtained from the March 1997 CPS to reflect family estimates as of December, 1997). The family level poststratification incorporated the following variables: poverty status (below poverty, from 100 to 125 percent of poverty, from 125 to 200 percent of poverty, from 200 to 400 percent of poverty, at least 400 percent of poverty); census region; MSA status; race/ethnicity of reference person (Hispanic, black but non Hispanic, and other); family type (reference person married, living with spouse; male reference person, unmarried or spouse not present; female reference person, unmarried or spouse not present); age of reference person; and family size as of December 31, 1997.

Overall, the weighted population estimate for the number of MEPS family units containing at least one member of the U.S. civilian, noninstitutionalized population on December 31, 1997 is

111,014,574 (those families whose members have WTFAMF97>0 and FMRS1231=1). The inclusion of families whose members left the inscope population prior to December 31, 1997 brought the estimated total number of families represented by the MEPS responding families to 112,165,786 (those families whose members have WTFAMF97>0). The estimated total number of CPS-like families is 113,295,487 (those families whose members have WTCFAM97>0).

Variance Estimation

To obtain estimates of variability (such as the standard error of sample estimates or corresponding confidence intervals) for estimates based on MEPS survey data, the complex sample design of MEPS for both person and family level analyses must be taken into account. Various approaches can be used to develop such estimates of variance including use of the Taylor series or replication methodologies. Replicate weights have not been developed for the MEPS 1997 data.

Using a Taylor Series approach, variance estimation strata and the variance estimation PSUs within these strata must be specified. The corresponding variables on the 1997 MEPS full year utilization database are VARSTR97 and VARPSU97, respectively. Specifying a “with replacement” design in a computer software package, such as SUDAAN, should provide standard errors appropriate for assessing the variability of MEPS survey estimates. It should be noted that the number of degrees of freedom associated with estimates of variability indicated by such a package may not appropriately reflect the actual number available. For MEPS sample estimates for characteristics generally distributed throughout the country (and thus the sample PSUs), there are over 100 degrees of freedom for the 1997 full year data associated with the corresponding estimates of variance.

D. Variable-Source Crosswalk

SURVEY ADMINISTRATION VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
DUID	DU ID	Assigned in Sampling
PID	Person Number (PN)	Assigned in Sampling or by CAPI
DUPERSID	Sample Person ID (DU+PN)	Assigned in Sampling
PANEL97	Panel Number	Constructed
FAMID31	Family Identifier (Student Merged In) – R31	CAPI Derived
FAMID42	Family Identifier (Student Merged In) – R42	CAPI Derived
FAMID53	Family Identifier (Student Merged In) – R53	CAPI Derived
FAMID97	Fam Identifier (Stud Merged In) - 12/31/97	CAPI Derived
FAMIDYR	Annual Family Identifier	Constructed
CPSFAMID	CPS-Like Family Identifier	Constructed
FCSZ1231	Family Size Responding 12/31 CPS Family	Constructed
FCRP1231	Ref Person of 12/31 CPS Family	Constructed
RULETR31	RU Letter – R31	CAPI Derived
RULETR42	RU Letter – R42	CAPI Derived
RULETR53	RU Letter – R53	CAPI Derived
RULETR97	RU Letter As of Dec 31	CAPI Derived
RUSIZE31	RU Size – R31	CAPI Derived
RUSIZE42	RU Size – R42	CAPI Derived
RUSIZE53	RU Size – R53	CAPI Derived
RUSIZE97	RU Size As of Dec 31	CAPI Derived
RUCLAS31	RU: Standard/New/Student – R31	CAPI Derived
RUCLAS42	RU: Standard/New/Student – R42	CAPI Derived
RUCLAS53	RU: Standard/New/Student – R53	CAPI Derived
RUCLAS97	RU: Standard/New/Student - 12/31/97	CAPI Derived
FAMSZE31	RU Size Including Students – R31	CAPI Derived
FAMSZE42	RU Size Including Students – R42	CAPI Derived
FAMSZE53	RU Size Including Students – R53	CAPI Derived
FAMSZE97	RU Size Including Students As of Dec 31	CAPI Derived
FMRS1231	Member of Responding 12/31 Family	Constructed
FAMS1231	Family Size of Responding 12/31 Family	Constructed
FAMSZEYR	Size of Responding Annualized Family	Constructed
FAMRFPYR	Reference Person of Annualized Family	Constructed
FYFAMTYP	CPS –Full Year Family Type	Constructed
INRU1231	Person Was In RU On 12/31/97	Constructed
REGION31	Census Region – R31	Assigned in Sampling
REGION42	Census Region – R42	Assigned in Sampling

VARIABLE	DESCRIPTION	SOURCE
REGION53	Census Region – R53	Assigned in Sampling
REGION97	Census Region As Of Dec 31	Assigned in Sampling
MSA53	MSA – R53	Assigned in Sampling
MSA97	MSA As Of Dec 31	Assigned in Sampling
REFPRS31	Reference Person At Round 31	RE 42-45
REFPRS42	Reference Person At Round 42	RE 42-45
REFPRS53	Reference Person At Round 53	RE 42-45
REFPRS97	Reference Person As Of Dec 31	RE 42-45
RESP31	1st Respondent Indicator For Rnd 31	RE 6, 8
RESP42	1st Respondent Indicator For Rnd 42	RE 6, 8
RESP53	1st Respondent Indicator For Rnd 42	RE 6, 8
RESP97	1st Respondent Indicator As Of 12/31/97	RE 6, 8
PROXY31	Was Respondent A Proxy In R31	RE 2
PROXY42	Was Respondent A Proxy In R42	RE 2
PROXY53	Was Respondent A Proxy In R53	RE 2
PROXY97	Was Respondent A Proxy As Of 12/31/97	RE 2
BEGRFD31	R31 Reference Period Begin Date: Day	CAPI Derived
BEGRFM31	R31 Reference Period Begin Date: Month	CAPI Derived
BEGRFY31	R31 Reference Period Begin Date: Year	CAPI Derived
ENDRFD31	Reference Period End Date: Day – R31	CAPI Derived
ENDRFM31	Reference Period End Date: Month – R31	CAPI Derived
ENDRFY31	Reference Period End Date: Year – R31	CAPI Derived
BEGRFD42	R42 Reference Period Begin Date: Day	CAPI Derived
BEGRFM42	R42 Reference Period Begin Date: Month	CAPI Derived
BEGRFY42	R42 Reference Period Begin Date: Year	CAPI Derived
ENDRFD42	Reference Period End Date: Day – R42	CAPI Derived
ENDRFM42	Reference Period End Date: Month – R42	CAPI Derived
ENDRFY42	Reference Period End Date: Year – R42	CAPI Derived
BEGRFD53	R53 Reference Period Begin Date: Day	CAPI Derived
BEGRFM53	R53 Reference Period Begin Date: Month	CAPI Derived
BEGRFY53	R53 Reference Period Begin Date: Year	CAPI Derived
ENDRFD53	Reference Period End Date: Day – R53	CAPI Derived
ENDRFM53	Reference Period End Date: Month – R53	CAPI Derived
ENDRFY53	Reference Period End Date: Year – R53	CAPI Derived
ENDRFD97	1997 Reference Period End Date: Day	RE Section
ENDRFM97	1997 Reference Period End Date: Month	RE Section
ENDRFY97	1997 Reference Period End Date: Year	RE Section
KEYNESS	Person Key Status	RE Section
INSCOP31	Inscope – R31	RE Section
INSCOP42	Inscope – R42	RE Section

VARIABLE	DESCRIPTION	SOURCE
INSCOP53	Inscope – R53	RE Section
INSCOP97	Inscope – R53 Start Through 12/31/97	RE Section
INSC1231	Inscope Status on 12/31/97	Constructed
INSCOPE	Was Person Ever Inscope In 1997	RE Section
ELGRND31	Eligibility – R31	RE Section
ELGRND42	Eligibility – R42	RE Section
ELGRND53	Eligibility – R53	RE Section
ELGRND97	Eligibility – R53 Start Through 12/31/97	RE Section
ELIGIBLE	Was Person Ever Eligible In 1997	RE Section
PSTATS31	Person Disposition Status – R31	RE Section
PSTATS42	Person Disposition Status – R42	RE Section
PSTATS53	Person Disposition Status – R53	RE Section
RURSLT31	RU Result – R31	Assigned by CAPI
RURSLT42	RU Result – R42	Assigned by CAPI
RURSLT53	RU Result – R53	Assigned by CAPI

DEMOGRAPHIC VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
AGE31X	Age – R31 (Edited/Imputed)	RE 12, 57-66
AGE42X	Age – R42 (Edited/Imputed)	RE 12, 57-66
AGE53X	Age – R53 (Edited/Imputed)	RE 12, 57-66
AGE97X	Age - 12/31/97 (Edited/Imputed)	RE 12, 57-66
DOBMM	Date of Birth: Month	RE 12, 57-66
DOBYY	Date of Birth: Year	RE 12, 57-66
SEX	Sex	RE 12, 57, 61
RACEX	Race (Edited/Imputed)	RE 101, 102
RACETHNX	Race/Ethnicity (Edited/Imputed)	RE 98-102
HISPANX	Hispanic Ethnicity (Edited/Imputed)	RE 98-100
HISPCAT	Specific Hispanic Ethnicity Group	RE 98-100
MARRY31X	Marital Status – R31 (Edited/Imputed)	RE 13, 97
MARRY42X	Marital Status – R42 (Edited/Imputed)	RE 13, 97
MARRY53X	Marital Status – R53 (Edited/Imputed)	RE 13, 97
MARRY97X	Marital Status - 12/31/97 (Edited/Imputed)	RE 13, 97
SPOUID31	Spouse ID – R31	RE 13, 76, 77, 97
SPOUID42	Spouse ID – R42	RE 13, 76, 77, 97
SPOUID53	Spouse ID – R53	RE 13, 76, 77, 97
SPOUID97	Spouse ID - 12/31/97	RE 13, 76, 77, 97
SPOUIN31	Marital Status W/ Spouse Present – R31	RE 13, 76, 77, 97
SPOUIN42	Marital Status W/ Spouse Present – R42	RE 13, 76, 77, 97
SPOUIN53	Marital Status W/ Spouse Present – R53	RE 13, 76, 77, 97
SPOUIN97	Marital Status W/Spouse Present - 12/31/97	RE 13, 76, 77, 97
EDUCYR31	Completed Years of Education – R31	RE 103-105
EDUCYR42	Completed Years of Education – R42	RE 103-105
EDUCYR53	Completed Years of Education – R53	RE 103-105
EDUCYR97	Completed Years of Education - 12/31/97	RE 103-105
HIDEG31	Highest Degree – R31	RE 103-105
HIDEG42	Highest Degree – R42	RE 103-105
HIDEG53	Highest Degree – R53	RE 103-105
HIDEG97	Highest Degree - 12/31/97	RE 103-105
FTSTU31X	Student Status If Ages 17-23 - Round 31	RE 11A, 106-108
FTSTU42X	Student Status If Ages 17-23 - Round 42	RE 11A, 106-108
FTSTU53X	Student Status If Ages 17-23 - Round 53	RE 11A, 106-108
FTSTU97X	Student Status If Ages 17-23 - 12/31/97	RE 11A, 106-108
ACTDTY31	Military Full-Time Active Duty – R31	RE14, 96A
ACTDTY42	Military Full-Time Active Duty – R42	RE 14, 96B1
ACTDTY53	Military Full-Time Active Duty – R53	RE 14, 96B1

VARIABLE	DESCRIPTION	SOURCE
DIDSERVE	Ever Served In Armed Forces	RE 18, 95
VETPVIET	Served In Post-Vietnam Era	RE 35, 94, 94A, 95, 96
VETVIET	Served In Vietnam War Era	RE 35, 94, 94A, 95, 96
VETKOR	Served In Korean War Era	RE 35, 94, 94A, 95, 96
VETWW	Served In WWI Or WW2 Era	RE 35, 94, 94A, 95, 96
VETOTH	Served In Other Period	RE 35, 94, 94A, 95, 96
RFREL31X	Relation To Ref Pers - R31 (Edited/Imputed)	RE 76-77
RFREL42X	Relation To Ref Pers - R42 (Edited/Imputed)	RE 76-77
RFREL53X	Relation To Ref Pers - R53 (Edited/Imputed)	RE 76-77
RFREL97X	Relation To Ref Pers 12/31/97 (Edit/Imp)	RE 76-77
MOPID31X	PID Of Person's Mom (Edited/Imputed)	RE 76-77
MOPID42X	PID Of Person's Mom (Edited/Imputed)	RE 76-77
MOPID53X	PID Of Person's Mom (Edited/Imputed)	RE 76-77
DAPID31X	PID Of Person's Dad (Edited/Imputed)	RE 76-77
DAPID42X	PID Of Person's Dad (Edited/Imputed)	RE 76-77
DAPID53X	PID Of Person's Dad (Edited/Imputed)	RE 76-77

INCOME VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
SSIDIS97	SSI RECEIPT DUE TO DISABILITY	IN 39
AFDC97	DID PERSON'S CHECK INCLUDE AFDC OR ADC	IN 44
FILEDR97	HAS PERSON FILED A FED INCOME TAX RETURN	IN 2
WILFIL97	WILL PERSON FILE FED INCOME TAX RETURN	IN 3
FLSTAT97	PERSON'S FILING STATUS	IN 4
FILER97	PRIMARY OR SECONDARY FILER	IN 4
JTINRU97	JOINT FILER'S MEMBERSHIP IN RU	IN 5
JNTPID97	PID OF SECONDARY FILER	IN 5
CLMDEP97	DID/WILL PERS CLAIM DEPENDENTS ON RETURN	IN 6
DEPDNT97	PERSON IS FLAGGED A DEPENDENT	IN 7
DPINRU97	DEPENDENTS IN/OUT OF RU	IN 7
DPOTSD97	HOW MANY DEPENDENTS LIVE OUTSIDE RU	IN 8
TAXFRM97	TAX FORM PERSON WILL FILE	IN 9
DEDUCT97	ITEMIZE OR STANDARD DEDUCTION	IN 10
ITMEXP97	WILL PERSON ITEMIZE MEDICAL EXPENSE	IN 11
MEXAMT97	TOTAL AMOUNT FOR MEDICAL EXPENSES	IN 12
NTMDED97	PERSON'S NET MEDICAL EXPENSE DEDUCTION	IN 13
TOTDED97	TOTAL OF ALL ITEMIZED DEDUCTIONS	IN 14
CLMHIP97	DID/WILL PERS DEDUCT HEALTH INSUR PREM	IN 15
ELDISC97	DID/WILL PERS RECEIVE ELDERLY/DISAB CRED	IN 16
EICRDT97	DID/WILL PERS RECEIVE EARNED INC CREDIT	IN 17
UNEMTX97	TAXABLE PERCENTAGE OF UNEMPLOYMENT	IN 30
INTRTX97	TAXABLE PERCENTAGE OF INTEREST	IN 19
SSECTX97	TAXABLE PERCENTAGE OF SOCIAL SECURITY	IN 31
IRATAX97	TAXABLE PERCENTAGE OF IRA INCOME	IN 25
FOODST97	DID ANYONE PURCHASE FOOD STAMPS	IN 55
FOODMN97	NUMBER OF FOOD STAMPS PURCHASED	IN 56
FOODCT97	AMOUNT FAMILY PAID FOR FOOD STAMPS	IN 57
FOODVL97	MONTHLY VALUE OF FOOD STAMPS	IN 58
TTLP97X	PERSON'S TOTAL INCOME	Constructed
POVCAT97	FAMILY INCOME AS PERCENT OF POVERTY LINE	Constructed
WAGEP97X	PERSON'S WAGE INCOME	Constructed
WAGIMP97	WAGEPN IMPUTATION FLAG	Constructed
BUSNP97X	PERSON'S BUSINESS INCOME	Constructed
BUSIMP97	BUSNPN IMPUTATION FLAG	Constructed
FARMP97X	PERSON'S FARM INCOME	Constructed
FARIMP97	FARMPN IMPUTATION FLAG	Constructed
UNEMP97X	PERSON'S UNEMPLOYMENT COMP INCOME	Constructed
UNEIMP97	UNEMPN IMPUTATION FLAG	Constructed

VARIABLE	DESCRIPTION	SOURCE
WCMPP97X	PERSON'S WORKMAN'S COMPENSATION	Constructed
WCPIMP97	WCMPPN IMPUTATION FLAG	Constructed
INTRP97X	PERSON'S INTEREST INCOME	Constructed
INTIMP97	INTRPN IMPUTATION FLAG	Constructed
DIVDP97X	PERSON'S DIVIDEND INCOME	Constructed
DIVIMP97	DIVDPN IMPUTATION FLAG	Constructed
SALEP97X	PERSON'S SALES INCOME	Constructed
SALIMP97	SALEPN IMPUTATION FLAG	Constructed
PENSP97X	PERSON'S PENSION INCOME	Constructed
PENIMP97	PENSPN IMPUTATION FLAG	Constructed
SSECP97X	PERSON'S SOCIAL SECURITY INCOME	Constructed
SSCIMP97	SSECPN IMPUTATION FLAG	Constructed
TRSTP97X	PERSON'S TRUST/RENT INCOME	Constructed
TRTIMP97	TRSTPN IMPUTATION FLAG	Constructed
VETSP97X	PERSON'S VETERAN'S INCOME	Constructed
VETIMP97	VETSPN IMPUTATION FLAG	Constructed
IRASP97X	PERSON'S IRA INCOME	Constructed
IRAIMP97	IRASPN IMPUTATION FLAG	Constructed
REFDP97X	PERSON'S REFUND INCOME	Constructed
REFIMP97	REFDPN IMPUTATION FLAG	Constructed
ALIMP97X	PERSON'S ALIMONY INCOME	Constructed
ALIIMP97	ALIMPN IMPUTATION FLAG	Constructed
CHLDP97X	PERSON'S CHILD SUPPORT	Constructed
CHLIMP97	CHLDPN IMPUTATION FLAG	Constructed
CASHP97X	PERSON'S OTHER REGULAR CASH CONTRIB	Constructed
CSHIMP97	CASHPN IMPUTATION FLAG	Constructed
SSIP97X	PERSON'S SSI	Constructed
SSIIMP97	SSIPN IMPUTATION FLAG	Constructed
PUBP97X	PERSON'S PUBLIC ASSISTANCE	Constructed
PUBIMP97	PUBPN IMPUTATION FLAG	Constructed
OTHRP97X	PERSON'S OTHER INCOME	Constructed
OTHIMP97	OTHRPN IMPUTATION FLAG	Constructed

