

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

MEPS – HC

Design and Estimation

Sadeq Chowdhury, PhD
Sadeq.Chowdhury@AHRQ.hhs.gov

Outline

- MEPS-HC Sample Design
- Estimation from MEPS-HC
  - ► Producing Estimates
  - ► Computing Standard Errors
- Analysis of Subpopulations
- Pooling Multiple Years of MEPS-HC Data

Sample Design

Features of MEPS Sample

- MEPS sample is a sub-sample of National Health Interview Survey (NHIS)
- Each year a new panel of sample is selected from responding households to the previous year’s NHIS
- Each Panel is followed for 2 years using 5 interview rounds
- MEPS full sample for each year is an overlap of 2 panels
- Subpopulations of interest are oversampled

MEPS Sample Design – Inherited from NHIS

- NHIS sample is based on complex stratified area sample design
- Hence MEPS is based on the same complex design
- Complexity of the sample design affects the accuracy of a survey estimate
- Why complex multistage design instead of simple design?

Simple Vs. Complex Design

- **Single Stage Simple Random Sampling**

  - List of all sampling units available

  - One stage selection

  - Equal Probability

  - Sample from all areas

Example: A sample of 10,000 persons selected directly from a list of all persons in the U.S.

- Efficient design i.e., estimates are more accurate

- Expensive to create frame and collect data

NHIS Stratified Multistage Area Sample Design Until 2016 (MEPS 2017)

- **First Stage or Primary Sampling Units (PSUs)**
  - **A PSU is a county or group of adjacent counties**
  - **Whole U.S. is partitioned into many PSUs**
  - **PSUs grouped into homogeneous design strata**
  - **PSUs sampled for NHIS, roughly half used in MEPS**

- **Second Stage Units (SSUs)**
  - **An SSU is a cluster of housing units (Census blocks or tracts)**
  - **Each sampled PSU is divided into SSUs**
  - **A sample of SSUs selected from each selected PSU**

NHIS Stratified Multistage Area Sample Design Until 2016 (MEPS 2017)

- **Final Stage Units**
  - ► **Sample of households from each selected SSUs**
  - ► **All families and persons within selected households are included**
- **Same PSUs and SSUs but different HHs**
  - ► **Every year the sample is selected from the same PSUs and SSUs but different households (hence different families and persons), unless a redesign of NHIS (roughly every 10 years)**

NHIS Sample Redesign 2016 (MEPS 2017)

