



# **Methodology Report #24**

**Estimation Procedures for the 2007  
Medical Expenditure Panel Survey  
Household Component**



## Table of Contents

1.0 Introduction	1
1.1 Summary of MEPS Sample Design	2
1.2 Terms Related to Sampling/Target Population	3
1.3 MEPS Analytic Files	4
1.4 Overview of Development of MEPS Weights	5
2.0 Point-in-Time (PIT) Weights	8
2.1 Development of Panel-Specific PIT Weights	9
2.2 Development of Combined-Panel PIT Weights	14
2.3 General Issues Related to Developing PIT Weights	15
3.0 Full Year (FY) Weights	17
3.1 Development of Panel-Specific FY Weights	17
3.2 Development of Combined-Panel Person Weights for FY Population Characteristics File	20
3.3 Development of Final FY Weights for Consolidated File	21
3.4 General Issues Related to Developing FY Weights	24
4.0 Weights for Survey Supplements	25
4.1 Self-Administered Questionnaire (SAQ)	25
4.2 Diabetes Care Supplement (DCS)	26
5.0 Longitudinal Weights	27
6.0 Variance Estimation	29
7.0 References	31
Figure 1. Overview of MEPS Household Component Panel Design, 2002–2007	2
Figure 2. Illustration of MEPS from Longitudinal Perspective, Panel 10 (2006–2007)	3
Figure 3. Person-Level PIT and FY Weight Development Process	6
Figure 4. Family-Level PIT and FY Weight Development Process	7
Figure 5. Pooling of Overlapping MEPS Panels for PIT File	8
Appendix A. Distribution of Weights at Different Stages of Weighting for 2007 MEPS	32
Appendix B. Variables Used in Nonresponse and Poststratification/Raking Adjustments	33
Appendix C. Glossary of Selected Terms	49

## ABSTRACT

The primary purpose of this report is to describe the procedures used to develop various MEPS analytic weights. A working knowledge of these procedures is important for researchers who wish to produce accurate and valid national estimates from the MEPS data. Following an introductory section that provides a brief overview of the survey design as well as the types of MEPS analytic files and weight variables, the report provides detailed descriptions of the procedures used to develop the various MEPS analytic weights (Point-in-Time, Full Year, Survey Supplements, and Longitudinal weights). The last section provides an overview of the variance estimation procedures used to analyze MEPS data. Although MEPS weighting procedures do not vary substantially from year to year, the details presented in this report apply specifically to the 2007 data year.

### **Suggested Citation:**

Machlin S.R., Chowdhury S.R., Ezzati-Rice T., DiGaetano R., Goksel H., Wun L.-M., Yu W., Kashiara D. *Estimation Procedures for the Medical Expenditure Panel Survey Household Component*. Methodology Report #24. September 2010. Agency for Healthcare Research and Quality, Rockville, MD.

[http://www.meps.ahrq.gov/mepsweb/data\\_files/publications/mr24/mr24.pdf](http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr24/mr24.pdf)

# **Estimation Procedures for the Medical Expenditure Panel Survey Household Component**

Steven Machlin, Sadeq Chowdhury, Trena Ezzati-Rice, Ralph DiGaetano, Huseyin Goksel, Lap-Ming Wun, William Yu, David Kashihara

## **1.0 Introduction**

The Medical Expenditure Panel Survey (MEPS) Household Component (HC) is a nationally representative sample of the U.S. civilian noninstitutionalized population. Sponsored by the Agency for Healthcare Research and Quality (AHRQ) of the U.S. Department of Health and Human Services (DHHS), MEPS has been conducted continuously since 1996. MEPS provides comprehensive data on health care use, expenditures, sources of payment, and health insurance coverage as well as information on survey respondents' health status, demographic/socio-economic characteristics, employment status, access to health care, and satisfaction with health care. Estimates can be produced for persons and families in the U.S. as well as subgroups of the population.

Because MEPS is a complex probability sample, analytic approaches based on data from a simple random sample are usually not appropriate. In particular, ignoring the complex design can lead to biased estimates and inaccurate significance levels. Sample weights and the stratification and clustering aspects of the design must be incorporated into analyses in order to produce appropriate estimates and standard errors of estimates.

The primary purposes of this report are to describe the procedures used to develop the various MEPS analytic weights and provide an overview of appropriate variance estimation procedures. A working knowledge of this information is important for researchers who wish to produce accurate, valid national estimates from the MEPS data. In this introduction, we provide a brief overview of the survey design as well as the types of MEPS analytic files and weighting variables. Sections 2–6 provide detailed descriptions of the procedures used to develop the various MEPS analytic weights while Section 7 provides an overview of the main variance estimation procedures used to analyze MEPS data. MEPS weighting procedures do not generally change substantially over time but there can be minor differences in implementation from year to year. The details presented in this report apply specifically to the 2007 data year. A glossary of various terms used in this report that have a technical definition for MEPS can be found in Appendix C.

## 1.1 Summary of MEPS Sample Design

The MEPS-HC is a complex national probability sample survey of the U.S. civilian noninstitutionalized population. Each year a new sample of households is selected from among those households that participated in the previous year’s National Health Interview Survey (NHIS), another large ongoing federal health survey conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). The NHIS is based on a probability sample of the U.S. civilian noninstitutionalized population selected through a complex multistage area sample design. The details of the NHIS sample design can be found in Botman et al. (2000). The MEPS sample of households is a subsample of NHIS responding households and reflects many of the features of the NHIS design. Ezzati-Rice et al. (2008) provides some details of the common features of the NHIS and MEPS designs.

Each new MEPS sample is referred to as a panel and data for each panel are collected through a series of five rounds of computer-assisted personal interviews (CAPI) that yield annual data for each of two consecutive calendar years. The first two interviews (Rounds 1–2) cover most of the first year, the last two interviews (Rounds 4–5) cover most of the second year, and the middle interview (Round 3) covers the end part of the first year and the beginning part of the second year.

A new MEPS panel of households has been selected and fielded every year since 1996. As illustrated in Figure 1 for 2002–2007, data are combined across two distinct nationally representative samples, making use of the MEPS overlapping panel design to increase the precision of annual estimates produced from MEPS.<sup>1</sup> More specifically, annual estimates are made by combining data from two consecutive panels—one from the first year of data collection and the other from the second year of data collection. For example, 2007 annual estimates are based on data collected for the second year of Panel 11 and data collected for the first year of Panel 12.

**Figure 1. Overview of MEPS Household Component Overlapping Panel Design, 2002–2007**

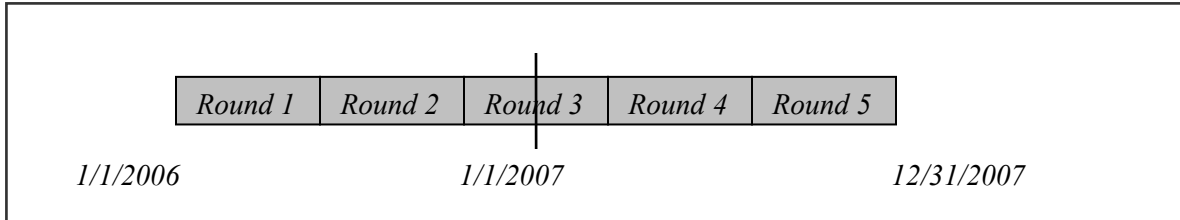
MEPS Panel	Year							
	□	2002	2003	2004	2005	2006	2007	→
6		R3 R4 R5						
7		R1 R2 R3 R4 R5						
8			R1 R2 R3 R4 R5					
9				R1 R2 R3 R4 R5				
10					R1 R2 R3 R4 R5			
11						R1 R2 R3 R4 R5		
12							R1 R2 R3	

In addition to annual estimates, the MEPS design structure permits longitudinal estimates over two consecutive calendar years, thus allowing examination of person-level changes

<sup>1</sup> In 1996, the first year of MEPS, only one panel (Panel 1) was fielded, thus the annual data for 1996 were based on this single panel of data.

in selected variables over a two year period for a single panel. For example, research analysts can assess the persistence of high health care expenditures by examining whether individuals with high expenditures in one year also have high expenditures in the subsequent year or shift to a lower expenditure level. Figure 2 provides an illustration of MEPS from a longitudinal perspective for Panel 11. Data for this panel can be used to analyze person-level changes in round-specific variables or changes in annual level variables (e.g., total health care expenditures) between 2006 and 2007.

**Figure 2. Illustration of MEPS from Longitudinal Perspective, Panel 10 (2006–2007)**



## 1.2 Terms Related to Sampling/Target Population

The terms "RU," "in-scope," "Key," and "eligible" are used at times in this report in the context of developing analytic weights for MEPS sample persons.

An *RU (Reporting Unit)* consists of all members of a family unit living together (and reporting as a single family for MEPS) or of a single person (if living without other family members). The definition of an RU and a MEPS family (see Section 3.3.2) are identical with the exception of student RUs. A student RU is a young adult family member living away from home to attend school who is linked back to his or her family for MEPS family level analysis. Student RUs are identified as an operational device to help obtain accurate data for the “student living away from home” component of the general population.

An individual is *in-scope* at a particular point in time if (s)he is a member of the civilian, noninstitutionalized population of the U.S. (i.e., target population for MEPS). Since the MEPS sample is a subsample of NHIS respondents, the chance of selection for MEPS is directly tied to the chance of selection for the NHIS. Individuals who were in-scope at the time of the NHIS are defined as Key persons if and only if they were members of an NHIS household sampled for MEPS in a given year. Persons who were not in-scope at the time of the NHIS (i.e., those living outside the U.S., in the military, in a nursing home, or newborns) are defined as Key if and only if they were in-scope at the time they join an RU participating in MEPS. Therefore, in-scope persons not selected as part of the original NHIS sample are always non-Key even if they subsequently join a MEPS household.

Finally, a person is *eligible* for data collection in MEPS if (s)he is a member of an RU containing at least one person classified as in-scope and Key. A glossary at the end of this

report (Appendix C) contains definitions of these and other selected terms and acronyms used in this report.

### **1.3 MEPS Analytic Files**

In general, three types of person-level MEPS public use data files with appropriate weight and variance estimation variables are released every year. These files, in sequential order of release during the year, are referred to as: Point-in-Time (PIT), Full Year (FY), and Longitudinal (L) data files. The FY data are released in two phases: a preliminary "Population Characteristics File" is released first which is superseded by a final "Consolidated File" several months later. These two files are also known as "Use File" and "Expenditure File" respectively, because the preliminary file is the first to provide annual data on health care utilization while the final file includes both utilization and associated expenditure data. Person-level weights are included on all files while family-level weights (attached to person-level records) are included on the PIT and FY Consolidated files only.

As a consequence of the MEPS overlapping panel design referenced above, data collected for the same calendar year from two consecutive MEPS sample panels are used to develop the PIT and FY files. For example, the 2007 PIT file comprises data from Round 1 for the Panel 12 sample and Round 3 (2007 portion) for the Panel 11 sample. Similarly, the 2007 FY files consist of data from Rounds 1, 2, and 3 (2007 portion) for the MEPS Panel 12 sample and from Rounds 3 (2007 portion), 4, and 5 for the Panel 11 sample. In contrast, longitudinal files contain data for the entire two-year MEPS survey period for one specific MEPS Panel.

