

Estimation Procedures for the 2023 Medical Expenditure Panel Survey Household Component

Methodology Report #38 | August 2026 | Sadeq Chowdhury, Ph.D., Tien-Huan Lin, M.S., Jay Clark, M.S., Pengyu Huang, M.S., Elizabeth Eisenhauer, Ph.D.

Abstract

The primary purpose of this report is to describe the procedures used to develop various Medical Expenditure Panel Survey (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 (Full Year, Survey Supplements, and Longitudinal weights). The final section provides an overview of the procedures used to estimate variances of MEPS estimates, which are necessary for the proper analysis of MEPS data. Although MEPS weighting procedures do not vary substantially from year to year, the details presented in this report apply specifically to the 2023 data year.

1. Introduction	1
1.1 Summary of MEPS Sample Design	1
1.2 Terms Related to Sampling/Target Population	2
1.3 MEPS Analytic Files	3
1.4 Full Year (FY) File and Weights	3
1.5 Overview of Development of MEPS Weights	4
2. Development of Panel-Specific Weights	6
2.1 Round 1 of New Panel	6
2.2 Rounds 1–3 of New Panel in Year 1	7
2.3 Rounds 3–5 of Preceding Panel in Year 2	8
3. Full Year (FY) Weights	9
3.1 Development of Combined-Panel Person Weights for FY Population Characteristics File	9
3.2 Development of Final FY Weights for Consolidated File	10
3.2.1 Person Level	10
3.2.2 Family Level	11
3.3 General Issues Related to Developing FY Weights	12
4. Weight for Survey Supplement	13
4.1 Self-Administered Questionnaire	13
5. Longitudinal Weights	14
5.1 SAQ Longitudinal Weights	15
6. Variance Estimation	16
References	17
Suggested Citation	17

Appendixes

Appendix A	Distribution of Weights at Different Stages of Weighting for 2023 MEPS	18
Appendix B	Variables Used in Nonresponse and Poststratification/Raking Adjustments	19
Appendix C	Glossary of Selected Terms	34

Table

Table 1.1.	Type of public-use file (PUF) and variable name in PUF	4
------------	--	---

Figures

Figure 1.1.	Overview of MEPS household component overlapping panel design, 2018–2023	2
Figure 1.2.	Illustration of MEPS from longitudinal perspective, Panel 27 (2022–2023)	2
Figure 1.3.	Weight development process	5

1. Introduction

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

Because MEPS is a complex probability sample design, analytic approaches that assume a simple random sample design are generally not appropriate. Ignoring the complex survey design can lead to biased estimates and inaccurate significance levels. To obtain valid estimates and standard errors, analyses must incorporate the MEPS sample weights as well as the design's stratification and clustering features.

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 2–5 provide detailed descriptions of the procedures used to develop the various MEPS analytic weights, and Section 6 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 2023 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 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 Moriarity et al. (2022). The MEPS sample of households is a subsample of NHIS-responding households and reflects many of the features of the NHIS design. Chowdhury et al. (2019) provide a detailed overview of the MEPS sample design, and specifics for 2023 are available in the documentation for MEPS HC-247: 2023 Full Year Population Characteristics File.

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 or computer-assisted video interviews 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 of the first year and the beginning of the second year.

Since 1996, MEPS has selected and fielded a new panel of households each year. Figure 1.1 shows the panel design from 2018 through 2023, excluding COVID-19–related adjustments in 2020–2022 that extended Panels 23 and 24 beyond their original timelines. As illustrated, MEPS typically combines data from two distinct, nationally representative panels to improve the precision of annual estimates.¹

¹ 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. Because of the COVID-19 pandemic, 2020 estimates combined Panel 23, 24, and 25 samples; 2021 estimates combined Panel 23, 24, 25, and 26 samples; and 2022 estimates combined Panel 24, 26, and 27 estimates. See documentation for *MEPS HC-243: 2022 Full Year Consolidated Data File* for details.

Specifically, estimates are based on data from two consecutive panels—one in its first year of data collection and the other in its second year. For example, 2023 estimates use data from the second year of Panel 27 and the first year of Panel 28.

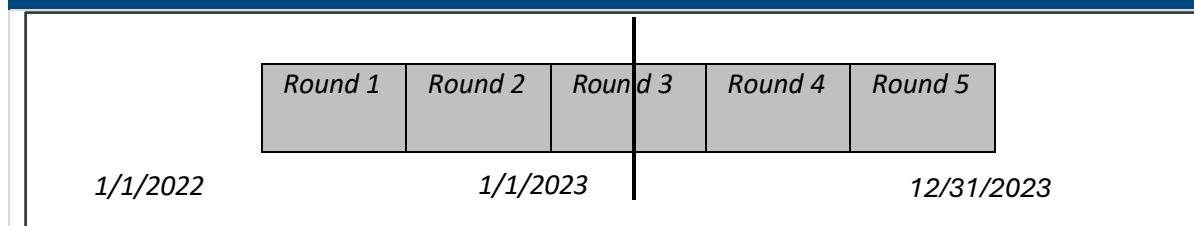
Figure 1.1. Overview of MEPS household component overlapping panel design, 2018–2023

MEP Panel	Year													
	2017	2018			2019		2020		2021		2022	2023		2024
22		R3	R4	R5										
23		R1	R2	R3	R4	R5	R6	R7	R8	R9				
24				R1	R2	R3	R4	R5	R6	R7	R8	R9		
25						R1	R2	R3	R4	R5				
26							R1	R2	R3	R4	R5			
27										R1	R2	R3	R4	R5
28												R1	R2	R3

* R = Round.

In addition to annual estimates, the MEPS design structure permits longitudinal estimates over 2 consecutive calendar years, thus allowing examination of person-level changes in selected variables over a 2-year period for a single panel. For example, research analysts can assess the persistence of high healthcare 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 1.2 provides an illustration of MEPS from a longitudinal perspective for Panel 27. 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 healthcare expenditures) between 2022 and 2023.

