

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

MEPS Data Tools and Programming Overview

Emily M. Mitchell, PhD

Table of Contents

Table of Contents

MEPS Data Tools

meps.ahrq.gov

- Survey Questionnaires

Data and Statistics

- Data Overview
- MEPS Topics
- Publications Search
- Summary Data Tables
- MEPSnet Query Tools
- Data Files
- Data Centers

Communication

MEPS Topics

- Access to Health Care
- Children's Health
- Children's Insurance Coverage
- Elderly Health Care
- Health Care Costs/Expenditures
- Health Care Disparities
- Health Insurance
- Medical Conditions
- Medicare/Medicaid/SCHIP
- Men's Health
- Mental Health
- Obesity

Click here for full topic list ...

What's New Highlights

Upcoming Events

Summary Data Tables

Household Component summary tables

Use, expenditures, and population

Utilization, spending, and population totals by demographic characteristics, event type, or source of payment.

Health insurance

Number and percentage of people by insurance coverage and demographic characteristics.

Accessibility and quality of care

Number and percentage of people with a usual source of care, difficulty accessing needed care, preventive care, diabetes care, and patient-reported quality of doctor's visits, by demographic characteristics.

Medical conditions, 1996–2015

Utilization, spending, and number of people with care for medical conditions by demographic characteristics, based on ICD-9 codes (1996-2015)

Medical conditions, 2016 and later

Utilization, spending, and number of people with care for medical conditions by demographic characteristics, based on ICD-10 codes (2016 and later)

Prescribed drugs

Purchases and spending by prescribed drug or therapeutic class.

Summary Data Tables

Select statistic:

Number of people

☑ Show standard errors

Select variable:

Insurance coverage, all ages

Select Levels

Select data view:

◉ Trends over time

○ Cross-sectional

Year: **to:**

1996 | 2017

Table | Plot | </> Code

Number of people in thousands (standard errors) by insurance coverage, all ages, United States, 1996-2017

Year	Any private, all ages	Public only, all ages	Uninsured, all ages
2017	220,527 (4,602)	82,755 (1,939)	21,498 (811)
2016	216,880 (5,962)	81,653 (2,279)	24,609 (1,152)
2015	214,446 (6,141)	80,828 (2,472)	26,149 (1,099)
2014	208,377 (5,913)	78,739 (2,549)	31,324 (1,309)
2013	201,609 (5,235)	73,576 (2,334)	40,537 (1,585)
2012	201,911 (5,328)	71,733 (2,377)	39,847 (1,496)
2011	203,056 (5,365)	69,113 (2,151)	38,957 (1,376)
2010	200,580 (5,089)	67,557 (2,118)	40,437 (1,614)
2009	201,395 (4,951)	63,769 (2,094)	41,497 (1,730)

Summary Data Tables

Select statistic:

Number of people

☑ Show standard errors

Select variable:

Insurance coverage, all ages

Select Levels

Select data view:

○ Trends over time

◉ Cross-sectional

Year:

2017

Group by:

Race/ethnicity

Table | Plot | </> Code

Number of people in thousands (standard errors) by race/ethnicity and insurance coverage, all ages, United States, 2017

Race/ethnicity	Any private, all ages	Public only, all ages	Uninsured, all ages
Hispanic	29,188 (1,402)	21,036 (974)	9,131 (591)
Black	22,725 (1,020)	14,357 (822)	2,874 (235)
White	147,462 (4,006)	39,149 (1,380)	7,918 (467)
Amer. Indian, AK Native, or mult. races	6,775 (487)	3,987 (408)	561 (119)
Asian, Hawaiian, or Pacific Islander	14,377 (863)	4,225 (407)	1,014 (161)

-- Estimates suppressed due to inadequate precision (see FAQs for details).

