
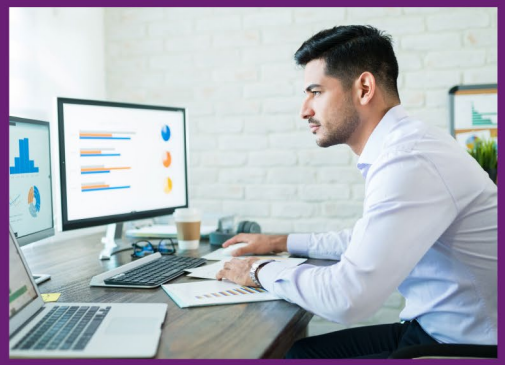


AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

35TH AHRQ ANNIVERSARY
Today's Research, Tomorrow's Healthcare

MEPS – HC

Estimation and Analysis

Sadeq Chowdhury, PhD

Outline

- Estimation from MEPS-HC
  - ► Computing/Producing Estimates
  - ► Computing Precision of Estimates
- Analysis of Subpopulations
- Family-Level Analysis
- Other Types of Analyses

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**
- **Weights must be used to produce unbiased estimates**
  - **Unweighted estimates are biased**

Distribution of Final Positive Person Weights

Distribution of Weight	Year		
	2019	2020	2021
Minimum	497	331	229
Average	10,574	12,238	12,119
Maximum	104,866	101,795	106,959
Variable Name	PERWT19F	PERWT20F	PERWT21F

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, 2021

- Sample Size 27,332
- Estimate = $6,934 (Average Expense per Capita)
- Standard Error = 189
- 95% Confidence Interval
  =($6,934 ± 1.96x189, i.e., $6,563 to $7,306)
- Relative Standard Error (RSE)
  = (189 ÷ 6,934) x 100 = 2.7%

Computing Variances of Estimates from Complex Sample Design

- Appropriate method must be used to compute standard errors to account for complex sample design
- Assuming simple random sampling usually underestimates standard errors

Computing Standard Error
(Precision of an Estimate)

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

Example: Average Total Expenditures, 2021

- **Weighted mean = $ 6,934 per capita**
  Unweighted mean = $ 7,870 (biased)
- **SE complex survey procedure = 189**
  - **SAS: PROC SURVEYMEANS**
  - **SUDAAN: PROC DESCRIPT**
  - **Stata: svy: mean**
  - **R: svymean**
- **SE assuming SRS = 167 (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 of the sample 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 is ok 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, 2021 (MEPS Annual Household File)

Number of Families	MEPS Anytime During Year	MEPS December 31	CPS December 31
Unweighted	13,014	12,910	13,461
Weighted (millions)	142.5	141.3	147.8
Family Weight Variable Name	FAMWT21F	FAMWT21F (subset to FMRS1231 = 1)	FAMWT21C

Family-Level Estimation

Example: Average total healthcare expenses per MEPS family by family size, 2021

- The estimation is based on MEPS families in scope at any time during the year
- Average number of persons per family is about 2.3

Family size	Estimate	SE
All	$15,526	$436
1	$10,719	$443
2	$18,772	$748
3	$18,121	$1,757
4	$17,741	$1,293
5+	$17,873	$1,371

Estimation and Analysis

- Event Level
- State Level
- MEPS-HC Supplements

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 2021,

- the mean expense per office visit to a physician was $365 (SE = $16)
- the mean expense per emergency room (ER) visit was $1,164 (SE = $48)
- the mean expense per inpatient stay was $16,809 (SE = $639)

State-Level Estimation

- Considerable interest in state-level estimates
- MEPS sample is not designed to produce estimates for all states
  - Small sample sizes
  - Insufficient number of PSUs for variance estimation
- 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***

MEPS-HC Supplements

- **Special supplement variables 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***

Thank you!

Sadeq.Chowdhury@ahrq.hhs.gov