Figure 1.2. Illustration of MEPS from longitudinal perspective, Panel 27 (2022–2023)



1.2 Terms Related to Sampling/Target Population

The terms “RU,” “in scope,” “Key,” and “eligible” are used 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 a single person if living without other family members. The definition of an RU and a MEPS family (see Section 3.2.2) are identical except for student RUs. A student RU refers to a young adult family member living away from home to attend school but is linked back to their 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, a special component of the general population.

An individual is in scope at a particular point in time if they are a member of the target population for MEPS - the civilian noninstitutionalized population of the United States.

“Key” in MEPS indicates a person’s inclusion status in the survey based on their sampling status in the NHIS. Because the MEPS sample is drawn as a subsample of NHIS respondents, the probability of selection for MEPS is directly tied to the probability of selection for the NHIS. An individual is defined as Key in MEPS if the person was in scope at the time of the NHIS and was a member of an NHIS responding household selected for MEPS. Individuals who were not in scope at the time of the NHIS (e.g., because they were living outside the United States, serving in the military, residing in a nursing home, or not yet born) are also defined as Key if they subsequently become in scope and join a household participating in MEPS. Thus, Keyness relates to an individual’s probability of being included in MEPS. Hence, a person who was in scope (i.e., a member of the civilian noninstitutionalized U.S. population) at the time of the NHIS but was not selected as part of an original NHIS sampled household is always classified as non-Key, even if that person subsequently joins a MEPS household.

Finally, a person is eligible for data collection in MEPS if they are a member of an RU containing at least one person classified as in scope and Key. In the typically rare cases where a person’s eligibility for data collection cannot be ascertained, these persons with unknown eligibility are treated like nonrespondents. 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, two types of person-level MEPS public-use data files (PUFs) with appropriate weight and variance estimation variables are released every year: the Full Year (FY) Population Characteristics PUF and the FY Consolidated PUF. The Population Characteristics PUF is the first to provide annual data on healthcare utilization, whereas the Consolidated PUF includes both utilization and associated expenditure data. Beginning in 2024, the Population Characteristics PUF will be discontinued and only the Consolidated PUF will be released.

Because of the MEPS overlapping panel design, data collected for the same calendar year from two consecutive MEPS sample panels are used to develop the files.² For example, the 2023 files consist of data from Rounds 1, 2, and 3 (2023 portion) for the MEPS Panel 28 sample and from Rounds 3 (2023 portion), 4, and 5 for the Panel 27 sample.

1.4 Full Year (FY) File and Weights

MEPS FY data are released in two phases. First, a preliminary FY file, the Population Characteristics PUF, is released. This file is then superseded a few months later by the final FY file, the Consolidated PUF. 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 PUF. Moreover, the weights are not identical in the two FY 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 PUF 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 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 PUF that can be used to produce family-level estimates based on two slightly different definitions of family units (Current Population Survey [CPS] versus MEPS).

² Due to the COVID-19 pandemic, 2020 and 2022 estimates combined data from three panels, and 2021 estimates combined data from four panels. However, this report focuses on 2023, where two panels are combined.

FY weights are the result of a composite of the following factors:

- Adjustments for person-level nonresponse (survey attrition)
- Raking/poststratification adjustments.

Weights are constructed separately for the two overlapping individual panels/rounds and then combined into a final weight for the FY PUF. The steps used to develop the FY person- and family-level weights are illustrated in Figure 1.3 and described in detail in the following sections.

1.5 Overview of Development of MEPS Weights

For most MEPS PUFs, data are pooled from two MEPS panels fielded during the same year. The weight construction process follows these general steps:

1. Establish a “base” or initial weight for each panel separately, using the probability of selection.
2. Adjust for nonresponse within each panel to account for potential differences between respondents and nonrespondents.
3. Calibrate each panel separately to independent population control figures such as the CPS (usually by raking, with consideration of trimming and re-raking) to account for undercoverage and differences between the sample and the population.
4. Composite the individual panel weights to produce a single weight for the pooled data from the two panels (achieved by multiplying the individual panel weights by factors representing the relative effective sample sizes of the two panels).
5. Recalibrate the composited weight to the same population control totals. Monitor presence of outliers and trim and recalibrate if necessary.

