

**MEPS HC-016A:
1997 Prescribed Medicines**

June 2001

**Agency for Healthcare Research and Quality
Center for Cost and Financing Studies**

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 Information	C-1
	1.0 General Information	C-1
	2.0 Data File Information	C-2
	2.1 Codebook Structure	C-3
	2.2 Reserved Codes	C-4
	2.3 Codebook Format	C-4
	2.4 Variable Naming	C-4
	2.4.1 General	C-5
	2.4.2 Expenditure and Source of Payment Variables	C-5
	2.5 Data Collection	C-6
	2.5.1 Methodology for Collecting Household Reported Variables ...	C-6
	2.5.2 Methodology for Collecting Pharmacy-Reported Variables ...	C-7
	2.6 File Contents	C-7
	2.6.1 Survey Administration Variables	C-7
	2.6.1.1 Person Identifier Variables (DUID, PID, DUPERSID)	C-7
	C-7
	2.6.1.2 Record Identifier Variables (RXRECIDX, LINKIDX)	C-7
	C-7
	2.6.1.3 Round Variable (PURCHRD)	C-8
	2.6.2 Characteristics of Prescribed Medicine Events	C-8
	2.6.2.1 Date When Prescribed Medicine Was First Taken	C-8
	(RXBEGDD-RXBEGYR)	C-8
	2.6.2.2 Prescribed Medicine Attributes (RXNAME-RXUNIT)	C-8
	C-8
	2.6.2.3 Type of Pharmacy (PHARTP1-PHARTP7)	C-9
	2.6.2.4 Analytic Flag Variables (RXFLG-DIABFLG)	C-10
	2.6.2.5 The Sample Variable (SAMPLE)	C-11
	2.6.2.6 Condition Codes (RXICD1X-RXICD3X) and Clinical	C-11
	Classification Codes (RXCCC1X-RXCCC3X)	C-11
	2.6.2.7 Record Count Variable (NUMCOND)	C-12
	2.6.3 Expenditure Variables (RXSF97X-RXXP97X)	C-13

	2.6.3.1 Definition of Expenditures	C-13
	2.6.3.2 Sources of Payment	C-13
2.6.4	Sample Weights and Variance Estimation Variables (WTDPER97- VARPSU97)	C-14
	2.6.4.1 Overview	C-14
	2.6.4.2 Details on Person Weights Construction	C-14
	2.6.4.3 The Final Weight for 1997	C-15
	2.6.4.4 Coverage	C-15
3.0	General Data Editing and Imputation Methodology	C-16
3.1	Rounding	C-17
3.2	Edited/Imputed Expenditure Variables (RXSF97X-RXXP97X)	C-17
4.0	Strategies for Estimation	C-18
4.1	Variables with Missing Values	C-18
4.2	Basic Estimates of Utilization, Expenditure and Sources of Payment	C-18
4.3	Estimates of the Number of Persons with Prescribed Medicine Events	C-20
4.4	Person-Based Ratio Estimates	C-20
	4.4.1 Person-Based Ratio Estimates Relative to Persons with Prescribed Medicine Events	C-20
	4.4.2 Person-Based Ratio Estimates Relative to the Entire Population	C-21
4.5	Sampling Weights for Merging Previous Releases of MEPS Household Data with this Event File	C-21
5.0	Variance Estimation	C-21
6.0	Merging/Linking MEPS Data Files	C-22
6.1	Linking a Person Level File to the Prescribed Medicines File	C-22
6.2	Linking the 1997 Conditions File and/or the Other 1997 MEPS Event Files to the 1997 Prescribed Medicines File	C-23
	6.2.1 Limitations/Caveats of RXLK and CLNK	C-23
References	C-25
Attachment 1	C-A1-1
D. Variable-Source Crosswalk	D-1

A. Data Use Agreement

Individual identifiers have been removed from the microdata 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 Healthcare Research and Quality (AHRQ) and/or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which they were 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.
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, AHRQ will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHRQ, 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 these 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 Healthcare Research and Quality requests that users cite AHRQ 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.

MEPS is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized 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) (formerly the Agency for Health Care Policy and Research (AHCPR)) 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, and the National Medical Expenditure Survey (NMES-2) was conducted 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 sampling frame for the MEPS HC is 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 noninstitutionalized 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 5 rounds of interviews over a 2½-year period. Using 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 subsequent 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 sampling frame for the MEPS HC is drawn from respondents to NHIS, conducted by NCHS. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized 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 contacting 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 all office-based physicians:

- Providing care for HC respondents receiving Medicaid.
- Associated with a 75-percent sample of HC households receiving care through an HMO (health maintenance organization) or managed care plan.
- Associated with 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) and medication name.
- 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 survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone follow-up 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 medicines, 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 3 rounds of data collection over a 1½-year period using the CAPI system. Community data were collected by telephone using computer-assisted telephone interviewing (CATI) technology. At the end of 3 rounds of data collection, the sample consisted of 815 responding facilities, 3,209 residents in the facility on January 1, and 2,690 eligible residents admitted during 1996.

5.0 Survey Management

MEPS data are collected under the authority of the U.S. 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 in this regard.

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.

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 MEPS web site:

<<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.

C. Technical Information

1.0 General Information

This documentation describes one in a series of public use event files from the 1997 Medical Expenditure Panel Survey Household Component (MEPS HC) and Medical Provider Component (MPC). Released as an ASCII data file and SAS transport file, this public use file provides detailed information on household-reported prescribed medicines for a nationally representative sample of the civilian noninstitutionalized population of the United States and can be used to make estimates of prescribed medicine utilization and expenditures for calendar year 1997. Each record on this event file represents a unique prescribed medicine event; that is, a prescribed medicine reported as being purchased or otherwise obtained by the household respondent. In addition to expenditures related to the prescribed medicine, each record contains household reported characteristics and medical conditions associated with the prescribed medicine.

Data from this event file can be merged with other 1997 MEPS HC data files, for purposes of appending person characteristics such as demographic or health insurance coverage to each prescribed medicine record.

Counts of prescribed medicine utilization are based entirely on household reports. Information from the Pharmacy Component (within the MEPS Medical Provider Component) was used to provide expenditure and payment data, as well as details of the medication (e.g., strength, quantity, etc.).

The file can be used to construct summary variables of expenditures, sources of payment, and other aspects of utilization of prescribed medicines. Aggregate annual person level information on the use of prescribed medicines and other health services use is provided on the 1997 Full Year Consolidated Data File, where each record represents a MEPS sampled person.

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

Sample Weights and Variance Estimation Variables

Merging MEPS Data Files

References

Variable to Source Crosswalk

For more information on MEPS HC survey design see S. Cohen, 1997; J. Cohen, 1997; and S. Cohen, 1996. For information on the MEPS MPC design, see S. Cohen, 1998. A copy of the survey instrument used to collect the information on this file is available on the MEPS web site at the following address: <<http://www.meps.ahrq.gov>>.