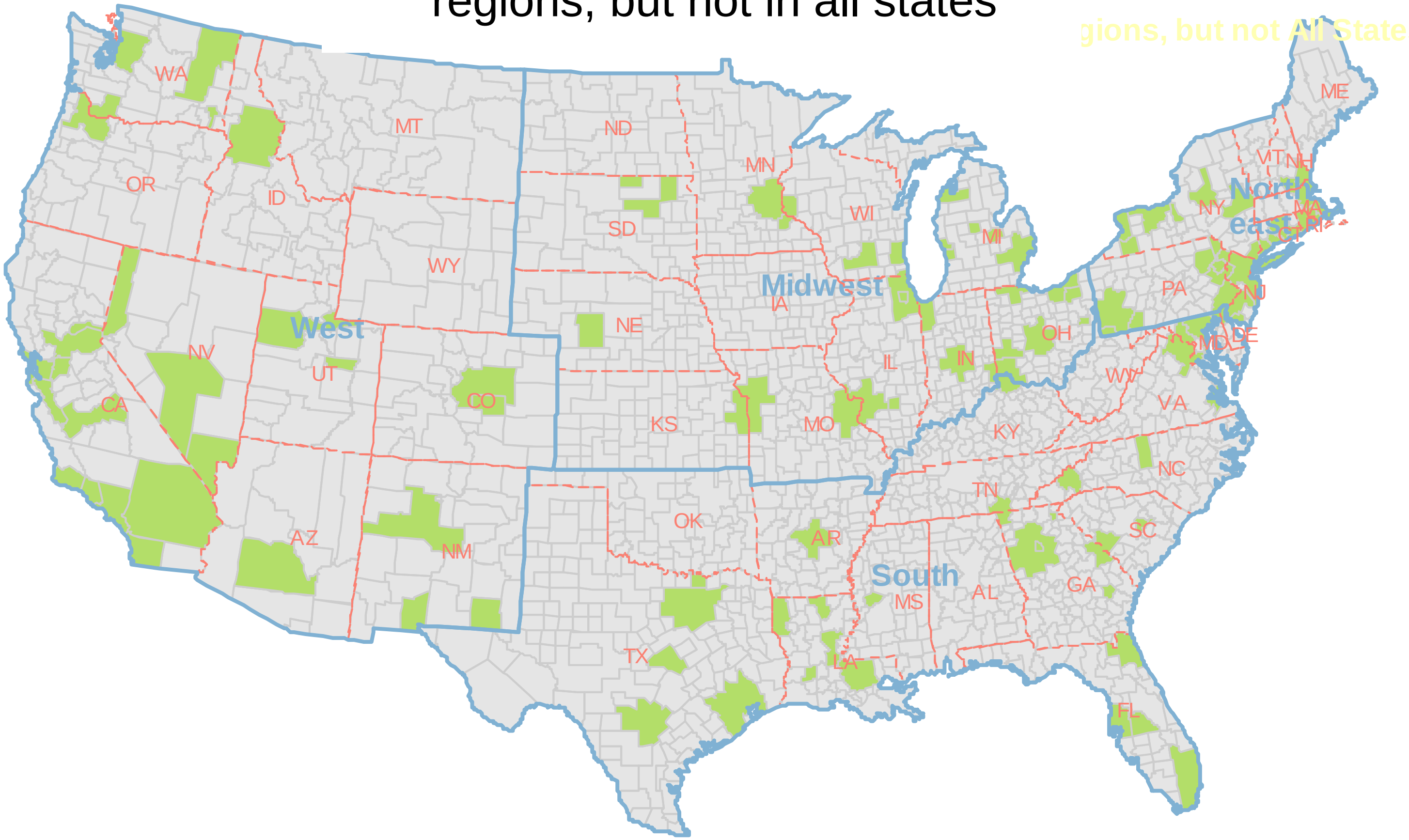
- NHIS sample is redesigned every 10 years
- A new design was introduced in 2016
- Stratification by State for State-level estimation
- PSUs formed and selected as before
- Households selected directly from USPS list of addresses within PSUs
  - USPS list available for most of the country
  - No need for listing of households
- Roughly 100 addresses (equal 1 cluster) selected from each PSU

NHIS Sample Redesign 2016 (MEPS 2017)

- Multiple clusters were selected from large PSUs
- A cluster includes many sub-clusters of 4 addresses
- Sub-clusters selected systematically from the PSU-wide list of addresses
- Traditional multistage design not needed anymore
- MEPS Panel 2017 based on the new design
- Same PSUs used for 10 years but different clusters every year

Illustration of Hypothetical 100 PSU Sample

Sample sufficient in all regions, but not in all states



Oversampling in MEPS

- To produce reliable estimates for subpopulations of interest
- Oversampled subpopulations
  - Asians
  - Blacks
  - Hispanics
  - Veterans (2018 panel)
- Increases variation in sampling weights

MEPS Overlapping Panel Design

<table>
<tr><th></th><th colspan="5">2016</th><th colspan="5">2017</th><th colspan="5">2018</th></tr>
<tr><td>Panel 21</td><td colspan="2">R1</td><td colspan="2">R2</td><td colspan="2">R3</td><td colspan="2">R4</td><td colspan="2">R5</td><td colspan="5"></td></tr>
<tr><td>Panel 22</td><td colspan="5"></td><td colspan="2">R1</td><td colspan="2">R2</td><td colspan="2">R3</td><td colspan="2">R4</td><td colspan="2">R5</td></tr>
</table>

FY 2017

- Panel 21: R3, R4, R5
- Panel 22: R1, R2, R3

MEPS Annual Files – Combination of Two Panels

Panel	Year			
	2014	2015	2016	2017
18	Yr2			
19	Yr1	Yr2		
20		Yr1	Yr2	
21			Yr1	Yr2
22				Yr1

Estimation From MEPS

(Producing Estimates &

Computing Standard Errors)

Producing Estimates - Weights Must be Used

- 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		
	2015	2016	2017
Minimum	637	572	497
Average	9,483	9,716	10,573
Maximum	98,104	99,173	104,865
Variable Name	PERWT15F	PERWT16F	PERWT17F

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, 2017

- Sample Size = 30,716
- Estimate = $5,306 (Average Expense per Capita)
- Standard Error = 126
- 95% Confidence Interval