## 1.4 Overview of Development of MEPS Weights

Figures 3 and 4 present flowcharts of the weight development processes for PIT and FY weights at the person-level and family-level, respectively. These flowcharts are designed to facilitate the descriptions of the weight development processes described in Section 2 (PIT) and Section 3 (FY) of this report. In general, the development of MEPS weights involves a series of derivations and adjustments, with the PIT weight for Round 1 developed first and then serving as baseline for deriving the FY weight for a new MEPS Panel. This FY weight serves as the starting point for developing the PIT weight associated with the second year of data collection (i.e., Round 3 weight) for the Panel. Adjustments for nonresponse/attrition and raking/poststratification to control totals are incorporated at different points in the process. Panel-specific weights for concurrent panels are ultimately combined to produce final PIT/FY weights.

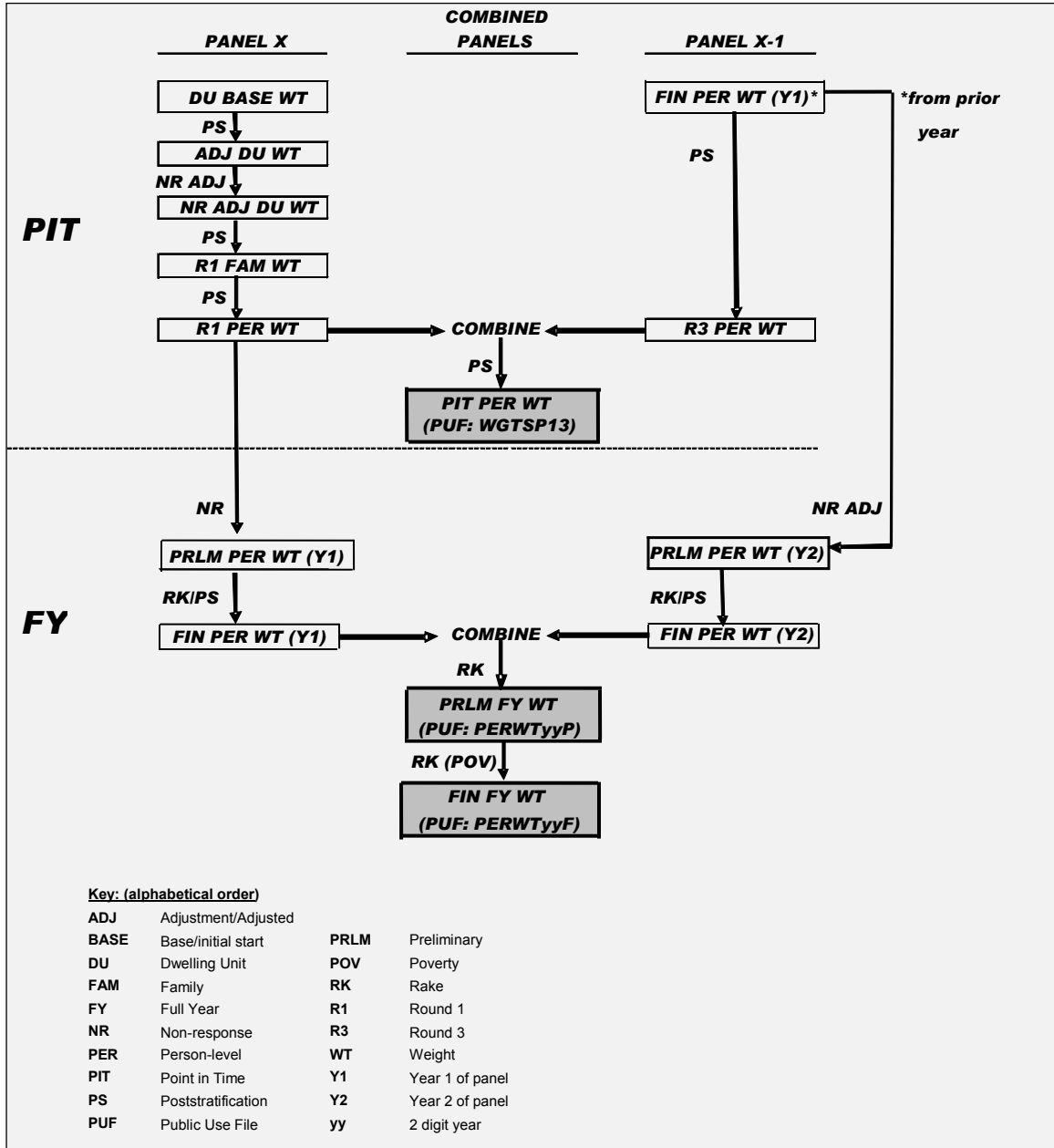
The table below lists the final weight variables provided on MEPS public use files (PUF) that are described in the following sections (2–6) of this report. These weights vary in derivations with respect to reference time periods and do not apply to exactly the same sample persons. Therefore, the pertinent sample weight should be used to produce appropriate estimates based on MEPS data. Appendix A provides summary distributions for 2007 of selected interim and PUF weights mentioned in this report.

Type of Public Use File (PUF)	Variable Name in PUF <sup>2</sup>	
	Person-Level	Family-Level
PIT	WGTSP13	WGTRU13
FY Preliminary SAQ Supplement DCS Supplement	PERWT <sub>yy</sub> P SAQWT <sub>yy</sub> P DIABW <sub>yy</sub> P	— — —
FY Final Consolidated SAQ Supplement DCS Supplement	PERWT <sub>yy</sub> F SAQWT <sub>yy</sub> F DIABW <sub>yy</sub> F	FAMWT <sub>yy</sub> F, FAMWT <sub>yy</sub> C — —
Longitudinal	LONGWT	—

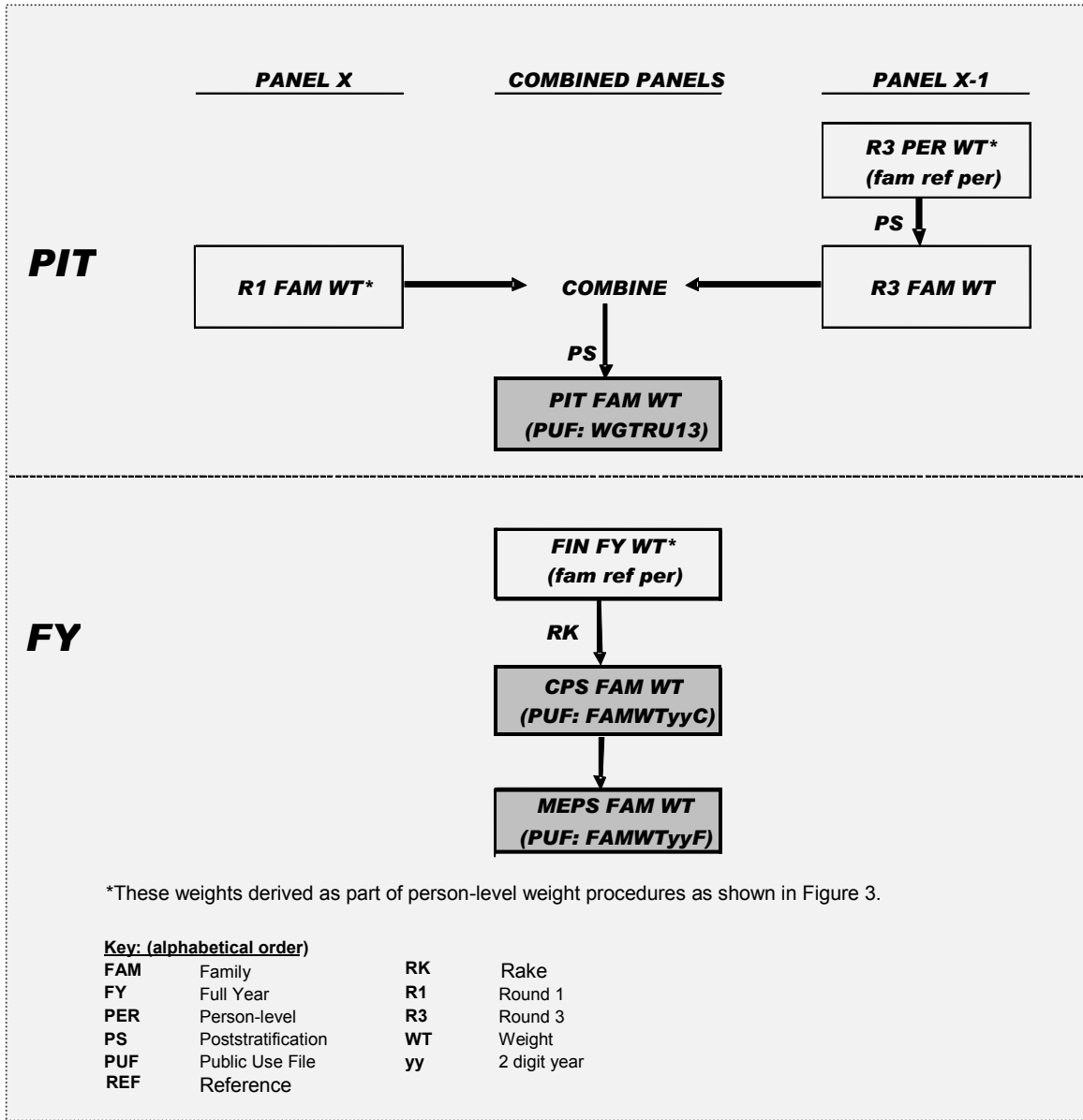
<sup>2</sup> "yy" in variable names indicates 2-digit year. PIT and longitudinal weight variable names do not have a "yy" component since PIT files cover the first part of the year while longitudinal files cover 2 years for a specific panel. The PIT/longitudinal weight variable names do not vary across years/panels, respectively.



**Figure 3. Person-Level PIT and FY Weight Development Process**



**Figure 4. Family-Level PIT and FY Weight Development Process**



## 2.0 Point-in-Time (PIT) Weights

The starting point for development of the FY weights is the PIT weights. A PIT file is produced every year to provide timely data for estimates related to health insurance and other selected variables of interest that are available from a single round and require minimal editing. Because two consecutive (overlapping) panels in MEPS are fielded concurrently, data from Round 1 of the current panel and Round 3 of the previous panel represent the same time period and the pooling of the two rounds increases the sample sizes for estimation. Thus, in the PIT files, Round 1 data for the most recent panel are combined with Round 3 data from the previous panel. PIT sample weights are constructed at both the person and family levels and can be used to produce estimates that reflect approximately the first half of the year and/or the date of first interview in the year. Moreover, the PIT weight serves as the building block for constructing FY weight variables (see Section 3.0 below).

As shown in Figures 3 and 4, the PIT weight is the result of a composite of several factors as follows:

- a base weight (based on an interim NHIS weight),
- probability of selection for MEPS from NHIS,
- adjustments for nonresponse, and
- poststratification using external control totals.

Weights are constructed separately for the two overlapping individual panels/rounds and then combined into a final analysis weight for the PUF. MEPS Panel 11 spans the two calendar years 2006 and 2007 while MEPS Panel 12 spans 2007 and 2008. The 2007 file consists of the subset of data from the eleventh and twelfth MEPS panels covering January 1 through, roughly, the spring of calendar year 2007. More specifically, data from the 2007 portion of the third round of data collection for the MEPS Panel 11 sample are pooled with data from the first round of data collection for the MEPS Panel 12 sample (see Figure 5).

**Figure 5. Pooling of Overlapping MEPS Panels for PIT File**

	2006			2007	
	<u>Jan</u>			<u>Jan</u>	
<b>Panel 11 2006–2007</b>	<b>Round 1</b>	<b>Round 2</b>		<b>Round 3</b>	
	<b>Panel 12 2007–2008</b>			<b>Round 1</b>	

The steps used to develop the PIT person and family-level weight variables are described in detail below.

