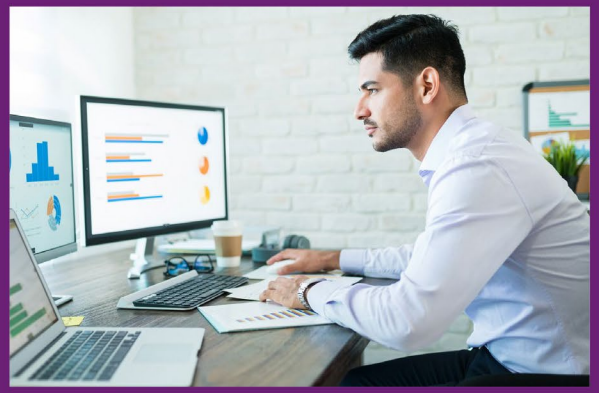




AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

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

MEPS-HC: Using Longitudinal Files & Pooling multiple years of data

Philippe Gwet, PhD

Cross-sectional vs. Longitudinal Analysis

- **Cross-sectional Analysis**
  - A snapshot of the US population at a given point in time
  - Example: The *2021 mean annual expense for the elderly (65+) with any healthcare expenses is $14,683 (SE=$527)*

- **Longitudinal Analysis**
  - Quantifies gross changes in the US population over 2 or more years.
  - Example: *4.9% of the US population had no insurance throughout 2020-2021*

Overview

- **Longitudinal Analysis / Panel Files**
  - Structure of the longitudinal files
  - Available variables
  - Types of analyses supported
  - Survey design variables to use
  - Using with other MEPS data files

- **Pooling Multiple Years of MEPS Data**
  - Full-year consolidated files

File Types

- **Full-Year (FY) Consolidated File**
  - Person-level files
  - FY2021 based on 4 MEPS panels, FY2020 based on 3 panels, FY2019 and prior are based on 2 panels.
  - Used to generate annual estimates for a given year
  - Used to examine trends over time / net changes
  - *Cannot be used to evaluate person-level changes over time*

- **Longitudinal (Panel) Files**
  - Person-level files
  - Respondent data for 2 years or more in one file.
  - *Useful for examining person-level changes over time / net changes*

Longitudinal Analysis using Panel Files

MEPS Longitudinal Weight Files

MEPS Panel	Years Covered	PUF Number
25	2020-21	HC-234
24	2019-21	HC-235
24	2019-20	HC-225
23	2018-21	HC-236
23	2018-20	HC-226
23	2018-19	HC-217
22	2017-18	HC-210
1	1996-97	HC-080

MEPS Panel Design

Available Variables

- **Insurance coverage**
  - Monthly indicators (24 measures per person)
    Examples:
    COV BY MEDICAID OR SCHIP IN JUN20
    COV BY MEDICAID OR SCHIP IN JUN21
  - Annual summary (2 measures per person)
    Examples:
    COV BY MEDICAID OR SCHIP - 12/31/20
    COV BY MEDICAID OR SCHIP - 12/31/21

- **Access to Care**
  - 2+ variables per person (Rounds 2, 4, …)
    Examples:
    AFRDCA2: COULD NOT AFFORD MED CARE-R2
    AFRDCA4: COULD NOT AFFORD MED CARE-R4

Available variables

- **Health status**
  - Each round (5+ measures for perceived general/mental health)
    Example: MNHLTH1: Perceived Mental Health Status-Rd 1
  - Rounds 1,3… (2+ measures activities of daily living)
    Example: LFTDIF1: Difficulty Lifting 10 Pounds-Rd 1
  - Rounds 2,4,… (2+ measures hearing, vision, & disability)
    Example. DFHEAR42=Serious Difficulty Hearing-Rd 4/2

- **Use and expenditures**
  - Annual (2+ measures per person)
    Examples:
    OBVEXPY1: 2020 Total Office-Based Expenditures
    OBVEXPY2: 2021 Total Office-Based Expenditures
    OBTOTVY1: # 2020 Office-Based Provider Visits
    OBTOTVY2: # 2021 Office-Based Provider Visits