  =($ 5,306 ± 1.96x126, i.e., $5,059 to $5,553)
- Relative Standard Error (RSE)

  = (126 ÷ 5,306) x 100 = 2.4%

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)
  - ► Other (SPSS and R)

Example:
Average Total Expenditures, 2017

- **Weighted mean = $ 5,306   per capita**
  Unweighted mean = $ 5,111   (biased)
- **SE complex survey procedure = 126**
  - ► **SAS:** **PROC SURVEYMEANS**
  - ► **SUDAAN:** **PROC DESCRIPT**
  - ► **Stata:** **svy: mean**
- **SE assuming SRS = 87   (too low)**
  - ► **SAS:** **PROC UNIVARIATE or MEANS**

Example Codes to Produce Estimates and SEs

- SAS V9.2
  proc surveymeans data=HC201 mean;
  stratum varstr;   cluster varpsu;
  weight perwt17f;   var totexp17;

- Stata
  svyset varpsu [pweight=perwt17f], strata(varstr)
  svy: mean 2

- SUDAAN (SAS-callable)
  First sort the file by varstr & varpsu
  proc descript data=HC201 filetype=SAS design=wr;
  nest varstr varpsu; weight perwt17f;
  var totexp17;

Computing Standard Errors for MEPS Estimates

- **Document on MEPS website**

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

Analysis of Subpopulations
(Domain Analysis)

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**

Example

proc surveymeans data=HC201 mean;
stratum varstr; cluster varpsu;
weight perwt17f; var totexp17;
domain racethnx;

References on Analysis of Subpopulations

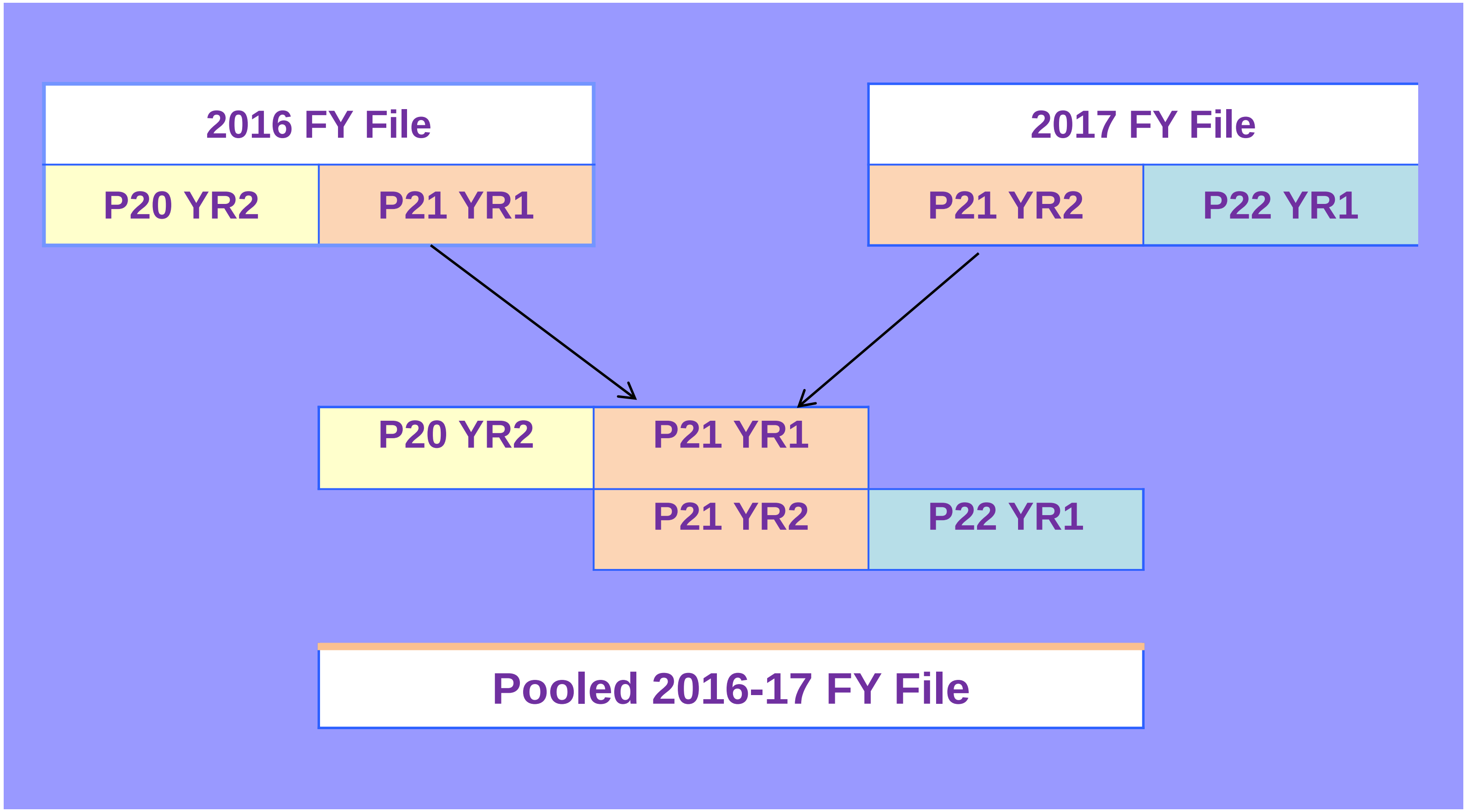
- **Computing Standard Errors for MEPS Estimates**
  - http://www.meps.ahrq.gov/mepsweb/survey_comp/standard_errors.jsp
- **Variance Estimation from MEPS Event Files**
  - http://meps.ahrq.gov/mepsweb/data_files/publications/mr26/mr26.pdf