## 2.1 Development of Panel-Specific PIT Weights

### 2.1.1 Most Recent Panel (Round 1)

**a. Constructing Dwelling Unit (DU) Base Weight.** The MEPS DU (e.g., household) base weight is calculated as the nonresponse adjusted NHIS household weight multiplied by the reciprocal of the MEPS subsampling rate used to select the DU (this rate varies by MEPS sample domain<sup>3</sup>). A constant factor is then applied to adjust for the number of quarter/panel combinations available for use as the MEPS sample frame from among the 16 NHIS quarter/panel combinations (Botman et al. 2000). This factor is applied to each household weight so that the sum of the weights can be viewed as a national estimate of households. In most years this factor is 16/6, reflecting the inclusion of households from 3 of 4 quarters in 2 of 4 NHIS panels.

The construction of the DU base weight can be expressed as:

$$WTDU_{0i} = WTDU_{Ni} \times \frac{1}{f_i} \times A_0$$

where, for DU  $i$ ,  $WTDU_{0i}$  is the MEPS DU base weight,  $WTDU_{Ni}$  is the NHIS household weight,  $f_i$  is the MEPS subsampling rate, and  $A_0$  is the constant adjustment factor (usually 16/6).

To reduce the impact of extremely large weights on the variances of MEPS estimates, the extreme base weights are trimmed by reviewing the distribution of the weights within each weight trimming class ( $c$ ) defined by the cross-classification of DU's minority status<sup>4</sup>, NHIS sampling strata defined at the segment level, and the MEPS subsampling domains. The trimmed base weight can be expressed as

$$WTDU_{1i} = WTDU_{0i} \times A_{1c} \quad \text{with } A_{1c} = \min\left(\frac{WTDU_{0c}^{\ominus}}{WTDU_{0i}}, 1\right), \quad i \in c$$

where  $WTDU_{1i}$  is the trimmed base weight for DU  $i$  that belongs to the trimming class  $c$ ,  $WTDU_{0c}^{\ominus}$  is the trimming cutoff value determined for class  $c$ , and  $A_{1c}$  is the weight trimming factor for class  $c$ . For almost all cases  $A_{1c} = 1$  and only for a few cases with very large weights  $A_{1c} < 1$ .

**b. Poststratification Ratio Adjustment of Trimmed Base Weight.** To improve the representativeness of the MEPS sample with respect to the NHIS full sample, a poststratification ratio adjustment is applied to the trimmed DU base weights using household level control totals, which are estimated using the household reference

<sup>3</sup> For more information regarding the MEPS sample design, see Ezzati-Rice et al.

<sup>4</sup> NHIS indicator for the black and Hispanic households (starting with Panel 12, Asian households are also included in minority status).

person’s final person weight in the NHIS full sample.<sup>5</sup> The poststratification cells are defined using the cross-classification of MSA status, family income, employment status of the NHIS reference person<sup>6</sup>, race/ethnicity of the NHIS reference person, and reported health status of household members. See Appendix B.1 for detailed definitions of these variables. The poststratified weight can be expressed as:

$$WTDU_{2i} = WTDU_{1i} \times A_{2c} \text{ with } A_{2c} = \frac{CTDU_c}{\sum_{i \in c} WTDU_{1i}}, i \in c,$$

where  $WTDU_{2i}$  is the poststratified weight for DU  $i$  that belongs to poststratification cell  $c$ <sup>7</sup>,  $A_{2c}$  is the poststratification adjustment factor, and  $CTDU_c$  is the DU level control total for cell  $c$  as derived from NHIS.

**c. Adjustment for DU Level Nonresponse in Round 1.** The poststratified DU level weight calculated above is adjusted to compensate for nonresponding DUs to the Round 1 interview. Using a broad range of 29 demographic, geographic, socioeconomic and health covariates from the NHIS (see Appendix B.2), a Chi-squared Automatic Interaction Detector (CHAID) analysis (Kass, 1980) is carried out to form MEPS nonresponse adjustment classes to adjust for the MEPS DU level nonresponse. Since the eligibility status<sup>8</sup> is known for a responding DU only and unknown for a nonresponding DU, both eligible and ineligible DUs are included in this adjustment. The nonresponse adjustment is applied by inflating the weights of the responding DUs in each adjustment cell as follows:

$$WTDU_{3i} = WTDU_{2i} \times A_{3c} \text{ if } i \text{ is a responding DU and } i \in c$$

$$\text{with } A_{3c} = \frac{\sum_{i \in c} WTDU_{2i}}{\sum_{i \in c \cap r} WTDU_{2i}}, i \in c,$$

where  $WTDU_{3i}$  is the poststratified weight for DU  $i$  which is a respondent and belongs to the nonresponse adjustment cell  $c$ ,  $A_{3c}$  is the nonresponse adjustment factor for cell  $c$ ,  $r$  represents the set of responding DUs, and therefore,  $c \cap r$  represents the set of responding DUs in cell  $c$ . That means the adjustment factor is the ratio of the sum of the weights of all DUs over the sum of the weights of the responding DUs in an adjustment cell.<sup>9</sup> As mentioned above, the numerator and the denominator include both eligible and ineligible DUs in this adjustment.

<sup>5</sup> The control totals for this poststratification adjustment were derived from NHIS for the adjustment of all MEPS panels except for Panel 13 when control totals were derived from the Current Population Survey (CPS) using a different set of poststratification cells based on region, MSA, and selected characteristics of a household reference person age, race/ethnicity, marital status, and education.

<sup>6</sup> Reference person is an individual identified as owning or renting the home.

<sup>7</sup> Note that the notation  $c$  is used to denote an adjustment cell throughout the report but the cell definition varies from step to step.

<sup>8</sup> A DU is classified as eligible if at least one member meets the eligibility criterion for MEPS.

<sup>9</sup> Note that the nonresponding cases drop out of the file after the nonresponse adjustment.

**d. Family-Level Poststratification Adjustment to Form Final Round 1 Family-Level Weight.** After calculation of the nonresponse adjusted DU weight above, a family-level weight is derived for responding families<sup>10</sup> by starting with the nonresponse adjusted DU weight as the base weight and then applying a poststratification adjustment at the family level.<sup>11</sup> The poststratification is carried out using the cross-classification of the following variables: family type, race/ethnicity of reference person<sup>12</sup>, region, MSA status, age category of reference person, and number of eligible family members (see Appendix B.3). The March Current Population Survey (CPS) family-level totals are used as control totals for this adjustment. This adjustment also serves as an adjustment for nonresponse at the family level to yield the final Round 1 family weight.

More specifically, the final family weight ( $WTFM_1$ ) for Round 1 can be expressed as:

$$WTFM_{1j} = WTFM_{0j} \times F_{1c}, \quad j \in c$$

$$\text{with } WTFM_{0j} = WTDU_{3i}, \quad j \in i$$

$$\text{and } F_{1c} = \frac{CTFM_c}{\sum_{j \in c \cap r} WTFM_{0j}},$$

where  $WTFM_{0j}$  is the base family weight and  $WTFM_{1j}$  is the Round 1 final weight for family  $j$ ;  $F_c$  is the poststratification adjustment factor for cell  $c$ ,  $c \cap r$  indicates the set of responding families within cell  $c$ , and  $CTFM_c$  is the family-level control total for cell  $c$ . The distribution of this weight is reviewed, and very large weights, if any, are trimmed to the level of highest acceptable weight, and the poststratification adjustment is repeated.

Only families that meet the following conditions are assigned an initial weight and included in poststratification adjustment to produce the final Round 1 family weight:

- at least one Key in-scope person in the family during Round 1,
- the reference person in the family is Key, and
- all Key in-scope family members have a positive person weight.

**e. Person-Level Poststratification Adjustment to Form Final Round 1 Person-Level Weight.** For reference persons and married persons with spouse present under the age of 65 (i.e., subgroup  $g_1$ ), the final family weight produced in step d above is assigned as their final person weight.<sup>13</sup> For the remaining persons (i.e., subgroup  $g_2$ ), poststratification at the person level of the family weight produced in step d above, is

<sup>10</sup> See glossary for definitions of families in NHIS and MEPS.

<sup>11</sup> A small proportion of DUs contain more than one family unit.

<sup>12</sup> A reference person is defined for each family in MEPS and each DU in NHIS. See glossary for definitions of reference persons in NHIS and MEPS.

<sup>13</sup> Because undercoverage was determined to be negligible for these persons, no poststratification adjustments were applied to this group in order to avoid unnecessary inflation of the variance of the weights.

carried out using the March CPS control totals to yield the final Round 1 person weight.<sup>14</sup> The variables used in the poststratification are: region, race/ethnicity, sex, and age category (see Appendix B.4).

The derivation of the final Round 1 person weight for person  $k$  ( $WTPR_{1k}$ ) can be expressed as follows:

$$\begin{aligned}
 WTPR_{1k} &= WTPR_{0k} \times P_{1c}, \quad k \in c \\
 &\text{with } WTPR_{0k} = WTFM_{1j} \quad (k \in j) \\
 &\text{and } P_{1c} = 1 \quad \text{if } k \in g_1 \\
 &= \frac{\left( CTPR_c - \sum_{k \in c \cap g_1} WTPR_{0k} \right)}{\sum_{k \in c \cap g_2} WTPR_{0k}} \quad \text{if } k \in c \cap g_2
 \end{aligned}$$

where,  $WTPR_{0k}$  is the base weight of person  $k$  who belongs to family  $j$ <sup>15</sup>,  $P_{1c}$  is the adjustment factor for cell  $c$ . For group  $g_1$ , since no poststratification adjustment is applied,  $P_{1c} = 1$ . For group  $g_2$ , the numerator of the adjustment factor is obtained by subtracting the weighted total for group  $g_1$  from the overall control total  $CTPR_c$  for both groups in cell  $c$  i.e., an estimated control total for group  $g_1$  in cell  $c$ , and the denominator of the adjustment factor is the sum of weights of persons in group  $g_2$  in cell  $c$ . Similar to the family weight, outlier weights are trimmed and the poststratification process is repeated if any trimming is done.

### 2.1.2 Preceding Panel (Round 3)

**a. Initial Person Weight for Round 3.** The initial person-level weight assigned to Round 3 respondents is the Full Year (FY) person weight from the previous year ( $WTPR_{FY_1}^*$ ) (see Section 3.1.1 below for general description of how this weight is derived); an asterisk is used to denote that the weight is for the prior year of the preceding panel (now in its second year). Persons in the FY file who were not in-scope during the Round 3 portion of the second year are not included in PIT estimation.

**b. Poststratification Adjustment to Compute Final Round 3 Person-Level Weight.** A poststratification adjustment of the initially assigned person weight from step c above is carried out using the current year March CPS control totals for the U.S. civilian noninstitutionalized population to yield the final Round 3 person weight. The variables

<sup>14</sup> This poststratification focuses on the portion of the U.S. population where undercoverage was determined to be most problematic.

<sup>15</sup> Responding members of nonresponding families (i.e., families without response for all members) are assigned their final DU weight as base weight.

used in forming poststratification cells are sex, age category, race/ethnicity, and Census region (see Appendix B.5). The poststratified weight for person  $k$  can be expressed as:

$$WTPR_{3k}^* = WTPR_{FY_1k}^* \times P_{3c} \quad \text{with } P_{3c} = \frac{CTPR_c}{\sum_{k \in c} WTPR_{FY_1k}^*}, \quad k \in c,$$

where  $WTPR_{3k}^*$  is the poststratified Round 3 final weight for person  $k$  who belongs to poststratification cell  $c$ ,  $WTPR_{FY_1k}^*$  is the FY weight for person  $k$  from the previous year as discussed above,  $P_{3c}$  is the poststratification adjustment factor, and  $CTPR_c$  is the March CPS control total for cell  $c$ . As usual after any poststratification adjustment, outlier weights are trimmed and the poststratification adjustment is repeated if any trimming is done.

