



AGENCY FOR HEALTHCARE RESEARCH AND QUALITY



# **MEPS-HC: Using the Longitudinal Files, pooling multiple years of data, and other analytic topics**

---

**Marc Zodet, MS**

# Overview

- **Longitudinal analysis using panel files**
  - Structure of the Longitudinal (Panel) files
  - Available variables
  - Types of analyses supported
  - Survey design variables to use
  - Using with other MEPS data files
  - Extending the longitudinal period
- **Pooling multiple years of MEPS data**
  - Full-year consolidated files
  - Longitudinal (Panel) files
- **Use of price indices**

# Longitudinal analysis using panel files

# Panel structure

- **Full-Year (FY) Consolidated file refresher**
  - Person-level files
  - Comprised of two MEPS Panels
  - Used to generate annual estimates for a given year
  - Examine trends in those estimates over time
  - *Not useful for examining individual person-level changes over time*
- **Longitudinal (Panel) files**
  - Person-level files
  - Respondent data for two years in one file rather than split across two annualized FY files
  - *Useful for examining individual person-level changes over time*

# Panel structure

## Example:

### Panel 24 (2019–2020)



# Available variables

- **Insurance coverage**

- Monthly indicators (24 measures per person)
- Annual summary (2 measures per person)

- **Health status**

- Each round (5 measures for perceived general/mental health)
- Rounds 1 & 3 (2 measures activities of daily living)
- Rounds 2 & 4 (2 measures hearing, vision, & disability)

- **Having a usual source of care**

- Rounds 2 & 4 (2 measures per person)

- **Use and expenditures**

- Annual (2 measures per person)

# Available variables: case selection

Variable	Description
YEARIND	1=both years, 2=in Year 1 only, and 3=in Year 2 only
ALL5RDS	In-scope and data collected in all 5 rounds (0=no, 1=yes)
DIED	Died during the two-year survey period (0=no, 1=yes)
INST	Institutionalized for some time during the two-year survey period (0=no, 1=yes)
MILITARY	Active-duty military for some time during the two-year survey period (0=no, 1=yes)
ENTRSRVY	Entered survey after beginning of panel (mainly births; also includes persons who had no initial chance of selection who moved into a MEPS sample household) (0=no, 1=yes)
LEFTUS	Moved out of the country after beginning of panel (0=no, 1=yes)
OTHER	Not identified in any of the above analytic groups (0=no, 1=yes)

# Types of analyses supported

- National estimates of person-level changes over 2-year period
- Examination of characteristics associated with changes over time

*Examples later*



# Survey design variables to use

- **As with FY files need to use survey design variables to derive estimates**
  - Analytic weight - to yield national estimates
  - Stratum and PSU - to yield proper standard errors
- **Stratum (*VARSTR*) and PSU (*VARPSU*) are same as on FY**
- **Longitudinal files utilize a different analytic weight than the FY: *LONGWT***

# Survey design variables to use

- **Why *LONGWT*?**

- Longitudinal files have only about half the records of FY
- Persons in the Panel who did not participate in the survey for the entire period they were in-scope are excluded; *LONGWT* adjusted for this nonresponse/attrition

- ***LONGWT* yields national estimates for persons in two consecutive years**

- **For Panel 24 (2019-2020) . . .**

<b>LONGWT &gt; 0</b>	<b>All 5 Rounds</b>	<b>Participated for entire period in-scope (not all 5 rounds)</b>
9,797	9,120 (93.1%)	677 (6.9%)

# Estimates from the longitudinal files

## Examples using Panel 24 (2019-20):

- Of those uninsured throughout 2019, an estimated 74.3% were also uninsured throughout 2020.
- An estimated 5.0% of the population had no insurance throughout 2019 / 2020.
- Of those with no expenses for health care in 2019, an estimated 42.8% had some expenses in 2020.