* Relative standard error is greater than 30%

Source: Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, Medical Expenditure Panel Survey, 2017

Notes

Race/ethnicity

Summary Data Tables

Select statistic:

Number of people

Show standard errors

Select variable:

Insurance coverage, all ages

Select Levels

Select data view:

Trends over time

Cross-sectional

Year:

2017

Group by:

Race/ethnicity

Table | Plot | </> Code

Select programming language:

R

To run the code, first download and unzip the required public use data files from the MEPS data files page, and save them to your computer. More information on downloading and analyzing MEPS data in R, SAS, and Stata can be found at the AHRQ GitHub site. Note that some standard error estimates may differ between R and SAS, since SAS doesn't support any options to adjust for lonely PSUs.

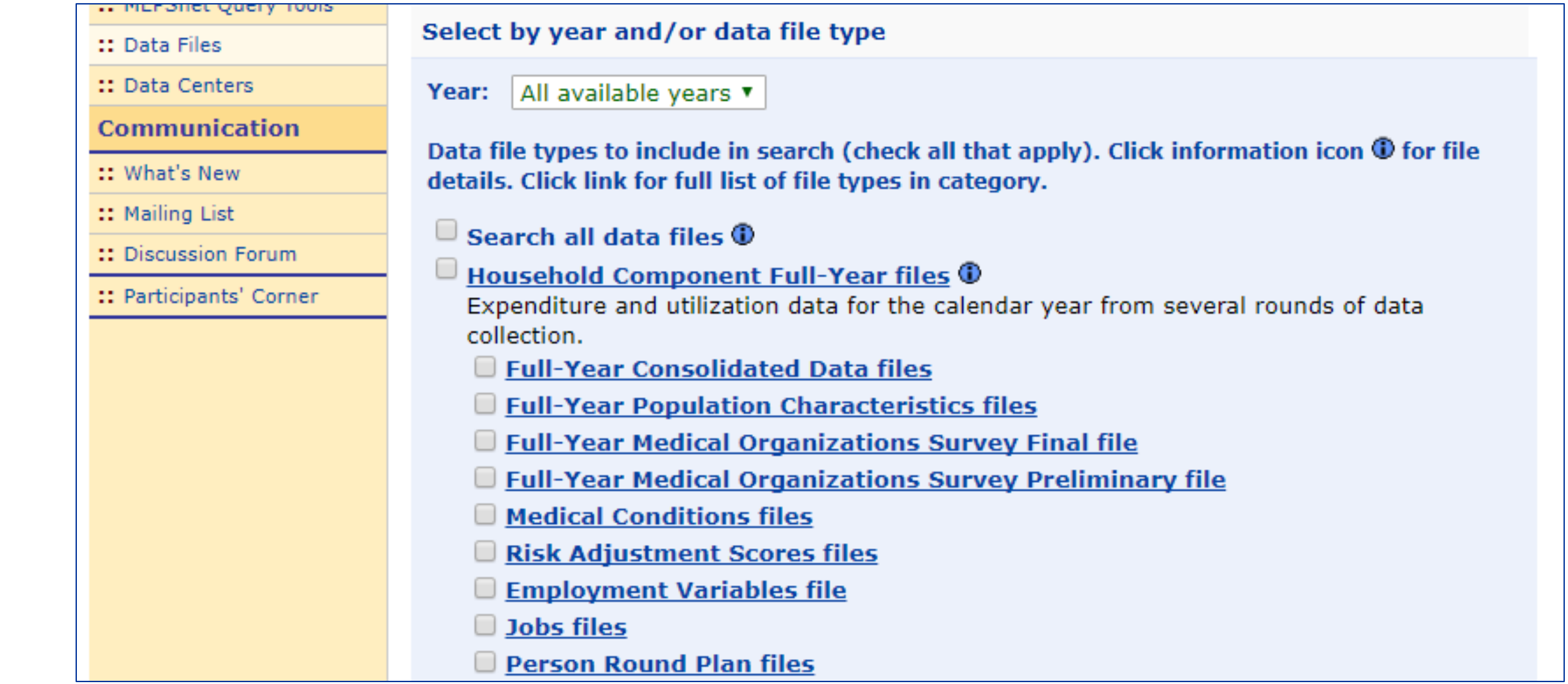
The following code can be used to generate the selected estimates, where the SAS transport data files (.ssp) have been saved to the folder 'C:\MEPS'. For trend estimates, example code is shown for the most recent year selected:

```
# Install and load packages
  package_names <- c("survey","dplyr","foreign","devtools")
  lapply(package_names, function(x) if(!x %in% installed.packages()) install.
  lapply(package_names, require, character.only=T)

  install_github("e-mitchell/meps_r_pkg/MEPS")
  library(MEPS)
```

Table of Contents

MEPS Public Use Data Files



MEPS Public Use Data Files

- Data Centers

Communication

- What's New
- Mailing List
- Discussion Forum
- Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.