EMPLOYMENT VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
EMPST31	Employment Status Rd 3/1	EM 1-3; RJ 1, 6
EMPST42	Employment Status Rd 4/2	EM 1-3; RJ 1, 6
EMPST53	Employment Status Rd 5/3	EM 1-3; RJ 1, 6
RNDFLG31	Data Collection Round for Rd 3/1 CMJ	Constructed
MORJOB31	Has More Than One Job Rd 3/1 Int Date	EM 1-4, 51; RJ 1, 6; Constructed
MORJOB42	Has More Than One Job Rd 4/2 Int Date	EM 1-4, 51; RJ 1, 6; Constructed
MORJOB53	Has More Than One Job Rd 5/3 Int Date	EM 1-4, 51; RJ 1, 6; Constructed
EVRWRK	Ever Worked For Pay in Life as of 12/31/97	EM 1-4, 51; RJ 1, 6; Constructed
HRWG31X	Hourly Wage Rd 3/1 CMJ	EW 5, 7, 11-13, 17-18, 24; EM 104, 111
HRWG42X	Hourly Wage Rd 4/2 CMJ	EW 5, 7, 11-13, 17-18, 24; EM 104, 111
HRWG53X	Hourly Wage Rd 5/3 CMJ	EW 5, 7, 11-13, 17-18, 24; EM 104, 111
HRWGIM31	HRWG31X Imputation Flag	Constructed
HRWGIM42	HRWG42X Imputation Flag	Constructed
HRWGIM53	HRWG53X Imputation Flag	Constructed
HRHOW31	How Hourly Wage Was Calculated R3/1	EM 2-3, 51, 104, 111; EW 2-24
HRHOW42	How Hourly Wage Was Calculated R4/2	EM 2-3, 51, 104, 111; EW 2-24
HRHOW53	How Hourly Wage Was Calculated R5/3	EM 2-3, 51, 104, 111; EW 2-24
HOUR31	Hours Per Week at RD 3/1 CMJ	EM 1-3, 51, 104-105, 111; EW 17
HOUR42	Hours Per Week at RD 4/2 CMJ	EM 1-3, 51, 104-105, 111; EW 17
HOUR53	Hours Per Week at RD 5/3 CMJ	EM 1-3, 51, 104-105, 111; EW 17
SELFCM31	Self-Employed at RD 3/1 CMJ	EM 1-3, 51; RJ 01

VARIABLE	DESCRIPTION	SOURCE
SELFCM42	Self-Employed at RD 4/2 CMJ	EM 1-3, 51; RJ 01
SELFCM53	Self-Employed at RD 5/3 CMJ	EM 1-3, 51; RJ 01
DISVW31X	Disavowed Health Insurance at R3/1 CMJ	EM113, 117; RJ07, 08, 08A; HX and OE Sections
DISVW42X	Disavowed Health Insurance at R4/2 CMJ	EM113, 117; RJ07, 08, 08A, HX and OE Sections
DISVW53X	Disavowed Health Insurance at R5/3 CMJ	EM113, 117; RJ07, 08, 08A; HX and OE Sections
CHOIC31	Choice of Health Plans at Rd 3/1 CMJ	EM 1-3, 51, 96, 113-115, 124; RJ08
CHOIC42	Choice of Health Plans at Rd 4/2 CMJ	EM 1-3, 51, 96, 113-115, 124; RJ08
CHOIC53	Choice of Health Plans at Rd 5/3 CMJ	EM 1-3, 51, 96, 113-115, 124; RJ08
CIND31	Condensed Industry Code Rd 3/1 CMJ	EM 97-100; RJ01; Constructed
CIND42	Condensed Industry Code Rd 4/2 CMJ	EM 97-100; RJ01; Constructed
CIND53	Condensed Industry Code Rd 5/3 CMJ	EM 97-100; RJ01; Constructed
NUMEMP31	Number of Employees at Rd 3/1 CMJ	EM 91-92, 124; RJ01
NUMEMP42	Number of Employees at Rd 4/2 CMJ	EM 91-92, 124; RJ01
NUMEMP53	Number of Employees at Rd 5/3 CMJ	EM 91-92, 124; RJ01
MORE31	Rd 3/1 CMJ Firm Has More Than One Location	EM 1-3, 51, 94; RJ01

VARIABLE	DESCRIPTION	SOURCE
MORE42	Rd 4/2 CMJ Firm Has More Than One Location	EM 1-3, 51, 94; RJ01
MORE53	Rd 5/3 CMJ Firm Has More Than One Location	EM 1-3, 51, 94; RJ01
UNION31	Union Status at Rd 3/1 CMJ	EM 1-3, 51, 96, 116; RJ01
UNION42	Union Status at Rd 4/2 CMJ	EM 1-3, 51, 96, 116; RJ01
UNION53	Union Status at Rd 5/3 CMJ	EM 1-3, 51, 96, 116; RJ01
NWK31	Reason Not Working During Rd 3/1	EM 1-3, 101-102, 126-127, 132-133, 138-139, 141, 141.OV; RJ10
NWK42	Reason Not Working During Rd 4/2	EM 1-3, 101-102, 126-127, 132-133, 138-139, 141, 141.OV; RJ10
NWK53	Reason Not Working During Rd 5/3	EM 1-3, 101-102, 126-127, 132-133, 138-139, 141, 141.OV; RJ10
CHGJ3142	Changed Job Between Rd 3/1 and Rd 4/2	RJ01, 01A
CHGJ4253	Changed Job Between Rd 4/2 and Rd 5/3	RJ01, 01A
YCHJ3142	Why Chngd Job Between Rd 3/1 and Rd 4/2	RJ10, 10.OV
YCHJ4253	Why Chngd Job Between Rd 4/2 and Rd 5/3	RJ10, 10.OV
STJBMM31	Month Started Rd 3/1 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBDD31	Day Started Rd 3/1 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBYY31	Year Started Rd 3/1 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBMM42	Month Started Rd 4/2 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A

VARIABLE	DESCRIPTION	SOURCE
STJBDD42	Day Started Rd 4/2 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBYY42	Year Started Rd 4/2 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBMM53	Month Started Rd 5/3 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBDD53	Day Started Rd 5/3 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
STJBYY53	Year Started Rd 5/3 CMJ	EM10, 10.OV, 10.OV2; RJ01, 01A
EVRETIRE	Person Has Ever Retired	EM 1-3, 101-102, 126-127, 132-133, 138-139, 141, 141.OV; RJ 01, 10
COCCP31	Condensed Occupation Code Rd 3/1 CMJ	EM99-100; RJ 01, 01A; Constructed
COCCP42	Condensed Occupation Code Rd 4/2 CMJ	EM99-100; RJ 01, 01A; Constructed
COCCP53	Condensed Occupation Code Rd 5/3 CMJ	EM99-100; RJ 01, 01A; Constructed
BGNWK31	Usual Start Time of Rd 3/1 CMJ	EM 105, 105A, 105OV; RJ 01, 02
BGNWK42	Usual Start Time of Rd 4/2 CMJ	EM 105, 105A, 105OV; RJ 01, 02
BGNWK53	Usual Start Time of Rd 5/3 CMJ	EM 105, 105A, 105OV; RJ 01, 02
ENDWK31	Usual End Time of Rd 3/1 CMJ	EM 105, 105A, 105OV; RJ 01, 02

VARIABLE	DESCRIPTION	SOURCE
ENDWK42	Usual End Time of Rd 4/2 CMJ	EM 105, 105A, 105OV; RJ 01, 02
ENDWK53	Usual End Time of Rd 5/3 CMJ	EM 105, 105A, 105OV; RJ 01, 02
PAYVAC31	Paid Vacation at Rd 3/1 CMJ	EM 1-3, 51, 109; RJ 01, 02
PAYVAC42	Paid Vacation at Rd 4/2 CMJ	EM 1-3, 51, 109; RJ 01, 02
PAYVAC53	Paid Vacation at Rd 5/3 CMJ	EM 1-3, 51, 109; RJ 01, 02
SICPAY31	Paid Sick Leave at Rd 3/1 CMJ	EM 1-3, 51, 107; RJ 01, 02
SICPAY42	Paid Sick Leave at Rd 4/2 CMJ	EM 1-3, 51, 107; RJ 01, 02
SICPAY53	Paid Sick Leave at Rd 5/3 CMJ	EM 1-3, 51, 107; RJ 01, 02
PAYDR31	Paid Leave to Visit Dr Rd 3/1 CMJ	EM 1-3, 51, 107-108; RJ 01, 02
PAYDR42	Paid Leave to Visit Dr Rd 4/2 CMJ	EM 1-3, 51, 107-108; RJ 01, 02
PAYDR53	Paid Leave to Visit Dr Rd 5/3 CMJ	EM 1-3, 51, 107-108; RJ 01, 02
RETPLN31	Pension Plan at Rd 3/1 CMJ	EM 1-3, 51, 110; RJ 01, 02
RETPLN42	Pension Plan at Rd 4/2 CMJ	EM 1-3, 51, 110; RJ 01, 02
RETPLN53	Pension Plan at Rd 5/3 CMJ	EM 1-3, 51, 110; RJ 01, 02
SHFTWK31	Irregular Work Shift at Rd 3/1 CMJ	EM 1-3, 51, 105; RJ 01, 02
SHFTWK42	Irregular Work Shift at Rd 4/2 CMJ	EM 1-3, 51, 105; RJ 01, 02
SHFTWK53	Irregular Work Shift at Rd 5/3 CMJ	EM 1-3, 51, 105; RJ 01, 02
BSNTY31	Sole Prop, Partner, Corp, Rd 3/1 CMJ	EM 1-3, 51, 94-95; RJ 01, 02
BSNTY42	Sole Prop, Partner, Corp, Rd 4/2 CMJ	EM 1-3, 51, 94-95; RJ 01, 02
BSNTY53	Sole Prop, Partner, Corp, Rd 5/3 CMJ	EM 1-3, 51, 94-

VARIABLE	DESCRIPTION	SOURCE
		95; RJ 01, 02
JOBORG31	Priv (Profit/Nonprofit) Gov Rd 3/1 CMJ	EM 1-3, 51, 96; RJ 01, 02
JOBORG42	Priv (Profit/Nonprofit) Gov Rd 4/2 CMJ	EM 1-3, 51, 96; RJ 01, 02
JOBORG53	Priv (Profit/Nonprofit) Gov Rd 5/3 CMJ	EM 1-3, 51, 96; RJ 01, 02
HELD31X	Health Insurance Held from Rd 3/1 CMJ	EM117; HX, HP and OE Sections
HELD42X	Health Insurance Held from Rd 4/2 CMJ	EM117; HX, HP and OE Sections
HELD53X	Health Insurance Held from Rd 5/3 CMJ	EM117; HX, HP and OE Sections
OFFER31X	Health Insurance Offered by Rd 3/1 CMJ	EM113, 114, 117; RJ and HX Sections
OFFER42X	Health Insurance Offered by Rd 4/2 CMJ	EM113, 114, 117; RJ and HX Sections
OFFER53X	Health Insurance Offered by Rd 5/3 CMJ	EM113, 114, 117; RJ and HX Sections

HEALTH INSURANCE VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
CHmm97X	Covered by Champus/Champva in mm 97 (Ed), where mm = JA – DE	HX12, 13, PR19-22, HQ Section, RE14, 96A, and age at interview date
MCRmm97	Covered by Medicare in mm 97, where mm = JA – DE	HX05-07, 27, 29, 29OV
MCRmm97X	Covered by Medicare in mm 97 (ED), where mm = JA – DE	HX05-07, 27, 29, 29OV, see documentation, section 2.5.8, for additional edit specifications
MCDmm97	Covered by Medicaid in mm 97, where mm = JA – DE	HX10-11, PR07-10 and HQ Section
MCDmm97X	Covered by Medicaid in mm 97 (ED), where mm = JA – DE	MCDJA97, HX14-16, 18-19, 41-43, 45, PR11-14, 23-32, 39-42
OPAmm97	Covered by Other Public A Ins in mm 97, where mm = JA – DE	HX14-15, 41-45, PR 23-32 and HQ Section
OPBmm97	Covered by Other Public B Ins in mm 97, where mm = JA – DE	HX14-15, 41-43, PR23-30 and HQ Section
STAmm97	Covered by Other State Prog in mm 97, where mm = JA – DE	HX16-19, PR35-38 and HQ Section
PUBmm97X	Covered by Any Public Ins in mm 97 (ED), where mm = JA – DE	CHJA97X, MCRJA97X, MCDJA97X, OPAJA97, OPBJA97
PEGmm97	Covered by Empl Union Ins in mm 97, where mm = JA – DE	HX2-4, 21-24, 48; HP, OE, HQ, EM, RJ Sections
PDKmm97	Covrd by Priv Ins (Source Unknwn) in mm 97, where mm = JA – DE	HX21-24, 48, HP, OE, and HQ Sections

VARIABLE	DESCRIPTION	SOURCE
PNGmm97	Covered by Nongroup Ins in mm 97, where mm = JA – DE	HX21-24, 48, HP, OE, and HQ Sections
POGmm97	Covered by Other Group Ins in mm 97, where mm = JA – DE	HX21-24, 48, HP, OE, and HQ Sections
PRSmm97	Covered by Self-Emp-1 Ins in mm 97, where mm = JA – DE	HX3, 4, 48, HQ, OE, RJ and EM sections
POUmm97	Covered by Holder Outside of RU in mm 97, where mm = JA – DE	HX21-24, 48, HP, OE, and HQ Sections
PRImm97	Covered by Private Ins in mm 97, where mm = JA – DE	POGJA97, PDKJA97, PEGJA97, PRSJA97, POUJA97, PNGJA97
HPEmm97	Holder of Empl Union Ins in mm 97, where mm = JA – DE	PEGJA97, HP9, 11
HPDmm97	Holder of Priv Ins (Source Unknwn) in mm 97, where mm = JA – DE	PDKJA97; HP11
HPNmm97	Holder of Nongroup Ins in mm 97, where mm = JA – DE	PNGJA97; HP11
HPOmm97	Holder of Other Group Ins in mm 97, where mm = JA – DE	POGJA97; HP11
HPSmm97	Holder of Self-Emp-1 Ins in mm 97, where mm = JA – DE	PRSJA97; HP9
HPRmm97	Holder of Private Insurance in mm 97, where mm = JA – DE	HPEJA97, HPSJA97, HPOJA97, HPNJA97, HRDJA97
INSmm97X	Covrd By Hosp/Med Ins in mm 97, where mm = JA – DE (Ed)	PUBJA97X, PRIJA97
PRVEV97	Ever have private insurance during 97	Constructed
CHPEV97	Ever have CHAMPUS/CHAMPVA during 97	Constructed
MCDEV97	Ever have Medicaid during 97	Constructed
MCREV97	Ever have Medicare during 97	Constructed
OPAEV97	Ever have other public A during 97	Constructed
OPBEV97	Ever have other public B during 97	Constructed
UNINS97	Uninsured all of 97	Constructed
INSCOV97	Health insurance coverage indicator 97	Constructed

HEALTH STATUS VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
RTHLTH31	Perceived Health Status (R3-R1)	CE 1
RTHLTH42	Perceived Health Status (R4-R2)	CE 1
RTHLTH53	Perceived Health Status (R5-R3)	CE 1
RTPROX31	Self/Proxy Rating Of Health (R3-R1)	CE 1OV
RTPROX42	Self/Proxy Rating Of Health (R4-R2)	CE 1OV
RTPROX53	Self/Proxy Rating Of Health (R5-R3)	CE 1OV
MNHLTH31	Perceived Mental Health Status (R3-R1)	CE 2
MNHLTH42	Perceived Mental Health Status (R4-R2)	CE 2
MNHLTH53	Perceived Mental Health Status (R5-R3)	CE 2
MNPROX31	Self/Proxy Rating Of Mental Health (R3-R1)	CE 2OV
MNPROX42	Self/Proxy Rating Of Mental Health (R4-R2)	CE 2OV
MNPROX53	Self/Proxy Rating Of Mental Health (R5-R3)	CE 2OV
IADLHP31	IADL Screener (R3-R1)	HE 2-4
IADLHP42	IADL Screener (R4-R2)	HE 2-4
IADLHP53	IADL Screener (R5-R3)	HE 2-4
ADLHLP31	ADL Screener (R3-R1)	HE 5-6
ADLHLP42	ADL Screener (R4-R2)	HE 5-6
ADLHLP53	ADL Screener (R5-R3)	HE 5-6
AIDHLP31	Used Assistive Devices (R3-R1)	HE 7-8
AIDHLP53	Used Assistive Devices (R5-R3)	HE 7-8
WLKLIM31	Limitations In Physical Functioning (R3-R1)	HE 9-18
WLKLIM53	Limitations In Physical Functioning (R5-R3)	HE 9-18
LFTDIF31	Difficulty Lifting 10 Pounds (R3-R1)	HE 11
LFTDIF53	Difficulty Lifting 10 Pounds (R5-R3)	HE 11
STPDIF31	Difficulty Walking Up 10 Steps (R3-R1)	HE 12
STPDIF53	Difficulty Walking Up 10 Steps (R5-R3)	HE 12
WLKDIF31	Difficulty Walking 3 Blocks (R3-R1)	HE 13
WLKDIF53	Difficulty Walking 3 Blocks (R5-R3)	HE 13
MILDIF31	Difficulty Walking A Mile (R3-R1)	HE 14
MILDIF53	Difficulty Walking A Mile (R5-R3)	HE 14
STNDIF31	Difficulty Standing 20 Minutes (R3-R1)	HE 15
STNDIF53	Difficulty Standing 20 Minutes (R5-R3)	HE 15
BENDIF31	Difficulty Bending/Stooping (R3-R1)	HE 16
BENDIF53	Difficulty Bending/Stooping (R5-R3)	HE 16
RCHDIF31	Difficulty Reaching Overhead (R3-R1)	HE 17
RCHDIF53	Difficulty Reaching Overhead (R5-R3)	HE 17
FNGRDF31	Difficulty Using Fingers To Grasp (R3-R1)	HE 18
FNGRDF53	Difficulty Using Fingers To Grasp (R5-R3)	HE 18