**c. Poststratification Adjustment to Form Final Round 3 Family-Level Weight.** First, a family-level initial weight is assigned using the person-level weight (from step 2.1.2b above) of the reference person of the family. Then a poststratification adjustment at the family level is carried out using the March CPS family-level control totals to yield the final Round 3 family weight. The variables used in the poststratification are family type, race/ethnicity of the reference person, region, MSA status, age category of the reference person, and number of eligible members in the family (see Appendix B.3).

The family-level weight for Round 3 can be expressed as follows:

$$WTFM_{2j}^* = WTPR_{3k}^*, \quad k \in j$$

where  $WTFM_{2j}^*$  is the initial weight for family  $j$  and  $WTPR_{3k}^*$  is the weight of person  $k$  from the previous step, where person  $k$  is the reference person of family  $j$ . Then the poststratified family weight for Round 3 can be expressed as:

$$WTFM_{3j}^* = WTFM_{2j}^* \times F_{3c}^* \quad \text{with } F_{3c}^* = \frac{CTFM_c}{\sum_{j \in c} WTFM_{2j}^*}, \quad j \in c$$

where  $F_{3c}^*$  is the poststratification adjustment factor for cell  $c$  and  $CTFM_c$  is the family-level control total for cell  $c$ . After the adjustment, extremely large weights are trimmed and the poststratification process is repeated if any trimming is done.

Similar to Round 1, only families that meet the following conditions are assigned an initial weight and included in poststratification adjustment to produce the final Round 3 family weight:

- at least one Key in-scope person in the family during the second year portion of Round 3,
- the reference person in the family is Key, and
- all Key in-scope family members have a positive person weight.



## 2.2 Development of Combined-Panel PIT Weights

### 2.2.1 Person-Level

#### *a. Assigning Compositing Factors to Panel-Specific Weights in the Combined Panel.*

The panel-specific files are put together to create a combined person-level file of both panels. A compositing factor is applied to the weights of each individual panel to derive the weight for the combined panel. The compositing factor is calculated so that it reflects the number of respondents in each individual panel relative to the total number of respondents in both panels combined. The PIT composite weight for person  $k$  in the combined panel can be expressed as follows:

$$\begin{aligned} PERWT_{0k} &= WTPR_{1k} \times \lambda && \text{if } k \in \text{recent panel} \\ &= WTPR_{3k}^* \times (1 - \lambda) && \text{if } k \in \text{preceding panel} \end{aligned}$$

where  $WTPR_{1k}$  is the final Round 1 person weight from the recent panel,  $WTPR_{3k}^*$  is the Round 3 final person weight from the preceding panel, and  $\lambda$  is the compositing factor which is the proportion of total persons in the combined panels who came from the recent panel. In recent years, the values of  $\lambda$  generally ranged from around 0.45 to 0.55.

*b. Poststratification Adjustment to Produce Final Person-Level PIT Weight.* A poststratification adjustment is applied to the composite person weight in the combined panel by using the March CPS control totals. The variables used for this poststratification are sex, race/ethnicity, age category, region, and MSA status (see Appendix B.5). The poststratified final PIT person weight ( $PERWT_{PITk}$ ) for person  $k$  can be expressed as:

$$PERWT_{PITk} = PERWT_{0k} \times P_{PITc} \text{ with } P_{PITc} = \frac{CTPR_c}{\sum_{k \in c} PERWT_{0k}}, k \in c,$$

where  $P_{PITc}$  is the poststratification adjustment factor and  $CTPR_c$  is the person-level control total based on the March CPS for poststratification cell  $c$  that includes person  $k$ . After the adjustment, outlier weights are trimmed and the poststratification process is repeated if any trimming is done. This poststratified weight is named WGTSP13 in the PUF.

### 2.2.2 Family-Level

#### *a. Assigning Compositing Factors to Panel-Specific Weights in the Combined Panel.*

Similar to the person level, the panel-specific family-level datasets are put together to create a family-level dataset of responding families of both panels combined. The same compositing factor used for the person-level is applied to the weights of each individual panel to derive the family-level weight for the combined panel. The PIT composite weight for family  $j$  in the combined panel can be expressed as:

$$FAMWT_{0j} = WTFM_{1j} \times \lambda \quad \text{if } j \in \text{recent panel}$$

$$= WTFM_{3j}^* \times (1 - \lambda) \quad \text{if } j \in \text{preceding panel}$$

where  $WTFM_{1j}$  is the Round 1 family weight from the recent panel,  $WTFM_{3j}^*$  is the Round 3 family weight from the preceding panel and  $\lambda$  is the compositing factor as derived for the person-level compositing.

**b. Poststratification Adjustment to Produce Final Family-Level PIT Weight.** A poststratification adjustment is applied to the family weight in the combined panel to the same March CPS family-level control totals used for the individual panel-specific family weights. Similar to the person-level poststratification adjustment described above, the cells available for the combined sample are more comprehensive and refined due to the increased sample size. The variables used for the poststratification are family type, race/ethnicity of the reference person, region, MSA status, age category of the reference person, and number of eligible members in the family (see Appendix B.3). The poststratified final PIT family weight ( $FAMWT_{PITj}$ ) for family  $j$  can be expressed as:

$$FAMWT_{PITj} = FAMWT_{0j} \times F_{PITc} \quad \text{with } F_{PITc} = \frac{CTFM_c}{\sum_{j \in c} FAMWT_{0j}}, \quad j \in c,$$

where  $F_{PITc}$  is the poststratification adjustment factor and  $CTFM_c$  is the family-level control total based on March CPS for cell  $c$  that includes family  $j$ . Also, at this stage, extremely large weights are trimmed and the poststratification process is repeated if any trimming is done.

This family-level weight is assigned to all members of the family who are eligible to receive a family weight and is named WGTRU13 in the PUF.

### **2.3 General Issues Related to Developing PIT Weights**

**a. Cell Collapsing.** The initial nonresponse or poststratification cells created in different steps are collapsed in some instances to avoid very small cell sizes or large adjustment factors. Generally, a cell size of at least 20 in the MEPS file and a cell size of at least 100 in the control total file (e.g., CPS) are considered acceptable for nonresponse or poststratification adjustments. In addition, cells with adjustment factors greater than two times the average adjustment factor are usually either truncated or collapsed with a neighboring cell.

**b. Weight Trimming.** The distribution of weights is examined after each round of poststratification adjustments to determine whether there are any inordinately large values. Then weights with inordinately large values are trimmed in a manner designed to reduce overall MSE (i.e., slight increases in bias more than offset by reductions in variance). If any trimming is done then the poststratification/raking procedure is repeated.

*c. Zero Weighted Cases.* Most persons in MEPS public use files have a positive value for both the person and family weight variables. However, a small proportion of cases are assigned a value of zero for either the person or family weight. The conditions when sample persons are assigned a value of 0 for the person or family-level weight are described below:

***Person Weight.*** A person is assigned a person weight of zero (but a positive family weight) if the following two conditions are met: (1) the person is either non-Key or in the military for the entire period but living at home; and (b) the person is a member of a family that has been assigned a positive family-level weight.

***Family Weight.*** A person is assigned a family weight of 0 (but a positive person weight) when either: (1) the person is a member of a family in which at least one Key in-scope member does not have a positive person weight (due to nonresponse); or (b) the reference person of the family is non-Key<sup>16</sup>.

Appendix A provides summary distributions for 2007 of selected interim and PUF weights discussed in this report.

---

<sup>16</sup> This removes the issue of multiple chances of selection being associated with a family.

### 3.0 Full Year (FY) Weights

MEPS full year data are released in two phases: a preliminary full year "Population Characteristics File" (a.k.a. Use File) is released first which is superseded a few months later by a final full year "Consolidated File" (a.k.a. Expenditure File). The preliminary file contains all annual survey data except medical expenditure and income/tax-related variables which require a longer time to process and thus are included in the final consolidated file. Moreover, the weights are not identical in the two full year files because income data that are not available at the initial release are used to refine the weight variables for the final release. More specifically, an adjustment using poverty status is incorporated into the final Consolidated File weight. The FY files allow analysts to produce person-level estimates for the U.S. civilian noninstitutionalized population (i.e., the in-scope or target population) at any time during the year and/or slightly more restricted cross-sectional estimates for the U.S. civilian noninstitutionalized population on December 31. In addition, weights are included in the FY Consolidated File that can be used to produce family-level estimates based on two slightly different definitions of family units (CPS versus MEPS).

Full year weights are the result of a composite of several factors as follows:

- the previously constructed PIT weight (see Section 2 above),
- adjustments for person level nonresponse (survey attrition), and
- raking/poststratification adjustments.

Weights are constructed separately for the two overlapping individual panels/rounds and then combined into a final weight for the PUF. The steps used to develop the FY person and family-level weight variables are illustrated in Figures 3 and 4 respectively and described in details below.

#### 3.1 Development of Panel-Specific FY Weights

##### 3.1.1 Most Recent Panel (Rounds 1–3)

**a. Initial Weight.** The final Round 1 person-level weight used to develop the PIT weight (see Section 2.1.1 above) is assigned as the initial weight for the FY person weight. Therefore, the initial FY weight for person  $k$  is  $WTPR_{1k}$ , where  $WTPR_{1k}$  is the final Round 1 weight for person  $k$ .

**b. Adjustment for Person-Level Nonresponse.** The initial weight is adjusted to compensate for person-level nonresponse over Rounds 2 and 3 (referred to as year 1 nonresponse). Only those individuals who are Key and were ever *in-scope* during the year are included in developing the adjustment factor. The respondents are those individuals who responded for their entire period of eligibility over Round 2 and the year 1 portion of Round 3, and the nonrespondents are those who did not respond for some

part of their eligibility over Round 2 and the year 1 portion of Round 3. To form nonresponse adjustment cells, a CHAID<sup>17</sup> analysis is carried out using a set of potential predictor variables. The set of 11 predictor variables used as input to the CHAID analysis to adjust for this year 1 attrition is detailed in Appendix B.7.

The nonresponse adjustment is applied to the weights of the responding persons in each adjustment cell  $c$  as follows.

$$WTPR_{2k} = WTPR_{1k} \times P_{2c} \quad \text{if person } k \text{ is a respondent in R2/R3}$$

$$\text{with } P_{2c} = \frac{\sum_{k \in c} WTPR_{1k}}{\sum_{k \in c \cap r} WTPR_{1k}}, \quad k \in c,$$

where  $WTPR_{2k}$  is the nonresponse adjusted weight for person  $k$  who belongs to nonresponse adjustment cell  $c$ ,  $P_{2c}$  is the nonresponse adjustment factor for cell  $c$ ,  $r$  represents the set of responding persons, and therefore,  $c \cap r$  represents the set of responding persons in cell  $c$ . The adjustment factor is the ratio of the sum of weights of all persons over the sum of the weights of the responding persons in cell  $c$ .