2.0 Data File Information

This public use data set contains 234,532 prescribed medicine records. Each record represents one household-reported prescribed medicine that was purchased or obtained during calendar year 1997. These data were collected during the 1997 portion of Round 3, and Rounds 4 and 5 for Panel 1, as well as Rounds 1, 2 and the 1997 portion of Round 3 for Panel 2. Of the 234,532 prescribed medicine records, 230,012 records are associated with persons having a positive person level weight (WTDPER97). The persons represented on this file had to meet either criteria a or b below:

- a) Be classified as a key in-scope person who responded for his or her entire period of 1997 eligibility (i.e., persons with a positive 1997 full-year person level sampling weight ($WTDPER97 > 0$)), or
- b) Be classified as either an eligible non-key person or an eligible out-of-scope person who responded for his or her entire period of 1997 eligibility, and belonged to a family (i.e., all persons with the same value for a particular FAMID variable) in which all eligible family members responded for their entire period of 1997 eligibility, and at least one family member has a positive 1997 full-year person weight (i.e., eligible non-key or eligible out-of-scope persons who are members of a family, all of whose members have a positive 1997 full-year MEPS family level weight ($WTFAM97 > 0$)).

Please refer to Attachment 1 for definitions of key, non-key, in-scope and eligible. Persons with no prescribed medicine use for 1997 are not included on this file (but are represented on MEPS person level files). A codebook for the data file is provided (in file H16ACB.PDF).

This file includes prescribed medicine records for all household survey respondents who resided in eligible responding households and reported at least one prescribed medicine. Only prescribed medicines that were purchased or otherwise obtained in calendar year 1997 are represented on this file. This file includes prescribed medicines identified in the Prescribed Medicines section of the Household Component survey instrument, as well as those prescribed medicines identified in association with medical events. Each record on this file represents a single acquisition of a prescribed medicine reported by household respondents. Some household respondents may have multiple acquisitions of prescribed medicines and thus will be represented in multiple records on this file. Other household respondents may have reported no acquisitions of prescribed medicines and thus will have no records on this file.

It should also be noted that refills are included on this file. The HC obtains information on the name of the prescribed medicine and the number of refills, if any, associated with it. The data collection design for the HC does not allow separate records to be created for multiple acquisitions of the same prescribed medicine. However, in the Pharmacy Component, each original purchase, as well as any refill, is considered a unique prescribed medicine event. Therefore, for the purposes of editing, imputation and analysis, all records in the Household

Component were “unfolded” to create separate records for each original purchase and each refill. Please note, MEPS did not collect information in the HC to distinguish multiple acquisitions of the same drug between the original purchase and refills. The survey only collected data on the number of times a prescribed medicine was acquired during a round. In some cases, all purchases may have been refills of an original purchase in a prior round or prior to the survey year. The file also includes a variable, (SAMPLE), which indicates whether or not the household received a free sample of that drug in that round. (To obtain more details on free samples, please see Section 2.6.2.5)

Each record on this file includes the following: an identifier for each unique prescribed medicine; detailed characteristics associated with the event (e.g., national drug code (NDC), medicine name, etc.); conditions, if any, associated with the medicine; the date on which the person first used the medicine; total expenditure and sources of payments; types of pharmacies that filled the household’s prescriptions; whether the prescription is one in which the household received a free sample of it during the round; and a full-year person level weight.

Data from this file can be merged with previously released MEPS HC person level data using the unique person identifier, DUPERSID, to append person characteristics such as demographic or health insurance coverage to each record. Data from this file can also be merged with the 1997 Full year Consolidated Data File to estimate expenditures for persons with prescribed medicines. The Prescribed Medicines event file can also be linked to the MEPS 1997 Medical Conditions File and additional MEPS 1997 event files. Please see the 1997 Appendix File for details on how to link MEPS data files.

2.1 Codebook Structure

For each variable on the file, both weighted and unweighted frequencies are provided. The codebook and data file sequence list variables in the following order:

- Unique person identifiers
- Unique prescribed medicine identifiers
- Other survey administration variables
- Prescribed medicine characteristics variables
- ICD-9 codes
- Clinical Classification Software codes
- 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.
-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.
-13 VALUE SUPPRESSED	Data suppressed.
-14 NOT YET TAKEN/USED	Respondent answered that the medicine has not yet been used.

Generally, values of -1, -7, -8 and -9 have not been edited on this file. The values of -1 and -9 can be edited by analysts by following the skip patterns in the questionnaire. The value of -13 was assigned when originally reported Household Component data were suppressed because imputed versions of the variable are available on the Public Use File. The value -14 was a valid value only for the variable representing the year the respondent reported having first used the medicine (RXBEGYR). RXBEGYR= -14 means that when the interviewer asked the respondent the year he/she first started using the medicine, he/she responded that he/she had not yet starting using the medicine.

2.3 Codebook Format

The codebook describes an ASCII data set (although the data are also being provided in a SAS transport file). The following codebook items are provided for each variable:

IDENTIFIER	DESCRIPTION
Name	Variable name (maximum of 8 characters)
Description	Variable descriptor (maximum of 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 8 character limitation. For

questions asked in a specific round, the end digit in the variable name reflects the round in which the question was asked. Generally, imputed/edited variables end with an “X”.

2.4.1 General

Variables contained on this file were derived from the HC questionnaire itself, the MPC data collection instrument, or from the CAPI. The source of each variable is identified in Section D, entitled “Variable-Source Crosswalk.” Sources for each variable are indicated in one of four ways: (1) variables which are derived from CAPI or assigned in sampling are so indicated; (2) variables which come from one or more specific questions have those numbers and the questionnaire section indicated in the “Source” column; (3) variables constructed from multiple questions using complex algorithms are labeled “Constructed” in the “Source” column; and (4) variables which have been imputed are so indicated.

2.4.2 Expenditure and Source of Payment Variables

Only imputed/edited versions of the expenditure variables are provided on the file. Expenditure variables on this event file follow a standard naming convention and are seven characters in length. The 12 source of payment variables and one sum of payments variable are named consistently in the following way:

The first two characters indicate the type of event:

IP - inpatient stay	OB - office-based visit
ER - emergency room visit	OP - outpatient visit
HH - home health visit	DV - dental visit
OM - other medical equipment	RX - prescribed medicine

In the case of the source of payment variables, the third and fourth characters indicate:

SF – self or family	OF – other Federal Government	XP – sum of payments
MR – Medicare	SL – State/local government	
MD – Medicaid	WC – Worker’s Compensation	
PV – private insurance	OT – other insurance	
VA – Veterans	OR – other private	
CH – CHAMPUS/CHAMPVA	OU – other public	

The fifth and sixth characters indicate the year (97). All imputed/edited expenditure variables end with an “X”.

For example, RXSF97X is the edited/imputed amount paid by self or family for the 1997 prescribed medicine expenditure.

2.5 Data Collection

Data regarding prescription drugs were obtained through the HC questionnaire and a pharmacy follow-back component (within the Medical Provider Component).

2.5.1 Methodology for Collecting Household Reported Variables

During each round of the MEPS HC, all respondents were asked to supply the name of any prescribed medicine they or their family members purchased or otherwise obtained during that round. For each medicine in each round, the following information was collected: whether any free samples of the medicine were received; the name(s) of any health problems the medicine was prescribed for; the number of times the prescription medicine was obtained or purchased; the year, month, and day on which the person first used the medicine; and a list of the names, addresses, and types of pharmacies that filled the household's prescriptions. In the HC, 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 follow-back component. However, charge and payment information was collected for self-filers in the household questionnaire, because it was thought that payments by private third-party payers for self-filers' purchases would not be available from pharmacies. Uninsured persons were treated in the same manner as non-self-filers.