# Using with other MEPS data files



- **Medical Conditions files**

- Can be used to identify persons with specific conditions of interest
- Directly linkable to Longitudinal files via *DUPERSID*

- **Event-level files**

- Payment amounts/sources already “rolled-up” on longitudinal files
- Other event characteristics can be obtained (e.g., number of office-based visits involving labs, prescribed medicines, etc.)
- Directly linkable to Longitudinal files via *DUPERSID*

- **More complex linking using CLNK and RXLK**

- MEPS data page under **Appendix to MEPS Event files**

# IDs used to link MEPS files

- **Longitudinal files** (*DUPERSID*)
- **Medical Conditions files** (*DUPERSID, CONDIDX*)
- **Event files** (*DUPERSID, EVNTIDX*)
- **CLNK** (*DUPERSID, CONDIDX, EVNTIDX*)
- **RXLK** (*DUPERSID, EVNTIDX, LINKIDX*)

# Example of generalized linking process



## Examine healthcare utilization/expenditures for persons with asthma over a two-year period

- ID persons w/ asthma in Medical conditions files (2 years needed)
- If data on Longitudinal files is sufficient merge asthma indicators directly onto the file (*DUPERSID*)
- If need event-level info (e.g., expenditures for services related to asthma), merge CLNK (*CONDIDX*) then desired event-level data (*EVNTIDX*; 2 years)
- Prescribed medicine events are not directly linked to conditions; link PMED event file to RXLK file (*LINKIDX*) then linkable to conditions/other event files via CLNK (*EVNTIDX/CONDIDX*)

**NOTES:** With all file merges, be sure to only keep the Panel of interest.  
See CLNK/RXLK doc for SAS and STATA programing examples.

# Extending the longitudinal period: MEPS-NHIS



- **MEPS-HC is a nationally representative subsample of responding households from the previous year's NHIS.**
  - Prior-year NHIS data available for many MEPS respondents
- **MEPS/NHIS link file**
  - Crosswalk to merge MEPS full-year public use data to NHIS person-level public use data
  - Crosswalk file not public use; available in AHRQ Data Center

# Extending the longitudinal period: MEPS-NHIS

2018		2019		2020
		MEPS Panel 23 Year 2		
NHIS 2018	➡	MEPS Panel 24 Year 1		MEPS Panel 24 Year 2
		NHIS 2019	➡	MEPS Panel 25 Year 1



# Linked files, weighting, and estimation



- **Linking MEPS-NHIS expands the analytic capabilities.**
- **Not all MEPS respondents link (birth, marriage, etc.)**
  - NHIS now collects data only on a sample adult / sample child
  - MEPS can only link to NHIS sample adult / child
- **Weighting adjustment for non-linkage is recommended**
  - Necessary since NHIS move to sample adult / child
- **Informational resources are available online:**

2012 American Statistical Association proceedings paper (Chowdhury, Machlin, and Wun)

[https://meps.ahrq.gov/mepsweb/data\\_stats/Pub\\_ProdResults\\_Details.jsp?pt=Conference+Proceedings&opt=3&id=1241](https://meps.ahrq.gov/mepsweb/data_stats/Pub_ProdResults_Details.jsp?pt=Conference+Proceedings&opt=3&id=1241)

2013 Federal Committee on Statistical Methodology proceedings paper (Mirel and Machlin)

[https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/H2\\_Mirel\\_2013FCSM.pdf](https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/H2_Mirel_2013FCSM.pdf)