**c. Person-Level Raking/Poststratification Adjustments.** This step depends on the in-scope status of a person on December 31 as follows:

- i. Raking Adjustment for Persons In-Scope on December 31.** The nonresponse adjusted weights of all Key in-scope responding persons on December 31 (about 99 percent of cases) are raked to December 31 control totals estimated based on the subsequent March CPS estimates of the U.S. civilian noninstitutionalized population and Census Bureau's population estimates for January 1. The raking dimensions used in the adjustment are based on various combinations of the following variables: race/ethnicity, sex, census region, MSA status, and age category (see Appendix B.6).
- ii. Poststratification Adjustments for Persons Not In-scope on December 31.**<sup>18</sup> The small proportion of cases who are not in-scope at the end of the year (total of only about 1 percent) are adjusted as follows:
  - Decedents.** A special poststratification is applied to the weights of respondents who died during the target year using control totals derived from vital statistics data that have been adjusted to eliminate estimated deaths among nursing home residents.<sup>19</sup> Separate decedent control totals are used for persons under age 65 and persons age 65 and over.
  - Nursing Home Entrants.** A second special poststratification adjustment is applied to the weights of those who entered a nursing home prior to December 31

<sup>17</sup> See glossary for a brief description of CHAID.

<sup>18</sup> The purpose of this poststratification is to compensate for potential underrepresentation of high medical expenditure groups.

<sup>19</sup> Adjustment based on data from the Medicare Current Beneficiary Survey.

and who were not members of the U.S. civilian noninstitutionalized population on December 31 using an estimated control total.<sup>20</sup>

**Other Not In-scope Persons on December 31.** For out-of-scope persons on December 31 who do not belong to the two special out-of-scope groups defined above (decedents and nursing home entrants), the FY weight is set equal to their nonresponse adjusted weight without any further adjustment.

The FY weight for person  $k$  in year 1 in the recent panel can be expressed as:

$$WTPR_{FY_1k} = WTPR_{2k} \times P_{FY_1k}$$

where,

- $P_{FY_1k}$  = raking adjustment factor<sup>21</sup> for person  $k$  who was in-scope on December 31
- = poststratification adjustment factors for the two special out-of-scope groups i.e., decedents and nursing home entrants
- = 1 for other out-of-scope persons on December 31.

The distribution of the raked/poststratified weights is examined to identify and trim extremely large values and the raking/poststratification process is repeated if any trimming is done.

### 3.1.2 Preceding Panel (Rounds 3–5)

**a. Initial Weight.** The year 1 full year person weight derived for this panel in the previous year after Rounds 1–3 is assigned as the initial weight for the second year for the panel. See Section 3.1.1 above for the derivation of year 1 FY weight for the recent panel. Therefore, the initial weight for person  $k$  at this step can be denoted by

$WTPR_{FY_1k}^*$ , an asterisk used to indicate the same weight for the preceding panel.

**b. Adjustment for Person-Level Nonresponse.** The initial weight from above is adjusted to compensate for year 2 nonresponse over Rounds 4 and 5. A CHAID analysis is again used to form nonresponse adjustment cells based on a set of potential predictor variables (see Appendix B.8 for list of variables). Since the nonresponse adjustment procedure is the same as described above for the other panel, without repeating the detailed steps here the nonresponse adjusted weight for person  $k$  in this panel will be denoted as  $WTPR_{4k}^*$ .

**c. Person-Level Raking/Poststratification Adjustments.** As for the other panel, this step depends on the in-scope status of a person on December 31 as follows:

- i. Raking Adjustment for Persons In-Scope on December 31.** The nonresponse adjusted weights of all Key in-scope responding persons on December 31 (about 99 percent of cases) are raked to December 31 control totals estimated based on the subsequent March CPS estimates of the U.S. civilian noninstitutionalized

<sup>20</sup> This control total is based on data from the 1996 MEPS Nursing Home component.

<sup>21</sup> Factors are derived using an iterative raking algorithm.

population and Census Bureau's population estimates for January 1. The raking dimensions used in the adjustment are based on various combinations of the following variables: race/ethnicity, sex, census region, MSA status, and age category (see Appendix B.6)<sup>22</sup>.

**ii. Poststratification Adjustments for Persons Not In-scope on December 31<sup>23</sup>.**

As for the other panel, special poststratification adjustments are applied to the weights of decedents and nursing home entrants and no adjustment is applied to the weights of the remaining persons who were not in-scope on December 31 (see section 3.1.1c (ii) above).

Therefore, the FY weight for person  $k$  in year 2 of the preceding panel can be expressed as:

$$WTPR_{FY_2, k}^* = WTPR_{4k}^* \times P_{FY_2, k}^*$$

where,

- $P_{FY_2, k}^*$  = raking adjustment factor for person  $k$  who was in-scope on December 31
- = poststratification adjustment factors for the two special out-of-scope groups i.e., decedents and nursing home entrants
- = 1 for other out-of-scope persons on December 31.

As for the most recent panel, the distribution of the raked/poststratified weights is examined to identify and trim extremely large values and the raking/poststratification process is repeated if any trimming is done.

### **3.2 Development of Combined-Panel Person Weights<sup>24</sup> for FY Population Characteristics (Use) File**

**a. Assigning Compositing Factors to Panel-Specific Weights in the Combined Panel.**

The panel-specific FY files are put together to create a combined person-level FY file and a compositing factor is applied to the panel-specific FY weight of each individual panel to derive the FY weight for the combined panel (a.k.a. use file weight). The compositing factor is calculated so that it reflects the number of respondents for each individual panel relative to the total number of respondents in both panels combined. The FY composite weight for person  $k$  in the combined panel can be expressed as:

$$\begin{aligned} PERWT_{1k} &= WTPR_{FY_1k} \times \theta && \text{if } k \in \text{recent panel} \\ &= WTPR_{FY_2k}^* \times (1 - \theta) && \text{if } k \in \text{preceding panel} \end{aligned}$$

<sup>22</sup> Note that poverty status is not included here but is included in raking for the final FY Consolidated File.

<sup>23</sup> The purpose of this poststratification is to compensate for potential underrepresentation of these high medical expenditure groups.

<sup>24</sup> A family-level weight is not constructed for the FY Population Characteristics file.

where  $WTPR_{FY_1k}$  is the FY weight for year 1 for the recent panel,  $WTPR_{FY_2k}^*$  is the FY weight for year 2 for the preceding panel, and  $\theta$  is the compositing factor—which is the proportion of total persons in the combined panels who came from the most recent panel. In recent years, the values of  $\theta$  generally ranged from around 0.45 to 0.55.

**b. Raking Adjustment of the FY Composite Weight.** The resulting composite person weights (excluding decedents and nursing home entrants) are raked to the same December CPS based control totals used for the individual panel FY weights. The raking dimensions used in this adjustment are based on various combinations of the following variables: race/ethnicity, sex, Census region, MSA status, and age category (see Appendix B.6).<sup>25</sup> The decedents and nursing home entrants are separately poststratified as described in Section 3.1.1.c above. These adjustments produce the PUF variable  $PERWT_{yyP}$  i.e., preliminary FY weight.

Since this step is the same as in Section 3.1.1.c, without going into the details, the preliminary FY weight for person  $k$  in the combined panel can be expressed as:

$$PERWT_{PFYk} = PERWT_{1k} \times P_{PFYk}$$

where  $P_{PFYk}$  is the raking/poststratification adjustment factor. The distribution of the adjusted weight is checked for outlier weights and the raking procedure is repeated if any weight trimming is required. This weight is called preliminary FY (PFY) weight because the poverty status is not included in the raking adjustment yet.

Note that no family weight is produced for the FY Population Characteristics File.

### 3.3 Development of Final FY Weights for Consolidated File

#### 3.3.1 Person-Level

To produce the final person-level FY consolidated file weight (a.k.a. expenditure file weight), the person-level weights produced for the Population Characteristics File are re-raked using poverty status<sup>26</sup> in addition to the same set of variables used before (see above and Appendix B.6) corresponding to the estimated control totals for December 31. Persons who are out of scope on December 31 are excluded from this raking so preliminary FY weights for these persons are carried forward from the previous step without any further adjustment. This produces the PUF variable  $PERWT_{yyF}$ . Therefore, the final FY weight for person  $k$  can be expressed as:

$$\begin{aligned} PERWT_{FYk} &= PERWT_{PFYk} \times P_{FYk} && \text{if } k \in \textit{inscope} \text{ on Dec. 31} \\ &= PERWT_{PFYk} && \text{if } k \in \textit{outscope} \text{ on Dec. 31} \end{aligned}$$

<sup>25</sup> Poverty status is not available at the time of development of the Population Characteristics file, but is incorporated in raking for the final weight in the subsequent FY Consolidated File.

<sup>26</sup> The poverty status categories are shown in Appendix B.6.



with  $P_{FYk}$  represents the complete iterative raking adjustment factor for person  $k$ . The distribution of the final weight is checked, and outlier weights, if any, are trimmed followed by reapplication of the raking procedure.

### 3.3.2 Family-Level

Two final family-level weights are derived for the FY consolidated file; one based on the CPS definition of a family (PUF variable FAMWTyyC) and the other based on the MEPS definition of a family (PUF variable FAMWTyyF). In addition to the difference in family definitions, there is also a difference in temporal scopes of these two weights. The CPS family weight is derived to produce estimates only for those families in-scope on December 31 of the year while the MEPS family weight is derived to produce estimates for all in-scope MEPS families that existed any time during the year. In other words, the MEPS family weight applies to families in-scope on December 31 plus the families whose members became out-of-scope prior to the end of the year.<sup>27</sup>

While the CPS and MEPS definitions of family units are identical in most instances, unmarried partners who identify themselves as a single family unit are considered to be a single family in MEPS but represent two separate families in the CPS. In addition, foster children are considered to be family members in MEPS but not in the CPS. The control totals for a family-level raking adjustment are obtained from the CPS and represent families in existence on December 31 of a given year. CPS-like families are formed from the MEPS families in existence on December 31. For those MEPS families that are split to form multiple CPS-like families, each CPS-like family is assigned a reference person. The initial CPS family weight assigned is the person weight of the reference person of the CPS-like family. Then, raking adjustments are applied to this initial family weight to obtain the final CPS family weight. For those MEPS families in-scope on December 31, the raked CPS family weight of the MEPS reference person is assigned as the final MEPS family weight. For persons in the small number of MEPS families that are out-of-scope at the end of the year, the person-level weight of their reference persons serves as their final MEPS family weight.

Only families that meet the following conditions are assigned an initial weight and included in poststratification adjustment to produce final FY family weights:

- at least one Key in-scope person in the family during the year,
- the reference person in the family is Key, and
- all Key in-scope family members have a positive person weight.

The following steps are used to derive the CPS and MEPS family weights:

**a. *Initial CPS Family Weight.*** After forming the CPS families, a family-level initial weight is assigned to all responding CPS families using the person-level weight of the

---

<sup>27</sup> Estimates can be restricted to MEPS families in-scope on December 31 if appropriate for analytic purposes.

reference person of the CPS family. Therefore, the initial family-level weight for the CPS family  $j_s$  can be defined as:

$$FAMWT_{1j_s} = PERWT_{FYk} \quad k \in j_s$$

where  $PERWT_{FYk}$  is the weight of person  $k$  who is the reference person of CPS family  $j_s$ .<sup>28</sup>

**b. Raking Adjustment to Produce CPS Family Weight.** A family-level raking adjustment, using various combinations of family type, race/ethnicity of the reference person, poverty status, region, MSA status, family size, and age category of the reference person as raking dimensions (see Appendix B.9), is then applied to the initial CPS family weight. The raking adjustment is applied only to the eligible CPS families (i.e., those with a Key reference person) that are in-scope on December 31 of the year. The control totals at the family level for December 31 are derived by calibrating the estimates obtained from the March CPS of the following year to December 31 using an estimated monthly average growth rate for the total population of families.