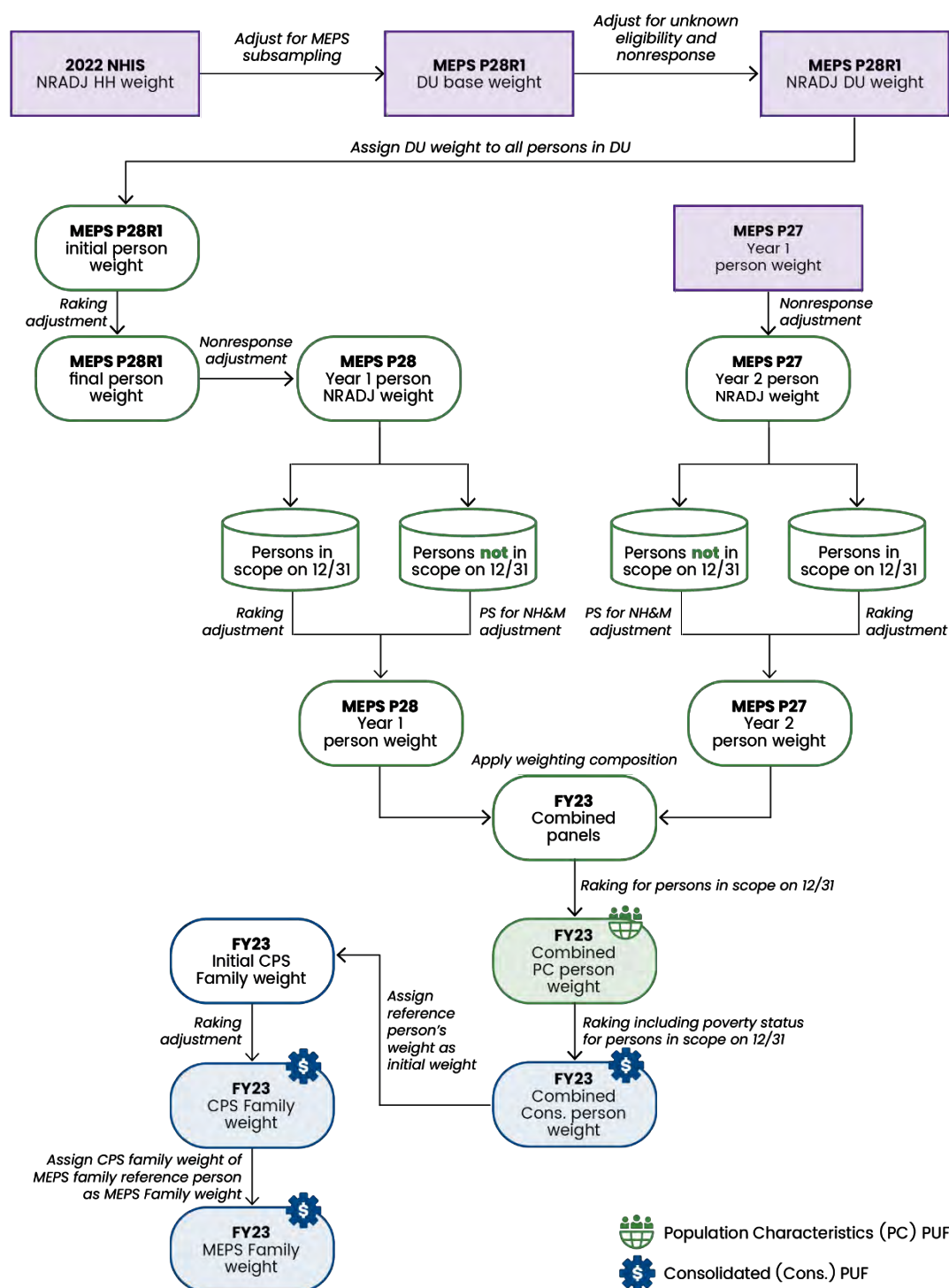
Figure 1.3 presents a flowchart of the weight development process described in Section 2 (Panel-Specific Weights) and Section 3 (Full Year Weights). The Round 1 weight is developed first and serves as the initial weight for deriving the FY weight for a MEPS Panel in its first year. Panel-specific weights for concurrent panels are combined to produce final FY weights.

The Population Characteristics PUF includes person weights (panel-specific and combined) and Self-Administered Questionnaire (SAQ) weights (combined), and the Consolidated PUF includes person weights (combined), SAQ weights (panel-specific and combined), the poverty-adjusted family weights (MEPS families and CPS-like families), and the food security weight. Table 1.1 lists the final weight variables provided on the MEPS PUFs that are described in Sections 2–5 of this report. Appendix A provides summary distributions for 2023 of selected interim and PUF weights mentioned in this report.

Table 1.1. Type of public-use file (PUF) and variable name in PUF		
Type of PUF	Variable name in PUF*	
	Person level	Family level
FY Population Characteristics	PPANWTyy, PERWTyyP	—
SAQ	SAQWTyyP	—
FY Consolidated	PERWTyyF	FAMWTyyF, FAMWTyyC
SAQ	PSAQWyyF, SAQWTyyF	—
Longitudinal	LONGWT	—

* “yy” in variable names indicates 2-digit year.

Figure 1.3. Weight development process



Additional Notes: CPS = Current Population Survey; DU = dwelling unit; FY = full year; HH = household; P27 = Panel 27; P28 = Panel 28; P28R1 = Panel 28 Round 1; NHIS = National Health Interview Survey; NH&M = nursing home and mortality; NRADJ = nonresponse adjusted; PS = Poststratification

2. Development of Panel-Specific Weights

2.1 Round 1 of New Panel

The starting point for development of the MEPS FY weights is the Round 1 weight. As shown in Figure 1.3, the Round 1 weight is the result of a composite of the following factors:

- 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
- a. Constructing the 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. (Note that this rate varies by MEPS sample domain.³)

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

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

where, for DU i , $WTDU_{0i}$ is the MEPS DU base weight, $WTDU_{Ni}$ is the NHIS household weight, and f_i is the MEPS subsampling rate.

- b. Adjustment for DU-level unknown eligibility and nonresponse in Round 1.** The DU-level base weight calculated previously is adjusted to compensate for unknown eligibility and nonresponding DUs to the Round 1 interview. First, an adjustment factor for unknown eligibility is calculated as the ratio of the weighted total of all cases to the weighted total of respondents, ineligible, and nonrespondents with known eligibility. This adjustment factor is computed in a simplistic fashion because the percentage of DUs with unknown eligibility is generally very low. The unknown eligibility adjustment is applied as follows:

$$WTDU_{1i} = WTDU_{0i} \times A_1, \text{ with } A_1 = \frac{\sum WTDU_{0i}}{\sum_{i \notin unk} WTDU_{0i}}$$

where $WTDU_{0i}$ is the DU base weight in Step a, A_1 is the overall unknown eligibility adjustment factor, unk represents unknown eligibility, and thus $i \notin unk$ indicates DUs with known eligibility (i.e., respondents, ineligible, and nonrespondents with known eligibility).