**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

MEPS Public Use Data Files

- Data Centers

Communication

- What's New
- Mailing List
- Discussion Forum
- Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.
**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

Codebook MEPS Public Use Data Files

Name	Start	End	Description
REGION17	79	80	CENSUS REGION AS OF 12/31/17
REGION31	73	74	CENSUS REGION - R3/1
REGION42	75	76	CENSUS REGION - R4/2
REGION53	77	78	CENSUS REGION - R5/3

Variable Name: REGION17

Description: CENSUS REGION AS OF 12/31/17

VALUE	UNWEIGHTED	WEIGHTED BY PERWT17F
-1 INAPPLICABLE	204	2,152,631
1 NORTHEAST	4,935	56,041,643
2 MIDWEST	6,406	67,551,951
3 SOUTH	12,266	122,086,667
4 WEST	8,069	76,947,017
TOTAL	31,880	324,779,909

MEPS Public Use Data Files

- Data Centers
- **Communication**
- What's New
- Mailing List
- Discussion Forum
- Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

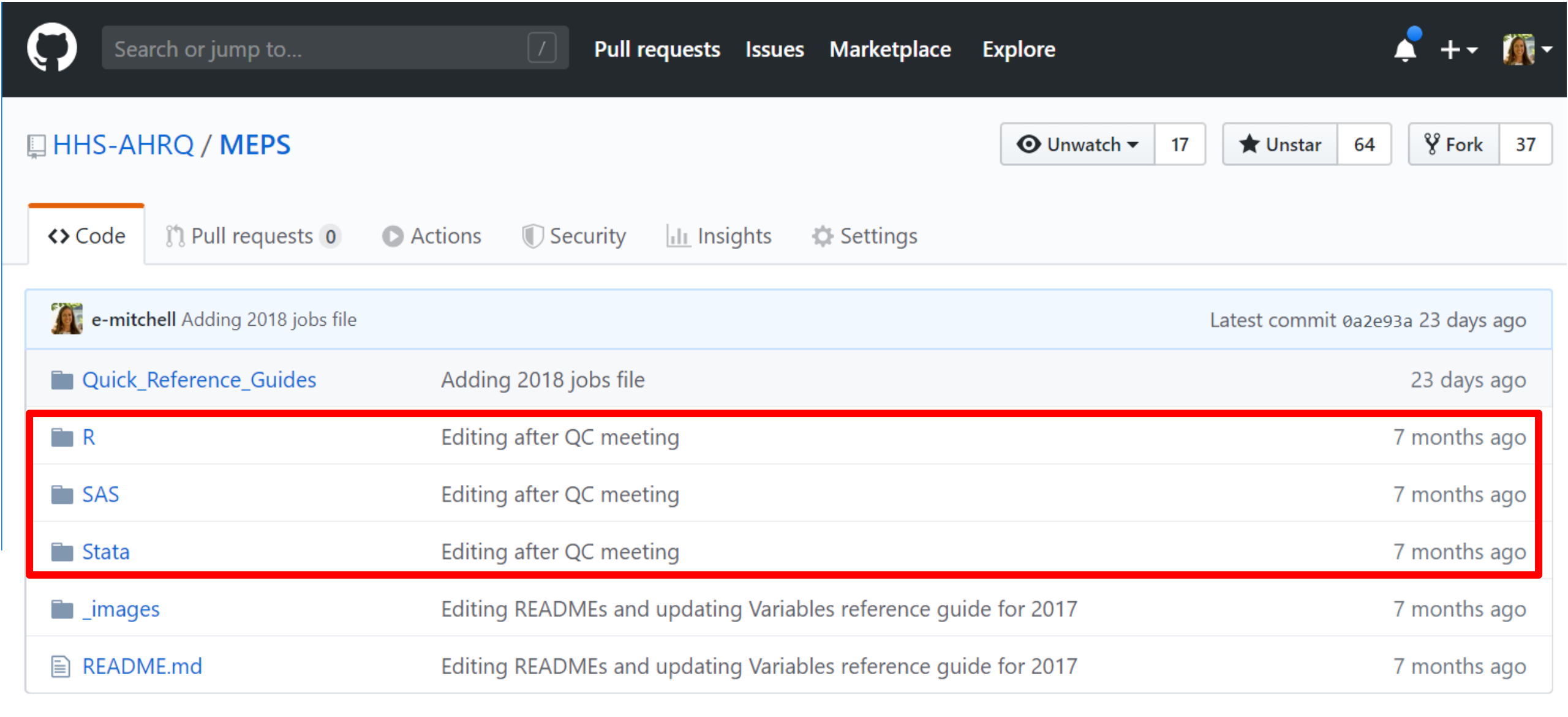
Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