Pooling Multiple Years of MEPS Data

Reasons for Pooling

- Increasing sample size
- Reducing standard errors of estimates
- Enhancing ability to analyze small subgroups

Example: Pooling 2016-2017



Pros and Cons of 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
- So 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 and PSU in variance estimation is sufficient to account for all stages of correlation
- https://meps.ahrq.gov/survey_comp/hc_clustering_faq.pdf

Example: Pooled Sample Sizes

For Adults age 18-64 with diabetes, by insurance status

	Sample Size		
Year	Privately Insured	Publicly Insured	Uninsured (all year)
2016	873	548	204
2017	844	553	137
2016-17 (Pooled)	1,717 person-yrs	1101 person-yrs	341 person-yrs

Example: Relative Standard Errors

of Avg. Annual Expenditures, Adults Age 18-64 with Diabetes, by Insurance Status

	Relative Standard Error (SE÷Estimate)		
Year	Privately Insured	Publicly Insured	Uninsured (all year)
2016	5.8%	8.0%	18.4%
2017	7.3%	6.6%	17.4%
2016-17 Pooled	5.1%	5.7%	14.6%

Computing Standard Errors from Pooled File

Use standardized stratum and PSU variables for variance estimation

- **Pooling annual data from 2002 onward**
  - ► Annual files already contain standardized stratum (varstr) and PSU (varpsu) variables
- **Pooling annual data from any year before 2002**
  - ► Use Pooled Estimation Linkage File (HC-036)
  - ► Stratum and PSU variables obtained from HC-036 for 1996-2017 (stra9617, psu9617)
  - ► Documentation for HC-036 provides instructions on how to properly create pooled analysis file

Creating Pooled Files
Summary of Important Steps

1. Rename analytic and weight variables from different years to common names. **Example:**
   - ► Expenditures: TOTEXP16 & TOTEXP17 = TOTEXP
   - ► Weights: PERWT16F & PERWT17F = POOLWT
2. Concatenate annual files
3. Divide weight by number of years pooled to produce estimates for "an average year" during the period.
   - ► Keep original weight if estimating total for the period
4. Merge variance estimation variables from HC-036 onto file (only if any year prior to 2002)
   - ► Strata variable: STRA9617
   - ► PSU variable: PSU9617

Estimation from Pooled Files

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

Example: Pooled 2016-17 data

The average annual per capita health care expenses in 2016-17 was $5,156 (SE=$96)

Inflating expenditures

- Analyses involving multiple years
  - Typically adjust expenditures to most current MEPS data year
- CFACT guidelines on appropriate indices
  - Varies by…
    1) purpose of the analysis
    2) type of expenditure
- Resource page

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

Crosswalk of price indices and MEPS analyses

	Recommended Index			
Objective of analysis	**GDP or PCE**	**CPI**	**PHCE or PCE-Health Total**	**PHCE Component**
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
Trends with income measures		x		