

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

MEPS-HC

Estimation and Analysis

Sadeq Chowdhury, PhD

Outline

- Estimation from MEPS-HC
  - Producing Person-level Estimates
  - Computing Standard Errors
- Analysis of Subpopulations
- Family-Level Analysis
- Other Types of Analyses
- Pooling Multiple Years of MEPS Data

Estimation From MEPS

(Producing Estimates & Computing Precision or Reliability)

Producing Estimates - Weights Must be Used

- **Sample weight**
  - ► **Indicates how many persons in the population a sample person represents**
- **Unequal sample weights due to**
  - ► **Oversampling of Blacks, Hispanics, Asians**
  - ► **Differential response rates**
- **Sample weights must be used to produce unbiased estimates**
  - ► **Unweighted estimates are biased**

Distribution of Final Positive Person Weights

Distribution of Weight	Year		
	2021	2022	2023
Minimum	229	342	502
Average	12,119	15,315	18,119
Maximum	106,959	117,276	131,657
Variable Name	PERWT21F	PERWT22F	PERWT23F

Final Person Weights - Positive versus Zero

- **Weight > 0 (i.e., positive)**
  - Persons Key and in-scope for survey
  - More than 95% cases
- **Weight = 0**
  - about 5% of cases every year
  - persons not key or in-scope for survey but living in households with in-scope person(s)
  - included for family analysis

Measures of Precision/ Reliability of Estimates

- Sampling error, Variance or Standard error
- Standard Error (SE) = $\sqrt{\text{Variance}}$
- Relative Standard Error (RSE)
  - SE of estimate ÷ estimate
  - also called Coefficient of Variation (CV)
- Confidence Interval (CI)
  - 95% CI: Estimate ± 1.96xSE

Example: Precision of Average Total Expenses, 2023

- **Sample Size 18,463**
- **Estimate = $7,487 (Average Expense per Capita)**
- **Standard Error = 173**
- **95% Confidence Interval**

  = ($7,487 ± 1.96x173, i.e., $7,145 to $7,830)
- **Relative Standard Error (RSE)**

  = (173 ÷ 7,487) x 100 = 2.3%

Computing Variances of Estimates from Complex Sample Design

- **Appropriate method must be used to compute standard errors to account for complex sample design**

  Software to account for complex design
  - SUDAAN (stand-alone or callable within SAS)
  - STATA (svy commands)
  - SAS 9.2 (survey procedures)
  - R (survey package)
  - Other (SPSS)

- **Assuming simple random sampling usually underestimates standard errors**

  Basic software procedures assume simple random sampling (SRS)
  - Estimates correct, if weighted
  - Standard errors usually smaller than actual

Example: Average Total Expenditures, 2023

- **Weighted mean = $7,487 per capita**
  Unweighted mean = $8,536 (biased)
- **SE assuming complex survey procedure = 173**
  - **SAS: PROC SURVEYMEANS**
  - **SUDAAN: PROC DESCRIPT**
  - **Stata: svy: mean**
  - **R: svymean**
- **SE assuming SRS = 146 (too low)**
  - **SAS: PROC UNIVARIATE or MEANS**

Example Codes to Produce Estimates and SEs

- **SAS V9.2**

```
proc surveymeans data=HC224 mean;
stratum varstr;   cluster varpsu;
weight perwt20f;
 var totexp20;
```

- **Stata**

```
svyset varpsu [pweight=perwt20f], strata(varstr)
svy: mean 2
```

- **SUDAAN (SAS-callable)**

```
First sort the file by varstr & varpsu
 proc descript data= HC224 filetype=SAS design=wr;
nest varstr varpsu; weight perwt20f;
var totexp20;
```

- **R**

```
mepsdsgn = svydesign(id = ~varpsu, strata = ~varstr, weights = ~perwt20f,
      data = HC224, nest = TRUE)
svymean(~ totexp20, design = mepsdsgn)
```

Computing Standard Errors for MEPS Estimates

- **Document on MEPS website**

http://www.meps.ahrq.gov/mepsweb/survey_comp/standard_errors.jsp

Domain Estimation
or
Analysis of Subpopulations

Analysis of Subpopulations – Special Procedure Needed

- **Analysis within specific subpopulation say within a Race-ethnicity, Poverty or Insurance status categories**

  Example: Asian 65+ years only OR Uninsured Hispanics

- **Special procedure or domain analysis must be used**

Analysis of Subpopulations – Avoid Subsetting the File

- Analyzing a subset file may produce incorrect standard errors
- A subset file may not contain all variance estimation information
- Software may give error messages in some situations
- Particularly important for analyzing small subpopulations that are not available in all PSUs
- Subsetting works for large subpopulations which are likely to be available in all PSUs such as males, females, children, elderly, etc.