Next, a nonresponse adjustment factor is computed as the ratio of the weighted total of respondents and nonrespondents with known eligibility to the weighted total of respondents within each nonresponse adjustment class. Using a broad range of 26 demographic, geographic, socioeconomic, and health covariates from the NHIS (see Appendix B, Section B1), a Chi-squared Automatic Interaction Detector (CHAID) analysis⁴ (Kass, 1980) is carried out to form MEPS nonresponse adjustment classes to adjust for the MEPS DU-level nonresponse. The nonresponse adjustment is applied by inflating the weights of the responding DUs in each adjustment cell as follows:

$$WTDU_{2i} = WTDU_{1i} \times A_{2c}, \quad \text{if } i \text{ is a responding DU}$$
$$\text{with } A_{2c} = \frac{\sum_{i \in cnr \& nr} WTDU_{1i}}{\sum_{i \in cnr} WTDU_{1i}}$$

³ For more information regarding the MEPS sample design, see Chowdhury et al., 2019.

⁴ See glossary for a brief description of CHAID.

where $WTDU_{2i}$ is the nonresponse adjusted weight, $WTDU_{1i}$ is the unknown eligibility adjusted weight calculated previously, A_{2c} is the nonresponse adjustment factor for cell c , r represents the set of responding DUs, and nr represents the set of nonresponding DUs; therefore, $i \in c \cap r$ represents the set of responding DUs in cell c , and $i \in c \cap r \cap nr$ represents the set of responding and nonresponding DUs in cell c .

- c. Person-level raking adjustment to form final Round 1 person-level weight.** A person-level file is created and the DU-level nonresponse adjusted weight is assigned as the initial Round 1 person weight for the MEPS respondents. The initial person weight is raked to the March CPS control totals to yield the final Round 1 person weight. The variables used in the raking are region, metropolitan statistical area (MSA) status, race/ethnicity, sex, and age category (see Appendix B, Section B2). The final Round 1 person weight can be expressed as follows:

$$WTPR_{1k} = WTDU_{2i} \times A_3$$

where $WTPR_{1k}$ is the final Round 1 weight for MEPS respondent k , $WTDU_{2i}$ is the DU nonresponse adjusted weight in Step b and assigned as the initial person weight for MEPS respondent k , and A_3 is the raking adjustment factor. The distribution of the adjusted weight is checked for extremely large weights and the raking procedure is repeated if any weight trimming is required.

2.2 Rounds 1–3 of New Panel in Year 1

- a. Initial weight.** The final Round 1 person-level weight 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 analysis is carried out using a set of potential predictor variables. The set of 16 predictor variables used as input to the CHAID analysis to adjust for this Year 1 attrition is detailed in Appendix B, Section B4.

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

$$WTPR_{2k} = WTPR_{1k} \times P_{2c} \quad \text{if person } k \text{ is a respondent in Rounds 2/3}$$

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

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 thus $c \cap r$ 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 .

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 the U.S. 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, age category, and education categories (see Appendix B, Section B3).

- ii. **Poststratification adjustments for persons not in scope on December 31.**⁵ The small proportion of cases that 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.⁶ 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 and who were not members of the U.S. civilian noninstitutionalized population on December 31 using an estimated control total of 500,000.⁷
 - **Other not in-scope persons on December 31.** For persons who are out of scope on December 31 and do not belong to the two special out-of-scope groups defined previously (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 factor⁸ 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.

2.3 Rounds 3–5 of Preceding Panel in Year 2

- a. **Initial weight.** The Year 1 FY 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 2.2 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 $WTPR_{FY_1k}^*$, using an asterisk to indicate the same weight for the preceding panel.
- b. **Adjustment for person-level nonresponse.** The initial weight is adjusted to compensate for Year 2 nonresponse over Rounds 4 and 5. A CHAID analysis is again used to form nonresponse adjustment

⁵ The purpose of this poststratification is to compensate for potential underrepresentation of high medical expenditure groups.

⁶ Adjustment based on Underlying Cause of Death, 1999–2020, data from the CDC’s Wide-ranging ONline Data for Epidemiologic Research (WONDER) database.

⁷ This control total is based on data from the Minimum Data Set (MDS) of the Centers for Medicare & Medicaid Services in 2018.

⁸ Factors are derived using an iterative raking algorithm.

cells using a set of potential predictor variables (see Appendix B, Section B5 for list of variables). Because the nonresponse adjustment procedure is the same as described previously for the other panel, there is therefore no need to repeat the detailed steps here; thus, 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 population and the U.S. 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, age category, and education categories (see Appendix B, Section B3).⁹
- ii. Poststratification adjustments for persons not in scope on December 31.**¹⁰ 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 2.2.c.ii).

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

$$WTPR_{FY_2k}^* = WTPR_{4k}^* \times P_{FY_2k}^*$$

where:

$$\begin{aligned} P_{FY_2k}^* &= \text{raking adjustment factor for person } k \text{ who was in scope on December 31} \\ &= \text{poststratification adjustment factors for the two special out-of-scope groups} \\ &\quad \text{(i.e., decedents and nursing home entrants)} \\ &= 1 \text{ for other out-of-scope persons on December 31} \end{aligned}$$

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. Full Year (FY) Weights

3.1 Development of Combined-Panel Person Weights¹¹ for FY Population Characteristics 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 (or the Population Characteristics file weight). The compositing factor is calculated so that it reflects the effective sample size for each individual panel relative to the total effective sample size in

⁹ Note that poverty status is not included here but is included in raking for the final FY Consolidated File.

¹⁰ The purpose of this poststratification is to compensate for potential underrepresentation of these high medical expenditure groups.

¹¹ A family-level weight is not constructed for the FY Population Characteristics File.

both panels combined. The FY composite weight for person k in the combined panel can be expressed as:

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

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 θ is the compositing factor, which is the proportion of effective sample size in the combined panels who came from the most recent panel (i.e., $\theta = \frac{\frac{n_1}{DEFF_1}}{\frac{n_1}{DEFF_1} + \frac{n_2}{DEFF_2}}$,

where $DEFF$ is the design effect of the single-panel weights after adjusting for nonresponse). In recent years, the values of θ generally ranged from around 0.4 to 0.6.