Questionnaires — see **Survey Questionnaires**

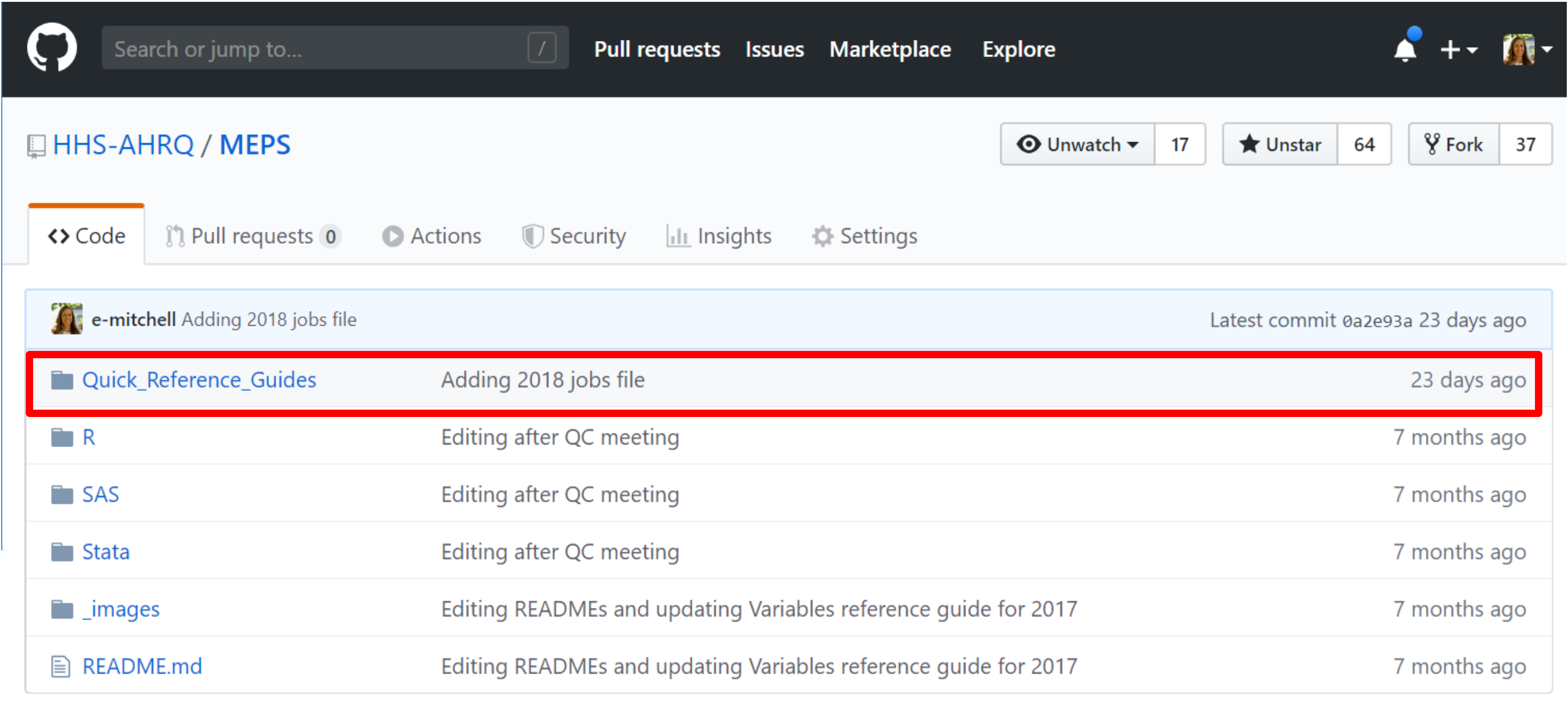
*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.