An inaccuracy in the number of times a household reported purchasing or otherwise obtaining a prescription drug in a particular round for a small percentage of household-reported medications was discovered. This inaccuracy was due to an instrument design flaw which caused interviewer error, and in isolated cases, resulted in mis-reported large numbers of prescription refills for a medicine in a given round. This inaccuracy was confined to only a very small percentage of unique drugs on the original data delivered. For some cases, it seems that the year that the person started taking the drug was recorded in the field that gives the number of times that the person purchased, or otherwise obtained the drug, during the round, as well as in the field that provides the year the person started taking the medicine. For example (in the round a specific drug was first mentioned), a person was reported to have first started taking the drug in 1997, a "97" was entered in the field for the year the person first started taking the drug. For a small percentage of the cases in which persons began taking a drug in 1997, a "97" appeared in the preceding field indicating the number of times the drug was purchased or otherwise obtained during the round, as well. Outlier values where this situation occurred (and similar instances) were determined by comparing the number of days a respondent was in the round and the number times the person reported having purchased or otherwise obtained the drug in the round, and were determined in consultation with an industry expert. For these events, a new value for the number of times a drug was purchased or otherwise obtained by a person in a round was imputed. In addition, the prescribed medicine events in which a household respondent did not know/remember the number of times a certain prescribed medicine was purchased or otherwise obtained were imputed a value for that variable.

For those rounds that spanned two years, drugs mentioned in that round were allocated between 1997 and 1998 based on the year the person started taking the drug, the length of the person's round, the dates of the person's round, and the number of drugs for that person in the round. In addition, a "folded" version of the pharmacy component on an event level, as opposed to an acquisition level, was used for these types of events to assist in determining whether an acquisition of the drug should be allocated to 1997 or 1998.

2.5.2 Methodology for Collecting Pharmacy-Reported Variables

If the respondent with the prescription gave written permission to release his or her pharmacy records, pharmacy providers identified by the household were contacted by mail or telephone for the pharmacy follow-back component. The signed permission forms were provided to the various establishments in the initial mailing to those contacted by mail or were faxed to those contacted by telephone. Each establishment was informed of all persons participating in the survey who had prescriptions filled at their place of business, and a computerized printout of all prescriptions filled for each person 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/amount dispensed); total charge; and payments by source.

2.6 File Contents

2.6.1 Survey Administration Variables

2.6.1.1 Person Identifier Variables (DUID, PID, DUPERSID)

The dwelling unit ID (DUID) is a 5-digit random number assigned after the case was sampled for MEPS. The 3-digit person number (PID) uniquely identifies each person within the dwelling unit. The 8-character variable DUPERSID uniquely identifies each person represented on the file and is the combination of the variables DUID and PID. For detailed information on dwelling units and families, please refer to the documentation of a public use file containing person level population characteristics.

2.6.1.2 Record Identifier Variables (RXRECIDX, LINKIDX)

The variable RXRECIDX uniquely identifies each record on the file. This 15-character variable is comprised of the following components: prescribed medicine event generated through the Household Component (positions 1-12) + enumeration number (positions 13-15). The prescribed medicine event generated through the Household Component (positions 1-12) can be used to link a prescribed medicine event to the conditions file and to other event files, via link files, and is provided on this file as the variable LINKIDX. (For more details on linking, please refer to Section 6.2 and to the 1997 Appendix File.)

The following hypothetical example illustrates the structure of these ID variables. This example illustrates a person in Round 1 of the household interview who reported having purchased

Amoxicillin three times. The following example shows three acquisition level records, all having the same RXNDC (00364021802), for one person (DUPERSID=00002026) in one round. Only one NDC is associated with a prescribed medicines event because matching was performed at an event level, as opposed to an acquisition level. (For more details on matching, please see Section 3.0.). The LINKIDX (000020260083) remains the same for all three records, whereas the RXRECIDX (000020260083001, 000020260083002, 000020260083003) differs for all three records.

DUPERSID	RXRECIDX	LINKIDX	RXNDC
00002026	000020260083001	000020260083	00364021802
00002026	000020260083002	000020260083	00364021802
00002026	000020260083003	000020260083	00364021802

2.6.1.3 Round Variable (PURCHRD)

The variable PURCHRD indicates the round in which the prescribed medicine was obtained/purchased and takes on the value of 1, 2, 3, 4, or 5.

2.6.2 Characteristics of Prescribed Medicine Events

2.6.2.1 Date When Prescribed Medicine Was First Taken (RXBEGDD-RXBEGYR)

There are three variables which indicate when a prescribed medicine was first taken (or obtained), as reported by the household. They are the following: RXBEGDD indicates the day a person first started taking a medicine, RXBEGMM denotes the month in which a person first started taking a medication, and RXBEGYR reflects the year in which a person first started taking a medicine. These “first taken” questions are only asked the first time a prescription is mentioned by the household. These questions are not asked of refills of the prescription for a person in subsequent rounds and result in a value of -1 being assigned to those types of events for these variables. These variables are unedited.

2.6.2.2 Prescribed Medicine Attributes (RXNAME-RXUNIT)

For each prescribed medicine included on this file, several data items collected describe in detail the medication obtained or purchased. These data items are the following:

- a. Medication name - pharmacy reported (RXNAME)
- b. Medication name - household reported (RXHHNAME)
- c. National drug code (RXNDC)
- d. Quantity of the prescribed medicine dispensed, e.g., number of tablets in the prescription (RXQUANTY)
- e. Form of the prescribed medicine; e.g., tablets (RXFORM and RXFORMOS) - for those forms of a drug not listed as a choice for the RXFORM variable, 91 OTHER

SPECIFY was chosen for the RXFORM variable and then the follow-up variable, RXFORMOS, allowed other forms of a drug to be entered

f. Strength of the dose of the medicine prescribed; e.g., 10 (RXSTRENG)

g. Unit of measurement for the dose of the prescribed medication; e.g., mg (RXUNIT and RXUNITOS) - for those units of a drug not listed as a choice for the RXUNIT variable, 91 OTHER SPECIFY was chosen for the RXUNIT variable and then the follow-up variable, RXUNITOS, allowed other units of a drug to be entered

The national drug code (NDC) generally is an 11-digit code. The first 5 digits indicate the manufacturer of the prescribed medicine. The next 4 digits indicate the form and strength of the prescription, and the last 2 digits indicate the package size from which the prescription was dispensed. NDC values were imputed from a proprietary database to certain Pharmacy Component (PC) prescriptions because the NDC reported by the pharmacy provider did not match to the proprietary database. These records are identified by RXFLG=3. AHRQ's licensing agreement for the proprietary database precludes the release of these imputed NDC values to the public, so for these prescriptions, the household-reported name of the prescription (RXHHNAME) and the original NDC (RXNDC) and prescription name (RXNAME) reported by the pharmacist are provided to allow users to do their own imputation. Otherwise, the imputed NDC values for the RXFLG=3 cases may be accessed through the MEPS Data Center. For those events not falling in the RXFLG=3 category, the reserve code (-13) is assigned to the household reported medication name (RXHHNAME). For information on accessing confidential data through the MEPS Data Center, contact the MEPS Project Director by email at: <mepspd@ahrq.gov>.