The raked family-level FY weight for the CPS family  $j_s$  can be expressed as:

$$FAMWT_{FYj_s} = FAMWT_{1j_s} \times F_{FYj_s}$$

where  $F_{FYj_s}$  is the raking adjustment factor for CPS family  $j_s$ . The raking procedure is repeated if any trimming of large weights is necessary.

This family-level FY weight is assigned to all persons within a CPS family who are eligible to receive a family weight<sup>29</sup>. All persons in families not in-scope on December 31 or persons in families with a non-Key reference person are assigned a zero CPS family weight<sup>30</sup>. This produces the PUF variable FAMWTyyC.

**c. Producing MEPS Family Weight.** Finally, the MEPS family weight at the person-level is assigned based on the MEPS definition of families. This is done by assigning the CPS family weight of the reference person of the MEPS family (as derived above) to all members of the MEPS family, which becomes FAMWTyyF. If the MEPS family is not in-scope on December 31 then the person-level weight (as derived in section 3.3.1) of the reference person of the MEPS family is assigned to all members of the MEPS family.

<sup>28</sup> For a small number of cases, when the weight of the reference person is not positive, the weight of the spouse of the reference person or the weight of the oldest member in the family is used.

<sup>29</sup> Family weights are assigned to all members of responding families whose reference person is Key whether the individual member is Key or not or in-scope or not (e.g., a person in the military living with a Key and in-scope family member is assigned a family weight).

<sup>30</sup> Sometimes partitioning of a MEPS family results in a CPS family where the reference person is non-Key.

### **3.4 General Issues Related to Developing FY Weights**

The general issues discussed in relation to the development of PIT weights in Section 2.3 also apply to the development of FY weights. In other words, the same principles are used for cell collapsing, weight trimming after each raking/poststratification adjustment, and assigning zero weights to persons and families. However, raking procedures are not used in the PIT weighting procedures but are used in the development of FY weights. Generally, a cell size of at least 100 in the MEPS file and a cell size of at least 500 in the control total file (e.g., CPS) for a category within a raking dimension are considered acceptable in the MEPS raking procedures.

Appendix A provides summary distributions for 2007 of selected interim and PUF weights discussed in this report.

## 4.0 Weights for Survey Supplements

In addition to the person and family-level weight variables included on FY files, two additional person-level weight variables are included for appropriate estimation using data collected in the MEPS Self-Administered Questionnaire (SAQ) or Diabetes Care Supplement (DCS). These supplements are self-administered paper and pencil questionnaires that are fielded in selected rounds of the survey to all adults and to adults identified as having diabetes, respectively. In general, the weight variables for these supplements are used in lieu of the FY person-level weight when the analysis involves data from the supplement.<sup>31</sup> As for the general survey FY weights described in Section 3.0 above, preliminary weights for supplements are provided on the preliminary FY file while final weights for supplements are provided on the final FY file. The following are overviews of procedures for developing these special supplement weight variables.

### 4.1 Self-Administered Questionnaire (SAQ)

An SAQ questionnaire is requested to be completed by each adult (persons aged 18 and older) family member in Round 2 (Panel X) and Round 4 (Panel X-1) to obtain additional information for measures of adult health status and health care quality<sup>32</sup>. Thus, the target population for the SAQ is the adult civilian, noninstitutionalized population at the time data were collected for Rounds 2/4. Following is a summary of the procedure for developing weights to be used when producing estimates of data collected in the SAQ:

- a. The SAQ weight is developed starting with initial weights used for developing the panel-specific FY weight for each panel (see Sections 3.1.1.a and 3.1.2.a). Based on these panel-specific initial weights, a composite weight is developed by applying the same factors used for combining the panel-specific FY weights (see Section 3.2.a). This composite weight is then adjusted for both nonresponse over the year<sup>33</sup> and to the SAQ. Variables used in the nonresponse adjustment process are region, MSA status, family size, marital status, level of education, health status, health insurance status, age category, sex, and race/ethnicity (see Appendix B.10).
- b. The nonresponse adjusted SAQ weight in step 1 above is raked to the CPS estimates corresponding to December of the analytic year (the same source of control figures used for the full year person weights) to produce the preliminary SAQ weight (SAQWTyyP) for the FY Population Characteristics File. The variables used in the raking adjustment of preliminary SAQ weights are region, MSA status, age category, sex, and race/ethnicity, as were used for the preliminary FY person weights. The only difference is that age categories are developed after excluding ages under 18, since only adults were eligible for the SAQ.

---

<sup>31</sup> The DCS weight would generally be applied if the analysis involves data from both the SAQ and DCS supplements.

<sup>32</sup> The SAQ includes questions from the Consumer Assessment of Health Plans (CAHPS®), the SF-12, the EuroQol 5D, and attitude items.

<sup>33</sup> Rounds 2–3 for the year 1 panel and Rounds 4–5 for the year 2 panel.

- c. For the final FY SAQ weight, the weight in step b above is raked again to CPS estimates based on the five variables mentioned in step b as well as poverty status to produce the final SAQ weight (SAQWTyyF) for the final FY consolidated data file.

As usual, for both preliminary and final SAQ weighting, after raking the distribution of the weights is checked for outliers and the raking procedure is repeated if any weight trimming is implemented.

## **4.2 Diabetes Care Supplement (DCS)**

The DCS is fielded during Round 3 (Panel X) and Round 5 (Panel X-1) to collect a variety of measures from adults reported to have been told by a health professional that they had diabetes. All adults who reported on the CAPI questionnaire to have been diagnosed with diabetes are administered the DCS questionnaire. Following are the steps used to construct the DCS weight for use when producing estimates for data collected in the DCS:

- a. To produce the DCS weight for the preliminary FY file, the initial weight assigned is the SAQ preliminary weight derived for the FY preliminary file (SAQWTyyP—see Section 4.1 above).
- b. The DCS initial weight from (a) above is then adjusted for nonresponse to the main CAPI diabetes question. The variables used in CHAID analysis as potential predictors of response propensity to form the nonresponse adjustment classes include region, MSA status, family size, marital status, level of education, health status, health insurance status, age category, sex, race/ethnicity. This is the same set of variables used for SAQ nonresponse adjustment as presented in Appendix B.10.
- c. Finally, the weight from (b) above is adjusted for nonresponse to the DCS questionnaire (generally about 10%) to produce the DCS weight for the preliminary FY file (DIABWyyP). The nonresponse adjustment classes for this stage are formed using race/ethnicity, sex, and age category. Note that the age categories (age at the date of the interview) used differ from those in stage (i) above. These are: 18–29; 30–44; 45–64; 65+.
- d. To produce the final DCS weight for the FY consolidated file (DIABWyyF), steps a to c above are repeated but substituting the final SAQ weight (SAQWTyyF) as the initial weight in step a.

For both preliminary and final DCS weighting, the nonresponse adjusted weights are checked for outliers and a weight trimming procedure is applied as necessary, followed by an adjustment of the sum of weights.

## 5.0 Longitudinal (L) Weights

In contrast to the PIT and FY files which include persons from two consecutive overlapping panels, the persons included in a longitudinal data file are from one specific sample panel and represent those who were in the MEPS population (U.S. civilian noninstitutionalized) for all or part of a given two-year period (e.g., Panel 11: 2006–2007). Although data are available for all five rounds for more than 90% of the cases in most longitudinal files, persons who were born, died, were in the military or an institution, or left the country during the two-year period do not have data for one or more rounds. In contrast, persons in the panel who participated in the survey only for part of the period they were in-scope are treated as nonrespondents and not included in this file. The longitudinal weight variable (described below) reflects adjustments for this attrition (nonresponse).

To derive the longitudinal person-level weight, a panel is divided into three non-overlapping groups as follows:

- a. All persons in their year 2 FY file who are in-scope on December 31<sup>st</sup> of year 2 or who are out-of-scope (OOS) due to death or entering a nursing home on December 31<sup>st</sup> of year 2.
- b. All persons in their year 2 FY file who are OOS on December 31<sup>st</sup> of year 2 for reasons other than death or entering a nursing home.
- c. All persons in their year 1 FY file from the panel who are OOS on December 31<sup>st</sup> of year 1 and not part of the year 2 FY file.

The longitudinal weight for person  $k$  is then assigned as follows:

$$\begin{aligned}
 LONGWT_k &= PERWT_{FY_2,k} && \text{if } k \in a \\
 &= PERWT_{FY_2,k} \times L_k && \text{if } k \in b \\
 &= PERWT_{FY_1,k} && \text{if } k \in c
 \end{aligned}$$

where  $a, b, \& c$  are three groups as defined above,  $PERWT_{FY_2,k}$  is the panel-specific final poverty adjusted year 2 annual weight for person  $k$ , and  $PERWT_{FY_1,k}$  is the panel-specific final poverty adjusted year 1 annual weight for person  $k$ . Both these weights<sup>34</sup> have been adjusted to compensate for survey attrition during the year.  $L_k$  is an adjustment factor designed to make the sum of the longitudinal weights of the persons in group  $b$  equal to the sum of the combined panel FY person weights of similar OOS persons from both panels in the year 2 FY file.<sup>35</sup>

<sup>34</sup> Panel-specific annual weights are not included on PUFs.

<sup>35</sup> This adjustment was approximately 11% for Panel 11.

The adjustment factor  $L_k$  for person  $k$  can be expressed as:

$$L_k = \frac{\sum_{k \in oos} PERWT_{FYk}}{\sum_{k \in oos} PERWT_{FY_2k}}$$

where  $\sum_{k \in oos} PERWT_{FYk}$  is the sum of combined panel FY weights of persons in both panels who are OOS on December 31<sup>st</sup> for reasons other than death or entering a nursing home and  $\sum_{k \in oos} WTPR_{FY_2k}$  is the sum of panel-specific year 2 FY weights of persons in group b of the longitudinal panel only. In other words, the numerator includes all OOS persons as specified above from both panels in the FY file and the denominator includes the same OOS persons from the longitudinal panel only.

Beginning with panel 14, the adjustment factor  $L_k$  will no longer be applied to group  $b$  (i.e.,  $L_k = 1$  will be used in the above expression for deriving the longitudinal panel weight).

## 6.0 Variance Estimation

The MEPS-HC data is collected using a complex multistage sample design that involves stratification, clustering, and unequal selection probabilities. Unlike the data obtained through a simple random sample design where all observations are independent with equal sampling weights, the MEPS data are correlated due to clustering and have unequal weights. Moreover, the MEPS weighting procedures employ differential adjustments for nonresponse, coverage, and poststratification/raking. These sample design and estimation complexities require special consideration in computing variances/standard errors of MEPS estimates. Therefore, a variance estimation method under the assumption of a simple random sample as employed by the most commonly used statistical packages is not appropriate for MEPS. These software packages will usually underestimate the variance of MEPS estimates. To obtain accurate estimates of the variances of MEPS person- or family-level estimates for either descriptive statistics or more sophisticated analyses based on multivariate models, the MEPS sample design complexities must be taken into account using special analysis approaches developed for complex surveys. A review of commonly used approaches to design-based estimation of the variances of estimates from complex survey data can be found in Wolter (2007) and Shao (1996). The two most commonly used variance estimation methods for complex survey data are the Taylor series linearization method and Balanced Repeated Replication (BRR) method. Various software packages that use these approaches to analyze data from complex surveys are SUDAAN, STATA, WESVAR, and SAS Proc Survey procedures.

The Taylor series variance estimation method is most commonly applied when analyzing MEPS data. MEPS annual public use files include the two necessary sample design variables for implementing this method. These variables identify the variance estimation strata (VARSTR) and variance estimation clusters (VARPSU). Specifying these variables in conjunction with a 'with replacement' design in software packages that employ the Taylor series approach (e.g. SUDAAN, STATA, or SAS Proc Survey procedures) will produce variances that reflect the complexities of the MEPS design.