VARIABLE	DESCRIPTION	SOURCE
ACTLIM31	Any Limitation Work/Housewrk/School (R3-R1)	HE 19-20
ACTLIM53	Any Limitation Work/Housewrk/School (R5-R3)	HE 19-20
WRKLIM31	Work Limitation (R3-R1)	HE 20A
WRKLIM53	Work Limitation (R5-R3)	HE 20A
HSELIM31	Housework Limitation (R3-R1)	HE 20A
HSELIM53	Housework Limitation (R5-R3)	HE 20A
SCHLIM31	School Limitation (R3-R1)	HE 20A
SCHLIM53	School Limitation (R5-R3)	HE 20A
UNABLE31	Completely Unable To Do Activity (R3-R1)	HE 21
UNABLE53	Completely Unable To Do Activity (R5-R3)	HE 21
SOCLIM31	Social Limitations (R3-R1)	HE 22-23
SOCLIM53	Social Limitations (R5-R3)	HE 22-23
COGLIM31	Cognitive Limitations (R3-R1)	HE 24-25
COGLIM53	Cognitive Limitations (R5-R3)	HE 24-25
WRGLAS42	Wears Glasses or Contacts (R2-R4)	HE 26-27
SEEDIF42	Difficulty Seeing (W/Glasses/Contacts)	HE 28-29
BLIND42	Person Is Blind	HE 30
READNW42	Can Read Newsprint (W/Glasses/Contacts)	HE 31
RECPEP42	Can Recognize People (W/Glasses/Contacts)	HE 32
VISION42	Vision Impairment (Summary)	Constructed
HEARAD42	Person Wears Hearing Aid	HE 33-34
HEARDI42	Any Difficulty Hearing (W/Hearing Aid)	HE 35-36
DEAF42	Person Is Deaf	HE 37
HEARMO42	Can Hear Most Conversation	HE 38
HEARSM42	Can Hear Some Conversation	HE 39
HEARNG42	Hearing Impairment (Summary)	Constructed
ANYLIM97	Any Limitation (RD 3/1, RD 4/2 & RD 5/3)	Constructed
LIMACT42	Limited In Any Activities (<5 Years)	HE 40-41
PLYLIM42	Limited In Play Activity (<5 Years)	HE 42
CANTPL42	Can’t Participate In Usual Play (<5 Yr)	HE 43
SPCPRO42	In Special Program (<5 Years)	HE 44
DPTSHT42	Immunization For Dpt Shots (<7 Years)	HE 45
NUMDPT42	One Or Several Dpt Shots (<7 Years)	HE 46
POLSHT42	Immunization For Polio (<7 Years)	HE 47
NUMPOL42	One Or Several Polio Shots (<7 Years)	HE 48
MMRSHT42	Immunization For Measles/Mumps/Rubella	HE 49
HEPSHT42	Immunization For Hepatitis (<7 Years)	HE 49A
MOMPRO42	Problem Getting Along With Mother (5-17)	HE 50
DADPRO42	Problem Getting Along With Father (5-17)	HE 50
UNHAP42	Problem Feeling Unhappy Or Sad (5-17)	HE 50

VARIABLE	DESCRIPTION	SOURCE
SCHLBH42	Problem Behavior At School (5-17)	HE 50
HAVFUN42	Problem Having Fun (5-17)	HE 50
ADUPRO42	Problem Getting Along With Adults (5-17)	HE 50
NERVAF42	Problem Feeling Nervous/Afraid (5-17)	HE 50
SIBPRO42	Problem Getting Along With Sibs (5-17)	HE 50
KIDPRO42	Problem Getting Along With Kids (5-17)	HE 50
SPRPRO42	Problem With Sports/Hobbies (5-17)	HE 50
SCHPRO42	Problem With Schoolwork (5-17)	HE 50
HOMEBH42	Problem With Behavior At Home (5-17)	HE 50
TRBLE42	Problem Staying Out Of Trouble	HE 50
SPCSCH42	Need Special School Program (5-17)	HE 51
SPECED42	In Special Education (5-17)	HE 52
SPCHTH42	Received Speech Therapy (5-17)	HE 52B
PSYCNS42	Psychological Counseling	HE 52B
OCUPTH42	Received Occupational Therapy (5-17)	HE 52B
VOCSVC42	Received Vocational Services (5-17)	HE 52B
TUTOR42	Received Tutoring (5-17)	HE 52B
READIN42	Uses Reader Or Interpreter (5-17)	HE 52B
PHYTHR42	Received Physical Therapy (5-17)	HE 52B
LIFSKL42	Received Life Skills Training (5-17)	HE 52B
FAMCNS42	Received Family Counseling (5-17)	HE 52B
RECTHR42	Received Recreational Therapy (5-17)	HE 52B
OTHSVC42	Received Other School Services (5-17)	HE 52B
CANTSC42	Limited/Unable To Go To School (5-17)	HE 53
LMOACT42	Limited In Non-School Activity (5-17)	HE 54
HLTHY42	Child Resists Illness Well (0-17)	HE 55
NTHLTH42	Less Healthy Than Same Age Kids (0-17)	HE 55
GETSIC42	Child Catches Things Going Around (0-17)	HE 55
HGTFT42	Child’s Height – Feet (0-17)	HE 56
HGTIN42	Child’s Height – Inches (0-17)	HE 56
WGTLB42	Child’s Weight – Pounds (0-17)	HE 57
WGTOZ42	Child’s Weight – Ounces (0-17)	HE 57
CHLIM42	Child Has Any Limitation (0-17)	Constructed

WEIGHTS VARIABLES - PUBLIC USE

VARIABLE	DESCRIPTION	SOURCE
WTDPER97	Poverty/Mortality Adj Person Weight	Constructed
WTFAMF97	Proverty Adjusted Family Weight	Constructed
WTCFAM97	Pov Adj Family Wgts-Cps Fam On 12/31/97	Constructed
VARSTR97	Variance Estimation Stratum-1997	Constructed
VARPSU97	Variance Estimation Psu-1997	Constructed

Appendix 1: Estimation Procedures in the Household Component of the 1996 Medical Expenditure Panel Survey

Steven B. Cohen
Center for Cost and Financing Studies
Agency for Healthcare Research and Quality

Ralph DiGaetano and Huseyin Goksel
Westat, Inc

The authors wish to thank Steven R. Machlin, D.E.B. Potter, and Joel W. Cohen, Center for Cost and Financing Studies, AHRQ, for their careful reviews of the manuscript and their insightful comments and suggested revisions.

Estimation Procedures in the Household Component of the 1996 Medical Expenditure Panel Survey

Steven B. Cohen, AHCPR, Ralph DiGaetano and Huseyin Goksel, Westat

1.0 Introduction

The Household Component of the 1996 Medical Expenditure Panel Survey (MEPS) was designed to produce national and regional estimates of the health care utilization, expenditures, sources of payment and insurance coverage of the U.S. civilian non-institutionalized population for calendar year 1996. The MEPS includes surveys of medical providers, employers and other health insurance providers to supplement the data provided by household respondents. The design of the MEPS survey permits both person based and family level estimates. The scope and depth of this data collection effort reflects the needs of government agencies, legislative bodies, and health professionals for the comprehensive national estimates needed in the formulation and analysis of national health policies. It is the third in a series of national probability surveys conducted by the Agency for Health Care Policy and Research (AHCPR) on the financing and utilization of medical care in the United States. Prior surveys include the 1977 National Medical Care Expenditure Survey (NMCES) and the 1987 National Medical Expenditure Survey (NMES).

The Medical Expenditure Panel Survey (MEPS) household component is an ongoing annual survey, with each sample panel collecting data over a 30 month period to obtain information that covers two consecutive calendar years. The MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services and how they are paid, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. MEPS is unparalleled for the degree of detail in its data, as well as its ability to link health service medical expenditures and health insurance data to the demographic, employment, economic, health status, utilization of health services, and other characteristics of survey respondents. Moreover, MEPS is the only federally sponsored national survey that provides a foundation for estimating the impact of changes in sources of payment and insurance coverage on different economic groups or special populations of interest, such as the poor, elderly families, veterans, the uninsured, and racial and ethnic minorities. The survey is co-sponsored by the Agency for Health Care Policy and Research and the National Center for Health Statistics. Westat and the National Opinion Research Center (NORC) are the data collection organizations for the 1996 MEPS Household Survey.

The sample design of the household component of the MEPS can be characterized as a stratified multi-stage area probability design with disproportionate sampling to facilitate the selection of an oversample of minorities (Cohen, 1997). This report provides a detailed summary of sample yields for the three rounds of data collection that cover calendar year 1996. An overview is also provided of the weighting strategies adopted to obtain national estimates of health care parameters for the U.S. civilian non-institutionalized population. In addition, survey design complexities which require special consideration for variance estimation and analysis are discussed.

2.0 The MEPS Household Component

The set of households selected for the 1996 MEPS is a subsample of those participating in the National Health Interview Survey (NHIS). The NHIS is an on-going annual household survey of approximately 42,000 households (109,000 individuals) conducted by the National Center for Health Statistics to obtain national estimates for the U.S. civilian non-institutionalized population on health care utilization, health conditions, health status, insurance coverage and access. In addition to the cost savings achieved by eliminating the need to independently list and screen households, selecting a subsample of NHIS participants has resulted in an enhancement in analytical capacity of the resultant survey data. Use of the 1995 NHIS data in concert with the data collected for the 1996 MEPS provides an additional capacity for longitudinal analyses not otherwise available. Furthermore, the greater number and dispersion of the primary sampling units (PSUs) that comprise the MEPS national sample has resulted in improvements in precision over prior expenditure survey designs (Arnett et al., 1996; Cohen, 1996).

The MEPS Household Component (HC) consists of an overlapping panel design in which any given sample panel is interviewed a total of 5 times over 30 months to yield annual use and expenditure data for two calendar years. Design specifications for the 1996 MEPS required that the full series of interviews necessary to acquire calendar year information for 1996 should be completed in approximately 9,000 households. The same panel of households were interviewed in person three times over the course of the survey to obtain data on their health care experience for 1996 (J. Cohen, 1997).

The 1996 MEPS Household Component sample was selected from households that responded to the 1995 National Health Interview Survey (NHIS). More specifically, the 1996 MEPS Household sample linked to the 1995 NHIS was selected from a nationally representative NHIS sub-sample from 2 NHIS panels out of 4 to represent the nation, and encompassed half of the participating households in the NHIS sample during the second and third quarters of 1995. It should be noted that the NHIS has been designed to permit nationally representative subsamples to be selected by restricting the sample to one of four distinct panels. Any combination of 1 to 4 panels will provide a nationally representative sample of households. Furthermore, each NHIS panel subsample for a given quarter of a calendar year is nationally representative.

The complete 1995 NHIS sample (panels 1-4) consists of 358 primary sampling units (PSUs: counties or groups of contiguous counties) with a targeted sample of approximately 42,000 responding households. The sample PSUs selected for the NHIS were stratified by geographic area (Census region and state), metropolitan statistical area, and socio-demographic measures (Judkins, Marker and Waksberg, 1994). Within sample PSUs, a sample of blocks (segments) were selected after being stratified by measures of minority population density, used to oversample areas with high population concentrations of blacks and Hispanics. A nationally representative sample of approximately 71,000 addresses within sampled blocks was selected and targeted for further screening to include an oversample of household containing blacks and Hispanics as part of the 1995 NHIS interview.

The nationally representative 1995 NHIS subsample selected for the 1996 MEPS consists of 195 PSUs, and in the two targeted quarters of 1995 these PSUs include 1,675 sample segments (second stage sampling units) and 10,597 responding NHIS households. This NHIS sample reflects an over-sample of households with Hispanics and blacks at the following approximate ratios of representation relative to the remaining households (Hispanics 2.0:1, blacks 1.5:1). The sample size for the 1996 MEPS was targeted at approximately 9,000 reporting units (generally families or single persons) yielding the complete series of core interviews (i.e., Rounds 1-3) to obtain use and expenditure data for calendar year 1996.

2.1 Procedures for Data Collection and MEPS Sampling Unit Definitions

Five in person interviews were conducted with each NHIS panel selected for the MEPS at three- to four-month intervals over an approximately 30-month field period. The first three of these rounds (Rounds 1-3) defined the 1996 MEPS Household survey, and serve to collect the main body of annual utilization and expenditure data for calendar year 1996. Rounds 3-5 for the 1996 MEPS panel cover 1997 and will be combined with Rounds 1-3 of the 1997 MEPS panel to yield the full sample base for the 1997 MEPS household survey and the source of annual estimates for that calendar year. All interviews were conducted in person using a computer assisted personal interview (CAPI) as the principal data collection mode. Round 1 asks about the period since January 1 of the MEPS year to the date of that interview; Round 2 asks about the time since the Round 1 interview through the date of the Round 2 interview; and Round 3 collects data from the date of the Round 2 interview in 1996 through the date of the Round 3 interview in 1997. Thus, Round 3 covers both calendar years, and the data are partitioned accordingly for estimation purposes.

The definitions for Dwelling Units and Group Quarters in the MEPS Household Component are generally consistent with the definitions employed for the National Health Interview Survey. While the MEPS sample is a subsample of NHIS dwelling units (referred to as households), a reporting unit for MEPS data collection purposes was defined as a person or group of persons in a sampled dwelling unit that are related by blood, marriage, adoption or other family associations, who were to be interviewed at the same time. Therefore, when unrelated persons were living in the same dwelling unit, sample households were split into multiple reporting units. Examples of the relationship between sample dwelling units and corresponding reporting units are:

1. A married daughter and her husband living with her parents in the same dwelling are considered one reporting unit.

2. A husband and wife and their unmarried daughter, age 18, who is living away from home at college constitute two reporting units operationally (though only one family unit analytically).

3. Three unrelated persons living in the same dwelling unit represent three reporting units.

In the first round of the 1996 MEPS, there was an average of 1.09 reporting units per sample dwelling unit. This low average reflects the fact that a substantial proportion of sample dwelling units contained only one reporting unit.

In MEPS, analyses are conducted at both the individual and family levels. Through the reenumeration section of the Round 1 questionnaire, the status of each individual sampled at the time of the NHIS interview is classified as “key or non-key” and “inscope or out-of-scope”. The “keyness” and “scope” indicators, together, define the target sample to be used for person level national estimates. They are discussed in detail below.

Inscope Persons: An individual is inscope whenever the person is a member of the civilian non-institutionalized population. Because a person’s eligibility for the survey may have changed since the NHIS interview, sampling reenumeration takes place in each subsequent reinterview for persons in all households selected into the core survey.

Key Persons: A “key” survey participant is one whose chance of selection for MEPS is linked to the sample of households originally selected for the NHIS (college students interviewed in dorms for the NHIS are not included in MEPS; college students living away from home are included in MEPS when identified by their parents during MEPS enumeration as living away from home as the students usual place of residence). A person must be key in order to be eligible to receive a person level weight (other conditions must be met as well).

Key survey participants include all civilian non-institutionalized individuals who resided in households that responded to the nationally representative NHIS subsample reserved for the MEPS. Members of the armed forces that are on full time active duty and reside in responding NHIS households, which include other family members who are civilian non-institutionalized individuals, are also considered key persons. However, they are considered out of scope for person level estimates derived for the survey unless they re-enter the civilian non-institutionalized population for some time during 1996.

Individuals who join the NHIS reporting units that define the 1996 MEPS household sample (in Round 1 or later MEPS rounds) and did not have an opportunity for selection during the time of the NHIS interview will also be considered key persons. These include newborns, individuals who were in an institution or outside the country moving to the United States, and military personnel previously residing on military bases who join MEPS reporting units to live in the community.

College students under 24 years of age interviewed at dormitories in the 1995 NHIS were considered ineligible for the 1996 MEPS sample and not included in that sample. Furthermore, any unmarried college students under 24 years of age who responded to the 1995 NHIS interview while living away at school (not in a dormitory) were excluded from the sample if it was determined in the MEPS Round 1 interview that the person was unmarried, under 24 years of age, and a student with parents living elsewhere who resided at his/her current housing only during the school year. If, on the other hand, the person’s status at the time of the MEPS Round 1 interview was no longer that of an unmarried student under 24 years of age living away from home, then the person was retained in the 1996 MEPS sample as a key person.

Alternatively, at the time of the MEPS Round 1 interview with NHIS sample respondents, a question was asked to determine if there are any related college students under 24 years of age who usually live in the sampled household, but are currently living away from home and going to school. These college students were considered key persons and were identified and interviewed at their college address, but linked to the sampled household for family analyses. Some of these college students living away from home at the time of the Round 1 interview were identified as living in sampled household at the time of the 1995 NHIS interview. The remainder were identified at the time of the MEPS Round 1 interview with the NHIS sampled households.

Non-key Persons: Persons who were not living in the original sampled dwelling unit at the time of the 1995 NHIS interview and where part of the civilian non-institutionalized population at that time will be considered non-key. If such persons happen to be living in a MEPS sampled household in Round 1 or later rounds, MEPS data, (e.g., utilization and income) will be collected for the period of time they are part of the sampled unit to permit family analyses. Non-key persons who leave a sample family without an accompanying key, inscope person will not be recontacted for subsequent interviews. Non-key individuals are not part of the target sample used to obtain person level national estimates.

In situations where key inscope MEPS participants move out (in Round 1 or later rounds) and join or create another family, data on all members who are related by blood, marriage, adoption or foster care to the key inscope MEPS participants will also be obtained from the point in time that the key inscope person(s) joined the family. Similarly, data will be collected (in Round 1 and later rounds) on all related persons who join families already participating in MEPS, whether the new persons are key or nonkey.

Key, inscope MEPS participants who entered a nursing home, thus leaving the civilian, non-institutionalized population of the United States, also had data collected during their stay in the nursing home. All other key inscope persons who left the civilian, non-institutionalized population of the United States did not require any data collection for this period. Upon their return to the U.S. civilian noninstitutional population, these persons were once again subject to data collection in MEPS.

Eligible Persons: A person is eligible for data collection in MEPS if they are key and inscope. In addition, individuals who are nonkey and inscope and a member of a family with at least one member who is key and inscope are also eligible for data collection. Out of scope individuals who are full time active duty members of the armed forces are also eligible for MEPS data collection for the time period they are a member of a family with at least one member who is key and inscope.