Imputed data on this event file (and not the other event files) may include missing data. This is because imputed data on this file are imputed from the Pharmacy Component or from a proprietary database. These sources did not always include complete information for each variable but did include an NDC which would typically enable an analyst to obtain any missing data items. For example, although there are a substantial number of missing values for the form and strength of the prescription that were not supplied by the pharmacist, these missing values were not imputed because this information is embedded in the NDC.

2.6.2.3 Type of Pharmacy (PHARTP1-PHARTP7)

Household respondents were asked to list the type of pharmacy from which their medications were purchased. A household could list multiple pharmacies associated with their prescriptions in a given round, or over the course of all rounds combined covering the survey year. As a result, this file contains, at most, seven of these household reported pharmacies, but there was no link in the survey or in the data file enabling users to know the type of pharmacy from which a specific prescription was obtained, if multiple pharmacies are listed. The set of variables (PHARTP1-PHARTP7) identify the types of pharmacy providers from which the person's prescribed medicines were purchased or otherwise obtained. The possible types of pharmacies include the following: (1) mail-order, (2) another store, (3) HMO/clinic/hospital, and (4) drug store. A -1 value for PHARTPn indicates that the household did not report an "nth" pharmacy.

2.6.2.4 Analytic Flag Variables (RXFLG-DIABFLG)

There are five flag variables included on this file (RXFLG, PCIMPFLG, SELFFLG, INPCFLG, and DIABFLG).

The variable RXFLG indicates how the NDC for a specific prescribed medicine event was imputed. This variable indicates whether or not there was any imputation performed on this record for the NDC variable, and if imputed, from what source the NDC was imputed. If no imputation was performed, RXFLG=1. If the imputation source was another pharmacy component record, RXFLG=2. Similarly, if the imputation source was a secondary, proprietary database and not the pharmacy component database, RXFLG=3. For these RXFLG=3 records, all the original data reported by the pharmacy is included on the record. Including only the original pharmacy reported data for these records was necessary in order to comply with legal restrictions associated with using the secondary data source as an imputation source. The imputed NDC value for the RXFLG=3 cases was used in the data editing, but is not available for public release. However, the imputed NDCs for the RXFLG=3 cases are available through the MEPS Data Center. Information on this topic can be obtained through the MEPS Project Director at <mepsdpd@ahrq.gov>.

PCIMPFLG indicates the type of match between a household reported event and a pharmacy component reported event. There are only two possible values for this variable (PCIMPFLG =1 or =3). These values indicate the possible “match-types” and are the following: =1 is an exact match for a specific event for a person between the pharmacy component and the household component and =3 is not an exact match between the pharmacy component and household component for a specific person (not an exact match means that a person’s household reported event did not have a matched counterpart in their corresponding pharmacy component records). PCIMPFLG assists analysts in determining which records have the strongest link to data reported by a pharmacy. It should be noted that whenever there are multiple purchases of a unique prescribed medication in a given round, MEPS did not collect information that would enable designating any single purchase as the “original” purchase at the time the prescription was first filled, and then designating other purchases as “refills”. The user needs to keep this in mind when the purchases of a medication are referred to as “refills” in the documentation. Because matching was performed at an event level as opposed to an acquisition level, the values for PCIMPFLG are either =1 or =3. Additionally, matching on an event versus acquisition level results in only one NDC being associated with a prescribed medicines event. (For more details on general data editing/imputation methodology, please see Section 3.0).

SELFFLG indicates whether or not an event was for a self-filer (SELFFLG=1) or a non-self-filer (SELFFLG=0). Self-filers are those respondents who reported that they submitted their own insurance claims directly to their insurance provider in a given round. Non-self-filers are those respondents who had their pharmacy provider submit their health insurance claim directly to their insurance carrier in a given round. SELFFLG is a round specific variable and the same person

may be both a self-filer and a non-self-filer during their period in the survey, but never in the same round.

INPCFLG denotes whether or not a household respondent had at least one prescription drug purchase in the pharmacy component (0=no, 1=yes).

When diabetic supplies, such as syringes and insulin, were mentioned in the Other Medical Equipment section of the MEPS HC, the interviewer was directed to collect information on these items in the Prescription Medicines section of the MEPS questionnaire. To the extent that these items are purchased without a prescription, they represent a non-prescription addition to the MEPS prescription drug expenditure and utilization data. Although these items may be purchased without a prescription, a prescription purchase may be required to obtain third party payments. For a majority of these types of events, third party payments were made, therefore, they are included on this file. Diabetic supplies can be identified in the file by using the variable, DIABFLG (0=not a diabetic supply/equipment or insulin, 1=is a diabetic supply/equipment or insulin). Diabetic supply/equipment and insulin events were identified by utilizing a proprietary database which assisted in assigning codes to each prescribed medicine event. This code assignment took into account the characteristics of the event. However, if desired, analysts are free to code and define diabetic supply/equipment and insulin events utilizing their own coding mechanism. If desired, DIABFLG can also be used by analysts to exclude diabetic supplies/equipment from their analyses.

2.6.2.5 The Sample Variable (SAMPLE)

SAMPLE indicates if a respondent reported receiving a free sample of the prescription medicine in the round (0=no, 1=yes). Each household respondent was asked in each round whether or not they received any free samples of a reported prescribed medicine during the round. However, respondents were not asked to report the number of free samples received, nor was it made clear that any free samples received were included in the count of the number of times that the respondent reported purchasing or otherwise obtaining the prescribed medicine during the round. Therefore, SAMPLE=1 for all acquisitions for a person for a specific prescription medicine during the round. This allows individual analysts to determine for themselves how free samples should be handled in their analysis.

2.6.2.6 Condition Codes (RXICD1X-RXICD3X) and Clinical Classification Codes (RXCCC1X-RXCCC3X)

Information on household reported medical conditions associated with each prescribed medicine event are provided on this file. There are up to three condition codes listed for each prescribed medicine event (99.7% of prescribed medicine events have 0-3 condition records linked). To obtain complete information associated with an event, the analyst must link to the 1997 Medical Conditions File. Details on how to link to the MEPS 1997 Medical Conditions File are provided in the 1997 Appendix File. The user should note that due to confidentiality restrictions, provider reported condition information are not publicly available.

The medical conditions reported by the Household Component respondent were recorded by the interviewer as verbatim text, which were then coded to fully-specified 1997 ICD-9-CM codes, including medical condition and V codes (see Health Care Financing Administration, 1980), by professional coders. Although codes were verified and error rates did not exceed 2.5 percent for any coder, analysts should not presume this level of precision in the data; the ability of household respondents to report condition data that can be coded accurately should not be assumed (see Cox and Cohen, 1985; Cox and Iachan, 1987; Edwards, et al, 1994; and Johnson and Sanchez, 1993). For detailed information on conditions, please refer to the documentation on the 1997 Medical Conditions File. For frequencies of conditions by event type, please see the 1997 Appendix File.