**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

https://github.com/HHS-AHRQ/MEPS

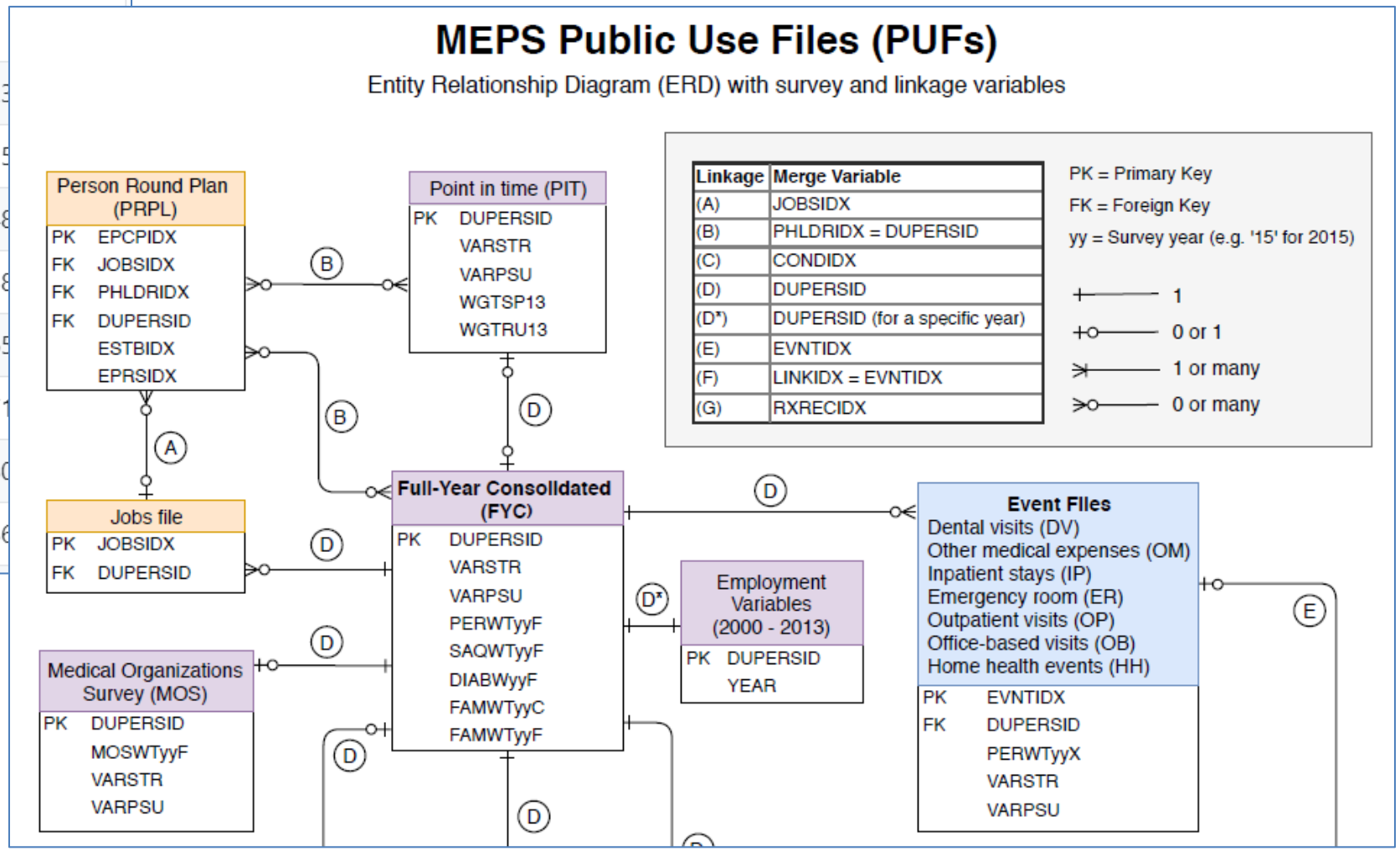


https://github.com/HHS-AHRQ/MEPS



Quick reference guides

FYC	Conditions	PMED Events	Events	Jobs	PRPL	Longitudinal
h12	h06r	h10a	h10*f1	h07	h24	-
h20	h18	h16a	h16*f1	h19	h47f1	h23
h28	h27	h26a	h26*f1	h25	h47f2	h35
h38	h37	h33a	h33*	h32	h47f3	h48
h50	h52	h51a	h51*	h40	h47f4	h58
h60	h61	h59a	h59*	h56	h57	h65
h70	h69	h67a	h67*	h63	h66	h71
h79	h78	h77a	h77*	h74	h76	h80
h89	h87	h85a	h85*	h83	h88	h86



Record-Level and Identifiers

Person-level

- Full-year consolidated file
- Longitudinal files

Event-level

- Event files: PMED, DN, OM, IP, ER, OP, OB, HH

Condition-level

- Medical conditions file

Job-level

- Jobs file

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	2000410131
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	20004101301
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Record-Level and Identifiers

Person-level

DUID	PID	DUPERSID
20004	101	20004101
20004	102	20004102
20004	103	20004103
20005	101	20005101

Event-level

DUPERSID	EVNTIDX
20004101	200041010011
20004101	200041010021
20005101	200051010151
20005101	200051010201

Conditions-level

DUPERSID	CONDN	CONDIDX
20004103	11	200041030011
20005101	11	200051010011
20005101	21	200051010021
20005101	51	200051010051

Jobs-level

DUPERSID	RN	JOBSN	JOBSIDX
20004101	3	1	20004101301
20004101	4	1	20004101401
20004101	5	1	20004101501
20004102	3	1	20004102301

Variable Naming Conventions

Edited Variables end in an "X"

RACE**X**

Year-specific variables use last two digits of year

TOTEXP**17**
PERWT**17**F

Round-specific variables use two-digit round