Available Variables: Case Selection

Variable	Description
YEARIND	Example (2020-2021, Panel 25): 1=both years, 2=in Year 1 only, and 3=in Year 2 only
ALL5RDS (Panel 25) (2020-2021)	In-scope and data collected in all 5 rounds (0=no, 1=yes)
ALL7RDS (Panel 24) (2019-2021)	In-scope and data collected in all 7 rounds (0=no, 1=yes)
ALL9RDS (Panel 23) (2018-2021)	In-scope and data collected in all 7 rounds (0=no, 1=yes)
SAQELIY1	Year 1 SAQ Eligibility Status
SAQELIY2	Year 2 SAQ Eligibility Status
SAQRDS24	SAQ Respondent in Both Rounds 2 and 4

Self-Administered Questionnaire (SAQ) includes core questions about health status, health care quality & preventive health care measures of adults.

Available Variables: Case Selection

Variable	Description
YEARIND	Example (2020-2021, Panel 25): 1=both years, 2=in Year 1 only, and 3=in Year 2 only
DIED	Died during the two-year survey period (0=no, 1=yes)
INST	Institutionalized during the 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 it began
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, 3-year, 4-year periods**
- **Examination of characteristics associated with changes over time**
- **Analysis of variables from the Self-Administered Questionnaire (SAQ)**

  Example

  ADSAD2: How often felt sad, R2

  ADSAD4: How often felt sad, R4

Conducting Longitudinal Analysis

- **Identify the Correct Longitudinal Dataset**
- **Select Analytic Variables of Interest**
- **Find the Correct Weight Variable:**
  - *LONGWT: Longitudinal Weight*
  - *LSAQWT: Longitudinal SAQ Weight*

  Weighting to obtain unbiased estimates
- **Identify the Design Variables**
  - 2 Design Variables needed for Standard Error calculation
  - Stratum Variable: **VARSTR**. PSU Variable: **VARPSU**

Using Longitudinal Weights

- **Why *LONGWT and LSAQWT?***
  - In-scope nonrespondents are excluded for the file.
  - The *LONGWT* and *LSAQWT* must adjust FY weights for nonresponse/attrition
- *LONGWT* and *LSAQWT* yield national estimates of gross changes in 2+ consecutive years
- **For Panel 25 (2020-2021)**

Weight	Weight > 0	Weight=0	All 5 Rounds
LONGWT	6,078 (100.0%)	0	5,681 (93.5%)
LSAQWT	3,455 (56.8%)	2,623 (43.2%)	

Examples: Longitudinal Estimates

- 72.2% of those without insurance in 2021 were uninsured in 2020 (SE=2.8%).
- An estimated 4.9% (SE=0.5%) of the population had no insurance throughout 2020-2021.
- Of those with no healthcare expenses in 2020, an estimated 53.4% (SE=3.0%) had some expenses in 2021.
- 32.2% of the top 5% with the highest healthcare expenses in 2021, had that position in 2020.

Linking Longitudinal & Other MEPS Data Files

- **Longitudinal Files Provide Various Total Annual Expenditures for each Individual Respondent**

- **No Disease-Specific Expenditures**
  - Example:
    A patient has $5,000 of annual office-based visit expenditures in 2021. What portion of it can be attributed to the treatment of mental illness?
- **Use the Condition-Event Link File (or CLNK File)**
- **Use the PMED-Event Link File (or RXLK File)**

Linking Longitudinal & Other MEPS Data Files

- **Medical Conditions files**
  - Can be used to identify persons with specific conditions of interest (e.g.: Arthritis, Asthma, High Cholesterol, ...)
  - 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*

IDs used to Link MEPS Files

- **Longitudinal files (*DUPERSID*)**
- **Medical Conditions files (*DUPERSID*, *CONDIDX*)**
- **Event files (*DUPERSID*, *EVNTIDX*)**
- **CLNK (*DUPERSID*, *CONDIDX*, *EVNTIDX*)**
  CLNK files link Condition files to Event files
- **RXLK (*DUPERSID*, *LINKIDX*,*EVNTIDX*)**
  PMED-event link files: Link Prescribed Medicines Event Files to other MEPS event files

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*)
  Link 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.