The ICD-9-CM condition codes were aggregated into clinically meaningful categories. These categories, included on the file as RXCCC1X-RXCCC3X, were generated using Clinical Classification Software (CCS) (formerly known as Clinical Classifications for Health Care Policy Research (CCHPR)), (Elixhauser, et al., 1998), which aggregates conditions and V-codes into 260 mutually exclusive categories, most of which are clinically homogeneous.

In order to preserve respondent confidentiality, nearly all of the condition codes provided on this file have been collapsed from fully-specified codes to 3-digit code categories. The reported ICD-9-CM code values were mapped to the appropriate clinical classification category prior to being collapsed to the 3-digit categories.

The condition codes (and clinical classification codes) linked to each prescribed medicine event are sequenced in the order in which the conditions were reported by the household respondent, which was in chronological order of reporting and not in order of importance or severity. Analysts who use the 1997 Medical Conditions file in conjunction with this prescribed medicines event file should note that the conditions on this file are sorted differently than they appear on the Medical Conditions file.

2.6.2.7 Record Count Variable (NUMCOND)

The variable NUMCOND indicates the total number of condition records which can be linked from the 1997 Medical Conditions File to each prescribed medicine event. For events with no condition records linked (NUMCOND=0), the condition and clinical classification code variables all have a value of -1 INAPPLICABLE. Similarly, for events without a linked second or third condition record, the corresponding second or third condition and clinical classification code variable was set to -1 INAPPLICABLE.

In order to obtain complete condition information for events with NUMCOND greater than 3, the analyst must link to the 1997 MEPS Condition File. Please see Section 6.0 for details on linking MEPS data files.

2.6.3 Expenditure Variables (RXSF97X-RXXP97X)

2.6.3.1 Definition of Expenditures

Expenditures on this file refer to what is paid for health care services. More specifically, expenditures in MEPS are defined as the sum of payments for care received, including out of pocket payments and payments made by private insurance, Medicaid, Medicare and other sources. The definition of expenditures used in MEPS differs slightly from its predecessors, 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. Although measuring expenditures as the sum of payments incorporates discounts in the MEPS expenditure estimates, the estimates do not incorporate any manufacturer or other rebates associated with Medicaid or other purchases. Another general 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. For details on expenditure definitions, please reference the following, “Informing American Health Care Policy” (Monheit, Wilson, Arnett, 1999).

2.6.3.2 Sources of Payment

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

1. Out of pocket by user or family
2. Medicare
3. Medicaid
4. Private Insurance
5. Veteran’s Administration, excluding CHAMPVA
6. CHAMPUS or CHAMPVA
7. Other Federal sources - includes Indian Health Service, Military Treatment Facilities, and other care by the Federal government
8. Other State and Local Source - includes community and neighborhood clinics, State and local health departments, and State programs other than Medicaid
9. Worker’s Compensation
10. Other Unclassified Sources - includes sources such as automobile, homeowner’s, liability, and other miscellaneous or unknown sources

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

11. Other Private - 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; and
12. Other Public - 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 these two additional sources of payment. While these payments stem from apparent inconsistent responses to 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 sampled 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 from persons who were not enrolled in Medicaid, but were presumed eligible by a provider who ultimately received payments from the program.

Please note, unlike the other events, the prescribed medicine events do have some remaining inconsistent responses between the insurance section of the Household Component and sources of payment from the PC (more specifically, discrepancies between Medicare only Household insurance responses and Medicaid sources 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 exact matches to events in the Pharmacy Component, and in addition, all of these types of events were purchases by persons with positive weights.

2.6.4 Sample Weights and Variance Estimation Variables (WTDPER97- VARPSU97)

2.6.4.1 Overview

There is a single full year person-level weight (WTDPER97) included on the file. A person-level weight was assigned to each prescribed medicines event reported by a key, in-scope person who responded to MEPS for the full period of time that he or she was in-scope during 1997. A key person either 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 interview (examples of the latter situation include newborns and persons returning from military service, an institution, or living outside the United States). A person is in-scope whenever he or she is a member of the civilian noninstitutionalized portion of the U.S. population.

2.6.4.2 Details on Person Weights Construction

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 variables. 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. Then a person level weight for Panel 1 was created, again including an adjustment for nonresponse over time and poststratification, again controlling to CPS population estimates based on the same five variables. When poverty status information derived from income variables became available, a 1997 composite 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 (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) as well as the original five poststratification variables in the establishment of control totals.

The panel specific weights described above are not available on the current file. This additional information is provided for your reference only. In order to determine which panel a sampled person was in, users must link to the 1997 Full Year Population Characteristics file to obtain the variable PANEL97.

2.6.4.3 The Final Weight for 1997

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 up 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.

2.6.4.4 Coverage

The target population for MEPS in this file 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.

3.0 General Data Editing and Imputation Methodology

The general approach to preparing the household prescription data for this file was to utilize the pharmacy follow-back prescription data to impute information collected from pharmacy providers 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 follow-back component 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 using a proprietary database on the basis of the medication names provided by the household and pharmacy, and, when available, the NDC provided in the pharmacy follow-back component. The matching process was done at an event level, as opposed to an acquisition level. Considerable editing was done prior to the matching to correct data inconsistencies in both data sets and to 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 secondary data file. In general, prescription drug unit prices were deemed to be outliers by comparing unit prices reported in the pharmacy database to the AWUP reported in the secondary data file and were edited, as necessary. Outlier thresholds were established in consultation with industry experts.

Drug matches between household drug mentions and pharmacy drug events for a person in the pharmacy follow-back were based on drug code, medication name, and the round in which the drug was reported. The matching of household drug mentions to pharmacy drug mentions was performed so that the most detailed and accurate information for each prescribed medicine event was obtained and was done at an event as opposed to an acquisition level. Exact dates of purchase were only available from the follow-back component. The matching program assigned scores to potential matches. Numeric variables required exact matches to receive a high score, while partial scores could be assigned to matches between character variables, such as prescription name, depending on the degree of similarity in the spelling and sound of the medication names. Household drug mentions that were deemed exact matches to pharmacy component drugs for the same person in the same round required sufficiently high scores to reflect a high quality match. Exact matches were used only once and were taken out of the donor pool from that point on (i.e., these matches were made without replacement). Any refill of a household drug mention that had been matched to a pharmacy drug event was also matched to the same pharmacy drug event. All remaining unmatched household drug mentions for persons either in or out of the pharmacy follow-back were statistically matched to the entire pharmacy donor base minus those PC records in which a NDC was imputed and was not an exact match on a generic product code applied to all records in the HC and PC, with replacement by medication name, drug code, type of third party coverage, health conditions, age, sex, and other characteristics of the individual.

3.1 Rounding

Expenditure variables on the 1997 Prescribed Medicines file have been rounded to the nearest penny. Person level expenditure information released on the 1997 Full Year Consolidated Data File were rounded to the nearest dollar. It should be noted that using the 1997 MEPS event files to create person level totals will yield slightly different totals than those found on the 1997 Full Year Consolidated Data File. These differences are due to rounding only. Moreover, in some instances, the number of persons having expenditures on the 1997 event files for a particular source of payment may differ from the number of persons with expenditures on the 1997 Full Year Consolidated Data File for that source of payment. This difference is also an artifact of rounding only. Please see the 1997 Appendix File for details on such rounding differences.