- Some questions only asked in certain rounds, e.g. the Self-Administered Questionnaire in rounds 2 and 4

AGE**31**X
AGE**42**X
AGE**53**X

Estimation Variables

Weight Variables

- Person-level (e.g. PERWT17F, DIABW17F, SAQWT17F)
- Family-level (e.g. FAMWT17F, FAMWT17C)
- Longitudinal (e.g. LONGWT)

Variance-Estimation Variables (Stratum and PSU):

- After 2002 FY data: VARSTR, VARPSU
- 1996-2001 FY data: VARSTRyy, VARPSUyy
  - When calculating variances with *pooled data*, use STRA9617, PSU9617 in data file HC-036

MEPS Reserve Codes

-1	Inapplicable	Question was not asked due to skip pattern
-7	Refused	Question was asked and respondent refused to answer question
-8	Don’t Know	Question was asked and respondent did not know answer
-9	Not Ascertained	Interviewer did not record the data
-10	Top-Coded	Variable was top-coded for confidentiality

MEPS Reserve Codes

-1 **Inapplicable**

-7 **Refused**

-8 **Don't Know**

-9 **Not Ascertained**

-10 **Top-Coded**

EXAMPLES

FYC file: Pregnancy

Event file: Expenditures for phone calls

Jobs file: Hourly Wage

Table of Contents

Programming Example

Compare average medical expenses for persons under age 65 vs. 65 and older in 2017.*

*** Not including people that have $0 in expenses**

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2017.*

1. Load datasets
2. Create new variables
3. Run survey procedure
4. Examine results

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2017.*

1. Load datasets

2. Create new variables

3. Run survey procedure

4. Examine results

2017 Full-Year Consolidated File
Person-level

Load datasets

- Data Centers
- **Communication**
- What's New
- Mailing List
- Discussion Forum
- Participants' Corner