3.0 MEPS Round 1 Field Results

The 1995 NHIS subsample eligible for the 1996 MEPS consisted of 10,639 responding NHIS dwelling units. Of these, 10,597 had sufficient information to permit MEPS data collection (99.6 percent). Table 1 summarizes response rates for MEPS (conditional on response to NHIS) at both the dwelling unit and reporting unit levels. The 10,509 sample dwelling units that had sufficient address information from the NHIS and were considered eligible for MEPS contained a total of 11,424 reporting units. Of these reporting units, 83.1 percent responded to the first MEPS interview, 2.2 percent could not be located, and 14.7 percent were located but did not participate in the MEPS interview.

In order for a reporting unit to be eligible for MEPS data collection, it had to include at least one "key" individual selected in the MEPS, who was a member of the civilian non-institutionalized population between 1/1/96 and the date of the MEPS interview. Within the 10,597 sampled dwelling units that constitute the MEPS Round 1 sample, 11,590 reporting units were identified and targeted for data collection. Of these, 166 reporting units were determined to be ineligible for the 1996 MEPS. MEPS sample ineligibility for Round 1 was a consequence of the following situations:

- All members of the reporting unit died prior to 1/1/96 (21);
- All members of the reporting unit were full time active duty members of the military prior to 1/1/96 (4);
- All members of the reporting unit were institutionalized prior to 1/1/96 (26);
- All members of the reporting unit left the U.S. prior to 1/1/96 (29);
- All members of the reporting unit ineligible for data collection (e.g. death and inst.) (86).

Of the 11,424 eligible reporting units targeted for interviews in Round 1, 9,488 responded to the first core MEPS interview (83.1 percent, Table 1). The remaining 1,936 eligible reporting units (16.9 percent) were classified nonrespondents due to the following reasons:

- refusal to complete the interview (1,506);
- unavailable during field period (43);
- unable to locate (251);
- illness (27); or
- other nonresponse (109).

Since the MEPS sample is a nationally representative sub-sample of households that were part of the 1995 NHIS, the response rate that has implications in the development of national estimates from MEPS is a function of the response rates to both surveys. Specifically, the overall Round 1 MEPS response rate can be derived as the product of the following three components:

1. the NHIS response rate achieved for the households eligible for the MEPS (93.9 percent)
2. the proportion of NHIS units selected that had sufficient information to permit inclusion in the MEPS data collection effort (99.6 percent)
3. the MEPS round 1 reporting unit conditional response rate (83.1 percent).

The combination of these factors resulted in a response rate of 77.7 percent (.939 x .996 x .831) for the 1996 MEPS Round 1 HC (Cohen and Machlin, 1998).

Table 1: Response Rates For 1996 MEPS Round 1

	Dwelling Units		Reporting Units	
	Number	**Percent**	**Number**	**Percent**
Sample Cases	10,597	--	11,590	--
Sample Eligibles[1]	10,509	100.0	11,424	100.0
Respondents[2]	8,793	83.7	9,488	83.1
Nonrespondents	1,716	16.3	1,936	16.9
Unable to Locate			251	2.2
Nonparticipants			1685	14.7

Note: Conditional on NHIS Response.

In Table 1, the conditional response rates for the 1996 MEPS Round 1 survey are shown at both the dwelling unit level and at the reporting unit level. Since there is generally a one-to-one correspondence between a dwelling unit and a reporting unit, the conditional response rates for both are very similar (83.7 vs. 83.1 percent, respectively). While the reporting unit level response rate is more meaningful from an operational perspective, the dwelling unit level response rate is also provided because it is at this level that the MEPS estimation weights are initially adjusted for nonresponse.

4.0 MEPS Sampling Weight Specifications

Due to the complex design of the MEPS Household Component, the MEPS sample data must be appropriately weighted to obtain approximately unbiased national estimates for the U.S. civilian non-institutionalized population. The sampling weights developed for this purpose reflect the disproportionate sampling adopted in the NHIS to oversample minority populations in addition to adjustments for the following:

---

[1]There were 88 sample dwelling units and 166 reporting units deemed ineligible for MEPS. To be eligible for MEPS, a unit must contain at least one individual from an NHIS sample household who was also a member of the civilian non-institutionalized population between 1/1/96 and the date of the MEPS interview. Most of these dwelling units consisted of individuals who died or were institutionalized prior to 1/1/96.

[2]A dwelling unit is classified as a respondent if at least one member reporting unit responded to the survey.

- Complete nonresponse of eligible sample units

- Partial response of survey participants providing data for only a portion of the time in 1996 during which they were eligible to respond

- Poststratification to more accurate population totals obtained from the Current Population Survey

The 1996 MEPS estimation weights are built from the estimation weights developed for the 1995 National Health Interview Survey. To reduce the impact of large sampling weights on resultant variances of survey estimates, the MEPS estimation weights reflect a weight trimming adjustment. The 1996 MEPS dwelling unit weights also include an initial ratio adjustment to population estimates for selected socio-economic measures derived from the full 1995 National Health Interview Survey and subsequent adjustments for nonresponse to the first round of the MEPS survey. In addition, the MEPS estimation weights developed at the person and family level reflect additional adjustments that poststratify the MEPS survey estimates to more accurate population totals obtained from the Current Population Survey. The details of the MEPS estimation weights development are described in this section.

4.1 Base weights for the 1996 MEPS

As a consequence of the survey linkage between the 1995 NHIS and the MEPS, the sampling weights developed for the NHIS serve as the base weights for the 1996 MEPS. More specifically, the base weight for the dwelling units selected in the 1996 MEPS is the nonresponse adjusted 1995 NHIS quarter-specific estimation weight of the reference person in the primary reporting unit of a sampled dwelling unit. The reference person is the person who owns or rents the house. This NHIS estimation weight reflects the household's probability for selection in the NHIS and adjustments for NHIS survey nonresponse.

More specifically, if

P(I) is the ith dwelling unit's probability of selection in the NHIS to represent the Qth Quarter of 1995, and includes disproportionate values associated with the oversampling of minorities, and

A(c) adjusts for NHIS nonresponse within nonresponse adjustment class c of which dwelling unit I is a member, then the NHIS estimation weight NHISWTQ(I), for the ith dwelling unit selected for the MEPS in quarter Q=2 or 3 would take the form

$$NHISWTQ(I) = (1/P(I)) * A(c)$$

The NHIS quarter specific base weight was obtained by using the final quarter basic NHIS weight on the 1995 NHIS analytical file delivered to AHCPR. The available estimation weight also included a first-stage ratio adjustment that adjusts the initial NHIS population estimates to Census estimates

for cross-classification of the population based on race/ethnicity (Hispanic, nonHispanic Black, other), Census region (East, Midwest, South and west) and MSA classification (MSA/nonMSA). This component needed to be factored out of the NHIS estimation weight, since the first-stage ratio adjustment was implemented in NHIS without reflecting the subsampling of NHIS PSUs for MEPS which occurred by MSA classification. Consequently, the initial MEPS base weight was specified as

$$WT.MEP.I = HIS.WT.BF/HIS.ADJ.$$

Use of the NHIS quarter-specific estimation weight across multiple quarters of 1995 to produce a national estimate required a division of the weight by the number of quarters being pooled. Since the MEPS sample was confined to quarters 2 and 3 of calendar year 1995, the NHIS quarter weight, HIS.WT.BF was initially divided by 2, HIS.WT.BF/2 . Since the MEPS sample was restricted to Panels 1 and 3 out of a 4 Panel NHIS design, it represented a 50 percent subsample of the NHIS. Consequently, the NHIS quarter weight, HIS.WT.BF/2, representing Quarters 2 and 3 needed to be further multiplied by 2 to reflect the 50 percent subsample considered for MEPS. Consequently, specification of the initial MEPS base weight as

$$WT.MEP.I = HIS.WT.BF/HIS.ADJ$$

reflects the restriction of the NHIS sample to quarters 2 and 3 and a 50 percent sample for the 1996 MEPS.

As mentioned previously, unmarried students between the ages of 17-23 living at dormitories who were respondents in the 1995 NHIS were not eligible for the 1996 MEPS. Furthermore, a very small set of NHIS dwelling units (54) that were determined to be eligible for the MEPS at the time of sample selection could not be linked back to the 1995 NHIS analytical file that was provided a year later. The following strategy was implemented to obtain a base weight for these dwelling units. Median values of WT.MEP.I were determined for the dwelling units eligible for MEPS that linked back to the NHIS analytical file, based on classes defined by a cross classification of the minority status of the dwelling unit (1. DU has a Hispanic or black member, 2 otherwise) and 20 mutually exclusive and exhaustive distinct sampling strata defined for NHIS at the segment level for oversampling purposes. MEPS base weight assignments for the nonlinked cases were made based on the median value of WT.MEP.I for the class with which they were associated.

4.2 Trimming MEPS Base Weights

An initial examination of the distribution of the MEPS base weights identified a high level of variability. To correct for the impact of large sampling weights on resultant MEPS variance estimates, the initial MEPS base weights were trimmed according to the following specifications:

In each of the 40 classes (c) determined by a cross-classification of the dwelling unit's minority status and the 20 NHIS sampling strata defined at the segment level for oversampling purposes, the mean of the initial MEPS base weight, MEANDUWT(c) = MEAN(WT.MEP.I(iεc)) was computed.

For the dwelling units within a given class c, if the initial MEPS base weight was greater than 3 times the mean of the base weights, the weight was truncated to that value. Otherwise, they retained their initial value. More specifically, for class c, where c= 1,..40,

IF WT.MEP.I(I) >3*MEANDUWT(c), then
TRIMFAC = 3* (MEANDUWT(c)/ WT.MEP.I(I));

IF WT.MEP.I(I) $\leq$ 3*MEANDUWT(c)
then TRIMFAC = 1

Consequently, the trimmed MEPS weights were specified as

TRIMDUWT(I) = TRIMFAC*WT.MEP.I(I).

The sum of the initial MEPS base weights, reflecting an adjustment for NHIS nonresponse, but no correction for undercoverage was 90,754,892. Subsequent to the trimming adjustment, the sum of the MEPS base weights at the dwelling unit level was 90,647,643. While only a few sampling weights were modified, as can be noted in the modest reduction in the sum of the sampling weights, the largest MEPS base weight was reduced by nearly 50 percent.

4.3 Ratio Adjustment of the Trimmed MEPS Base Weights

To improve the accuracy of the MEPS estimates, the trimmed dwelling unit weights were subsequently ratio-adjusted to population estimates derived from the full 1995 NHIS, using data from the first 3 quarters of the 1995 NHIS (all of the 1995 NHIS that was available at the time of MEPS sampling weights development). The following measures were used in the specification of the ratio adjustment cells to facilitate the adjustment at the dwelling unit level:

1. MSA Status (MSA: Central City, MSA: Not Central City, Non-MSA)
2. Family Income classification of reference person (Below Poverty, Under $35,000 but above poverty, Equal or greater than $35,000, Unknown)
3. Employment status of reference person (employed, unemployed or not in labor force)
4. Race/Ethnicity of reference person (Hispanic, black Non-Hispanic, other)
5. Dwelling unit level measure of activity limitations (At least 1 person in DU either can’t perform major activity or is limited in kind and amount in major or other activities, No member in DU has an activity limitation).

These measures were selected to represent a set of measures that related to the oversampling done in NHIS (DU minority status and MSA status), and socio-economic and health specific measures potentially associated with health care use, expenditure and insurance coverage behavior that would not be used to define the family and person level poststratification adjustments in MEPS. For dwelling units associated with more than one reporting unit, the reference person and family income of the primary reporting unit were used for classification purposes. Cross-classification of these

measures yielded 144 weighting class cells to implement the ratio adjustment to more accurate national estimates at the household level based on the entire 1995 NHIS sample for quarters 1-3.

$$A(c) = \frac{NHISDU(c)}{\sum_{i \in c} TRIMDUWT(i)}$$

More specifically, the DU-level ratio adjustment for the c-th weighting class takes the form:

where iεc represents all NHIS dwelling units in c selected for the 1996 MEPS , TRIMDUWT(I) represents the trimmed initial NHIS base weight for the ith dwelling unit selected for MEPS, NHISDU(c) represents the national population estimate at the dwelling unit level for weighting class c, derived from the 1995 NHIS, using data from Quarters 1-3. Consequently, the ratio adjusted MEPS dwelling unit weight DUPSWT(I), for the ith dwelling unit associated with class c, adjusted to population estimates derived from the full 1995 NHIS, takes form:

$$DUPSWT(I) = A(c) \times TRIMDUWT(I).$$

The sum of the ratio adjusted and trimmed MEPS dwelling unit weights was 104,002,800 for the 10,597 NHIS dwelling units fielded for the 1996 MEPS.

4.4 MEPS Nonresponse Adjusted Dwelling Unit Weights: Round 1

Of the 10,509 dwelling units eligible for data collection in the first Round of the 1996 MEPS, 8,793 (83.7 percent) contained at least one reporting unit that responded to the MEPS interview. Since survey nonresponse is potentially a significant source of bias in survey estimates, the MEPS dwelling unit sampling weights included an adjustment for survey nonresponse to help reduce the potential for bias due to survey nonresponse. In general, the greater the difference among subgroups in response rates and the analytic characteristic(s) of interest, the greater the need to adjust survey weights for nonresponse. In the MEPS, a weighting class nonresponse adjustment was implemented, under the assumption that nonresponding sampling units would have responded in a similar manner as respondents with similar socio-demographic and economic characteristics within the same adjustment class. Properly designed, a weighting class nonresponse adjustment strategy can result in a reduction in nonresponse bias. The technique requires a partitioning of the sample into mutually exclusive classes, with classification information available for both responding and nonresponding units (Cox and Cohen, 1985).

Analyses of characteristics associated with differential nonresponse in MEPS were conducted to identify the most important measures to employ in the development of a nonresponse adjustment to the MEPS sampling weights to correct for potential nonresponse bias at the dwelling unit level (DiGaetano and Goksel, 1996; Cohen and Machlin, 1997). To facilitate these comparisons, the demographic, socio-economic, health related and interview specific profiles of respondents and nonrespondents were examined, based on available data for both groups from the 1995 National Health Interview Survey.

Based on the results of these analyses, weighting classes were specified for the MEPS Round 1 dwelling unit nonresponse adjustments, defined by cross-classifications of the following measures:

- Family income of primary reporting unit (<10K, 10-19K, 20-34K, 35K+, unknown)
- Size of dwelling unit (1,2,3,4,5+)
- MSA size (MSA-population 500K+; MSA- population under 500K; non-MSA)
- Region (northeast, midwest, south, west)
- Employment classification of reference person (Gov't, private sector, not in labor force/never worked/worked without pay, unknown or <18)
- DU level personal help measure (at least one member unable to perform personal care activities or other routine needs; remaining units with person 70 or older; remaining units with no limitations)
- Propensity to cooperate: Phone number provided during NHIS (phone number provided, phone with no number provided, no phone, unknown)
- Age of reference person (<24, 25-34, 35-44, 45-64, 65+)
- Race/ethnicity of reference person (Hispanic, black/nonHispanic, other)
- Sex of reference person
- Marital status (married-spouse present, other)

Overall, 49 cells were identified based on cross-classifications of these measures with cell collapsing specified according to a hierarchy determined by significance level. The reference person was defined as the person within the dwelling unit who owns or rents the sampled residence.

More specifically, the nonresponse adjustment for the c-th weighting class takes the form:

$$B(c) = \frac{\sum_{i \varepsilon c} E(i)DUPSWT(i)}{\sum_{i \varepsilon c} R(i)DUPSWT(i)}$$

where DUPSWT(I) is the initial MEPS Round 1 dwelling unit weight for the ith sample dwelling unit, which reflects the reciprocal of the dwelling unit's selection probability for MEPS and a poststratification adjustment to 1995 NHIS population totals,

E(I)=1 for all eligible MEPS dwelling units, E(I)=0 otherwise;

R(I)=1 for all eligible MEPS dwelling units responding in Round 1, R(I)=0 otherwise;

and I ε c represents eligible dwelling units classified in weighting class c.

Consequently, the estimation weight adjusted for MEPS Round 1 dwelling unit nonresponse, WGTDU1(I), for the ith dwelling unit associated with class c, takes the form:

WGTDU1(I) = B(c) x DUPSWT(I).

The sum of the nonresponse adjusted MEPS dwelling unit weights was 102,892,600 for the 8,793 eligible dwelling units with at least one responding reporting unit in Round 1 of the MEPS.

4.5 MEPS Family Level Estimation Weights: Round 1

In MEPS, a family was defined as a person or group of persons who are living together and are related by blood, marriage (or partnerships that are viewed as such), adoption or other family associations. Any related college students under 24 years of age who usually live in the sampled household, but are currently living away from home and going to school full time are considered to be members of the family. These college students are considered key persons in MEPS and are interviewed at their college address, but linked to the sampled household for family analyses. Families in MEPS without college students living away from home were identified as single reporting units. Families in MEPS with college students living away from home were identified by linking the student reporting unit(s) back to their parent(s) reporting unit.

In order to be considered a responding family in MEPS for the Round 1 interview, the family needed to include at least one person who was key, inscope and eligible for data collection. Furthermore, all such key, inscope and eligible persons had to have responded for their entire period of eligibility in 1996 covered by the Round 1 interview. Each family in MEPS characterized as responding was assigned the weight of its corresponding dwelling unit that had been adjusted for nonresponse:

$$WGTFAM(I) = WGTDU1(I)$$

Overall, 9,488 reporting units responded to the first round of the MEPS, which translated to 9,388 responding families after linking the responding student reporting units back to their parent(s) family.

The initial weights at the family level were then further poststratified to reflect population estimates obtained from the March 1996 Current Population Survey for unrelated individuals plus families. This poststratification also served as an adjustment for nonresponse at the family level. The weighting classes that were considered for the family level poststratification adjustment were defined by a cross-classification of the following variables defined at the time of the MEPS Round 1 interview:

- Family type: reference person married (oldest person when there is no reference person) and spouse present, male reference person and spouse not present, female reference person and spouse not present;

- Race/ethnicity of reference person (oldest person when there is no reference person) in family: Hispanic, Black-nonHispanic, Other-nonHispanic;

- Region: Northeast, Midwest, South and West;

- Metropolitan Statistical Area status: MSA, Non-MSA:

- Number of persons in family: 1, 2, 3, 4, 5+;

- Age of reference person (oldest person when there is no reference person) in family: <35, 35-44, 45-64, 65+.