- 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, age category, and education category (see Appendix B, Section B3).¹² The decedents and nursing home entrants are separately poststratified as described in Section 2.2.c. These adjustments produce the PUF variable $PERWT_{yyP}$ (i.e., preliminary FY weight).

Because this step is the same as the one described in detail in Section 2.2.c, there is therefore no need to repeat the detailed steps here; thus, 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.

3.2 Development of Final FY Weights for Consolidated File

3.2.1 Person Level

To produce the final person-level FY Consolidated File weight, the person-level weights produced for the Population Characteristics File are re-raked using poverty status¹³ in addition to the same set of variables used before (see previous section and Appendix B, Section B3) that correspond 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:

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

¹² 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.

¹³ The poverty status categories are shown in Appendix B, Section B3.

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.

3.2.2 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, whereas 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 families whose members became out of scope prior to the end of the year.¹⁴

Although 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. 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 final person-level and poverty-adjusted weight of the reference person of the CPS-like family. Raking adjustments are then 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 final 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 raking adjustment to produce final FY family weights: (1) at least one Key in-scope person in the family during the year; (2) the reference person in the family is Key; and (3) 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 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 .¹⁵

- 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, household poverty status, region, MSA status, family size, and age category of the reference person as raking dimensions (see Appendix B, Section B6), 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

¹⁴ Estimates can be restricted to MEPS families in scope on December 31 if appropriate for analytic purposes.

¹⁵ 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.

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.¹⁶ 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.¹⁷ This produces the PUF variable $FAMWT_{yyC}$.

- 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 previously) to all members of the MEPS family, which becomes $FAMWT_{yyF}$. If the MEPS family is not in scope on December 31, then the final person-level weight (as derived in Section 3.2.1) of the reference person of the MEPS family is assigned to all members of the MEPS family.

3.3 General Issues Related to Developing FY Weights

- a. Cell collapsing.** The initial nonresponse or poststratification/raking 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 (i.e., CPS) are considered acceptable for nonresponse or poststratification/raking adjustments. In addition, cells with adjustment factors greater than two are usually either truncated or collapsed with a neighboring cell.
- b. Weight trimming.** The distribution of weights is examined after each round of poststratification/raking 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 Mean Square Error (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 PUFs have positive values for both the person-level and family-level weight. 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 zero value for the person- or family-level weight are described as follows:
 - **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 (2) 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

¹⁶ Family weights are assigned to all members of responding families whose reference person is Key, regardless of whether the individual member is Key/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).

¹⁷ Sometimes, partitioning of a MEPS family results in a CPS family where the reference person is non-Key.

not have a positive person weight (because of nonresponse); or (2) the reference person of the family is non-Key.¹⁸

Appendix A provides summary distributions for 2023 of selected interim and PUF weights discussed in this report.

4. Weight for Survey Supplement

In addition to the person- and family-level weight variables included on FY files, an additional person-level weight variable is included for appropriate estimation using data collected in the MEPS SAQ. This supplement is a self-administered paper-and-pencil questionnaire that is fielded in selected rounds of the survey to all adults. In general, the weight variable for this supplement is used in lieu of the FY person-level weight when the analysis involves data from the supplement. As for the general survey FY weights described in Section 3, the preliminary weight for the supplement is provided on the preliminary FY file and the final weight for the supplement is provided on the final FY file. The following is the overview of procedures for developing the special supplement weight variable.

4.1 Self-Administered Questionnaire

A SAQ is requested to be completed by each adult (i.e., 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 healthcare quality.¹⁹ Thus, the target population for the SAQ is the adult civilian noninstitutionalized population at the time data were collected for Rounds 2/4. The 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 the nonresponse adjusted weight used for developing the panel-specific FY weight for each panel (see Sections 2.2.b and 2.3.b). This weight is then adjusted to compensate for those persons who were FY respondents, eligible for SAQ, and in scope at the end of Rounds 2/4 but did not respond to the SAQ. The 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, race/ethnicity, NHIS quarter, and NHIS panel. For the second-year panel, poverty status collected in the previous year is also used (see Appendix B, Section B7).
- b. The nonresponse adjusted SAQ weight in Step a is raked to the CPS estimates corresponding to December of the analytic year (the same source of control figures used for the FY 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 the same as those used for the preliminary FY person weights: region, MSA status, age category, sex, race/ethnicity, and education category. The main difference is that age categories here are developed after excluding ages under 18, because only adults were eligible for the SAQ (see Appendix B, Section B8).
- c. The panel-specific files are put together to create a combined person-level file, and a compositing factor (with the same rationale as Section 3.1.a) is applied to the panel-specific nonresponse adjusted SAQ weight of each individual panel in Step b to derive the SAQ weight for the combined panel.
- d. The SAQ weight for the combined panel in Step c is raked again to CPS estimates based on the first 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.

¹⁸ This removes the issue of multiple chances of selection being associated with a family.

¹⁹ The SAQ includes questions from the Consumer Assessment of Health Plans (CAHPS®), the SF-12, the EuroQol 5D, and attitude items.

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.

5. Longitudinal Weights

In contrast to FY files that 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 27: 2022–2023). Although data are available for all five rounds for more than 90 percent 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 2-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 in the following) reflects adjustments for this attrition (nonresponse).