# Longitudinal files

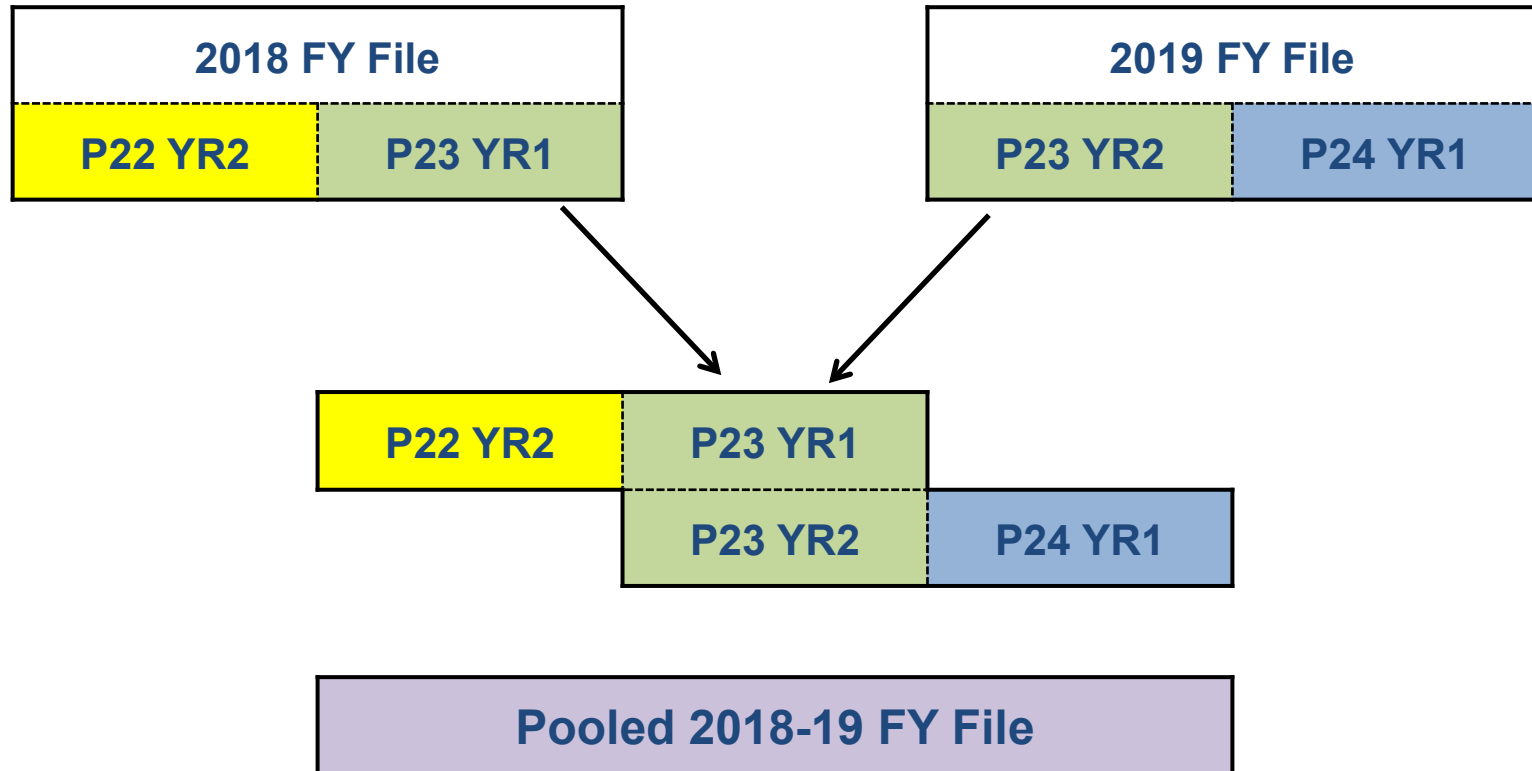
Panel Number	Years	File Number	Number of Persons
24	2019-20	HC-225	9,797
23	2018-19	HC-217	14,067
22	2017-18	HC-210	15,541
21	2016-17	HC-202	15,617
20	2015-16	HC-193	17,017
19	2014-15	HC-183	15,898
18	2013-14	HC-172	16,714
17	2012-13	HC-164	17,293
16	2011-12	HC-156	18,512
15	2010-11	HC-148	14,541
14	2009-10	HC-139	16,221