More specifically, the family level poststratification adjustment for the c-th weighting class takes the form:

$$C(c) = \frac{CPSFAM(c)}{\sum_{i \in c} WGTFAM(i)}$$

where *CPSFAM*(c) represents the national population estimate at the family level for weighting class c, derived from the March 1996 Current Population Survey, iεc represents all MEPS family units classified in c that responded to the Round 1 interview, and *WGTFAM(I)* represents the initial MEPS family level weight for the ith family unit responding in the 1996 MEPS. Consequently, the Round 1 poststratified MEPS family unit weight *WGTRU1(I)*, for the ith family unit associated with class c, adjusted to population estimates derived from the March 1996 Current Population Survey, takes the form:

$$WGTRU1(I) = C(c) \times WGTFAM(I).$$

The weighted estimate of the number of family units (including single person units) containing at least one member of the U.S. civilian non-institutionalized population is 110,206,950. It is obtained by summing the poststratified MEPS family unit weights for the 9,388 MEPS family units that responded to the Round 1 interview. In the development of family level attributes in MEPS, it should be noted that all eligible and responding persons who are inscope, consisting of both key and nonkey individuals, should be included when deriving family level estimates.

4.6 MEPS Person Level Estimation Weights: Round 1

Key, inscope MEPS survey participants in a responding Round 1 reporting unit, for whom data were obtained for their entire Round 1 period of eligibility in 1996, were assigned estimation weights.

In order to be considered a responding survey participant in MEPS for the Round 1 interview, the person needed to be in scope with data provided for their entire Round 1 period of eligibility in 1996. All key, inscope and eligible sample participants in MEPS that satisfied this criterion for survey response were assigned estimation weights. The initial person level estimation weight assigned to these MEPS survey respondents was the corresponding poststratified family unit estimation of which they were a member,

$$WGTPER(I) = WGTRU1(I)$$

Overall, 23,612 key, inscope and eligible individuals were classified as survey respondents in the first round of the MEPS. The Round 1 MEPS person level weights were then poststratified to population totals obtained from the March 1996 Current Population Survey.

To establish consistency between family level and person level estimates in the MEPS, the reference person for each family (oldest person when there is no reference person), in addition to married couples living together, retained the value of the MEPS family unit weight as their final person weight:

$$WGTSP1(I) = WGTRU1(I)$$

The person level estimation weights of all other MEPS key, inscope and eligible survey respondents (e.g. child of reference person) were poststratified to population totals obtained from the March 1996 CPS within weighting classes defined by a cross-classification of the following variables:

- Region: Northeast, Midwest, South and West;
- Race/ethnicity: Hispanic, Black-nonHispanic, Other-nonHispanic;
- Sex: Male, Female;
- Age at interview date: <1, 1-4, 5-9, 10-14, 15-19, 20-29, 30-34, 35-44, 45-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+.

This adjustment served as both a nonresponse and poststratification adjustment at the person level.

The person level poststratification adjustment factor for the c-th weighting class takes the form

$$D(c) = 1$$

for the reference person for each family (oldest person when there is no reference person), in addition to married couples living together (denoted by *L(I) = 1; L(I) = 0 otherwise)* and for others, this takes the form

$$D(c) = \frac{CPSPER(c) - \sum_{i \in c} WGTPER(i)L(i)}{\sum_{i \in c} WGTPER(i)(1 - L(i))}$$

where *CPSPER*(c) represents the national population estimate at the person level for weighting class c, derived from the March 1996 Current Population Survey, iεc represents all MEPS key and inscope survey participants classified in c that responded to the Round 1 interview, and *WGTPER(I)* represents the initial MEPS person level weight for the ith person responding in the 1996 MEPS. Consequently, the Round 1 poststratified MEPS person weight *WGTSP1(I)*, for the ith person

associated with class c, adjusted to population estimates derived from the March 1996 Current Population Survey, takes the form:

$$WGTSP1(I) = D(c) \times WGTPER(I).$$

The weighted estimate of the number of persons who are members of the U.S. civilian non-institutionalized population as of the Spring of 1996 is 263,515,813. It can be derived by summing the poststratified MEPS person weights for the 23,612 MEPS key and inscope survey participants classified as respondents for the Round 1 interview , as indicated on the MEPS Household Component Public Use File HC-001:1996 Panel, Round 1 Population Characteristics.

4.7 MEPS Full Year 1996 Person Level Estimation Weights: Part-Year Nonresponse Adjustment

The MEPS Round 1 person level weight was developed to make estimates of the health care experience and insurance coverage profiles of the civilian non-institutionalized population for the first half of 1996. In order to facilitate the derivation of person level estimates that cover all of calendar year 1996, an annual person level weight for 1996 was also developed. Application of this weight will permit the derivation of national estimates of the health care use, expenditures, insurance coverage and sources of payment profiles for the civilian non-institutionalized population for calendar year 1996.

In order to be considered a responding survey participant in MEPS for the purpose of deriving annual 1996 estimates, the person had to be key and inscope with data provided for their entire period of eligibility in 1996. If all the key, inscope and eligible sample participants in MEPS with positive values for the MEPS Round 1 person level weight, in addition to new key and inscope respondents who joined a responding household in 1996 after Round 1 (here, the new respondent acquiring the sampling weight of the family they joined), responded for their entire period of eligibility in 1996, no additional adjustment for part year survey nonresponse over the course of Rounds 1-3 would be necessary. Of 23,881 sample participants identified in MEPS, 21,571 or 90.33 percent provided data for their entire period of eligibility in 1996. Consequently, the overall MEPS person level response rate for deriving annual estimates was 70.2 percent (.777 x .903), after factoring in the impact of survey attrition.

Since survey nonresponse is potentially a significant source of error or bias in survey estimates, the MEPS full year sampling weights included an adjustment for survey attrition to help reduce the potential impact of bias. An analysis of the characteristics that distinguish MEPS respondents with survey response for their entire period of eligibility in 1996, relative to the Round 1 participants that discontinued survey participation, was conducted to identify the most important variables to incorporate into the nonresponse adjustments to the MEPS sampling weights to correct for part-year survey nonresponse. This analysis was based on data from the first two rounds of the survey, due to the unavailability of relevant Round 3 data for 1996 at the time of this analysis. The results of a logistic regression analysis identified the most important measures to include in the specification of

a nonresponse adjustment to the MEPS sampling weights to correct for part-year response for calendar 1996 at the person level. Based on the results of these analyses, weighting classes were specified for the MEPS full year person level nonresponse adjustments, defined by cross-classifications of the following measures as of Round 1, or the first eligible Round in MEPS for key and inscope respondents who joined a household after Round 1 :

- Round 1 Interview Classification (No Initial Refusal, Initial Refusal)
- Size of MEPS family (1,2,3,4,5+)
- Metropolitan statistical area (MSA, nonMSA)
- Age (<20, 20-29, 30-44, 45-64, 65+)
- Marital Status of Reference Person (Married, Widowed, Divorced, Separated, Never Married)

Overall, 218 cells were identified based on cross-classifications of these measures with cell collapsing specified according to a hierarchy determined by significance level. The nonresponse adjustment for the c-th weighting class takes the form:

$$F(c) = \frac{\sum_{i \varepsilon c} E(i)WGSP1(i)}{\sum_{i \varepsilon c} R(i)WGTSP1(i)}$$

where WGTSP1(I) is the MEPS Round 1 poststratified person level weight for the ith Round 1 respondent, and WGTSP1(I) = WGTRU(1) for key and inscope respondents who joined a household in 1996 after Round 1 (here, the new respondent acquires the sampling weight of the family joined):

E(I)=1 for all MEPS Round 1 respondents with positive values of WGTSP1(I), and for key and inscope respondents who joined a responding household in 1996 after Round 1 with positive values of WGTSP1(I), E(I)=0 otherwise;

R(I)=1 for all persons with E(I) = 1 who responded for their entire period of eligibility in 1996, R(I)=0 otherwise;

and I ε c represents all key and inscope MEPS full and part-year respondents classified in weighting class c.

Consequently, the estimation weight adjusted for survey attrition in MEPS covering calendar year 1996, WGTSP2(I), for the ith person associated with class c, takes the form:

WGTSP2(I) = F(c) x WGTSP1(I).

for the 21,571 key and inscope survey participants that responded for their entire period of eligibility in 1996.

4.8 MEPS Full Year 1996 Person Level Estimation Weights

The subset of the 21,571 key and inscope survey participants that responded for their entire period of eligibility in 1996, who were also inscope on December 31, 1996, had their part-year nonresponse adjusted annual estimation weights further poststratified to Census Bureau population estimates as of December 1996. The person level estimation weights, WGTSP2(I), of the 21,326 sample participants that meet this criteria were poststratified to population totals obtained from the March 1997 CPS and further scaled to reflect Census Bureau population estimates as of December, 1996, within weighting classes defined by a cross-classification of the following variables:

- Sex: Male, Female;
- Age as of 12/31/96: <1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-44, 45-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+.
- Race/ethnicity: Hispanic, Black-nonHispanic, Other-nonHispanic;
- Region: Northeast, Midwest, South and West;
- Metropolitan statistical area (MSA, nonMSA)

Within each of the weighting classes (cεC) associated with a given age by sex cross-classification, the population totals derived from the March 1997 CPS were further adjusted by the factor, SCALE(C), which was defined as the ratio of the December, 1996 Census Bureau population estimate to the March 1997 population estimate derived from the CPS (see Table 2 on next page).

More specifically, the person level poststratification adjustment for the c-th weighting class takes the form:

$$G(c) = \frac{MAR97CPS(c) * SCALE(C)}{\sum_{i \in c} WGTSP2(i)}$$

where *MAR97CPS*(c) represents the national population estimate at the person level for weighting class c, derived from the March 1997 Current Population Survey; *SCALE*(c) represents the ratio of the December, 1996 Census Bureau population estimate to the March 1997 population estimate derived from the CPS for the specific cross-classification of age and sex associated with cell C; iεc represents all key and inscope survey participants associated with cell c (cεC) that responded for their entire period of eligibility in 1996, and were also inscope on December 31, 1996; and *WGTSP2(I)* represents the annual person level estimation weight adjusted for survey attrition in MEPS covering calendar year 1996. Consequently, the MEPS full year 1996 person level weight *WGTSP96(I)*, for the ith key, full year survey participant in scope as of 12/31/96 who is associated with class c, adjusted to population estimates derived from the March 1997 Current Population Survey and further scaled to Census Bureau estimates for December 1996, takes the form:

$$WGTSP96(I) = G(c) \times WGTSP2(I).$$

The remaining 245 key, inscope MEPS survey participants who responded for their entire period of eligibility in 1996, but were not inscope as of December 31, 1996 (e.g., persons who died during the survey year), maintained their estimation weight adjusted for survey attrition. Consequently, their MEPS full year person level weight was specified as

$$WGTSP96(I) = WGTSP2(I).$$

The weighted estimate of the number of persons who are members of the U.S. civilian non-institutionalized population as of December 1996 is 265,439,511. It can be derived by summing the poststratified MEPS person weights for the 21,326 MEPS key and inscope survey participants classified as respondents and inscope as of December 31, 1996, as indicated on the MEPS Household Component Public Use File HC-003:1996 Panel, Full Year Utilization Estimates.

Analysts who desire to produce cross-sectional national insurance coverage estimates as of December, 1996 with the MEPS data are advised to restrict their sample to this set of survey participants who were in scope as of 12/31/96.

Similarly, the weighted estimate of the number of persons who are members of the U.S. civilian non-institutionalized population over the course of 1996 is 268,130,477. It can be derived by summing the final poststratified MEPS person weights for the 21,571 MEPS key and inscope survey participants who responded for their entire period of eligibility in 1996. Analysts who desire to produce annual 1996 health care utilization and expenditure estimates are advised to include all of these 21,571 key and inscope MEPS survey participants for the purposes of estimation. A future MEPS data release will include estimation weights to support annual 1996 family-level health care use and expenditure estimation.

Table 2. Population estimates by sex and age for December 1996 and March 1997

Age Categories	December 1996[1/] Male	December 1996[1/] Female	March 1997[2/] Male	March 1997[2/] Female
0	1,989,676	1,902,075	1,959,414	1,856,886
1-4	8,162,745	7,787,372	8,118,381	7,846,632
5-9	10,322,287	9,848,773	10,428,712	9,842,699
10-14	9,990,073	9,526,892	9,970,793	9,534,281
15-19	9,723,221	9,327,806	9,766,326	9,377,121
20-24	8,728,774	8,730,416	8,635,045	8,719,707
25-29	9,354,544	9,674,812	9,450,287	9,636,289
30-34	10,339,061	10,652,556	10,242,979	10,548,081
35-44	21,458,434	22,046,082	21,546,881	22,147,277
45-54	16,001,003	16,759,899	16,063,859	16,891,429
55-59	5,446,069	5,892,934	5,569,213	6,003,373
60-64	4,644,237	5,191,637	4,686,964	5,206,925
65-69	4,415,868	5,236,207	4,321,136	5,179,642
70-74	3,721,987	4,787,277	3,764,159	4,750,254
75 or older	5,280,322	8,496,472	5,318,511	8,543,637
Total	129,578,301	135,861,210	129,842,460	136,084,232

Notes: [1/] December 1996 estimates are obtained from the Bureau of the Census.
[2/] March 1997 estimates are computed from the CPS, March 1997 data file.

5.0 Variance Estimation

The sample design of the MEPS Household Component includes stratification, clustering, multiple stages of selection, and disproportionate sampling. This complex sample design results in serious departures from simple random sampling assumptions. Furthermore, the MEPS sampling weights reflect differential adjustments for survey nonresponse and poststratification. These survey design and estimation complexities requires special consideration with regard to variance estimation and analysis. To obtain accurate estimates of the standard errors associated with MEPS person and family level survey estimates, for either descriptive statistics or more sophisticated multivariate model-based analyses, the MEPS survey design complexities need to be taken into account. Several methods for estimating sampling variances which adjust for survey design complexities have been developed that are appropriate for analytical applications tied to MEPS (Cohen, 1997). These variance estimation strategies include the Taylor series linearization method, balanced repeated replication and the jack-knife method.

Variables necessary for implementing a Taylor series variance estimation approach for survey estimates have been included on the MEPS public use files. Using such an approach, the sampling strata and associated primary sampling units (PSU) that define the MEPS survey design need to be specified. The corresponding variables on the MEPS Round 1 data base are VARSTRT1 and VARPSU1, respectively. Similarly, the corresponding variables on the MEPS Full Year 1996 Utilization Estimates data base are VARSTR96 and VARPSU96. Specifying a “with replacement” design in a variance estimation software package appropriate for the analysis of complex survey data that utilizes the Taylor series approach, such as SUDAAN or Stata, will yield standard errors that have been appropriately adjusted for survey design complexities (Shah et al., 1996).

It should be noted that the number of degrees of freedom associated with estimates of variability obtained by application of these statistical software packages may not appropriately reflect the actual number. For MEPS sample estimates for the general population derived at the national level, it is estimated that there are approximately 170 degrees of freedom for purposes of variance estimation.

6.0 Summary

As a consequence of the national scope and depth of the MEPS data collection effort, and the use of resultant survey estimates to inform national health policies, the adoption of estimation strategies that improve the quality and accuracy of survey estimates is of particular importance. Research was conducted to help ascertain potential sources of nonresponse bias that were attributable to MEPS dwelling unit nonresponse and to incorporate the findings in the specification of the MEPS nonresponse adjustment strategy to help reduce the impact of nonresponse bias. As a consequence of the MEPS sample linkage to the NHIS, detailed information on the socio-demographic and health characteristics of the eligible MEPS sample was available to inform the investigation. The results of this investigation revealed that the dwelling units responding to the first round of the MEPS household survey differed from the nonrespondents on a number of dimensions. Based on the results of the multivariate analysis, the effects of family income, dwelling unit size, health status of household members (as measured by personal help needs), phone availability, MSA size, and item

nonresponse for employment classification, were significant factors in distinguishing MEPS respondents. The measures most significant in differentiating MEPS survey response status were used in the specification of the MEPS Round 1 dwelling unit nonresponse adjustments. Through the identification of weighting classes in MEPS that capture the greatest variation across subgroups in response rates, a reduction in the bias attributable to survey nonresponse should be achieved.

An analysis of the characteristics that distinguish MEPS respondents with survey response for their entire period of eligibility in 1996, relative to the Round 1 participants that discontinued survey participation, was also conducted to identify the most important measures to include in the specification of a nonresponse adjustment to the MEPS estimation weights to correct for part-year survey nonresponse. The results of a logistic regression analysis that identified characteristics that distinguished the MEPS full year respondents from their part-year respondent counterparts identified the most important measures to include in the specification of a nonresponse adjustment to the MEPS sampling weights to correct for part-year response for calendar 1996 at the person level. Family size, residence by MSA classification, age, marital status and reluctance to participate were found to be important factors in distinguishing the MEPS full year respondents from their part year counterparts.

The overall MEPS person level response rate for deriving annual 1996 estimates was 70.2 percent, after adjusting for the multiplicative effects of nonresponse to the NHIS, nonresponse to the first round of the MEPS and the impact of survey attrition. Additional poststratification adjustments were incorporated in the development of the annual MEPS estimation weights, to further improve the accuracy of resultant MEPS survey estimates. The poststratification adjustments relied on population estimates derived from the Current Population Survey and other Census Bureau sources. A detailed summary of the MEPS estimation weight specifications has been provided in this report to ensure a better understanding of the estimation procedures that were adopted.

7.0 References

Arnett, R.A., E. Hunter, S. Cohen, J. Madans, and J. Feldman (1996). The Department of Health and Human Services' Survey Integration Plan. *American Statistical Association, Proceedings of the Section on Government Statistics*.

Berk, M.L. and A.B. Bernstein (1984). Interviewer Characteristics and Performance on a Complex Health Survey. *American Statistical Association, Proceedings of the Section on Survey Research Methods*.

Cohen, S.B. and S.R. Machlin (1998). "Characteristics of Nonrespondents in the MEPS Household Component". Submitted for publication consideration in *The Journal of Economic and Social Measurement*.

Cohen, S. B. (1997). "An Evaluation of Alternative PC-Based Software Packages Developed for the Analysis of Complex Survey Data". *The American Statistician*. Vol. 51, No. 3. 285-292.