Update notes

Documentation	File type
Documentation	PDF (500 KB) / HTML
Codebook	PDF (121 KB) / HTML*
SAS Programming Statements	TXT (59 KB)
SPSS Programming Statements	TXT (343 KB)
STATA Programming Statements	TXT (348 KB)

Data	File type**
Data File, ASCII format	ZIP (8.3 MB) / EXE (8.8 MB)
Data File, SAS transport format	ZIP (9.3 MB) / EXE (9.9 MB)

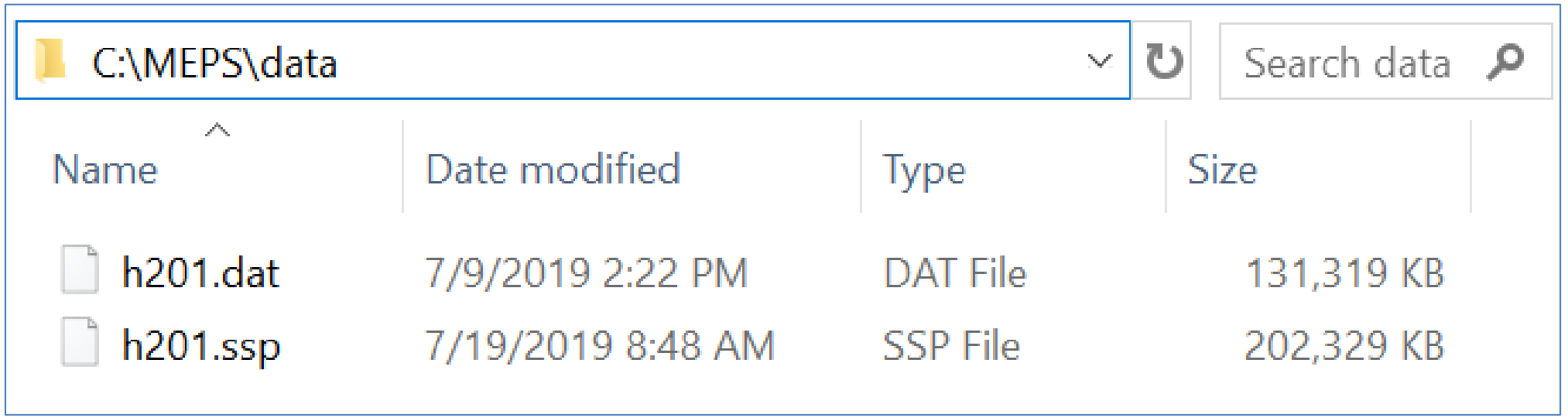
Questionnaires — see **Survey Questionnaires**

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and lets you navigate quickly to details on each variable.

**Right-click on the data file link, then select Save Target As or Save Link As to download the file.

Load datasets

Store .dat or .ssp file in a local directory:



Load datasets

SAS

```
FILENAME in_h201 'C:\MEPS\data\h201.ssp';

proc xcopy in = in_h201 out = WORK IMPORT;
run;
```

Stata

```
import sasxport "C:\MEPS\data\h201.ssp"
```

R

```
install.packages("foreign"); library(foreign);

h201 = read.xport("C:/MEPS/data/h201.ssp")
```

Load datasets -- shortcut

SAS

```
%load_MEPS(h201);
```

Stata

```
copy "https://meps.ahrq.gov/mepsweb/data_files/pufs/h201ssp.zip" "h201ssp.zip"
unzipfile "h201ssp.zip"

import sasxport "h201.ssp", clear
```

R

```
install.packages("foreign"); library(foreign);

download.file("https://meps.ahrq.gov/mepsweb/data_files/pufs/h201ssp.zip", temp <- tempfile())

h201 = read.xport(unzip(temp))
unlink(temp)  # Unlink to delete temporary file
```