Keywords for Specifying Subpopulations

- **Each software has special facility for subpopulation analysis using the entire file**
  - **SAS: domain**
  - **SUDAAN: subpopn**
  - **Stata: subpop**
  - **R: subset**

Example:

```
proc surveymeans data=HC233 mean;
stratum varstr; cluster varpsu;
weight perwt20f; var totexp20;
domain racethnx;
```

References on Analysis of Subpopulations

- **Variance Estimation from MEPS Event Files**

http://meps.ahrq.gov/mepsweb/data_files/publications/mr26/mr26.pdf

Family-Level Estimation

Family-Level Estimation

- Create a family-level file containing one record per family from person-level file (see instructions in the PUF documentation)
- Two family type options:
  - MEPS: Includes unmarried couples (starting with 2017, foster children excluded from MEPS)
  - Current Population Survey (CPS): Unmarried couples not family unit
- Two timeframe options:
  - December 31 (MEPS, CPS)
  - Any time during year (MEPS only)

Family Sample Sizes, 2020 (MEPS Annual Household File)

Number of Families	MEPS Anytime During Year	MEPS December 31	CPS December 31
Unweighted	12,290	12,176	12,733
Weighted (millions)	140.9	139.7	146.2
Family Weight Variable Name	FAMWT20F	FAMWT20F (subset to FMRS1231 = 1)	FAMWT20C

Family-Level Estimation

Example: Average total healthcare expenses per MEPS family by family size, 2020

- The estimation is based on MEPS families in scope at any time during the year
- Average number of persons per family is about 2.2

Family size	Estimate	SE
All	$13,960	$357
1	$9,909	$400
2	$16,926	$629
3	$15,617	$960
4	$15,573	$1,104
5+	$15,960	$1,214

Other Types of Estimations / Analyses

- Event-Level
- State-Level
- MEPS-HC Supplements

Event-Level
Medical Event as Unit of Analysis

- Event files can be used to estimate average expense per event
- Some estimates are available in HC summary tables on MEPS website

Example:
In 2020,

- Mean expense per office visit to a physician was $335 (SE = $15)
- Mean expense per emergency room (ER) visit was $1,150 (SE = $40)
- Mean expense per inpatient stay was $17,706 (SE = $898)

State-Level Estimation

- **MEPS sample is not designed to produce estimates for all states**
  - **Small sample sizes**
  - **Insufficient number of PSUs for variance estimation**
  - **Possible for larger states only**
- **To protect confidentiality, public use files (PUFs) do not include state identifiers**
  - **Tables and reports available on website for larger states (selected estimates)**
  - **Access to identifiers in AHRQ Data Center**
    - ***Need to use state-level sample design and state identifiers***

(Temporarily Suspended)