Cohen, S. B. (1997). “Sample Design of the 1996 Medical Expenditure Panel Survey Household Component”. Rockville (MD): Agency for Health Care Policy and Research; 1997. *MEPS Methodology Report, No. 2*. AHCPR Pub. No. 97-0027.

Cohen, J. W. (1997). “Design and Methods of the Medical Expenditure Panel Survey Household Component”. Rockville (MD): Agency for Health Care Policy and Research; 1997. *MEPS Methodology Report, No. 1*. AHCPR Pub. No. 97-0026.

Cohen, S. B. (1996). The Redesign of the Medical Expenditure Panel Survey: A Component of the DHHS Survey Integration Plan. *Proceedings of the COPAFS Seminar on Statistical Methodology in the Public Service.*.

Cohen, S.B. and B.L. Carlson (1995). Characteristics of Reluctant Respondents in the national Medical Expenditure Survey. *Journal of Economic and Social Measurement*. 21: 269-296.

Cox, B.G. and S.B. Cohen (1985). *Methodological Issues for Health Care Surveys*. Marcel Dekker, New York.

DiGaetano, R. And H. Goksel (1996). Internal Report on Factors Associated With MEPS Round 1 Dwelling Unit Nonresponse. Westat, Inc.

Judkins, D., D. Marker and J. Waksberg (1994). National Health Interview Survey: Research for the 1995 Redesign, Draft Report. Westat Report prepared for the National Center for Health Statistics.

Shah, B.V., B. G. Barnwell, G.S. Bieler, K.E. Boyle, R.E. Folsom, L. Lavange, S.C. Wheeless, and R. Williams (1996). *Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0*, Research Triangle Park, NC: Research Triangle Institute.

Singer, E. M. Frankel and M. Glassman (1983). The Effect of Interviewer Characteristics and Expectations on Response. *Public Opinion Quarterly*. 47:68-83.

Appendix 2: Sample Design of the 1997 Medical Expenditure Panel Survey: Household Component

Steven B. Cohen, Ph.D.
Division of Statistical Research and Methods
Center for Cost and Financing Studies
Agency for Healthcare Research and Quality

1.0 Introduction

The Household Component of the 1997 Medical Expenditure Panel Survey (MEPS) was designed to produce national and regional estimates of the health care use, expenditures, sources of payment, and insurance coverage of the U.S. civilian noninstitutionalized population. MEPS includes surveys of medical providers, employers, and other health insurance providers to supplement the data provided by household respondents. The MEPS design permits both person-based and family-level estimates. Government agencies, legislative bodies, and health professionals need comprehensive national estimates to use in formulating and analyzing national health policies. The scope and depth of this data collection effort reflect this need. MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of private health insurance held by and available to the U.S. population. MEPS is unparalleled for the degree of detail in its data. In addition, through MEPS, the medical expenditures and health insurance data of survey respondents can be linked to other characteristics such as demographic variables, employment status, economic status, health status, and use of health services. Moreover, MEPS is the only national survey that provides a foundation for estimating the impact of changes in sources of payment for health services and insurance coverage on different economic groups or special populations of interest, such as the poor, the elderly, veterans, the uninsured, and racial and ethnic minorities.

The MEPS reflects the first stage of implementation of the Department of Health and Human Services (DHHS) Survey Integration Plan, which provides directives targeted to improve the analytic capacity of programs, fill major data gaps, and establish a framework in which DHHS data activities are streamlined and rationalized. Through this effort, specifically through a linkage to NHIS, MEPS has achieved a number of significant design improvements and analytic enhancements (S. Cohen, 1997, J. Cohen, 1997, Arnett et al., 1996, Hunter et al., 1997).

In this report, the sample design of the Household Component of the 1997 MEPS is described in detail. Particular emphasis is given to a description of the overlapping panel design that characterizes the survey. Attention is given to the sample selection scheme implemented to facilitate an oversample of the functionally impaired, children with activity limitations, individuals predicted to incur high medical expenditures and individuals predicted to have incomes less than 200% of poverty level. The report also includes a summary of sample size specifications, survey response rates and targeted precision levels for national population estimates and health care expenditure estimates for policy-relevant population subgroups.

2.0 Sample Design

To fill major data gaps identified by the Department of Health and Human Services, the Medical Expenditure Panel Survey was specified as a continuous survey. The sample design of the 1997 MEPS Household Component is an overlapping panel design. Health care data are collected for each new MEPS sample (Panel) to cover a two-year period, with the first two MEPS Panels spanning 1996-97 and 1997-98, respectively. To produce health care estimates for calendar year 1997, the data are pooled across the two distinct nationally representative MEPS samples. More specifically, the

1997 design combines the second year of the first MEPS panel and the first year of the second MEPS panel. The National Health Interview Survey (NHIS) serves as the sampling frame for the MEPS. The NHIS is an ongoing annual household survey of approximately 42,000 households (109,000 individuals) conducted by the National Center for Health Statistics (NCHS) to obtain national estimates on health care use, health conditions, health status, insurance coverage, and access for the U.S. civilian noninstitutionalized population. In addition to the cost savings achieved by substituting the NHIS as the MEPS sample frame, relative to the use of an independent national screener sample, this design feature enhances the analytic capacity of the resultant survey data. Use of the NHIS data in concert with the MEPS data provides an additional capacity for longitudinal analyses not available in the prior national medical expenditure surveys sponsored by the Agency for Healthcare Research and Quality (AHRQ) (S.Cohen, 1996).

The analytical goals of the 1997 Medical Expenditure Panel Survey and budget constraints required that the sample design for the Household Component meet the following requirements:

- The full series of interviews for the pooled MEPS samples covering calendar year 1997 should be completed in approximately 13,500 households.

- The sample should be spread over 195 separate areas to represent the civilian noninstitutionalized population of the 50 States and the District of Columbia.

- The sample should yield approximately unbiased national estimates of the health care parameters under study and estimates of adequate precision for the four census regions.

- The sample should meet predesignated precision specifications for the following population subgroups of analytical interest: blacks, Hispanics, the functionally impaired, children with activity limitations, individuals predicted to have high medical expenditures, and persons predicted to have family income less than 200 percent of the poverty level.

The 1996 MEPS Household Component sample was selected from households that responded to the 1995 NHIS. NHIS is designed to permit the selection of nationally representative subsamples from any one of four panels. Furthermore, any combination of one to four panels will provide a nationally representative sample of households. Each NHIS panel subsample for a given quarter of a calendar year is nationally representative. The 1996 MEPS household sample was selected from two of the four 1995 NHIS panels during the second and third quarters of 1995. Consequently, the MEPS sample is an approximately 1/4 sub-sample of the overall 1995 NHIS sample.

The complete 1995 NHIS sample (panels 1-4) consists of 358 primary sampling units, or PSUs (which are counties or groups of contiguous counties), and approximately 42,000 responding households. The NHIS sample design is characterized by a stratified multi-stage area probability design, where the sample PSUs are stratified by geographic area (Census region and State), metropolitan status, and sociodemographic measures (Judkins, Marker, and Waksberg, 1994). Within sample PSUs, a sample of blocks (segments) was selected after the blocks were stratified by measures of minority population density that allowed for an oversample of areas with high

population concentrations of blacks and Hispanics. A nationally representative sample of approximately 71,000 addresses within sampled blocks was selected and targeted for further screening to facilitate an oversample of blacks and Hispanics as part of the 1995 NHIS interview.

The 1995 NHIS subsample selected for the 1996 MEPS consists of 195 PSUs. In the two targeted quarters of 1995, these PSUs included 1,675 sample segments (second-stage sampling units) and 10,597 responding households. This NHIS sample reflects oversampling of households with Hispanics and blacks at a ratio of approximately 2.0:1 for Hispanics and 1.5:1 for blacks. This 1996 MEPS sample constitutes a panel that was surveyed to collect annual data for 2 consecutive years (S. Cohen, 1997).

A new 1997 MEPS panel sample was selected as a nationally representative subsample of households responding to the 1996 NHIS. More specifically, this 1997 MEPS sample was selected from the same two NHIS panels used for the 1996 MEPS, using a nationally representative subsample of the 1996 NHIS that also reflected an oversample of Hispanics and blacks at the same ratios as the 1995 NHIS (Hispanics, 2.0:1; blacks, 1.5:1). It should be noted that in 1996, the National Health Interview Survey was undergoing a transition from a paper and pencil survey administration design to a computer assisted personal interview. The nationally representative subsample of the NHIS reserved for the 1997 MEPS sample selection retained the paper and pencil survey administration mode, to allow for a smoother transition between the integration of the two national surveys.

The new 1997 MEPS sample was selected from the first three quarters of the 1996 NHIS subsample within the two panels reserved for the MEPS. This NHIS nationally representative subsample was concentrated within the same 195 PSUs selected for the 1996 MEPS household sample and consisted of 14,706 responding NHIS dwelling units. A nationally representative subsample of 6,300 NHIS responding dwelling units (consisting of 6,480 reporting units) was selected to serve as the new 1997 MEPS sample. In addition to retaining the oversample of minorities that characterized the NHIS sample design, the 1997 MEPS was designed to oversample the following policy-relevant subgroups: functionally impaired adults, children limited in activities, adults predicted to have high medical expenditures, and persons predicted to have family incomes less than 200 percent of the poverty level. The new 1997 MEPS panel was designed to collect annual data for 2 consecutive years. Consequently, the full 1997 MEPS Household Component sample consists of the first year of the 1997 MEPS panel pooled with the second year of the 1996 MEPS sample.

Sampling Unit Definitions and Eligibility Criteria

The definitions for dwelling units and group quarters in the MEPS Household Component are generally consistent with the definitions employed for NHIS. More specifically, a dwelling unit is a house, apartment, group of rooms, or single room occupied as separate civilian non-institutional living quarters or vacant but intended for occupancy as separate living quarters. Group quarters consist of a single civilian noninstitutional dwelling or structure in which nine or more unrelated persons reside and where inhabitants are not considered a part of any other dwelling unit. A reporting unit is a person or group of persons in the sampled dwelling unit that are related by blood, marriage,

adoption, or other family associations, and are to be interviewed at the same time in MEPS. Examples of discrete reporting units follow.

- A married daughter and her husband living with her parents in the same dwelling are considered one reporting unit.
- A husband and wife and their unmarried daughter, age 18, who is living away from home at college constitute one family, but two reporting units.
- Three unrelated persons living in the same dwelling unit would be three reporting units.

College students under 24 years of age who usually live in the sampled household but are currently living away from home and going to school are treated as separate reporting units for the purpose of data collection.

The new 1997 MEPS sample consisted of households (dwelling units) that responded to the 1996 NHIS in the two panels reserved for MEPS, with the basic unit of analysis defined as the person, which mirrored the 1996 MEPS design. Analysis is planned using both the individual and the family as units. Through the reenumeration section of the Round 1 questionnaire for each MEPS panel, the status of each individual sampled at the time of the NHIS interview is classified as “key” or “non-key,” “inscope” or “out-of-scope,” and “eligible” or “ineligible” for MEPS data collection. For an individual to be inscope and eligible for person-level estimates derived from the MEPS household survey, he or she must be a member of the civilian noninstitutionalized population for some period of time in the calendar year of analytic interest. Because a person’s eligibility for the survey may change after the NHIS interview, sampling reenumeration takes place in each subsequent reinterview for persons in all households selected into the core survey. The keyness, inscope, and eligibility indicators, together, define the target sample to be used for person-level national estimates. Only persons who are key, inscope, and eligible for data collection are considered in the derivation of person-level national estimates from MEPS.

Key Persons

Key survey participants are defined as all civilian noninstitutionalized individuals who resided in households that responded to the nationally representative NHIS subsample reserved for MEPS (e.g., 6,300 households from the 1996 NHIS), with the exception of college students interviewed at dormitories. Members of the Armed Forces who are on full-time active duty are also defined as key persons if they reside in responding NHIS households that include other family members who are civilian noninstitutionalized individuals. However, they are out of scope for person-level estimates derived from the survey.

All individuals who join the NHIS reporting units that define the 1997 MEPS household sample (in Round 1 or later MEPS rounds) and were not available for selection during the time of the NHIS interview are also considered key persons. These include newborn babies, individuals who were in an institution or outside the country, and military personnel previously residing on military bases.

College students under 24 years of age interviewed at dormitories in the 1996 NHIS are not included in the 1997 MEPS sample, since this population subgroup will be targeted through their parents during the MEPS interview. The same rule applied for the sample selected for the 1996 MEPS sample selected from the 1995 NHIS. Furthermore, any unmarried college student under 24 years of age who responded to the 1996 NHIS interview while living away at school (not in a dormitory) is excluded from the sample if it is determined in the MEPS Round 1 interview that the person is unmarried, under 24 years of age, and a student who has parents living elsewhere and who resides at his or her current housing only during the school year. If, on the other hand, the person's status at the time of the MEPS Round 1 interview is no longer that of an unmarried student under 24 years of age living away from home, then the person is retained as a key person.

Additionally, during the MEPS Round 1 interview with NHIS sample respondents, a determination is made whether there are any related college students under 24 years of age who usually live in the sampled household but are currently living away from home and going to school. These college students are considered key persons and are identified and interviewed at their college address but linked to the sampled household for family analyses. Some of these college students will have been identified as living in the sampled household at the time of the 1995 NHIS interview. The remainder are identified at the time of the MEPS Round 1 interview.

Non-Key Persons

Persons who were not living in the original sampled dwelling unit at the time of the 1996 NHIS for the 1997 new MEPS sample interview (the 1995 NHIS for the original 1996 MEPS sample) and who had a nonzero probability of selection for that survey are considered non-key. If such persons happen to be living in sampled households in Round 1 or later rounds, MEPS data are collected for the period of time they are part of the sampled unit to permit family analyses. Non-key persons who leave any sampled household are not re-contacted for subsequent interviews. Non-key individuals are not part of the target sample used to obtain person-level national estimates.

A key person from the NHIS sampled household selected for MEPS may move out in Round 1 or later rounds and join or create another family. Data on all members of this new household who are related by blood, marriage, adoption, or foster care to the person from the NHIS sampled household are obtained from the time that the sampled person joined the household. Keyness status is determined for these new members based on their probability of selection for the NHIS. If it is positive, they are classified as non-key. Similarly, data are collected in Round 1 and later rounds on all related persons who join NHIS sampled households selected into MEPS.

Persons in NHIS sampled households selected in MEPS may subsequently enter an institution, thus no longer qualifying as a member of the U.S. civilian noninstitutionalized population. For those who enter nursing homes, data collection continues during the nursing home stay. For those who enter other institutions, data collection is suspended while they are instiutionalized, but their whereabouts are monitored during the field period. If they rejoin the U.S. civilian noninstitutionalized population, HC data collection resumes. (This is also the procedure for those entering military service away from home or moving out of the United States.)

MEPS Data Collection Eligibility

In order for a MEPS reporting unit to be eligible for data collection, it must include at least one individual who is key and inscope for some period of time during the reference period for a given round of data collection. If this condition holds, the persons who are key and inscope and all other individuals who are members of the reporting unit (living together and related by blood, marriage, adoption, or other family associations) are eligible for data collection in a given round of MEPS.

3.0 Sample Selection of the 1997 MEPS Panel

Sample Size Targets and Precision Requirements

An overall precision requirement for the 1997 MEPS survey was the achievement of an average design effect of 1.7 for the survey estimates of the policy relevant population subgroups. The precision requirements for the 1997 HC are presented in Table 1. They are presented in terms of relative standard errors for the following survey estimates:

1) a 20 percent population estimate at the person level for each specified domain (e.g. a percent population estimate such as the rate of the uninsured for the population under age 65); and

2) mean estimates of the following measures of health care utilization and expenditures at the person level (precision requirement specified as an average relative standard error):

a. total health expenditures;
b. utilization and expenditure estimates for inpatient hospital stays;
c. utilization and expenditure estimates for ambulatory physician visits;
d. utilization and expenditure estimates for dental visits;
e. utilization and expenditure estimates for prescribed medicines.

Table 1:Targeted average relative standard errors (RSE's) for subpopulations of analytic interest in the 1997 MEPS

Subpopulation	Average RSE for a population estimate of 20% (e.g. % ninsured)	Average RSE for mean use and expenditure estimates
Persons with family incomes less than 200% of poverty level	.020	.035
Persons predicted to incur high medical expenditures	.040	.070
Persons 65 years or older	.042	.070
Adults (18+) with functional impairments (1 or more activities of daily living (ADLs))	.080	.135
Adults (18+) with other impairments (1 or more instrumental activities of daily living (IADLs)),	.080	.135
Children with limitations (age 17 or younger)	.080	.135
Overall population	.015	.023

Source: 1997 Medical Expenditure Panel Survey, AHRQ

The 1997 MEPS person level precision requirements are specified for estimates derived from individuals that are considered full year respondents (individuals with responses for their entire period of eligibility in 1997). Consequently, in the determination of sample sizes necessary to achieve the precision requirements, additional adjustments must be made for survey nonresponse to obtain the targeted number of full year respondents. Approximately 34,000 persons completing the three core MEPS household interviews to cover calendar year 1997 (Rounds 1-3 for the new 1997

MEPS sample; Rounds 3-5 for the carry-over 1996 MEPS sample) were targeted for sample selection to achieve the desired precision specifications for national population estimates. Assuming 2.5 persons per original sampled reporting unit, approximately 13,600 families completing the three rounds in 1997 were estimated as the necessary sample yield to meet precision specifications. Table 2 indicates the desired number of persons in the various subpopulations of interest for analysis necessary to satisfy the survey precision requirements for the pooled 1996 and 1997 MEPS samples to permit 1997 population estimates.

Table 2. Targeted sample yields at the end of three core data collection rounds for 1997 for subpopulations of analytic interest.