3.2 Edited/Imputed Expenditure Variables (RXSF97X-RXXP97X)

There are 13 expenditure variables included on this event file. All of these expenditures have gone through an editing and imputation process and have been rounded to the second decimal place. There is a sum of payments variable (RXXP97X) which for each prescribed medicine event sums all the expenditures from the various sources of payment. The 12 sources of payment expenditure variables for each prescribed medicine event are the following: amount paid by self or family (RXSF97X), amount paid by Medicare (RXMR97X), amount paid by Medicaid (RXMD97X), amount paid by private insurance (RXPV97X), amount paid by the Veterans Administration (RXVA97X), amount paid by CHAMPUS/CHAMPVA (RXCH97X), amount paid by other federal sources (RXOF97X), amount paid by state and local (non-federal) government sources (RXSL97X), amount paid by Worker's Compensation (RXWC97X), and amount paid by some other source of insurance (RXOT97X). As mentioned previously, there are two additional expenditure variables called RXOR97X and RXOU97X (other private and other public, respectively). These two expenditure variables were created to maintain consistency between what the household reported as their private and public insurance status for hospitalization and physician coverage and third party prescription payments from other private and public sources (such as a separate private prescription policy or prescription coverage from the Veterans Administration, the Indian Health Service, or a State assistance program other than Medicaid). Users should exercise caution when interpreting the expenditures associated with these two additional sources of payment. While these payments stem from apparent inconsistent responses to health insurance and source of payment questions in the survey, some of these inconsistencies may have logical explanations. Please note the Prescribed Medicines file is the only file on which some of these inconsistencies remain. Please see Section 2.6.3 for details on these and all other source of payment variables.

4.0 Strategies for Estimation

This file is constructed for efficient estimation of utilization, expenditure, and sources of payment for prescribed medicines and to allow for estimates of number of persons with prescribed medicines for 1997.

4.1 Variables with Missing Values

It is essential that the analyst examine all variables for the presence of negative values used to represent missing values. For example, a record with a value of -8 for the first ICD9 condition code (RXICD1X) indicates that the condition was reported as unknown.

For continuous or discrete variables, where means or totals may be taken, it may be necessary to set minus values to values appropriate to the analytic needs. That is, the analyst should either impute a value or set the value to one that will be interpreted as missing by the computing language used. For categorical and dichotomous variables, the analyst may want to consider whether to recode or impute a value for cases with negative values or whether to exclude or include such cases in the numerator and/or denominator when calculating proportions.

Methodologies used for the editing/imputation of expenditure variables (e.g., sources of payment, etc.) are described in Section 3.0.

4.2 Basic Estimates of Utilization, Expenditure and Sources of Payment

While the examples described below illustrate the use of event level data in constructing person level total expenditures, these estimates can also be derived from the person level expenditure file unless the characteristic of interest is event specific.

In order to produce national estimates related to prescribed medicines utilization, expenditure and sources of payment, the value in each record contributing to the estimates must be multiplied by the weight (WTDPER97) contained on that record.

Example 1

For example, the total number of prescription medicines including diabetic supply/equipment or insulin, for the civilian non-institutionalized population of the U.S. in 1997 is estimated as the sum of the weight (WTDPER97) across all prescription records including diabetic supply/equipment or insulin. That is,

$$\sum W_j = 1,873,742,048 \quad \text{for all records} \quad (1)$$

Various estimates can be produced based on specific variables and subsets of records.

Example 2

For example, the estimate for the mean out-of-pocket payment per prescription medicine purchase, including diabetic supply/equipment or insulin, should be calculated as the weighted average of amount paid by self/family. That is,

$$\bar{X} = (\sum W_j X_j) / (\sum W_j) = \$18.19, \quad (2)$$

where

$$\sum W_j = 1,873,742,048$$

and

$$X_j = \text{RXSF97X}_j \text{ for all prescription records}$$

This gives \$18.19 as the estimated mean amount of out-of-pocket payment of expenditures associated with prescribed medicines including diabetic supply/equipment or insulin and 1,873,742,048 as an estimate of the total number of prescription medicine purchases including diabetic supply/equipment or insulin. Both of these estimates are for the civilian non-institutionalized population of the U.S. in 1997.

Example 3

Another example would be to estimate the average proportion of total expenditures paid by private insurance per prescription medicine purchase (including diabetic supply/equipment or insulin). This should be calculated as the weighted mean of the proportion of the total prescription purchase paid by private insurance at the prescribed medicine, including diabetic supply/equipment or insulin, level. That is,

$$\bar{Y} = (\sum W_j Y_j) / (\sum W_j) = 0.2603, \quad (3)$$

where

$$\sum W_j = 1,873,742,048$$

and

$$Y_j = \text{RXPV97X}_j / \text{RXXP97X}_j \text{ for all prescription records}$$

This gives 0.2603 as the estimated mean proportion of each prescription including diabetic supply/equipment or insulin paid by private insurance for the civilian non-institutionalized population of the U.S. in 1997.

4.3 Estimates of the Number of Persons with Prescribed Medicine Events

When calculating an estimate of the total number of persons with prescribed medicine events, users can use a person-level file (1997 Full Year Consolidated Data File) or the current file. However, this event file must be used when the measure of interest is defined at the event level. For example, to estimate the number of persons in the civilian non-institutionalized population of the U.S. with a prescribed medicine purchase in 1997 with an RXNDC = "00093310905" (Amoxicillin), this event file must be used. This would be estimated as

$$\sum W_i X_i \quad \text{across all unique persons } i \text{ on this file,} \quad (4)$$

where

W_i is the sampling weight (WTDPER97) for person i

and

$$X_i = \begin{cases} 1 & \text{if RXNDC = "00093310905" for any purchase of person } i \\ 0 & \text{otherwise} \end{cases}$$

4.4 Person-Based Ratio Estimates

4.4.1 Person-Based Ratio Estimates Relative to Persons with Prescribed Medicine Events

This file may be used to derive person-based ratio estimates. However, when calculating ratio estimates where the denominator is persons, care should be taken to properly define and estimate the unit of analysis up to person level. For example, the mean expense for persons with prescribed medicine purchases is estimated as,

$$(\sum W_i Z_i) / (\sum W_i) \quad \text{across all unique persons } i \text{ on this file,} \quad (5)$$

where

W_i is the sampling weight (WTDPER97) for person i

and

$Z_i = \sum RXXP97X_j$ across all prescription purchases for person i

4.4.2 Person-Based Ratio Estimates Relative to the Entire Population

If the ratio relates to the entire population, this file cannot be used to calculate the denominator, as only those persons with at least one prescribed medicine event are represented on this data file. In this case the 1997 Full Year Consolidated Data File, which has data for all sampled persons, must be used to estimate the total number of persons (i.e., those with use and those without use). For example, to estimate the proportion of civilian non-institutionalized population of the U.S. with at least one prescribed medicine event with RXNDC = "00093310905" (Amoxicillin), the numerator would be derived from data on this event file, and the denominator would be derived from data on the 1997 Full Year Consolidated Data File. That is,

$$\frac{(\sum W_i Z_i)}{(\sum W_i)} \text{ across all unique persons } i \text{ on the 1997 Full Year Consolidated Data File,} \quad (6)$$

where

W_i is the sampling weight (WTDPER97) for person i

and

$$Z_i = \begin{cases} 1 & \text{if RXNDC}_j = \text{"00093310905"} \text{ for any event of person } i \text{ on the event-level file} \\ 0 & \text{otherwise for all remaining persons on the 1997 Full Year Consolidated Data File} \end{cases}$$

4.5 Sampling Weights for Merging Previous Releases of MEPS Household Data with this Event File

There have been several previous releases of MEPS Household Survey public use data. Unless a variable name common to several tapes is provided, the sampling weights contained on these data files are file-specific. The file-specific weights reflect minor adjustments to eligibility and response indicators due to birth, death, or institutionalization among respondents.