MEPS-HC Supplements

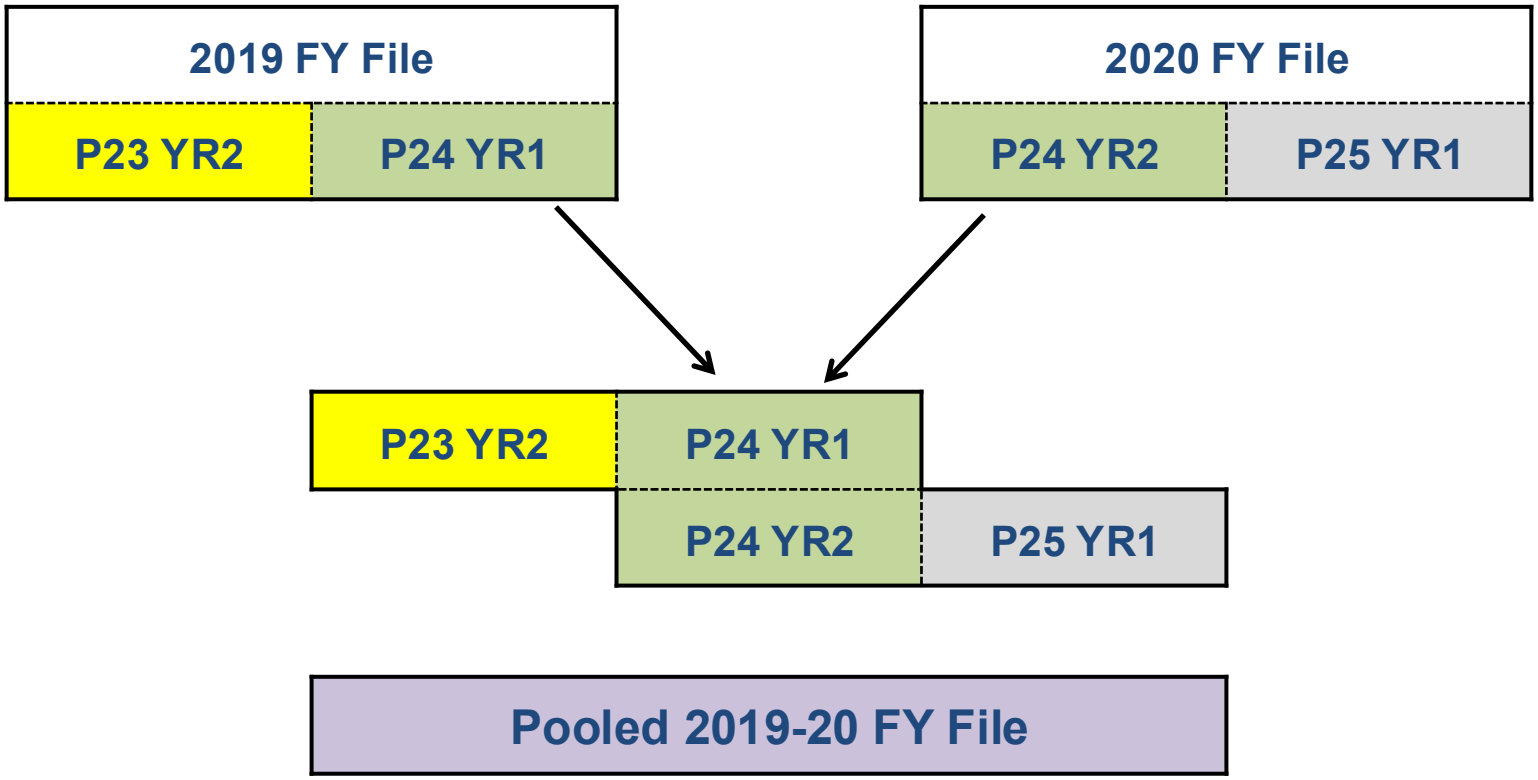
- **Special supplement variables available on person-level files**
- **Consult documentation for appropriate weight**
  - **Self-Administered Questionnaire (SAQ) → SAQWT*yy*F**
  - **Diabetes Care Survey (DCS) → DIABW*yy*F**
  - **Cancer SAQ (CSAQ) → CSAQW*yy*F**
    - ***MEPS full-year files 2011, 2016, and 2017 only***
  - **Veteran SAQ (VSAQ) → VSAQW*yy*F**
    - ***MEPS full-year files 2018 and 2019 only***

Pooling Multiple Years of MEPS Data

Reasons for Pooling

- Increase sample size
- Reduce standard errors of estimates
- Enhance ability to analyze small subgroups

Example of Pooling FY Files 2019-2020



Things to be Mindful of When Pooling

- Persons in the common panel are included twice
- Although correlated, data for the same person usually differ from year to year
- Each year represents nationally representative sample for that year
- Pooling produces average estimates across the pooled years
- Lack of independence diminishes the gain in precision from pooling