Subpopulation	Targeted sample yield
Persons under 200% of poverty level	15,000
Persons under age 65 with predicted high medical expenditures (top 15 % of the expenditure distribution)	4,000
Persons 65 years or older	3,700
Adults (18+) with functional impairments (1 or more ADLs)	1,000
Adults (18+) with other impairments (1 or more IADLs)	1,000
Children with limitations (age 17 or younger)	1,000
Overall population	34,000

Source: 1997 Medical Expenditure Panel Survey, AHRQ

Precision requirements for the 1997 MEPS Household Survey were stated in terms of national estimates at the person level. To meet these requirements, the survey must include a minimum number of persons in each domain of interest. The prior 1996 MEPS sample was also selected to satisfy distinct precision requirements at the person level for overall population estimates and for subgroup analyses of blacks and Hispanics for calendar year 1996 (S. Cohen, 1997). Projected yields in 1997 from the first MEPS Panel were derived to inform the specification of the final selection rates for the new 1997 MEPS sample in order to satisfy precision requirements for the pooled 1997 MEPS sample. Based on these projected sample yields (actual sample yields presented in Table 4), the additional sample sizes necessary to satisfy the precision requirements for the analytical domains were determined, after adjusting for expected survey response rates, and sampling rates were specified for the new 1997 sample to achieve these targets.

For both the 1996 and the 1997 MEPS, the unit of interviewing and subsampling was the household. To facilitate the sample selection of the new 1997 MEPS sample, the 1996 NHIS households were selected on the basis of the characteristics of the persons they included. There were seven sample domains of interest to which a NHIS dwelling unit could be assigned based on its composition with at least one member having the characteristic of interest:

- adults (age 18 and above) with functional impairments (at least 1 ADL requiring personal assistance);
- children with limitations in activity (under age 18);
- individuals 18-64 years old with predicted high medical expenditures (predicted probability is greater than or equal to .4, using the MEPS prediction model to identify likely high expenditure individuals);
- individuals with family incomes likely to be below 200% of poverty level (predicted probability is greater than or equal to .3, using the MEPS prediction model to identify low income households);
- Adults with other impairments (ages18-69 and at least 1 IADL and unable to work , age 70 and above and at least 1 IADL);
- elderly individuals (age 65 and above); and
- all remaining individuals.

These sampling domains were not mutually exclusive, but their order reflects the hierarchy of their sampling priority. For purposes of sampling, dwelling units containing members having the above characteristics were hierarchically classified based on the above ordering to form seven mutually exclusive and exhaustive sampling strata (DiGaetano, 1994).

Using Predictive Models for Domain Assignments

Poverty Status Model

Since a reporting unit's poverty status classification in 1997 was unknown at the time of the administration of the 1996 NHIS interview, a prediction model was used to determine whether a household was to be oversampled. More specifically, a logistic regression model was developed to estimate the probability that a reporting unit would have a family income less than 1.25 times the poverty level in a subsequent year based on the poverty status classification and other predictive measures obtained during the NHIS interview. Households with predicted probabilities above a certain threshold value were to be oversampled. In addition to facilitating an oversample of individuals with family incomes less than 125 percent of the poverty level, use of this prediction model was expected to facilitate an oversample of individuals with family incomes less than 200 percent of the poverty level. Consequently, all reporting units with a predicted probability of .3 or greater were classified as households predicted to have family incomes less than 200 percent of the poverty level.

The results listed below were observed based on an evaluation of the model's performance at the reporting unit level, using data from the prior 1987 National Medical Expenditure Survey (NMES2), and using a predicted probability of .3 or greater (derived from the logistic regression prediction model) as the criterion to target reporting units most likely to have members with family income less than 200 percent of the poverty level in 1996:

- Based on the NMES2 experience, the expected prediction rate for true positives (family income less than 200 percent of the poverty level) is 83.1 percent among the 19.5 percent of reporting units predicted to have members with family income less than 200 percent of the poverty level.
- The expected prediction rate for false negatives is 17.1 percent among the 80.5 percent of reporting units predicted to have family income equal to or greater than 200% of the poverty level.

Among the 30 percent of reporting units with family income less than 200 percent of the poverty level, 54 percent were predicted to have members with family income less than 200 percent of the poverty level. Alternatively, among the 70 percent of reporting units with family income above 200 percent of the poverty level, 95.3 percent were predicted to have members with family income above 200 percent of the poverty level.

The logistic regression model that was adopted was specified at the reporting unit level and required data on the following measures obtained in the NHIS interview (Moeller and Mathiowetz, 1994):

- Age of reference person;
- Home ownership;
- Reporting Unit size;
- Whether children of specific ages (under age 6, 6-15) are present in the RU;

- Whether someone in the RU other than the reference person is at least 65 years of age;
- Health status of reference person;
- Race/ethnicity of reference person;
- Census Division;
- Metropolitan Statistical Area (MSA) status of PSU;
- Education of reference person;
- Marital status and gender of reference person;
- Whether reference person or spouse was employed in the previous 3 months;
- Whether the family income of the reporting unit was less than 1.25 times the poverty level; and
- Whether anyone in the RU was covered by Medicaid.

High Expenditure Prediction Model

Among the sample domains to be oversampled in the main survey are individuals between the ages 18-64 who are predicted as likely to incur high medical expenditures. An individual's medical care expenditures in a future year will be unknown at the time of the administration of the 1996 NHIS interview; therefore, a prediction model based on NMES2 data was used to determine whether a household is to be oversampled as part of the high medical expenditures group because one or more of the family members are expected to incur high medical expenditures in the subsequent year. More specifically, a logistic regression model has been developed that estimates the expected probability an individual who is between the ages of 18-64 will incur high medical expenditures (top 15 percent of the health expenditure distribution) in a subsequent year based on predictive measures obtained during the NHIS interview. Households with at least one such person with a predicted probability above a certain threshold value were oversampled. The group was restricted to individuals who were between the ages 18-64, since the persons 65 or older that were functionally impaired were separately targeted to be oversampled. For purposes of sampling, all individuals with a predicted probability of .4 or greater were classified as likely to incur high medical expenditures in the subsequent year. This threshold was selected as the value that was expected to best limit prediction errors.

The logistic regression model under consideration was specified at the person level and requires data on the following measures obtained in the NHIS interview (Moeller and Mathiowetz, 1994):

- Gender
- Health status;
- Marital status;
- Poverty status;
- Whether the person lives alone;
- Age;
- Whether the person's health keeps him/her from working at a job, doing work around the house or going to school;

- Whether the person is unable to do certain kinds or amounts of work, housework, or schoolwork because of his/her health;
- The number of visits to a medical doctor or other medical care provider the person has had during the last 6 months;
- The number of times prescribed medicines were purchased or obtained for the person's use in the last 6 months (an imputation strategy was used to derive this measure since data were unavailable from the 1996 NHIS);
- Census Division; and
- MSA status of PSU.

The results listed below were observed based on an evaluation of the model's performance at the individual level, using data from NMES2, and using a predicted probability of .4 or greater (derived from the logistic regression prediction model) as the criterion to target individuals who are between the ages 18-64 and considered likely to incur high medical expenditures in the subsequent year:

- Based on the NMES2 experience, the expected prediction rate for true positives is 37.7 percent among the 14.1 percent of individuals in reporting units (computed at the reporting unit level) with members between the ages 18-64 who are predicted to incur high medical expenditures in the subsequent year. It should be noted that when restricting the evaluation to the subset of individuals (8.1 percent) that are predicted to incur high medical expenditures, the expected prediction rate for true positives is 65.3 percent (computed at the person level).

- The expected prediction rate for false negatives is 11.3 percent among the 85.9 percent of individuals in reporting units (computed at the reporting unit level) with members between the ages 18-64 who are predicted to not incur high medical expenditures in the subsequent year.

Sample Composition of 1996 NHIS Available for the 1997 MEPS Sample

In order to provide the 1997 MEPS sample to Westat and NORC (the MEPS data collection organizations) in the time frame specified to field the survey in February of 1997, it was necessary to restrict the sample selection from a nationally representative NHIS subsample confined to the first three quarters of 1996. This NHIS sample of 14,706 responding dwelling units was then classified into seven mutually exclusive and exhaustive sampling categories based on the demographic characteristic of its "highest priority" individual. This was the household member requiring the highest sampling rate to meet sample size targets. The sampling classes presented in Table 3A are arranged in order of highest priority. The table provides a distribution of the 14,706 responding NHIS dwelling units according to their sampling classes, the MEPS sampling rates and the resultant sample of dwelling units selected for the 1997 MEPS. It should be noted that a dwelling unit with a higher order sampling classification may include members with a characteristic of interest that defines a lower sampling classification. More specifically, a dwelling unit with a sampling

classification of 1, which indicates the dwelling unit includes an adult (age 18 and above) with functional impairments (at least 1 ADL requiring personal assistance), may also include a member with any of the other characteristics targeted for oversampling: children with limitations in activity (under age 17); individuals 18-64 years old with predicted high medical expenditures; individuals with family incomes likely to be below 200% of poverty level; adults with other impairments (ages 18-69 and at least 1 IADL and unable to work , age 70 and above and at least 1 IADL). However, dwelling units assigned to sampling classes with lower priority do not include members with a characteristic that defines a higher order classification.

For sampling purposes, a person was classified as having at least 1 ADL requiring personal assistance if there was an affirmative answer to the following question in the 1996 NHIS, "***Because of any impairment or health problem, does___need the help of other persons with personal care needs, such as eating, bathing, dressing, or getting around this home?"***. Similarly, a person was classified as having at least 1 IADL requiring assistance if there was an affirmative answer to the following question in the 1996 NHIS, ***"Because of any impairment or health problem, does___ need the help of others in handling routine needs , such as everyday household chores, doing necessary business, shopping or getting around for other purposes?"***

All NHIS dwelling units assigned to the first three sampling classes ordered by sampling priority were selected with certainty for inclusion for the 1997 MEPS sample. This rate of selection was specified to satisfy sample size targets for the pooled 1997 sample for individuals with one of the following characteristics: adults with functional impairments (at least 1 ADL requiring personal assistance), children with limitations in activity (under age 17), or individuals 18-64 years old predicted to incur high levels of medical expenditures. Dwelling units associated with the next highest priority sampling classes were then selected at a sampling rate of 0.6 designed to meet sample size requirements for the survey. This rate of selection was specified to satisfy sample size targets for the pooled 1997 sample for individuals with one of the following characteristics: individuals with family incomes predicted to be below 200% of poverty level, or adults with other impairments (at least one IADL). All remaining dwelling units associated with the remaining sampling classes were selected with a rate of 0.3 , again to satisfy sample size targets for the 1997 MEPS.

Prior to sample selection, dwelling units within each of the sampling classes were hierarchically sorted by the following measures:

- Quarter of 1996 based on calendar year
- Week within respective calendar quarter of 1996
- Census division
- State
- MSA classification
- NHIS primary sampling unit
- NHIS segment within primary sampling unit
- Minority classification of dwelling unit (Hispanic; Black-Non-Hispanic; Other).

A random systematic sample of dwelling units was then selected from the respective sampling class, using the specified sample selection rate (Table 3A). Table 3B provides a distribution of the 15,067 responding NHIS reporting units within the dwelling units according to these sampling classes, in addition to the MEPS sampling rates and the resultant sample of 6,480 reporting units selected for the 1997 MEPS. In addition, Table 3C provides a distribution of the 38,418 responding NHIS individuals within the dwelling units assigned to the hierarchically defined sampling classes, in addition to the subsample of 17,063 individuals selected for the new 1997 MEPS sample. Since individuals may be classified in more than one category based on the sampling domains under consideration, the sample yields for the new 1997 MEPS sample, allowing for multiple classifications is presented in Table 3D (Note that the sample distributions presented in Tables 3 A-D are confined to the new panel of MEPS introduced in 1997).

Table 3A: NHIS dwelling unit sample classification available for MEPS

Available Sample		1997 MEPS Subsample	
Dwelling Units with at least one member	NHIS 1996 Frequency	Frequency	Sampling Rate
1. Functionally impaired adults	478	478	1.0
2. Children with activity limitations	601	601	1.0
3. Individuals predicted to incur high expenditures	596	596	1.0
4. Low income	2,064	1,238	0.6
5.Adults with other limitations	324	194	0.6
6. Adults aged 65 and older	2,157	647	0.3
7. Other	8,486	2,546	0.3
Total	14,706	6,300	

Source: 1996 National Health Interview Survey, National Center for Health Statistics, CDC

Table 3B: 1996 NHIS reporting unit sample classification available for MEPS

Available Sample		1997 MEPS Subsample
Reporting Units in dwelling units with at least one member	1996 NHIS Frequency	Frequency
1. Functionally impaired adults	481	481
2. Children with activity limitations	601	601
3. Individuals predicted to incur high expenditures	600	600
4. Low income	2,126	1,274
5.Adults with other limitations	326	194
6. Adults aged 65 and older	2,163	652
7. Other	8,770	2,678
Total	15,067	6,480

Source: 1996 National Health Interview Survey, National Center for Health Statistics, CDC

Table 3C: 1996 NHIS person level sample classification available for MEPS

Available Sample		1997 MEPS Subsample
Persons in dwelling units (hierarchically classified) with at least one member	1996 NHIS Frequency	Frequency
1. Functionally impaired adults	506	506
2. Children with activity limitations	723	723
3. Individuals predicted to incur high expenditures	701	701
4. Low income	6,304	4,181
5.Adults with other limitations	393	253
6. Adults aged 65 and older	3,234	1,109
7. Other	26,557	9,590
Total	38,418	17,063

Source: 1996 National Health Interview Survey, National Center for Health Statistics, CDC

Table 3D 1996 NHIS person level sample selected for MEPS

	1997 MEPS Subsample
Individuals with the following characteristics (a person may be classified in more than one category)	Frequency
1. Functionally impaired adults	506
2. Children with activity limitations	723
3. Individuals predicted to incur high expenditures	755
4. Low income	7,990
5.Adults with other limitations	900
6. Adults aged 65 and older	4,600

Source: 1996 National Health Interview Survey, National Center for Health Statistics, CDC

4.0 Procedures for Data Collection

The preliminary contact with households responding to NHIS and subsampled as part of a MEPS panel is described in S. Cohen (1997). Procedures in the rounds of data collection are described below.

Rounds 1-5

Five interviews are conducted with each NHIS panel selected for MEPS at 4- to 5-month intervals over an approximately 24-month field period. The first three rounds (Panel 1:Rounds 1-3) define the 1996 MEPS Household Component and collect the main body of annual use and expenditure data for calendar year 1996. Rounds 3-5 of the 1996 MEPS panel (Panel 1: Rounds 3-5) are combined with Rounds 1-3 of the 1997 MEPS panel (Panel 2:Rounds 1-3) to yield the sample base for the 1997 MEPS Household Component and the source of annual estimates for that calendar year. All interviews are conducted in person through a computer-assisted personal interview (CAPI). Round 1 asks about the period from January 1 of the MEPS year to the date of that interview; Round 2 will ask about the time from the Round 1 interview through the date of the Round 2 interview, and Round 3 asks about the time from the date of the Round 2 interview through the date of the Round 3 interview in 1997.

Questionnaires for these field rounds parallel those used in the 1987 NMES but include some modifications implemented for a 1992 feasibility study and further changes stemming from the feasibility study and the NMES-3 pretest. The instruments contain items that are asked once in the life of the study, items that are asked repeatedly in each round, and items that are updated in later rounds. Questions asked only once include basic sociodemographic characteristics. Core questions asked repeatedly include health status, health insurance coverage, employment status, days of restricted activity due to health problems, medical use, hospital admissions, and purchase of medicines. For each health encounter identified, data are obtained on the nature of health conditions, characteristics of the provider, services provided, associated charges, and sources and amounts of payment.

Permission forms for medical are collected in the field. A sample of medical providers identified by MEPS respondents is contacted in the survey of medical providers (the Medical Provider Component (MPC) of MEPS), to verify and supplement information provided by the family respondent in the household interview. Employers and other health insurance providers are contacted in the survey of health insurance providers (the Insurance Component (IC) of the MEPS), to collect other information on insurance characteristics that household respondents would not typically know.

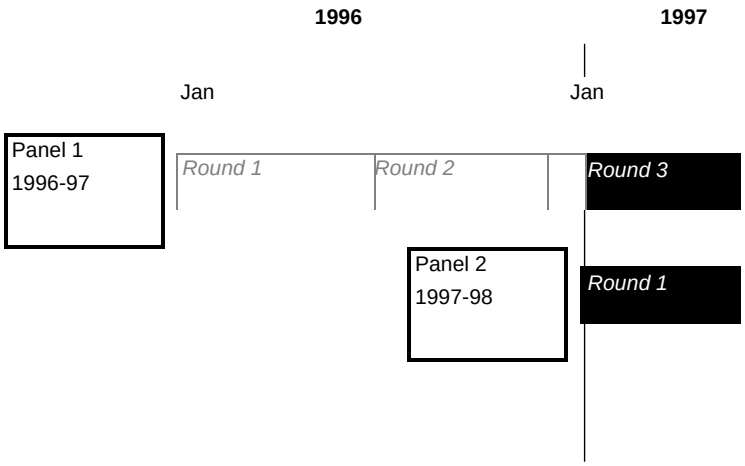
5.0 Sample Yields for the 1997 MEPS and Survey Response Rates

Data are collected for each MEPS panel to cover a two-year period, with the first two MEPS panels spanning 1996-97 and 1997-98, respectively. This section provides a summary of the sample yields for the deriving national person based estimates from the 1997 MEPS, for both point in time estimates (first part of calendar year 1997) and annual estimates. Attention will first be given to the

point in time estimation capacity of the survey, followed by an emphasis on the sample yields for producing calendar year health care estimates from the survey.

To produce point in time health care estimates for the first part of 1997 based on the MEPS sample design, data will need to be pooled from the first two MEPS national samples, with data covering approximately the first half of calendar year 1997. More specifically, data from the 1997 portion of the third round of data collection for the MEPS Panel 1 sample are pooled with data from the first round of data collection for the MEPS Panel 2 sample (illustrated below). This feature of the MEPS design supports the derivation of health insurance coverage estimates covering the first half of calendar year 1997.

Point in Time Estimates Covering First Part of 1997



MEPS Panel 1

The MEPS Panel 1 sample initially consisted of a sample of 10,639 households in 1996, a nationally representative subsample of the households responding to the 1995 National Health Interview Survey (NHIS). The 1995 NHIS sampled households with Hispanic members and households with Black members at approximately 2.0 and 1.5 times the rate of other households, respectively. These oversampling rates are also reflected in the MEPS sample of households. The 1995 NHIS response rate achieved for MEPS-eligible households was 94 percent. Of 10,639 responding NHIS dwelling units eligible for MEPS, 99.6 percent were identified with enough information to allow MEPS data collection. Of the 11,424 eligible reporting units targeted for interviews in Round 1, 9,488 (83.1 percent) responded. Overall, the joint NHIS-Round 1 response rate for the 1996 MEPS household survey was 77.7 percent (.939 x .996 x .831). Conditioned on participation in the MEPS, 90.33 percent of the sample participants provided data for their entire period of eligibility in 1996 and through the early part of 1997 (Round 3). Consequently, the overall MEPS Panel 1 response rate at the end of round 3 (which collects data for the first part of 1997) was 70.2 percent, reflecting response to the 1995 NHIS interview and the MEPS interviews for rounds 1-3 (S. Cohen, 1997). Overall, the Round 3 MEPS Panel 1 sample consisted of 21,411 survey participants.