Because it can be extremely difficult to use Taylor series to calculate the variances of complex estimators not readily available in complex survey software packages (e.g., two-part model of health expenditures, ratio between two medians), a linkage file containing a BRR replication structure (in the form of a set of half sample indicators) is also available for variance estimation

([http://www.meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files\\_detail.jsp?cboPufNumber=HC-036BRR](http://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-036BRR)) (accessed September 16, 2010). The file is called 'Replicates for Calculating Variances File' and includes all cases from 1996 to the current year. The half sample indicators in this file (+1 and -1) can be used to form BRR replicate weights to compute variances of MEPS estimates using either BRR or Fay's BRR (Fay 1989) methods.

To facilitate analysis of subpopulations and/or low prevalence events, it may be desirable to pool together more than one year of MEPS-HC data to yield sample sizes large enough



to generate reliable estimates. MEPS-HC samples from year to year are not completely independent because households are drawn from the same sample geographic areas and many persons are sample respondents for two consecutive years (see MEPS-HC Methodology Reports for more details at <http://www.meps.ahrq.gov>). Despite this lack of independence, it is valid to pool multiple years of MEPS-HC data and keep all observations in the analysis because each year of MEPS-HC is designed to be nationally representative. However, to obtain appropriate standard errors when pooling years of MEPS-HC data, it is necessary to insure a variance structure that consistently specifies MEPS geographic sampling units across years.

Starting in 2002, the annual MEPS public use files were released with a common variance structure that allows users to seamlessly pool annual files from 2002 onward. Prior to 2002, however, each annual MEPS public use file was released with a variance structure unique to the particular MEPS sample in that year. Therefore, when one or more years of data being pooled precede 2002, it is necessary to obtain a common variance structure from the Pooled Estimation Linkage File ([http://www.meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files\\_detail.jsp?cboPufNumber=HC-036](http://www.meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-036)) when producing pooled estimates using the Taylor series method. This file provides a common variance structure (i.e., consistent specification of MEPS geographic sampling units) across all years since the inception of MEPS in 1996. In addition, the Replicates for Calculating Variances File provides standardized replicates across all panels to facilitate appropriate BRR variance estimation from pooled data. However, pooling 1999 or 2000 with other years using the BRR file is inadvisable due to inconsistencies resulting from some primary sampling units (PSUs) that were dropped in those years.

## 7.0 References

Botman S.L., Moore T.F., Moriarity C.L., Parsons V.L. Design and Estimation for the National Health Interview Survey, 1995–2004. National Center for Health Statistics. Vital Health Stat 2(130). 2000.

Ezzati-Rice, T.M., Rohde, F., Greenblatt, J., (2008). *Sample Design of the Medical Expenditure Panel Survey Household Component, 1998–2007*. Methodology Report No. 22. March 2008. Agency for Healthcare Research and Quality, Rockville, MD.  
[http://www.meps.ahrq.gov/mepsweb/data\\_files/publications/mr22/mr22.pdf](http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr22/mr22.pdf)

Kass, G. V. (1980). "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of Applied Statistics*, Vol. 29, No. 2 (1980), pp. 119–127.

Korn E. and Graubard B. (1999). *Analysis of Health Surveys*. John Wiley and Sons Inc. New York.

Shao, J. (1996). "Resampling Methods in Sample Surveys (with discussion)." *Statistics*, 27, 203–254.

Wolter, K.M. (2007). *Introduction to Variance Estimation*. 2<sup>nd</sup> Edition. New York: Springer-Verlag.

## Appendix A

### Distribution of Weights at Different Stages of Weighting for 2007 MEPS

Panel/Year	Weight	PUF Variable	Number of Records	Sum of Weights	Minimum Weight	Median Weight	Average Weight	Maximum Weight	CV Weights (%)
Panel 12, 2007	DU Final Weight	-	5,525	116,942,394	4,406	20,345	21,166	116,377	48.6
	Round 1 Family Weight	-	5,736	129,215,974	4,266	20,491	22,527	104,048	54.0
	Round 1 Person Weight	-	14,819	296,056,836	3,152	17,823	19,978	106,987	58.9
Panel 11, 2007	Round 3 Person Weight	-	17,008	296,056,836	679	15,183	17,407	118,825	69.1
	Round 3 Family Weight	-	6,680	129,215,974	1,036	17,208	19,344	109,389	66.2
Panels 11 & 12 Combined, 2007	PIT Person Weight	WGTS13	31,827	296,056,836	329	8,170	9,302	63,811	65.3
	PIT Family Weight	WGTRU13	12,416	129,215,974	417	9,296	10,407	57,102	62.2
Panel 12, 2007	FY Person Weight*	-	13,015	301,170,737	2,338	20,447	23,140	131,419	60.5
Panel 11, 2007	FY Person Weight*	-	16,355	300,972,329	1,154	16,032	18,402	120,178	68.8
Panels 11 & 12 Combined, 2007	FY Person Weight (Preliminary)	PERWT07P	29,370	301,309,149	651	8,994	10,259	67,646	65.4
	FY Person Weight (Final)	PERWT07F	29,370	301,309,149	550	8,902	10,259	67,154	67.7
	FY Family Weight (CPS)	FAMWT07C	11,873	130,346,831	540	9,644	10,978	61,163	65.1
	FY Family Weight (MEPS)	FAMWT07F	11,615	127,885,890	540	9,682	11,010	61,163	64.9
Panel 11, 2006-07	Longitudinal Weight	LONGWT	16,533	304,831,607	958	15,908	18,438	120,341	71.0

\*The weight distributions for these weights do not reflect the special weight adjustments applied to persons not in-scope on December 31 (i.e., for decedents and nursing home entrants).

## **Appendix B**

### **Variables Used in Nonresponse and Poststratification/Raking Adjustments<sup>36</sup>**

#### **B.1 Variables Used in DU-level Poststratification Using NHIS Control Totals**

**1. MSA status:**

- In MSA in the principal city
- In MSA but not in the principal city
- Not in MSA

**2. Race/ethnicity of the DU reference person:**

- Hispanic
- Black, non-Hispanic
- Asian
- Other

**3. Income category of the DU:**

- Income < \$20,000
- Income between \$20,000 to 34,999
- Income > \$34,999
- Income > \$20,000 but exact value is unknown
- Income Unknown

**4. Reported health condition:**

- All members of the DU with good to excellent health
- Other

**5. Employment status of the DU reference person:**

- Employed
- Not employed

---

<sup>36</sup> Although these variables and categories are fairly stable across years, there are some instances in certain years where variables and/or categories used differ slightly.

**B.2 NHIS Variables Used in CHAID Analysis to Form Cells for DU-level NR Adjustment (\* indicates significant variables in 2007)****1. Any Asian in household (indicator for MEPS oversampling):**

- no Asian in the household
- at least one person is Asian in the household

**2. Predicted poverty for household (indicator for MEPS oversampling):**

- no RU in the DU with predicted poverty > 0.3
- at least one RU in the DU with predicted poverty > 0.3

**3. Any Black in household (indicator for MEPS oversampling):**

- Household is not Asian, not poor, at least one Black person
- Otherwise

**4. Age categories for DU reference person:**

- Less than 25 years old
- 25–34 years old
- 35–44 years old
- 45–64 years old
- 65 years or older

**5. Gender of DU reference person:**

- Male
- Female

**6. Race/ethnicity of DU reference person:**

- Hispanic
- Black, non-Hispanic
- Asian (non-Hispanic and non-black)
- Otherwise

**7. Marital status of DU reference person:**

- Married—spouse in household
- Otherwise

**8. Education level of DU reference person:**

- No school
- Elementary
- Some high school
- High school graduate or GED
- Some college
- Bachelor's degree
- Graduate school
- Unknown

**9. Family income of DU reference person:**

- Less than \$20,000 and less than \$20,000 with no detail given
- \$20,000–\$34,999
- \$35,000 or greater
- \$20,000 or higher but no detail given
- Unknown

**10. Employment status of DU reference person:**

- Employed/working
- Unemployed
- Not working
- Refusal, not ascertained, don't know

**11. DU reference person needs help with personal care:**

- Yes
- No or refused, not ascertained, don't know

**12. DU health status:**

- All members of the DU in good to excellent health
- At least one member in fair health, none with poor health
- At least one member in poor health

**13. Number of persons in DU:**

- One person
- Two persons
- Three persons
- Four persons
- Five or more persons

**14. Census region of DU:**

- Northeast
- Midwest
- South
- West

**15. CBSA/MSA status of DU:**

- In principal city of CBSA/MSA
- In CBSA/MSA but not principal city
- Not in CBSA/MSA

**16. CBSA size of DU:**

- In CBSA with population 500,000 or more
- In CBSA with population less than 500,000
- Not in CBSA

**17. Urban/Rural status of DU (1990 Census block designation):**

Urban  
Rural

**18. PSU Type:**

Self-representing  
Non-self-representing

**19. Telephone number status in NHIS:**

Has working telephone and gave phone number  
Has working telephone but refused phone number  
No working phone  
Unknown

**20. Type of home in NHIS:**

House/apt/flat  
Other type of dwelling  
Missing (not ascertained)

**21. Time without telephone in NHIS:**

3 weeks or less  
More than 3 weeks (including no telephone service or no working phone)  
No service interruption  
Missing (refused, not ascertained, don't know)

**22. Interview language in NHIS:**

English only  
Not English only  
Missing (Not ascertained or non-interview household)

**23. U.S. citizen status in NHIS:**

Citizen  
Not citizen  
Missing (Refused, not ascertained, don't know)

**24. Family medical expenses amount in NHIS:**

\$0  
\$1 to \$499  
\$500 to \$1,999  
\$2,000 or more  
Missing (Refused, not ascertained, don't know)

**25. Homeowner status in NHIS:**

Owned or being bought  
Rented or other  
Missing (Refused, not ascertained, don't know)

**26. Born in U.S.:**

Born in U.S.

Not born in U.S.

Missing (Refused, not ascertained, don't know)

**27. Reason did not work last week in NHIS:**

Not working for health reasons or disabled

Not working: Retired

Not working other (looking for work, keeping home, school, maternity leave, vacation, layoff, off-season, other, unknown why not working) or working not for pay

Working for pay or under 18 years old

Missing (refused, not ascertained, don't know)

**28. Number of nights in the hospital last year in NHIS:**

Zero nights

1 to 7 nights

8 or more nights

Missing (Refused, not ascertained, don't know)

**29. Insurance coverage status in NHIS:**

Insured

Uninsured

Missing (refused, not ascertained, don't know)



**B.3 Variables Used in Rounds 1 and 3 PIT Family-Level Poststratification****1. Family type:**

- Reference person married and spouse present
- Male reference person spouse not present
- Female reference person spouse not present

**2. Race/ethnicity of the reference person:**

- Hispanic
- Black, non-Hispanic
- Asian
- Other

**3. Region:**

- Northeast
- Midwest
- South
- West

**4. MSA status:**

- MSA
- Non-MSA

**5. Age categories for the reference person:**

- <35
- 35–44
- 45–64
- 65 or older

**6. Number of eligible members of the family:**

- 1 person
- 2 persons
- 3 persons
- 4 persons
- 5 or more persons

**B.4 Variables Used in Round 1 PIT Person-Level Poststratification****1. Region containing person (assumed to be the same as for family):**

Northeast  
Midwest  
South  
West

**2. Race/ethnicity of person (assumed to be the same as for family):**

Hispanic  
Black, non-Hispanic  
Asian  
Other

**3. Sex of person:**

Male  
Female

**4. Age categories of person:**

under 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  
75–80  
80+

**B.5 Variables Used in PIT Round 3 and PIT Final Poststratification****1. Race/ethnicity of person:**

Hispanic  
Black, non-Hispanic  
Asian  
Other

**2. Sex of person:**

Male  
Female

**3. Region containing person:**

Northeast  
Midwest  
South  
West

**4. MSA status:**

MSA  
Non-MSA

**5. Age categories of person:**

under 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  
75–80  
80+

**B.6 Variables Used in All FY Person Level Rakings****1. Race/ethnicity of person:**

Hispanic  
Black, non-Hispanic  
Asian  
Other

**2. Sex of person:**

Male  
Female

**3. Region containing person:**

Northeast  
Midwest  
South  
West

**4. MSA status:**

MSA  
Non MSA

**5. Age categories of person:**

Under 1  
1–19  
20–29  
30–44  
45–64  
65+

In addition, the following variable is used in the final raking adjustment to produce the FY consolidated person weight.