Accounting for Lack of Independence

- **MEPS panels are selected from the same sample PSUs and SSUs**
  - Correlation is not only at the person level, but persons within a PSU (segment/block) are also correlated
- **In multistage sampling, since PSU is the unit of sampling, specifying stratum (*VARSTR*) and PSU (*VARPSU*) in variance estimation is sufficient to account for all stages of correlation**

https://meps.ahrq.gov/survey_comp/hc_clustering_faq.pdf

Example of Pooled Sample Size & RSE

Adults 18-64 years old w/diabetes, by insurance status

	Sample Size		
Year	Privately Insured	Publicly Insured	Uninsured (all year)
2019	758	504	97
2020	777	588	120
2019-20 (Pooled)	1,535 person-years	1,092 person-years	217 person-years

	Relative Standard Error (RSE) (Standard error ÷ Point estimate)		
2019	8.3%	7.3%	26.4%
2020	9.0%	9.6%	32.7%
2019-20 (Pooled)	7.3%	6.8%	23.9%

Computing Standard Errors from Pooled Files

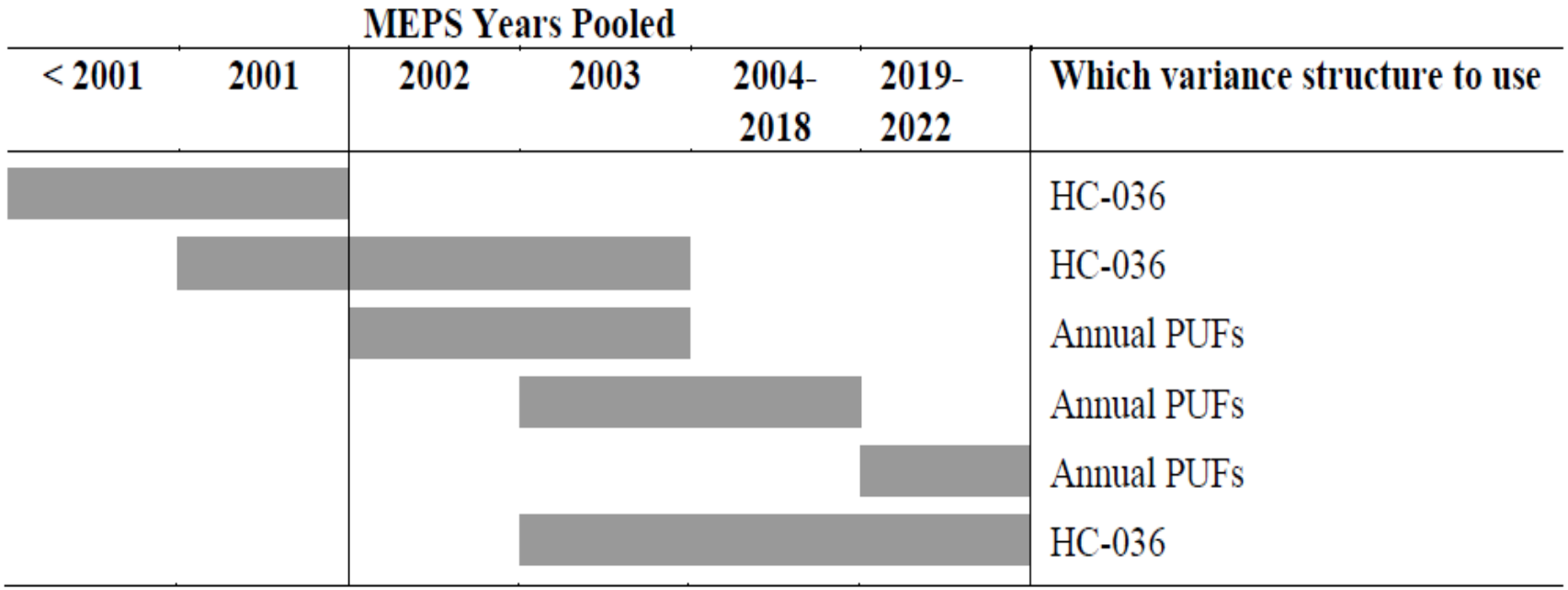
- **Variance structure (Stratum & PSU) not standardized across all years**
- **Pooled Estimation Linkage File (HC-036)**
  - Contains standardized stratum and PSU variables by DU ID
  - Stratum and PSU variables for 1996-2023 in HC-036 named