Panel 2

The 1997 MEPS Panel 2 sample initially consisted of a sample of 6,300 eligible NHIS dwelling units serving as a nationally representative subsample of the households responding to the NHIS. As for Panel 1, the Panel 2 sample reflects the oversampling of Hispanic and Black households in the NHIS. However, the sample allocation for Panel 2 of the MEPS differed from that for Panel 1 because of the additional oversampling of the targeted policy relevant groups. The 1996 NHIS response rate achieved for MEPS-eligible households was 93.8 percent. Of the 6,196 eligible dwelling units targeted for interviews in Round 1, 5,182 responded, consisting of 5,536 responding reporting units (.831). The overall MEPS Panel 2 response rate at the end of round 1 (when data were collected for the first part of 1997) was 77.9 percent. This overall rate reflects response to both the 1996 NHIS interview and the MEPS round 1 interview, consisting of a total sample of 14,505 survey participants.

Combined MEPS Response Rates for Point in Time 1997 Estimates (first half of year)

Each panel was given equal weight in the development of sampling weights to produce national estimates. Therefore, a pooled response rate for the survey respondents in this data set can be obtained by taking an average of the panel specific response rates. This pooled response rate for the combined panels is 74.1 percent, consisting of a total of 35,916 survey participants within 14,147 family and single person analytical units in MEPS. The weighted MEPS population estimate for the civilian non-institutionalized population as of March 1997 was 265,926,692, based on poststratification to population estimates produced from the March 1997 Current Population Survey. The weighted estimate of the number of family units (family and single person units) as of March 1997 was 112,106,153, based on data from the same source.

Using data from the 1997 MEPS Panel 1 Round 3/Panel 2 Round 1 Public Use file, population estimates of the proportion of the population that was uninsured was produced for the overall population and for a representative set of analytical domains, which included several of the population subgroups targeted for oversampling (Table 4). The low income and high expenditure population subgroups were not included in this analyses as a consequence of the unavailability of the 1997 MEPS full year data at the time of this publication. The table includes sample yields for the full 1997 MEPS sample, in addition to the level of precision achieved for the survey estimates as measured by the relative standard error and the respective survey design effects.

The 1997 MEPS point in time sample includes an oversample of minorities, with 7,960 Hispanic sample participants and 5,301 Black, Non-Hispanic sample participants, which reflects the oversampling rates for minorities inherent in the NHIS sample (Hispanics, 2.0:1; blacks, 1.5:1). Alternatively, the overall sample yield for the elderly, consisting of 4,104 sample participants and 11.4 percent of the sample, is quite consistent with their proportional representation in the population, as anticipated by the MEPS sample selection rates applied to the eligible NHIS sample.

As can be observed in Table 4, the sample yields achieved for the pooled 1997 MEPS point in time sample were consistent with targeted sample yields for the full year 1997 MEPS after adjusting for survey. After factoring in the anticipated sample size reductions attributable to survey attrition for the point in time 1997 estimates of the uninsured, it is evident that precision levels for the full year 1997 MEPS were largely realized when considering comparable population estimates. It should be noted that some of the estimates of the uninsured obtained from the 1997 MEPS were less than 20 percent (the value used to set precision targets), which would partially explain some of the observed differentials from the precision targets, which were also specified as average relative standard errors.

Table 4: 1997 MEPS Sample Yields	**1996**	**1997**	**Pooled (Panels 1 and 2)**				
Demographic Subgroup	Unweighted Sample	Unweighted Sample	Unweighted Sample	% Uninsured	Standard Error (%)	Relative SE (%)	Design Effect
Overall Population	21,411	14,505	35,916	16.76	0.389	2.321	3.89
Sex							
Male	10,191	6,842	17,033	18.51	0.472	2.550	2.52
Female	11,220	7,663	18,883	15.09	0.414	2.744	2.53
Race/ethnicity							
Hispanic	4,610	3,350	7,960	32.94	1.118	3.394	4.51
Black-nonHispanic	2,879	2,422	5,301	21.45	0.985	4.592	3.05
Other	13,922	8,733	22,655	13.63	0.392	2.876	2.95
Age							
Under 6	1,989	1,450	3,439	14.27	0.895	6.272	2.25
6-17	4,265	3,160	7,425	15.72	0.679	4.319	2.58
18-44	8,296	5,428	13,724	23.36	0.579	2.479	2.57
45-64	4,392	2,832	7,224	14.84	0.587	3.956	1.97
65 and older	2,469	1,635	4,104	1.02	0.194	19.020	1.53
Activity Limitations							
1+ADL (18 yrs and older)	292	346	638	4.81	4.137	23.638	1.80
1+IADL (18 yrs and older)	528	603	1,131	5.97	0.937	15.695	1.77
Region							
Northeast	4,238	2,752	6,990	14.20	0.639	4.500	2.34
Midwest	4,637	2,941	7,578	12.51	0.728	5.819	3.67
South	7,442	5,150	12,592	19.62	0.745	3.797	4.43
West	5,094	3,662	8,759	18.93	0.921	4.865	4.84

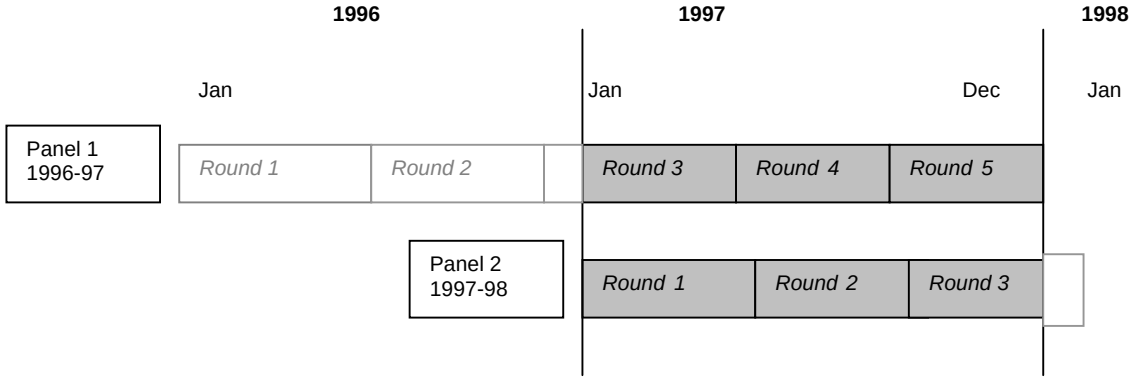
Source: 1997 MEPS, Center for Cost and Financing Studies, Agency for Healthcare Research and Quality.

More specifically, there were 638 adults in the MEPS who received help or supervision with activities of daily living (ADLs), which included bathing, dressing or getting around the house, because of an impairment or a physical or mental problem (Table 4). This subset of activities of daily

living is less inclusive than the set of ADLs considered in the NHIS for oversampling purposes, and indicates the lower bound in terms of sample size yields for this target population. In addition, a design effect of 1.8 was achieved for the survey estimate of the uninsured that characterizes this policy relevant population subgroup. There were also 1,131 adults in the MEPS who received help or supervision with instrumental activities of daily living (IADLs), which included using the telephone, paying bills, taking medications, preparing light meals, doing laundry, or going shopping, because of an impairment or a physical or mental problem (Table 4). This sample yield is convergent with sample size targets for this policy relevant population subgroup. A survey design effect of 1.8 was achieved for the associated survey estimate of the percent of the population. An examination of the efficacy of the sample design to achieve design goals for children with physical impairments, households with low incomes and individuals with high levels of medical expenditures will also be undertaken, to better inform future oversampling efforts in MEPS for these target population subgroups, once the expenditure and income data for the MEPS 1997 are available. A similar analysis will be conducted to examine the level of precision realized for survey estimates of health care utilization and expenditures, once the use and expenditure data for the 1997 MEPS are available.

Annual Estimates for Calendar Year 1997

In order to produce annual health care estimates for calendar year 1997 based on the full MEPS sample, data will also need to be pooled across the first two MEPS national samples. More specifically, full calendar year 1997 data collected in Rounds 3 through 5 for the MEPS Panel 1 sample are pooled with data from the first three rounds of data collection for the MEPS Panel 2 sample (illustrated below). Overall, the full 1997 MEPS household sample will consist of approximately 13,000 reporting units which include 32,636 individuals that completed the full series of MEPS interviews for their entire period of eligibility, providing the necessary information to produce national use and expenditure estimates for calendar year 1997.



Panel 1

Conditioned on response to Rounds 1-3 of the Panel 1 MEPS, of 21,696 key and inscope individuals eligible for data collection in 1997, 19,622 (90.44 percent) provided data for their entire period of eligibility. Consequently, after factoring in the impact of survey attrition, the overall Panel 1 MEPS person level response rate for deriving annual estimates was 63.5 percent (.702 x .9044). Of these full year respondents for calendar year 1997, 19,407 were in scope on December 31, 1997.

Panel 2

Conditioned on response to Round 1 of the Panel 2 MEPS, of 14, 644 key and inscope individuals eligible for data collection in 1997, 13,014 (88.87 percent) provided data for their entire period of eligibility. Consequently, after factoring in the impact of survey attrition, the overall Panel 2 MEPS person level response rate for deriving annual estimates was 69.2 percent (.779 x .8887). Of these full year respondents for calendar year 1997, 12,819 were in scope on December 31, 1997.

Combined MEPS Panels: Response Rate for Annual 1997 Estimates

Each panel was given equal weight in the development of sampling weights to produce annual national estimates. Therefore, a pooled response rate for the survey respondents in this data set can be obtained by taking an average of the panel specific response rates. This pooled response rate for the combined panels is 66.4 percent, consisting of a total of 32,636 survey participants. The weighted MEPS population estimate for the civilian non-institutionalized population as of December 31, 1997 was 267,704,802, based on poststratification to population estimates produced from the December 1997 Current Population Survey. Sample yields for the subset of the 32,636 survey participants that were in scope as of 12/31/97 (32,226) are presented in Table 5, controlling for gender, race/ethnicity, region, MSA status and age.

Table 5: 1997 MEPS, Panels I and II combined, sample yields for full year respondents as of 12/31/97

	Unweighted MEPS count	Weighted CPS count
Sex		
Male	15,239	130,734,620
Female	16,987	136,970,181
Race/Ethnicity		
Hispanic	7,440	30,680,491
Non-Hispanic black	4,743	33,578,472
Other	20,043	203,445,838
Region		
Northeast	6,144	51,118,238
Midwest	6,763	62,426,339
South	11,309	93,820,483
West	8,010	60,339,742
Metropolitan Status		
MSA	25,119	215,387,710
NonMSA	7107	52,317,091
Age		
Under 1	479	3,819,437
1 to 4 years old	2,028	15,840,700
5 to 9 years old	2,854	20,404,149
10 to 14 years old	2,807	19,563,172
15 to 19 years old	2,532	19,452,449
20 to 24 years old	1,913	17,531,979
25 to 29 years old	1,998	18,827,116
30 to 34 years old	2,335	20,322,814
35 to 44 years old	4,963	44,120,234
45 to 54 years old	3,963	33,907,056
55 to 59 years old	1,419	11,896,295
60 to 64 years old	1,198	9,956,233
65 to 69 years old	1,046	9,413,817
70 to 74 years old	1017	8,532,698
75 to 79 years old	776	6,842,152
80 years old and older	898	7,274,502
Major Age Categories		
Under 1	479	3,819,437
1 to 19 years old	10,221	75,260,469
20 to 29 years old	3,911	36,359,095
30 to 44 years old	7,298	64,443,048
45 to 64 years old	6,580	55,759,584
65 years old and older	3,737	32,063,169
Total	32,226	267,704,802

Note: The MEPS sample size yields presented in this table represents the population as of 12/31/97.
When the full year MEPS sample of all persons with positive person weight are included, the total MEPS respondent sample is 32,636

6.0 Summary

This report has provided a summary of the sample design features of the 1997 Household Component of the Medical Expenditure Panel Survey. Particular attention has been given to the sample selection scheme implemented for the new 1997 MEPS sample panel. The report also provides a summary of the precision specifications for the survey, sample yields and the level of precision in survey estimates. The details of the probabilistic models that were used to select an expected oversample of low income households and for individuals likely to incur high levels of medical expenditures in 1997 were also presented, in addition to measures of the predictive capacity of the respective models. Both the panel specific and pooled survey response rates were also summarized for the 1997 MEPS Panel 1 Round 3/Panel 2 Round 1 point in time sample (first half of 1997) and for annual estimates.

The MEPS data also serve as the primary source of information for research efforts examining how health care use and expenditures vary among different sectors of the population (such as the elderly, veterans, children, disabled persons, minorities, the poor, and the uninsured) and how the health insurance of individuals varies by demographic characteristics, employment status and characteristics, geographic locale, and other factors. The MEPS data are and will continue to provide answers to questions about private health insurance costs and coverage, and help evaluate the growing impact of managed care on health care expenditures and enrollment in different types of managed care plans.

7.0 References

Arnett RA, Hunter E, Cohen S, et al. The Department of Health and Human Services' Survey Integration Plan. In: Proceedings of the American Statistical Association (ASA). Section on Government Statistics. Chicago: 1996 Aug.

Cohen J. Design and Methods of the Medical Expenditure Panel Survey Household Component. Rockville (MD): Agency for Health Care Policy and Research; 1997. *MEPS Methodology Report No. 1.* AHCPR Pub. No. 97-0026.

Cohen, S.B. Sample Design of the 1996 Medical Expenditure Panel Survey Household Component. Rockville (MD): Agency for Health Care Policy and Research, 1997. MEPS Methodology Report No. 2. AHCPR Pub. No. 97-0027.

Cohen SB. The redesign of the Medical Expenditure Panel Survey, a component of the DHHS Survey Integration Plan. Proceedings of the Council of Professional Associations on Federal Statistics (COPAFS) Seminar on Statistical Methodology in the Public Service, Bethesda (MD); 1996 Nov.

DiGaetano R. Sample design of the Household Component of the National Medical Expenditure Survey (NMES-3). Unpublished report. Rockville (MD): Westat, Inc.; 1994. Contract No. 282-94-200.

Hunter E, Arnett R, Cohen S, et al. HHS Survey Integration Plan: Background materials. Agency for Health Care Policy and Research, Rockville (MD), and National Center for Health Statistics, Hyattsville (MD); 1995.

Judkins D, Marker D, Waksberg J. National Health Interview Survey: research for the 1995 redesign. Unpublished report. Prepared under contract for the National Center for Health Statistics. Rockville (MD): Westat, Inc.; 1994.

Moeller, J and Mathiowetz, N. Internal Memos on Low Income and High Expenditure Prediction Models; 1994.

Appendix 3: Summary of Utilization and Expenditure Variables by Health Service Category

HEALTH SERVICE CATEGORY	UTILIZATION VARIABLE(S)	EXPENDITURE VARIABLE(S)[1]
All Health Services	--	TOT***97
Office Based Visits		
Total Office Based Visits (Physician + Non-physician + Unknown)	OBTOTV97	OBV***97
Office Based Visits to Physicians	OBDRV97	OBD***97
Office Based Visits to Non-Physicians	OBOTHV97	OBO***97
Office Based Visits to Chiropractors	OBCHIR97	OBC***97
Office Based Nurse or Nurse Practitioner Visits	OBNURS97	OBN***97
Office Based Visits to Optometrists	OBOPTO97	OBE***97
Office Based Physician Assistant Visits	OBASST97	OBA***97
Office Based Physical or Occupational Therapist Visits	OBTHER97	OBT***97
Hospital Outpatient Visits		
Total Outpatient Visits (Physician + Non-physician + Unknown)	OPTOTV97	--
Facility Expense	--	OPF***97
SBD Expense	--	OPD***97
Outpatient Visits to Physicians	OPDRV97	--
Facility Expense	--	OPV***97
SBD Expense	--	OPS***97
Outpatient Visits to Non-Physicians	OPOTHV97	--
Facility Expense	--	OPO***97
SBD Expense	--	OPP***97

1 See key at end of table for specific categories for ***.

HEALTH SERVICE CATEGORY	UTILIZATION VARIABLE(S)	EXPENDITURE VARIABLE(S)
Emergency Room Visits		
Total Emergency Room Visits	ERTOT97	--
Facility Expense	--	ERF***97
SBD Expense	--	ERD***97

Inpatient Hospital Stays (Including Zero Night Stays)		
Total Inpatient Stays (Including Zero Night Stays)	IPDIS97, IPNGTD97	--
Facility Expense	--	IPF***97
SBD Expense	--	IPD***97
Zero night Hospital Stays	IPZERO97	--
Facility Expense	--	ZIF***97
SBD Expense	--	ZID***97

Dental Visits		
Total Dental Visits	DVTOT97	DVT***97
General Dental Visits	DVGEN97	DVG***97
Orthodontist Visits	DVORTH97	DVO***97

Home Health Care		
Total Home Health Care	HHTOTD97	--
Agency Sponsored	HHAGD97	HHA***97
Paid Independent Providers	HHINDD97	HHN***97
Informal	HHINFD97	--

Other		
Vision Aids	--	VIS***97
Other Medical Supplies and Equipment	--	OTH***97
Prescription Medicines[2]	RXTOT97	RX***97

KEY: To complete variable name, replace *** with a particular source of payment category as identified in the following table:

Source of Payment Category	***
Total payments (sum of all sources)	EXP
Out of Pocket	SLF
Medicare	MCR
Medicaid	MCD
Private Insurance	PRV
Veteran's Administration	VA
CHAMPUS or CHAMPVA	CHM
Other Federal Sources	OFD
Other State and Local Sources	STL
Workers' Compensation	WCP
Other Private	OPR
Other Public	OPU
Other Unclassified Sources	OSR
Total charges[2]	TCH

[2] No charge variables on file for prescription medicines.