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

- All persons in their Year 2 FY file who are in scope on December 31 of Year 2
- All persons in their Year 2 FY file who are out of scope on December 31 of Year 2 because of death or entering a nursing home
- All persons in their Year 2 FY file who are out of scope on December 31 of Year 2 for reasons other than death or entering a nursing home
- All persons in their Year 1 FY file from the panel who are out of scope on December 31 of Year 1 and not part of the Year 2 FY file

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

$$\begin{aligned}
 LONGWT_k &= PERWT_{FY_2k} && \text{if } k \in a \\
 &= PERWT_{FY_2k} \times L_{bk} && \text{if } k \in b \\
 &= PERWT_{FY_2k} \times L_{ck} && \text{if } k \in c \\
 &= PERWT_{FY_1k} && \text{if } k \in d
 \end{aligned}$$

where a, b, c & d are four groups as defined previously, $PERWT_{FY_2k}$ is the panel-specific final poverty-adjusted Year 2 annual weight for person k , and $PERWT_{FY_1k}$ is the panel-specific final poverty-adjusted Year 1 annual weight for person k . Both these weights²⁰ have been adjusted to compensate for survey attrition during the year. L_{bk} and L_{ck} are adjustment factors designed to make the sum of the longitudinal weights of the persons in group b or c equal to the sum of the combined panel FY person weights of similar out-of-scope persons from both panels in the Year 2 FY file.²¹

²⁰ Panel-specific annual weights are not included on PUFs.

²¹ This adjustment was approximately 11 percent for Panel 11.

The adjustment factor L_{bk} for person k can be expressed as:

$$L_{bk} = \frac{\sum_{k \in OOS} PERWT_{FYk}}{\sum_{k \in OOS} PERWT_{FY_2k}}$$

where $\sum_{k \in OOS} PERWT_{FYk}$ is the sum of combined panel FY weights of persons in both panels who are out of scope on December 31 because of 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.

The adjustment factor L_{ck} for person k can be expressed as:

$$L_{ck} = \frac{\sum_{k \in OOS} PERWT_{FYk}}{\sum_{k \in OOS} PERWT_{FY_2k}}$$

where $\sum_{k \in OOS} PERWT_{FYk}$ is the sum of combined panel FY weights of persons in both panels who are out of scope on December 31 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 c of the longitudinal panel only.

In other words, for L_{bk} and L_{ck} , the numerator includes all out-of-scope persons as specified previously from both panels in the FY file and the denominator includes the same out-of-scope persons from the longitudinal panel only.

5.1 SAQ Longitudinal Weights

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

- a. All SAQ eligible persons in their Year 2 FY file who are in scope or out of scope for any reason (death or entering a nursing home or any other reason) on December 31 of Year 2.
- b. All SAQ eligible persons in their Year 1 FY file from the panel who are out of scope on December 31 of Year 1 and not part of the Year 2 FY file.

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

$$\begin{aligned} LSAQWT_k &= SAQWT_{Y_2k} & \text{if } k \in a \\ &= SAQWT_{Y_1k} & \text{if } k \in b \end{aligned}$$

where,

$SAQWT_{Y_2k}$ = Year 2 SAQ weight for person k

$SAQWT_{Y_1k}$ = Year 1 SAQ weight for person k

6. 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 because of 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 or 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, R, and SAS survey procedures.

The Taylor series method is often used when analyzing MEPS data. MEPS annual PUFs include the two sample design variables necessary for implementing this method: the variance estimation strata (VARSTR) and the variance estimation clusters (VARPSU). Specifying these variables in conjunction with a ‘with replacement’ design in software packages that employ the Taylor series approach will produce variances that reflect the complexities of the MEPS design.

Because the Taylor series method can be difficult to use for calculating the variances of complex estimators that are not readily available in complex survey software packages (e.g., two-part model of health expenditures, ratios of 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. The file is called Replicate File for BRR Variance Estimation (AHRQ, 2025b) 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 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 2 consecutive years.²² 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 ensure a variance structure that consistently specifies MEPS geographic sampling units across years.

Before 2002, each annual MEPS PUF was released with a variance structure unique to the particular MEPS sample in that year. Between 2002 and 2018, the annual MEPS PUFs were released with a common variance structure that allows users to seamlessly pool annual files from 2002 to 2018. For 2019 and later years, the MEPS public use files were released with a different common variance structure that allows users to pool data from 2019 and later years. The common variance structure for 2002 to 2018 and the

²² See the MEPS-HC Methodology Reports at <https://meps.ahrq.gov/mepsweb> for more details.

common variance structure for 2019 and later years are not compatible with each other nor the structure on the annual PUFs released prior to 2002. Therefore, when pooling (a) any year from 1996 to 2001 with any year from 2002 or later, or (b) any year from 2019 or later with any earlier year, it is necessary to obtain a common variance structure from the Pooled Linkage File for Common Variance Structure (AHRQ, 2025a) 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 Replicate File for BRR Variance Estimation 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 because of inconsistencies resulting from some primary sampling units (PSUs) that were dropped in those years.

References

- Agency for Healthcare Research and Quality (AHRQ). 1996-2023 Pooled Linkage File for Common Variance Structure. *MEPS HC-036*. Updated August 2025a. Accessed February 24, 2026. https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-036
- Agency for Healthcare Research and Quality (AHRQ). MEPS 1996-2023 Replicate File for BRR Variance Estimation. *MEPS HC-036BRR*. Updated August 2025b. Accessed October 28, 2025. https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-036BRR
- Chowdhury SR, Machlin SR, Gwet KL. Sample Designs of the Medical Expenditure Panel Survey Household Component, 1996–2006 and 2007–2016. *Methodology Report #33*. Agency for Healthcare Research and Quality; 2019. https://meps.ahrq.gov/data_files/publications/mr33/mr33.pdf
- Fay RE. Theory and application of replicate weighting for variance calculations. In: *Proceedings of the Survey Research Methods Section*. American Statistical Association; 1989:212–217.
- Kass GV. An exploratory technique for investigating large quantities of categorical data. *Journal of Applied Statistics*. 1980;29(2):119–127.
- Moriarty C, Parsons VL, Jonas K, Schar BG, Bose J, Bramlett MD. Sample design and estimation structures for the National Health Interview Survey, 2016–2025. *National Center for Health Statistics; Vital Health Stat*. 2022;2(191). <https://dx.doi.org/10.15620/cdc:115394>
- Shao J. Resampling methods in sample surveys (with discussion). *Statistics*. 1996;27:203–254.
- Wolter KM. *Introduction to Variance Estimation*. 2nd Edition. Springer-Verlag; 2007.