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2017.*

1. Load datasets
2. **Create new variables**
3. Run survey procedure
4. Examine results

Age groups:
AGELAST < 65
AGELAST >= 65

Any expenditures:
TOTEXP17 > 0

Create new variables

SAS

```
data h201;
  set h201;

  if 0 <= AGELAST <= 64 then agecat = 1;
  else if AGELAST > 64 then agecat = 2;

  if TOTEXP17 > 0 then has_exp = 1;
  else if TOTEXP17 = 0 then has_exp = 0;
run;
```

Stata

```
gen agecat = 1
replace agecat = 2 if agelast > 64

gen has_exp = 1
replace has_exp = 0 if (totexp17 <= 0)
```

R

```
install.packages("dplyr")
library(dplyr)

h201 = h201 %>% mutate(
   agecat = ifelse(AGELAST > 64, 2, 1),
   has_exp = ifelse(TOTEXP17 <= 0, 0, 1) )
```

Create new variables

Quality check on new variables

agecat	agelast		
	Min	Mean	Max
1 (< 65)	0	31.2	64
2 (65+)	65	73.9	85

has_exp	totexp17		
	Min	Mean	Max
0	0	0	0
1	1	6,111	552,898

SAS

proc means
proc freq

Stata

bys
sum

R

group_by
summarise

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2017.*

1. Load datasets
2. Create new variables
3. **Run survey procedure**
4. Examine results

Mean TOTEXP17
- by Age groups
- if has_exp == 1

Run survey procedure

SAS

```
proc surveymeans data = h201 mean;
        stratum VARSTR;
        cluster VARPSU;
        weight PERWT17F;
        var TOTEXP17;
        domain has_exp * AGECAT;
run;
```

R

```
library(survey); options(survey.lonely.psu='adjust');

mepsdsgn = svydesign(
  id = ~VARPSU,
  strata = ~VARSTR,
  weights = ~PERWT17F,
  data = h201,
  nest = TRUE)

svyby(~TOTEXP17, by = ~agecat, FUN = svymean,
     design = subset(mepsdsgn, has_exp==1))
```

Stata

```
svyset [pweight=perwt17f], strata(varstr) psu(varpsu) vce(linearized) singleunit(missing)

svy, subpop(if has_exp==1): mean totexp17, over(agecat)
```

Run survey procedure

has_exp	agecat	totexp17	
		Mean	Std. Err.
1	1 (< 65)	4,939	142.6
	2 (65+)	11,947	371.3

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,939	142.6
	2 (65+)	11,947	371.3

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,939	142.6
	2 (65+)	11,947	371.3

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,939	**126.6**
	2 (65+)	11,947	**352.6**

Why survey procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,939	142.6
	2 (65+)	11,947	371.3

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	4,939	**126.6**
	2 (65+)	11,947	**352.6**

Ignoring VARSTR, VARPSU, and PERWT

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	**4,762**	**100.7**
	2 (65+)	**12,153**	**331.1**

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2017.*

1. Load datasets
2. Create new variables
3. Run survey procedure
4. **Examine results**

Examine results

Does output make sense?

- Population estimates
- Inflation adjustment?

Are estimates reliable?

- Sample size (n > 60)
- Standard errors (RSE < 0.3)

Consistent with other published results?

- Stat briefs
- Summary tables

Programming checklist

- ❑ Well-defined question
- ❑ Checked documentation
- ❑ Reserve codes addressed (-1, -9, etc.)
- ❑ Datasets merged correctly
- ❑ Adequate sample size / precision (PERWT17F)
- ❑ Survey procedures
  - ❑ PERWT, VARSTR, VARPSU
  - ❑ Using correct weights (PERWT / FAMWT / LONGWT)
  - ❑ 'domain' analysis for subsets (SAS)
- ❑ Results make sense

Exercises (⭐ difficulty level)

SAS / Stata

1. National health care expenses ⭐
2. Purchases and expenses for narcotic analgesics ⭐⭐
3. Pooling multiple years of MEPS data ⭐⭐⭐
4. Pooling longitudinal files ⭐⭐⭐

https://github.com/HHS-AHRQ/MEPS-workshop

Thank you!

Emily.Mitchell@ahrq.hhs.gov