# Pooling multiple years of MEPS data

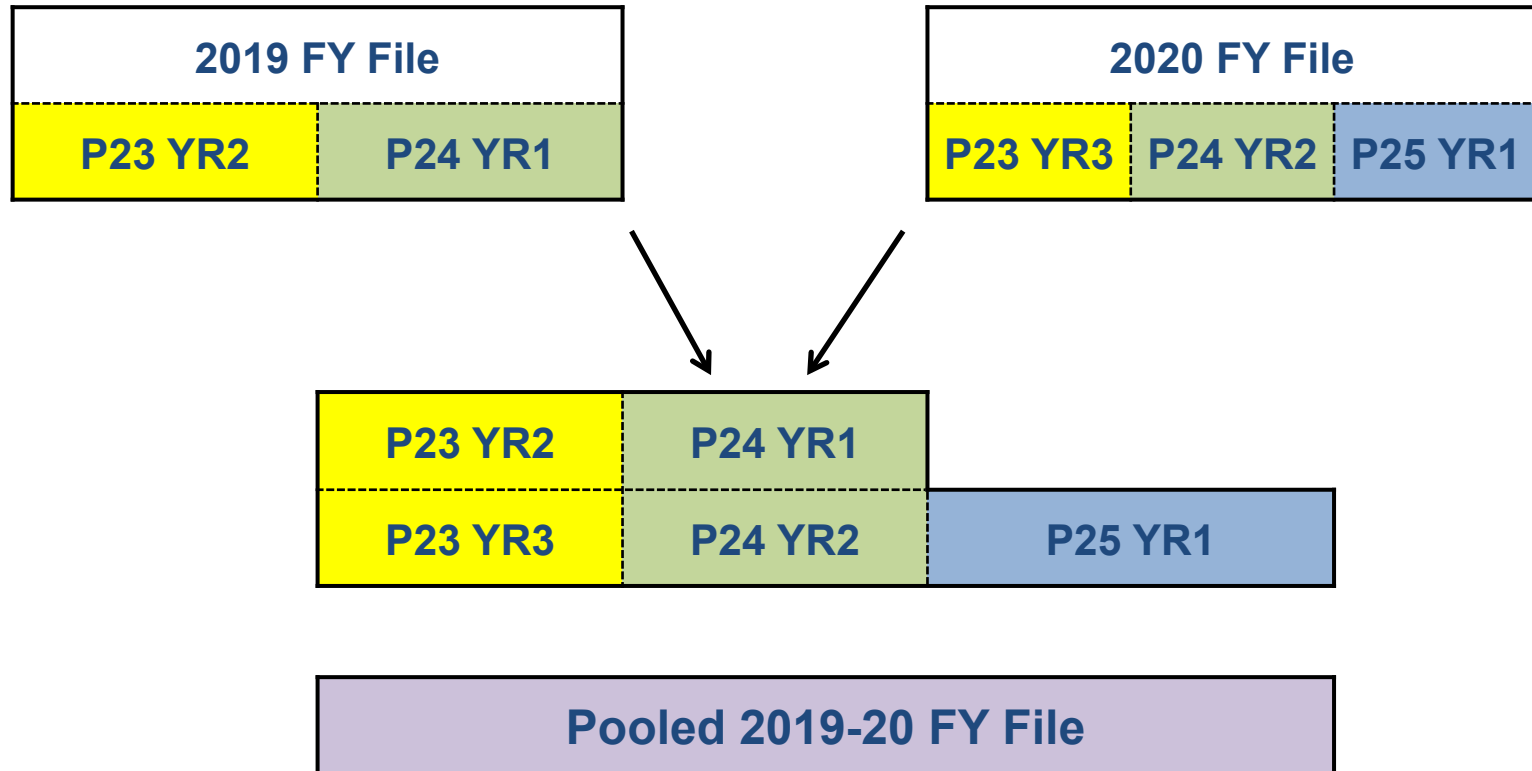
# Reasons for pooling

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

# Example of pooling FY files 2018 & 2019



# 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](https://meps.ahrq.gov/survey_comp/hc_clustering_faq.pdf)



# Example of pooled sample sizes

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

# Example of pooled RSEs of mean annual expenditures

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

	<b>Relative Standard Error (RSE)</b> (Standard error ÷ Point estimate)		
<b>Year</b>	<b>Privately Insured</b>	<b>Publicly Insured</b>	<b>Uninsured (all year)</b>
<b>2019</b>	<b>8.3%</b>	<b>7.3%</b>	<b>26.4%</b>
<b>2020</b>	<b>9.0%</b>	<b>9.6%</b>	<b>32.7%</b>
<b>2019-20 (Pooled)</b>	<b>7.3%</b>	<b>6.8%</b>	<b>23.9%</b>

# Caveat to computing standard errors from pooled files



- **Variance structure not standardized for all years**
- **Pooled Estimation Linkage File (HC-036)**
  - Contains standardized stratum and PSU
  - Stratum and PSU variables obtained from HC-036 for 1996-2020  
*STRA9620, PSU9620*
  - Documentation for HC-036 provides instructions on how to properly create pooled analysis file
- **Need to be mindful of what years you intend to pool and understand which stratum and PSU variables to use**

# Caveat to computing standard errors from pooled files



- **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 or 2019 and 2020**
- When pooling restricted to these years use stratum/PSU variables from annual files
- When pooling with any years prior to 2002 or with 2019/2020 use standardized stratum and PSU identifiers from Pooled Estimation Linkage File (HC-036)

- **2019, 2020**

- Stratum and PSU variables on annual files are standardized between these two years, but **not with preceding years**
- When pooling 2019 and 2020 use stratum and PSU variables from annual files.
- When pooling either 2019/2020 with any preceding year, use HC-036

# Steps for creating FY pooled files

- 1) **Rename analytic and weight variables from different years to common names. For example,**
  - Expenditures: *TOTEXP19* & *TOTEXP20* = TOTEXP