In general, for estimates from a MEPS data file that do not require merging with variables from other MEPS data files, the sampling weight(s) provided on that data file are the appropriate weight(s). When merging a MEPS Household data file to another, the major analytical variable (i.e., the dependent variable) determines the correct sampling weight to use.

5.0 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, one needs to take into account the complex sample design of MEPS. Various approaches can be used to develop such estimates of variance including use of the Taylor series or various replication methodologies. Replicate weights have not been developed for the MEPS 1997 data. Variables needed to implement a Taylor series estimation approach are described in the paragraph below.

Using a Taylor Series approach, variance estimation strata and the variance estimation PSUs within these strata must be specified. The corresponding variables on the MEPS full year utilization database are VARSTR97 and VARPSU97, respectively. Specifying a “with replacement” design in a computer software package such as SUDAAN (Shah, 1996) 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 associated with the corresponding estimates of variance. The following illustrates these concepts using two examples from Section 4.2.

Example 2 from Section 4.2

Using a Taylor Series approach, specifying VARSTR97 and VARPSU97 as the variance estimation strata and PSUs (within these strata) respectively and specifying a “with replacement” design in a computer software package SUDAAN will yield the standard error estimate of \$0.255 for the estimated mean of out-of-pocket payment.

Example 3 from Section 4.2

Using a Taylor Series approach, specifying VARSTR97 and VARPSU97 as the variance estimation strata and PSUs (within these strata) respectively and specifying a “with replacement” design in a computer software package SUDAAN will yield the standard error estimate of 0.006 for the weighted mean proportion of total expenditures paid by private insurance.

6.0 Merging/Linking MEPS Data Files

Data from this event file can be used alone or in conjunction with other files. This section provides instructions for linking the 1997 prescribed medicines file with other 1997 MEPS public use files, including a 1997 person level file, the 1997 conditions file, and the other 1997 event files.

6.1 Linking a Person Level File to the Prescribed Medicines File

Merging characteristics of interest from other 1997 MEPS files (e.g., the 1997 Full Year Consolidated File or the 1997 Office Based Provider File) expands the scope of potential estimates. For example, to estimate the total number of prescribed medicines purchased or otherwise obtained by persons with specific characteristics (e.g., age, race, and sex), population characteristics from a person level file need to be merged onto the prescribed medicines file. This procedure is illustrated below. The 1997 Appendix File provides additional details on how to merge 1997 MEPS data files.

1. Create data set PERSX by sorting a 1997 Full Year Population Characteristics File (file HCXXX), by the person identifier, DUPERSID. Keep only variables to be merged on to the prescribed medicines file and DUPERSID.
2. Create data set PMEDS by sorting the prescribed medicines file by person identifier, DUPERSID.
3. Create final data set NEWPMEDS by merging these two files by DUPERSID, keeping only records on the prescribed medicines file.

The following is an example of SAS code which completes these steps:

```
PROC SORT DATA=HCXXX(KEEP=DUPERSID AGE SEX EDUC)
OUT=PERSX;
  BY DUPERSID;
RUN;
```

```
PROC SORT DATA=HC016A OUT=PMEDS;
  BY DUPERSID;
RUN;
```

```
DATA NEWPMEDS;
  MERGE PMEDS (IN=A) PERSX(IN=B);
  BY DUPERSID;
  IF A;
RUN;
```

6.2 Linking the 1997 Conditions File and/or the Other 1997 MEPS Event Files to the 1997 Prescribed Medicines File

Due to survey design issues, there are limitations/caveats that an analyst must keep in mind when linking the different files. Those limitations/caveats are listed below. For detailed linking examples, including SAS code, analysts should refer to the 1997 Appendix File.

6.2.1 Limitations/Caveats of RXLK and CLNK

The RXLK file provides a link between the 1997 prescribed medicine records and the other 1997 MEPS event files. When using RXLK, analysts should keep in mind that a prescribed medicine event may link to more than one medical event. When this occurs, it is up to the analyst to determine how the prescribed medicine expenditures should be allocated among those events. In order to obtain complete information about those other event files, the analyst must link to the other public use event files.

The CLNK provides a link between the 1997 Conditions File and the 1997 Prescribed Medicines file. When using the CLNK, analysts should keep in mind that (1) conditions are self reported and (2) there may be multiple conditions associated with a drug purchase. Analysts need to verify that a particular medication is indeed an appropriate medication in treating the condition. Moreover, there may be some drugs that were purchased to treat a specific health condition for which there is no such link to the condition file.

References

- Cohen, S.B. (1998). Sample Design of the 1996 Medical Expenditure Panel Survey Medical Provider Component. Journal of Economic and Social Measurement. Vol 24, 25-53.
- 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*.
- Cox, B.G. and Cohen, S.B. (1985). Chapter 8: Imputation Procedures to Compensate for Missing Responses to Data Items. In *Methodological Issues for Health Care Surveys*. Marcel Dekker, New York.
- Moeller J.F., Stagnitti, M., Horan, E., et al. Outpatient Prescription Drugs: Data Collection and Editing in the 1996 Medical Expenditure Panel Survey. Rockville (MD): Agency for Healthcare Research and Quality; 2000. MEPS Methodology Report (forthcoming).
- Monheit, A.C., Wilson, R., and Arnett, III, R.H. (Editors). *Informing American Health Care Policy*. (1999). Jossey-Bass Inc, San Francisco.
- Shah, B.V., Barnwell, B.G., Bieler, G.S., Boyle, K.E., Folsom, R.E., Lavange, L., Wheelless, S.C., and Williams, R. (1996). *Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0*, Research Triangle Park, NC: Research Triangle Institute.

Attachment 1

Definitions

Dwelling Units, Reporting Units, Families, and Persons – The definitions of Dwelling Units (DUs) and Group Quarters 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 a group of persons in the sampled dwelling unit who is related by blood, marriage, adoption or other family association, and who is to be interviewed as a group in MEPS. Thus, the RU serves chiefly as a family-based “survey operations” unit rather than an analytic unit. 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.

Unmarried college students under 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 1 MEPS interview, were treated as a Reporting Unit separate from that of their parents for the purpose of data collection. These variables can be found on MEPS person-level files.

In-Scope—A person was classified as in-scope (INSCOPE) if he or she was a member of the U.S. civilian, non-institutionalized population at some time during the Round 1 interview. This variable can be found on MEPS person-level files.