Pooling Longitudinal Files

- **Options for pooling multiple panels to increase the sample size are limited**
  - 2019-2020 (2 years) longitudinal analysis can be based on data from panels 23 and 24
  - 2019-2021 (3 years) longitudinal analysis can be based on data from panels 23 and 24
  - 2020-2021 (2 years) longitudinal analysis can be based on data from 3 panels 23, 24, and 25
- **Pooling leads to more accurate estimates**

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
  - Linking MEPS-NHIS expands the analytic capabilities
- **MEPS / NHIS link file**
  - Crosswalk to merge MEPS data to NHIS person-level public use data
  - Crosswalk file not public use; available in AHRQ Data Center

MEPS/NHIS Linked files, weighting, and Estimation

- **Not all MEPS respondents link (birth, marriage, etc.)**
  - Since 2019 NHIS collects data only on a sample adult (18+) and a sample child (17-)
  - Since 2019, no more NHIS family questionnaire completed by one family member.
- **Weighting adjustment for non-linkage is recommended**
  - Necessary since NHIS move to sample adult & child
  - Reference:
    2013 Federal Committee on Statistical Methodology proceedings paper, by Mirel & Machlin.
    https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/H2_Mirel_2013FCSM.pdf

Pooling Multiple Years of MEPS Data in Cross Sectional Analysis

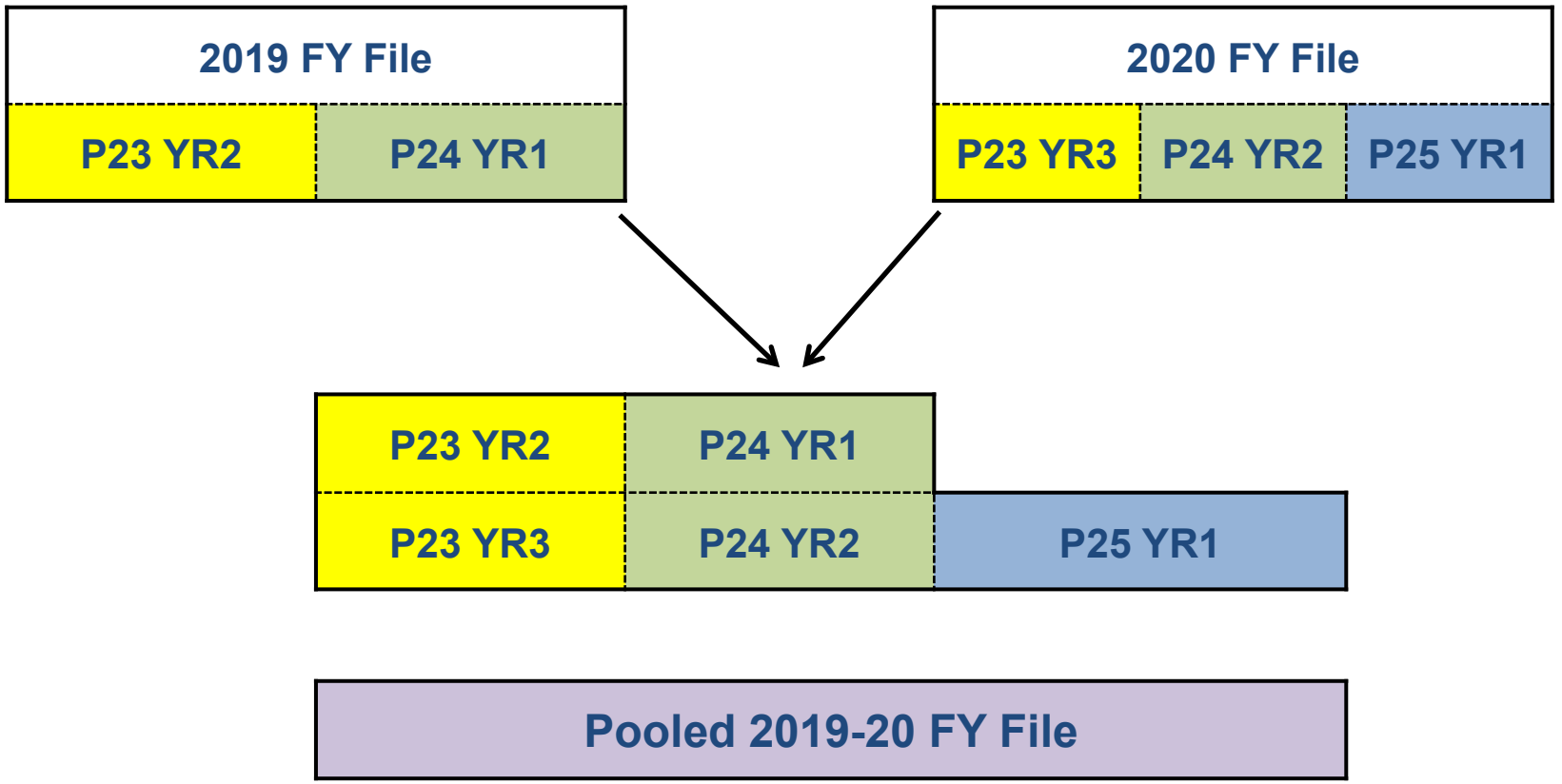
Reasons for pooling

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

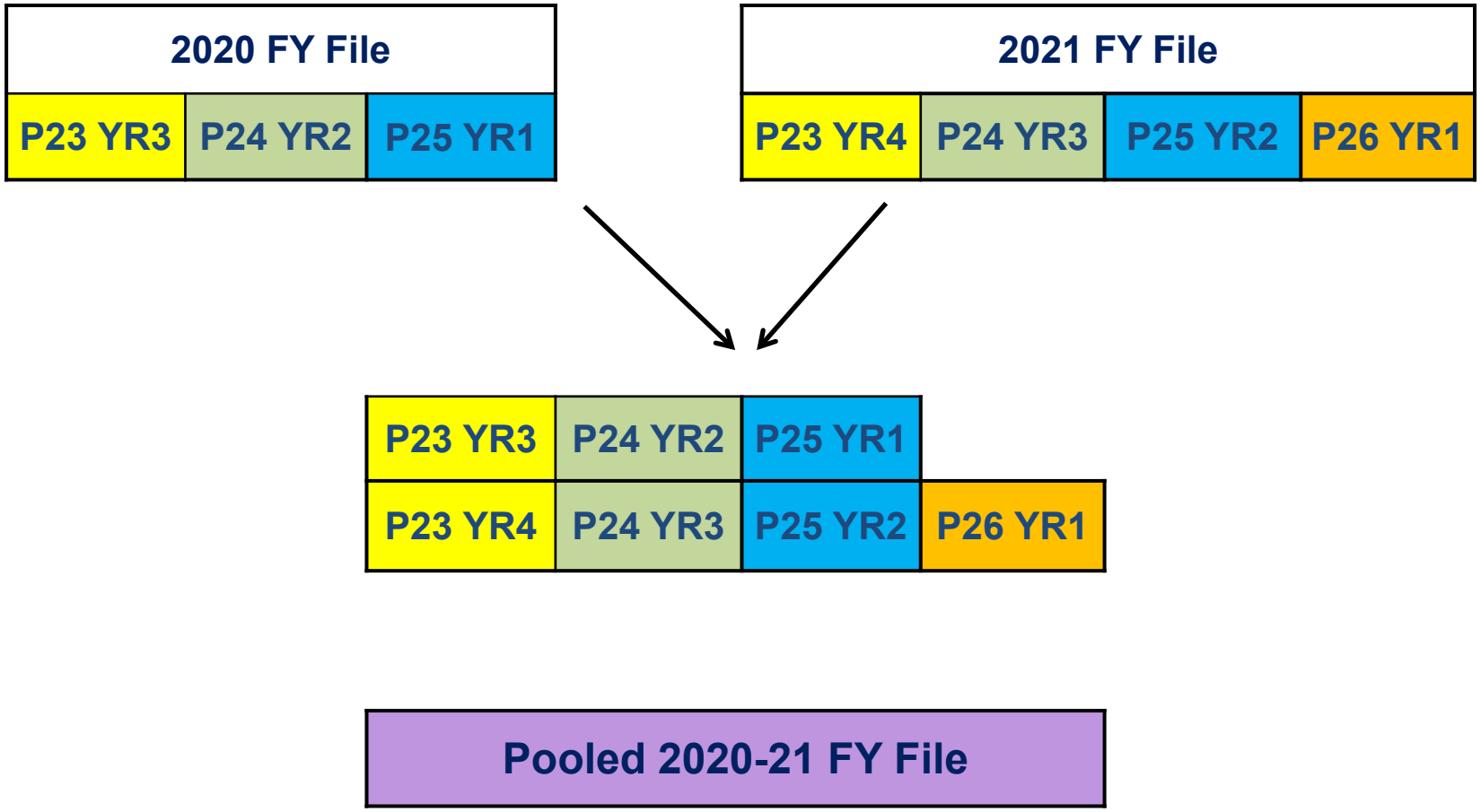
FY Consolidated Data Files

Year	File Number	Number of Persons
2021	HC-233	28,336
2020	HC-224	27,805
2019	HC-216	28,512
2018	HC-209	30,461
2017	HC-201	31,880
2016	HC-192	34,655
2015	HC-181	35,427
2014	HC-171	34,875
2013	HC-163	36,940
2012	HC-155	38,974
2011	HC-147	35,313

Example of pooling FY files 2019 & 2020



Example of Pooling FY files 2020 & 2021



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
- Lack of independence diminishes the gain in precision from pooling

Accounting for lack of independence

- **Repeated observations for the same persons does not affect the validity of variance estimates.**
- **Specify the stratum variable (*VARSTR*) and the PSU variable (*VARPSU*) when computing variances**
- **For more on this topic, read the document:**

  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

Year	Sample Size		
	Privately Insured	Publicly Insured	Uninsured (all year)
2019	758	504	97
2020	777	588	120
2021	772	603	105
2019-21 (Pooled)	2,307 person-years	1,695 person-years	322 person-years

Example of Pooled RSEs of mean annual expenditures

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

Year	Relative Standard Error (RSE) (Standard error / Point estimate)		
	Privately Insured	Publicly Insured	Uninsured (all year)
2019	8.5%	7.3%	26.4%
2020	9.0%	9.3%	33.0%
2021	8.1%	8.9%	56.9%
2019-21 (Pooled)	6.0%	6.4%	34.7%

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 1996-2021

    STRA9621, *PSU9621*
  - 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 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)
- **2019-2021**
  - Stratum and PSU variables on annual files are standardized between these two years, but **not with preceding years**
  - When pooling 2019-2021 use stratum and PSU variables from annual files.
  - When pooling 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* & *TOTEXP21* = TOTEXP
   - Weights: *PERWT19F*, *PERWT20F* & *PERWT21F* = 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 as for individual years**
- **Estimates interpreted as "average annual" for pooled period**

  For example, the average annual per capita health care expenses in 2019-21 was $6,486.

  Note: Per capita expenses were $6,252 in 2019, $6,266 in 2020, and $6,934 in 2021.

- **Adjust expenditure/income estimates with a price index for comparison across multiple years. For more information:**

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

Thank you

kilem.gwet@ahrq.hhs.gov