Suggested Citation

Chowdhury S, Lin TH, Clark J, Huang P, Eisenhauer E. Estimation Procedures for the 2023 Medical Expenditure Panel Survey Household Component. Methodology Report #38. AHRQ Publication No. 26-0018. August 2026. Agency for Healthcare Research and Quality, Rockville, MD. https://meps.ahrq.gov/mepsweb/data_files/publications/mr38/mr38.pdf

Appendix A

Distribution of Weights at Different Stages of Weighting for 2023 MEPS

Panel/year	Weight	PUF variable	Number of records	Sum of weights	Minimum weight	Median weight	Average weight	Maximum weight	CV weights (%)
Panel 28, 2023	DU Final Weight	-	6,543	128,148,914	1,898	18,216	19,586	102,236	44.5
	Round 1 Person Weight	-	14,373	327,339,467	2,335	20,584	22,775	113,379	50.2
Panel 28, 2023	FY Person Weight*	-	10,288	333,748,948	2,555	27,887	32,441	218,738	61.2
Panel 27, 2023	FY Person Weight*	-	8,175	333,805,592	1,991	32,777	40,832	300,572	75.0
Panels 27 & 28 Combined, 2023	FY Person Weight (Preliminary)	PERWT23P	18,463	334,530,273	797	15,136	18,119	131,243	67.2
	FY Person Weight (Final)	PERWT23F	18,463	334,530,273	502	15,100	18,119	131,657	69.1
	FY Family Weight (CPS)	FAMWT23C	8,438	149,273,967	639	14,423	17,691	134,738	72.3
	FY Family Weight (MEPS)	FAMWT23F	8,133	142,088,938	639	14,283	17,471	134,738	72.1
Panel 27, 2022–23	Longitudinal Weight	LONGWT	8,292	338,597,320	1,163	32,795	40,834	307,845	75.9

*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 Adjustments²³

B1. NHIS Variables Used in CHAID Analysis to Form Cells for DU-Level Nonresponse (NR) Adjustment

1. Completion status of the interview:

Complete
Partial
Roster only

2. Age categories for the household respondent:

Less than 25 years old
25–29 years old
30–34 years old
35–44 years old
45–64 years old
65 years or older

3. Sex of the household respondent:

Male
Female

4. Race/ethnicity of the household respondent:

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

5. U.S. citizen status of sample adult:

Citizen
Not citizen
Unknown

6. U.S. born status of sample adult:

Born in U.S.
Not born in U.S.
Unknown

²³ Although these variables and categories are fairly stable across years, there are some instances in certain years where the variables or categories used differ slightly.

7. Education level of the household respondent:

Less than high school
High school graduate or GED
Some college
Bachelor's or a higher degree
Unknown

8. Work status of the household respondent:

Working
Not working
Unknown

9. Family income of the sample adult:

Less than \$35,000
\$35,000–\$49,999
\$50,000–\$74,999
\$75,000–\$99,999
\$100,000 or higher
Unknown

10. House tenure:

Owned or being bought
Rented or other
Unknown

11. Telephone number reporting status:

Cell/home/work phone number reported
No telephone number was reported

12. NHIS interview language:

At least one interview in Spanish or Spanish/English mix
At least one interview in other language, none Spanish or English
English only, no other language
Unknown

13. NHIS interview proxy:

Sample adult interview is completed with a proxy
Sample adult interview is not completed with a proxy
No complete sample adult interview

14. Household type:

One adult household, male
One adult household, female
Two adult household, married spouse present
Two adult household, not married spouse present
All others

15. Household with a child:

At least one person under 18 in household
Otherwise

16. Household size:

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

17. House type:

House/apartment/flat/condo
Other type of dwelling
Unknown

18. DU health status:

Either sample adult or sample child is in fair or poor health
Both sample adult and sample child are in good to excellent health
Unknown health status

19. DU insurance coverage status:

Either sample adult or sample child has no health insurance coverage
Both sample adult and sample child have health insurance coverage
Unknown

20. DU hospitalization status (newborns under age 1 are excluded from hospitalization status):

Either sample adult or sample child was hospitalized within the last 12 months
Both sample adult and sample child were not hospitalized within the last 12 months
Unknown

21. Sampling domain:

Asian household
Hispanic household
Black household
Other household, complete
Other household, partial, roster

22. PSU sampling strata:

Self-representing PSU stratum
Non-self-representing PSU stratum

23. MSA type:

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

24. Urban/Rural (block level):

Urban
Rural

25. NHIS Quarter from which the MEPS sample is taken:

1
2
3

26. NHIS Panel from which the MEPS sample is taken:

01
03

B2. Variables Used in Round 1 Person-Level Raking

1. Region containing person (assumed to be the same as for family): Northeast

Midwest
South
West

2. MSA status:

MSA
Non-MSA

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

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

4. Sex of person:

Male
Female

5. Age categories of person:

under 1
1–4
5–9
10–14
15–17
18
19–25
26–29
30–34
35–44
45–54
55–59
60–64
65–69
70–74
75–80
80+

B3. Variables Used in All FY Person-Level Raking

1. Race/ethnicity:

Hispanic
Black, non-Hispanic
Other (including Asian), non-Hispanic

2. Sex of person:

Male
Female

3. Region:

Northeast
Midwest
South
West

4. MSA status:

MSA
Non-MSA

5. Age categories:

under 1
1–4
5–9
10–14
15–17
18
19–25
26–29
30–34
35–44
45–54
55–59
60–64
65–69
70–74
75–80
80+