    STRA9623, *PSU9623*
  - Documentation for HC-036 provides instructions on how to properly create pooled analysis file
- **Use of standardized stratum/PSU from HC-036 file not always essential**
  - Not necessary when pooling involves certain years (but HC-036 can always be used)

When Use of HC-036 File is Essential

- **1996 – 2001**
  - Stratum/PSU variables on annual files are **not standardized across range or with later years**
  - Must always use standardized stratum/PSU identifiers from HC-036
- **2002 – 2018**
  - Stratum/PSU variables on annual files are standardized across range, but **not with preceding years (1996-2001) or succeeding years (2019 or later)**
  - When pooling restricted to these years, Stratum/PSU variables can be used from annual files
  - When pooling any year (2002-2018) with any years prior to 2002 or after 2018 use standardized stratum and PSU identifiers from Pooled Estimation Linkage File (HC-036)
- **2019 – 2023**
  - Stratum/PSU variables on annual files are standardized between these years, but **not with preceding years**
  - When pooling years 2019 or later, use stratum/PSU variables from annual files.
  - When pooling any year 2019 to 2023 with any preceding year, use HC-036

When to Use Pooled Linkage File (HC-036)



Steps for Creating Pooled FY Files

1) **Rename analytic and weight variables from different years to common names. For example,**
   - Expenditures: *TOTEXP22* & *TOTEXP23* = TOTEXP
   - Weights: *PERWT22F* & *PERWT23F* = POOLWT
2) **Concatenate annual files**
3) **Divide weight by number of years pooled to produce estimates for "an average year" during the period, but keep original weight if estimating total for the period**
4) **Merge standardized stratum/PSU variables from HC-036 onto file if necessary**
   - see previous slide or HC-036 documentation for guidance

Estimation from Pooled Files

- **Produce estimates in analogous fashion as for individual years**
- **Estimates interpreted as "average annual" for pooled period**

Example using 2019 & 2020 pooled data:

The average annual per capita health care expenses in 2019-20 was $6,259

Note: Per capita expenses were $6,252 in 2019 and $6,266 in 2020

Inflating Expenditures

- **Analyses involving multiple years**
  - Typically, adjust expenditures to most current MEPS data year (i.e., inflate previous year expenditures)
- **CFACT guidelines on appropriate indices varies by:**
  - Purpose of the analysis
  - Type of expenditure
- **Resource page (updated bi-annually)**

http://www.meps.ahrq.gov/mepsweb/about_meps/Price_Index.shtml

Guidance for Choosing Index

Objective of analysis	Recommended Index				
	GDP or PCE	CPI	PHCE or PCE-Health Total	PHCE Component	CPI-M
Trends in expenditures	X				
Trends in out-of-pocket expenditures only		X			
Pooling total expenditures			X		
Pooling expenditures by type of service (e.g., prescription meds)				X	
Pooling out-of-pocket expenditures					X
Projecting total expenditures			X		
Projecting total expenditures by service type				X	
Projecting out-of-pocket expenditures					X
Trends with income measures		X			

Notes: CPI = Consumer Price Index; GDP = Gross Domestic Product;
PCE = Personal Consumption Expenditures; PHCE = Personal Health Care Expenditures

Example of Inflating Expenditures

- **Nominal total expenditures for the U.S. civilian noninstitutionalized population**
  - 2010: $1,263 billion
  - 2020: $2,059 billion

- **Gross Domestic Product (GDP) to put in constant dollars**
  - Use yearly indices from resource page to determine inflation factor
  - 2020 index / 2010 index → 113.740 / 96.164 = 1.1827711

- **Inflation adjusted total expenditures**
  - 2010: $1,494 billion
  - 2020: $2,059 billion *(reference year)*
  - about $564 billion more was spent on health care in 2020 than 2010, after accounting for inflation.

Thank you

Sadeq.Chowdhury@ahrq.hhs.gov