**6. Poverty status:**

Below poverty  
100–124 percent  
125–199 percent  
200–399 percent  
400 or more percent

**B.7 Variables Used in CHAID Analysis to Form Cells for Year 1 FY Person-Level NR Adjustment (\* indicates significant variables in 2007)****1. Reluctance to respond in Round 1 interview\*:**

RU was reluctant to respond  
Otherwise

**2. Age category\*:**

0–19  
20–29  
30–44  
45–64  
65+

**3. Marital status of family reference person\*:**

Currently married  
Widowed  
Divorced  
Separated  
Never married (includes inapplicable, under 16)

**4. Family size\*:**

1 person  
2 persons  
3 persons  
4 persons  
5 or more persons

**5. MSA status\*:**

MSA  
Non-MSA

**6. Sex\*:**

Male  
Female

**7. Race/ethnicity\*:**

Hispanic  
Black, non-Hispanic  
Asian  
Other

**8. Employment status of reference person\*:**

Employed  
Not employed  
Inapplicable, Unknown

**9. Education level of reference person\*:**

- No degree
- High school degree or GED with no college
- Some college
- College or a higher degree
- Inapplicable, Unknown

**10. Health insurance coverage status of person on interview date\*:**

- Yes
- No
- Inapplicable

**11. Census region\*:**

- Northeast
- Midwest
- South
- West

**B.8 Variables Used in CHAID Analysis to Form Cells for Year 2 FY Person-Level NR Adjustment (\* indicates significant variables in 2007)****1. Reluctance to respond in Round 1 interview\*:**

RU was reluctant to respond  
Otherwise

**2. Age category\*:**

0–19  
20–29  
30–44  
45–64  
65+

**3. Marital status of family reference person\*:**

Currently married  
Widowed  
Divorced  
Separated  
Never married (includes inapplicable, under 16)

**4. Family size\*:**

1 person  
2 persons  
3 persons  
4 persons  
5 or more persons

**5. MSA status\*:**

MSA  
Non-MSA

**6. Sex\*:**

Male  
Female

**7. Race/ethnicity\*:**

Hispanic  
Black, non-Hispanic  
Asian  
Other

**8. Education level of reference person\*:**

No degree  
High school degree or GED with no college  
Some college

College or a higher degree  
Inapplicable, Unknown

**9. Health insurance coverage status of person on interview date:**

Yes  
No  
Inapplicable

**10. Census region\*:**

Northeast  
Midwest  
South  
West

**11. First respondent indicator:**

Yes, First respondent  
No, Not first respondent  
Nonresponse

**12. PROXY, proxy respondent indicator:**

Respondent is an RU member  
Respondent is a proxy  
Nonresponse

**13. Self-perceived health status of the person in year 1\*:**

Excellent  
Very good  
Good  
Fair  
Poor  
Unknown/inapplicable

**14. Total healthcare expenditures of the person in year 1\*:**

\$ 0  
\$ 1 – 300  
\$ 301 – 950  
\$ 951 – 3150  
Over \$ 3150  
Unknown

**15. Number of office-based provider visits of the person in year 1\*:**

0  
1  
2 – 4  
5 +  
Unknown



**B.9 Variables Used in FY Family-level Raking****1. Family type on December 31:**

- Reference person is married, spouse present
- Male reference person with no spouse present
- Female reference person with no spouse present

**2. Race/ethnicity of the reference person:**

- Hispanic
- Black, non-Hispanic
- Asian
- Other

**3. Poverty status:**

- Below poverty
- 100–124 percent
- 125–199 percent
- 200–399 percent
- 400 or more percent

**4. Region:**

- Northeast
- Midwest
- South
- West

**5. MSA status:**

- MSA
- Non MSA

**6. Number of eligible members of the family on December 31:**

- 1 person
- 2 persons
- 3 persons
- 4 persons
- 5 or more persons

**7. Age categories for the reference person:**

- 34 or younger
- 35–44
- 45–64
- 65 or older

**B.10 Variables Used for SAQ Nonresponse Raking Adjustment****1. Region:**

Northeast  
Midwest  
South  
West

**2. MSA status:**

MSA  
Non MSA

**3. Family size:**

1 person  
2 persons  
3 or more persons

**4. Marital status:**

Married  
Widowed  
Divorced/separated  
Never married  
Under age 16  
Missing

**5. Education:**

Less than high school  
High school degree  
Some college  
College degree  
Missing

**6. Health status:**

Excellent  
Very good  
Good, fair, or poor  
Other

**7. Health insurance status:**

Yes  
No  
Missing

**8. Age on the day of the interview (only individuals 18 or older are eligible for SAQ):**

18–24  
25–34  
35–44  
45–54  
55–64  
65+

**9. Sex:**

Male  
Female

**10. Race/ethnicity:**

Hispanic  
Black, non-Hispanic  
Asian  
Other

## Appendix C

### Glossary of Selected Terms

*Chi-Squared Automatic Interaction Detection (CHAID):* CHAID is a software program with a stepwise statistical procedure that is commonly used to identify subgroups that differ substantially in their propensity to respond to a survey.

*Dwelling Unit (DU):* A DU is a house, an apartment, group of rooms, or single room occupied as separate U.S. civilian noninstitutionalized living quarters or vacant but intended for occupancy as separate living quarters. An occupied DU corresponds to a household, using the terminology of the National Health Interview Survey (NHIS) or the Current Population Survey (CPS). A sample of NHIS responding households serves as the MEPS sample of DUs (technically, occupied DUs). A household (DU) can contain one or more families and/or unrelated individuals.

*Eligible:* The eligibility of a person for MEPS pertains to whether or not data were to be collected for that person. All of the Key in-scope persons of a sampled RU are eligible for data collection. The only non-Key persons eligible for data collection are those who live in an RU with at least one Key, in-scope person. The eligibility of a non-Key person continues only for the time that they live with such a person. The only out-of-scope persons eligible for data collection are those who live with a Key in-scope person, again only for the time that they live with such a person. Only military persons can meet this description (for example, a person on full-time active duty military may live with a spouse who is Key and in-scope).

*Family:*

- 1) *CPS:* A CPS family consists of two or more persons living together in the same household who are related by blood, marriage, or adoption. Foster children are not considered a family member in the CPS.
- 2) *MEPS:* The definition of a MEPS family is a bit more expansive than that of the CPS. A MEPS family consists of two or more persons living together in the same household who are related by blood, marriage, or adoption, as well as foster children (foster children are not included as members under the CPS definition of a family) and partners (plus persons related to partners). Partners are unmarried persons living together who regard themselves as a family unit (these are not families under the CPS definition). Generally, MEPS RUs with two or more members are MEPS families. Student RUs are considered a member of the family of their parent's RU and thus assigned the same MEPS family ID as their parents (this is consistent with how the CPS handles students living away from home).

*Full Year (FY):* MEPS FY data are released in two phases: a preliminary full year –Population Characteristics File” is released first which is superseded a few months later by a final full year –Consolidated File.” The first file release contains all annual survey data except medical expenditure and income/tax-related variables which require a longer time to process and thus are included in the final consolidated file. Moreover, the analytic

weights are not identical in the two separate FY files because income data that are not available at the initial release are used to refine the weight variables for the final release. The FY files allow analysts to produce estimates for the U.S. civilian noninstitutionalized population (i.e., the MEPS target population) at the person and/or family levels.

*In-scope:* A member of the civilian, noninstitutionalized population of the U.S. (MEPS target population) is considered to be in-scope for the survey. This status can vary within a round since a person may, for example, enter or return from the military or enter or return from a nursing home during the reference period of a survey round.

*Key:* Keyness relates to an individual's chance of being included in MEPS. A person is Key if they are linked for sampling purposes to an NHIS responding household subsampled for inclusion in MEPS. More specifically, a person is Key if s/he was a member of an NHIS household at the time of the NHIS interview or became in-scope upon joining an RU with at least one Key member (examples of the latter situation include newborns and persons returning from military service, an institution, or living outside the United States). A person who was in-scope (a member of the civilian, noninstitutionalized U.S. population) at the time of the NHIS but was not a member of a responding NHIS household sampled for MEPS is "non-Key".

*Mean Square Error (MSE):* The MSE of an estimator  $\theta^*$  with respect to the estimated parameter  $\theta$  is defined as  $MSE = E[(\theta^* - \theta)^2]$ . The MSE is equal to the sum of the variance and the squared bias of the estimator i.e.,  $MSE = Var(\theta^*) + Bias(\theta^*)$ .

*Point-in-Time (PIT):* MEPS PIT files are produced every year to provide timely data for estimates related to health insurance and other selected variables of interest that are available from a single round and require minimal editing. In the PIT files, Round 1 data for the most recent panel are combined with Round 3 data from the previous panel. The PIT files can be used to produce estimates that reflect approximately the first half of the year and/or the date of first interview in the year.

*Poststratification:* Poststratification is a procedure used to adjust the sample weights of responding units so that the totals over various demographic categories match population totals from another data source.

*Raking:* Raking procedures involve adjusting the sample weights iteratively so that the sums over specified margins of the full cross-classification of characteristics match population sizes from another data source. In contrast to poststratification, the sum of the sample weights within each cross-classification cell may not equal the corresponding control totals from the other data source.

*Reference Person:*

1) *MEPS:* A reference person is defined for each MEPS family within a DU. This reference person is usually an RU member 16 years of age or older who either owns or rents the home. If more than one person meets this description, the RU respondent

identifies one from among them. This is consistent with the approach used for the Current Population Survey (CPS). If the respondent is unable to identify a person fitting this definition, the questionnaire asks for the head of family and this person is then considered the reference person for that RU.

2) *NHIS*: The NHIS reference person is defined as the household member 18 years of age or older who owns or rents the home. If more than one household member owns or rents the sample unit, or if none of the household members owns or rents the sample unit, the oldest household member is designated as the reference person.

*Reporting Unit (RU)*: A MEPS RU is an individual person or a group of persons in a sampled DU who are related by blood, marriage, adoption, or other familial association. Regardless of the legal status of their relationship, two persons living together are treated in MEPS as a single RU if they choose to be identified as a family unit. Most households contain a single family or a single individual, although this is not always the case. Examples of households that may be of the more unusual variety: a single RU consisting of married daughter and her husband living in the same house (DU) with her parents; two unrelated persons living in the same apartment (DU) who consider themselves as two independent entities represent two distinct RUs; and a pair of unmarried people living in a condominium (DU) who characterize themselves as a single family are treated as a single RU.