Keyness—The term “keyness” is related to an individual’s chance of being included in MEPS. A person is key if that person is appropriately linked to the set of NHIS sampled households designated for inclusion in MEPS. Specifically, a key person either 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 prior to joining that household (examples of the latter situation include newborns and persons returning from military service, persons returning from an institution, or persons 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 that was eligible but not sampled for the NHIS, who happened to have become a member of a MEPS reporting unit by the time of the MEPS Round 1 interview. MEPS data, (e.g., utilization and income) were collected for the period of time a non-key person was part of the sampled unit to permit family level analyses. However, non-key persons who leave a sample household would not be recontacted for subsequent interviews. Non-key individuals are not part of the target sample used to obtain person-level national estimates.

It should be pointed out that a person may 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 be living with his or her civilian spouse and children in a household sampled for the NHIS. The person in the military would be considered a key person for MEPS. However, such a person would not receive a person-level sample weight so long as he or she was in the military. All key persons who participated in the first round of a MEPS panel received a person-level sample weight except those who were in the military. The variable indicating “keyness” is KEYNESS. This variable can be found on MEPS person-level files.

Eligibility—The eligibility of a person for MEPS pertains to whether or not data were to be collected for that person. All key, in-scope 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 the same RU as one or more key persons, and their eligibility continued only for the time that they were living with a key person. The only out-of-scope persons eligible for data collection were those who were living with key in-scope persons, again only for the time they were living with a key person. Only military persons meet this description. A person was considered eligible if they were eligible at any time during Round 1. The variable indicating “eligibility” is ELIGRND1, where 1 is coded for persons eligible for data collection for at least a portion of the Round 1 reference period, and 2 is coded for persons not eligible for data collection at any time during the first round reference period. This variable can be found on MEPS person-level files.

Pre-imputed - This means that only a series of logical edits were applied to the HC data to correct for several problems including outliers, co-payments or charges reported as total payments, and reimbursed amounts counted as out-of-pocket payments. Missing data remains.

Unimputed - This means that only a series of logical edits were applied to the MPC data to correct for several problems including outliers, co-payments or charges reported as total payments, and reimbursed amounts counted as out-of-pocket payments. These data were used as the imputation source to account for missing HC data.

Imputation - A method of estimating values for cases with missing data. Hot-deck imputation creates a data set with complete data for all nonrespondent cases, by substituting the data from a respondent case that resembles the nonrespondent on certain known variables.

Household Reported Drug (mention) - A household reported drug is a unique prescribed medication reported by a household respondent. A household reported drug is checked on the prescribed medicines roster as being created during that round or selected from a roster from a previous round. Associated with each household reported drug mention in a given round may be multiple acquisitions of the medication during that round. Thus, what originally was reported as a single medication in the Household Component may appear as multiple unique medications on the prescribed medicines event file.

D. Variable-Source Crosswalk

D. VARIABLE-SOURCE CROSSWALK

MEPS HC016A: 1997 PRESCRIBED MEDICINES EVENTS

Survey Administration Variables

Variable	Description	Source
DUID	Dwelling unit ID	Assigned in sampling
PID	Person number	Assigned in sampling
DUPERSID	Sample person ID (DUID + PID)	Assigned in sampling
RXRECIDX	Record ID – Unique Prescribed Medicine Identifier	Constructed
LINKIDX	Link to condition and other event files	CAPI derived
PURCHRD	Round in which the Rx/prescribed medicine was obtained/purchased	Constructed

Prescribed Medicines Events Variables

Variable	Description	Source
RXBEGDD	Day person first used medicine	PM11OV1
RXBEGMM	Month person first used medicine	PM11OV2
RXBEGYR	Year person first used medicine	PM11
RXNAME	Medication name (Imputed)	Imputed
RXHHNAME	Household reported medication name	PM05
RXNDC	National drug code (Imputed)	Imputed
RXQUANTY	Quantity of Rx/prescribed medicine (Imputed)	Imputed
RXFORM	Form of Rx/prescribed medicine (Imputed)	Imputed
RXFORMOS	Other specify form of Rx/prescribed medicine (Imputed)	Imputed
RXSTRENG	Strength of Rx/prescribed medicine dose (Imputed)	Imputed
RXUNIT	Unit of measurement for Rx/prescribed medicine dose (Imputed)	Imputed
RXUNITOS	Other specify unit of measurement for Rx/prescribed medicine dose (Imputed)	Imputed

PHARTP1- PHARTP7	Type of pharmacy provider – (1st-7th)	PM16
RXFLG	Flag variable indicating imputation source for NDC on pharmacy donor record (Imputed)	Constructed
PCIMPFLG	Flag indicating type of household to pharmacy prescription match	Constructed
SELFFLG	Flag indicating whether or not the event is a self-filer event	CP01/Constructed
INPCFLG	Flag indicating if the person has at least one record in the pharmacy component	Constructed
DIABFLG	Flag indicating whether or not prescribed medicine was classified as insulin or diabetic supply/equipment	Constructed
SAMPLE	Flag indicating if a respondent received a free sample of this drug in the round	Constructed
RXICD1X	3 digit ICD-9 condition code	PM09
RXICD2X	3 digit ICD-9 condition code	PM09
RXICD3X	3 digit ICD-9 condition code	PM09
RXCCC1X	Modified Clinical Classification Code	Constructed/Edited
RXCCC2X	Modified Clinical Classification Code	Constructed/Edited
RXCCC3X	Modified Clinical Classification Code	Constructed/Edited
NUMCOND	Total number of conditions associated with a prescribed medicine event	Constructed
RXSF97X	Amount paid, self or family (Imputed)	CP11/Edited/ Imputed
RXMR97X	Amount paid, Medicare (Imputed)	CP12/CP13/Edited/ Imputed
RXMD97X	Amount paid, Medicaid (Imputed)	CP12/CP13/Edited/ Imputed
RXPV97X	Amount paid, private insurance (Imputed)	CP12/CP13/Edited/ Imputed
RXVA97X	Amount paid, Veteran’s Administration (Imputed)	CP12/CP13/Edited/ Imputed
RXCH97X	Amount paid, CHAMPUS/CHAMPVA (Imputed)	CP12/CP13/Edited/ Imputed
RXOF97X	Amount paid, other Federal (Imputed)	CP12/CP13/Edited/ Imputed
RXSL97X	Amount paid, state and local gov’t (Imputed)	CP12/CP13/Edited/ Imputed
RXWC97X	Amount paid, Worker’s Compensation (Imputed)	CP12/CP13/Edited/ Imputed

RXOT97X	Amount paid, other insurance (Imputed)	CP12/CP13/Edited/ Imputed
RXOR97X	Amount paid, other private (Imputed)	Constructed/Imputed
RXOU97X	Amount paid, other public (Imputed)	Constructed/Imputed
RXXP97X	Sum of payments RXSF97X – RXOU97X (Imputed)	CP12/CP13/Edited/ Imputed

Weights

Variable	Description	Source
WTDPER97	Poverty/mortality adjusted person level weight	Constructed
VARSTR97	Variance estimation stratum, 1997	Constructed
VARPSU97	Variance estimation PSU,1997	Constructed