  - Weights: *PERWT19F* & *PERWT20F* = 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 if necessary**
  - see previous slide / 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*

# Steps for creating Panel pooled files

- **The objective is the same**
  - Increase sample size
  - Improve precision of estimates (i.e., reduce standard errors)
  - Enable the analysis of smaller subgroups
- **Process**
  - Generally, no need to rename variables from different Panel years  
*caveat: some variables may not be present all years*
  - Append/concatenate/stack multiple panel files
  - If including Panels 1-6 must use HC-036 Pooled Estimation File
  - Standardize expenditure dollars to a reference year
  - Decide if need to divide estimates by the number of Panels pooled.

# Estimates from pooled Panel files

- **When pooling multiple panels temporal comparisons can still be made (e.g., Y1 vs. Y2)**
- **Averages/proportions – direct results/output fine**
- **Totals – divide by number of panels pooled;**
  - average annual
- **All references should be relative to temporal constructs; not specific to particular year or range of years.**



# Estimates from pooled Panel files



- **Hypothetical pooling of five most recent panel files**
  - P24 (2019-20), P23 (2018-19), P22 (2017-18), P21 (2016-17), P20 (2015-16)
- **Possible statements**
  - During 2015 to 2019, when considering two consecutive years, of those uninsured throughout the first year, an estimated  $x\%$  were also uninsured the subsequent year.
  - During 2015 to 2019, when considering two consecutive years, there was an average annual increase/decrease in total healthcare expenditures of \$ $x$  from year one to year two

# Use of price indices

# 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](http://www.meps.ahrq.gov/mepsweb/about_meps/Price_Index.shtml)

# Guidance for choosing index

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

**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  $\rightarrow 113.740 / 96.164 = 1.1827711$
- **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



[marc.zodet@ahrq.hhs.gov](mailto:marc.zodet@ahrq.hhs.gov)