6. Education categories:

No degree
High school degree, GED with no college, or some college
Bachelor's or a higher degree
Inapplicable or unknown

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

7. Poverty status:

Less than 100 percent of poverty level or negative income
100–124 percent of poverty level
125–199 percent of poverty level
200–399 percent of poverty level
400 percent or greater of poverty level

B4. Variables Used in CHAID Analysis to Form Cells for Year 1 FY Person-Level NR Adjustment

1. Reluctance to respond in Round 1 interview:

RU was reluctant to respond
Otherwise

2. Age category:

0–18
19–25
26–34
35–44
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, non-Hispanic
Other, non-Hispanic

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
Bachelor's or a higher degree
Inapplicable, Unknown

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

Private
Public only
Uninsured

11. Census region:

Northeast
Midwest
South
West

12. Self-perceived health status of person:

Excellent
Very good
Good
Fair
Poor
Unknown/inapplicable

13. Number of office-based provider visits of person:

0 visit
1 visit
2–4 visits
5+ visits
Unknown number of visits

14. Number of nights in the hospital for discharges:

0 nights
1+ nights
Refused, not ascertained, or don't know

15. NHIS Quarter from which the MEPS sample is taken:

1
2
3

16. NHIS Panel from which the MEPS sample is taken:

01
03

B5. Variables Used in CHAID Analysis to Form Cells for Year 2 FY Person-Level NR Adjustment

1. Reluctance to respond in Round 1 interview:

RU was reluctant to respond
Otherwise

2. Age category:

0–18
19–25
26–34
35–44
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, non-Hispanic
Other, non-Hispanic

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
Bachelor's or a higher degree
Inapplicable, Unknown

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

Private
Public only
Uninsured

11. Census region:

Northeast
Midwest
South
West

12. First respondent indicator:

Yes, First respondent
No, Not first respondent
Nonresponse

13. Proxy respondent indicator:

Respondent is an RU member
Respondent is a proxy
Nonresponse

14. Self-perceived health status of the person in Year 1:

Excellent
Very good
Good
Fair
Poor
Unknown/inapplicable

15. Total healthcare expenditures of the person in Year 1:

\$0
\$1–300
\$301–950
\$951–3,150
Over \$3,150
Unknown

16. Number of office-based provider visits of the person in Year 1:

0 visit
1 visit
2–4 visits
5+ visits
Unknown number of visits

17. Number of nights in the hospital for discharges:

0 nights
1+ nights
Refused, not ascertained, or don't know

18. NHIS Quarter from which the MEPS sample is taken:

1
2
3

19. NHIS Panel from which the MEPS sample is taken:

1
3

20. Poverty status:

Less than 100 percent of poverty level or negative income
100–124 percent of poverty level
125–199 percent of poverty level
200–399 percent of poverty level
400 percent or greater of poverty level

21. Ever in scope in Year 1:

In scope
Not in scope

B6. 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, non-Hispanic
Other, non-Hispanic

3. Poverty status:

Less than 100 percent of poverty level or negative income
100–124 percent of poverty level
125–199 percent of poverty level
200–399 percent of poverty level
400 percent or greater of poverty level

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

B7. Variables Used for SAQ Nonresponse 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 or unknown

5. Education:

No degree
High school degree or GED with no college
Some college
Bachelor's or a higher degree
Under 16 - inapplicable

6. Health status:

Excellent
Very good
Good
Fair
Poor
Unknown

7. Health insurance status:

Yes
No
Unknown

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, non-Hispanic
Other, non-Hispanic

11. NHIS Quarter from which the MEPS sample is taken:

1
2
3

12. NHIS Panel from which the MEPS sample is taken:

01
03

In addition, the following variable is used for SAQ nonresponse adjustment in the second-year panel.

13. Poverty status:

Less than 100 percent of poverty level or negative income
100–124 percent of poverty level
125–199 percent of poverty level
200–399 percent of poverty level
400 percent or greater of poverty level

B8. Variables Used in SAQ Raking

1. Race/ethnicity:

Hispanic
Black, non-Hispanic
Other (including Asian), non-Hispanic

2. Sex of person:

Male
Female

3. Region:

Northeast
Midwest
South
West

4. MSA status:

MSA
Non-MSA

5. Age categories:

18–19
20–25
26–29
30–34
35–44
45–54
55–59
60–64
65–69
70–74
75–80
80+

6. Education categories:

No degree
High school degree
GED with no college
Some college
Bachelor's or a higher degree
Inapplicable or unknown

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

7. Poverty status:

Less than 100 percent of poverty level or negative income
100–124 percent of poverty level
125–199 percent of poverty level
200–399 percent of poverty level
400 percent or greater of poverty level

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, a group of rooms, or a 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 data were to be collected for that person. All 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, and, 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 who are living together in the same household and 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 who are living together in the same household and are related by blood, marriage, or adoption, 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 parents' 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 FY Population Characteristics File is released first, which is superseded a few months later by a final FY 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 because 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. This 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 they were a member of an NHIS household at the time of the NHIS

interview or became in scope on 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 (i.e., 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 considered non-Key.

Mean Square Error (MSE). The MSE of an estimator θ^* with respect to the estimated parameter θ 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: $MSE = Var(\theta^*) + Bias(\theta^*)$.

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 CPS. If the respondent is unable to identify a person fitting this definition, the questionnaire asks for the head of the family, who 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 own or rent 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 include the following:

- A married daughter and her husband living in the same house (DU) with her parents comprise a single RU.
- Two unrelated persons living in the same apartment (DU) who consider themselves to be two independent entities represent two distinct RUs.
- A pair of unmarried people living in a condominium (DU) who characterize themselves as a single family are treated as a single RU.