

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

MEPS Data Tools and Programming Overview

Emily M. Mitchell, PhD

Table of Contents

Table of Contents (cont.)

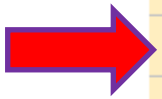
Data Tools

Public Use Files (PUFs)

Programming Example (SAS, Stata, R)

MEPS Data Tools

https://datatools.ahrq.gov



- MEPS Home
- **About MEPS**
  - :: Survey Background
  - :: Workshops & Events
  - :: Data Release Schedule
- **Survey Components**
  - :: Household
  - :: Insurance/Employer
  - :: Medical Provider
  - :: Survey Questionnaires
- **Data and Statistics**
  - :: Data Overview
  - :: MEPS Topics
  - :: Publications Search
  - :: MEPS Data Tools - **NEW!**
  - :: Data Files
  - :: Data Centers
- **Communication**

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage. Learn more about MEPS.

Contact MEPS

New to MEPS?

Select a profile:
- General user
- Researcher
- Policymaker
- Media
- Survey participant

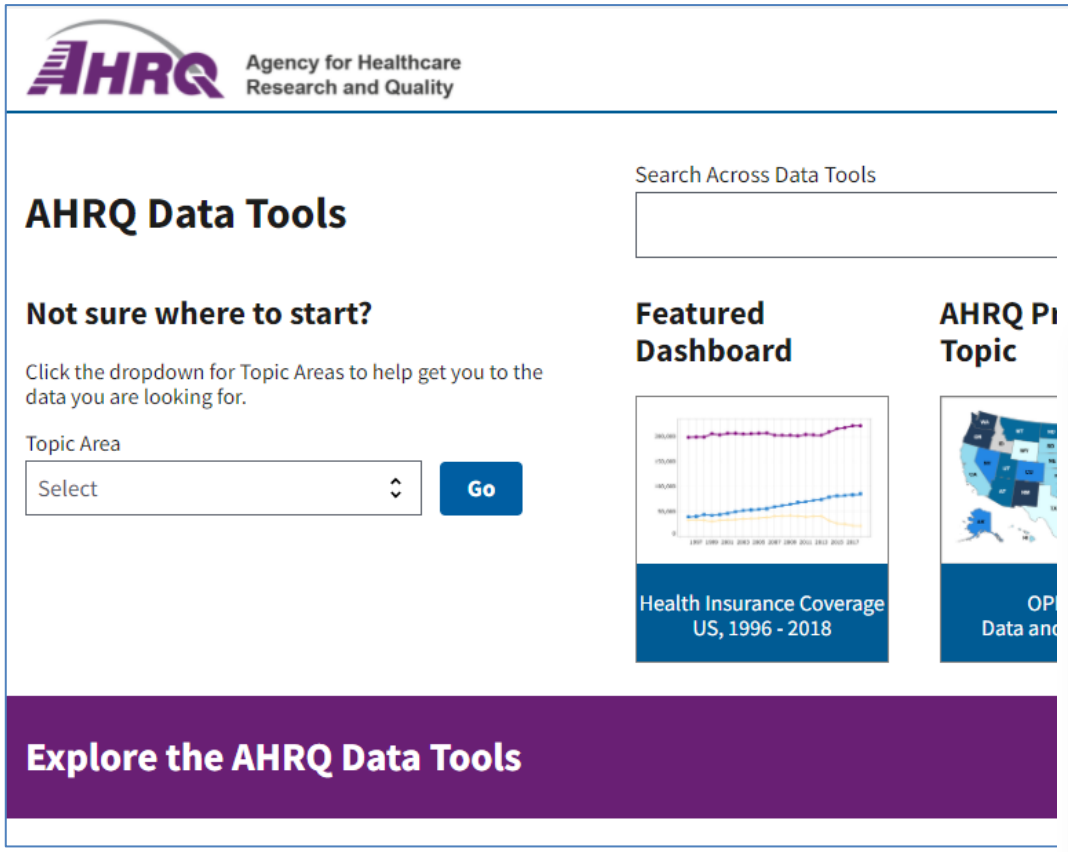
MEPS Topics

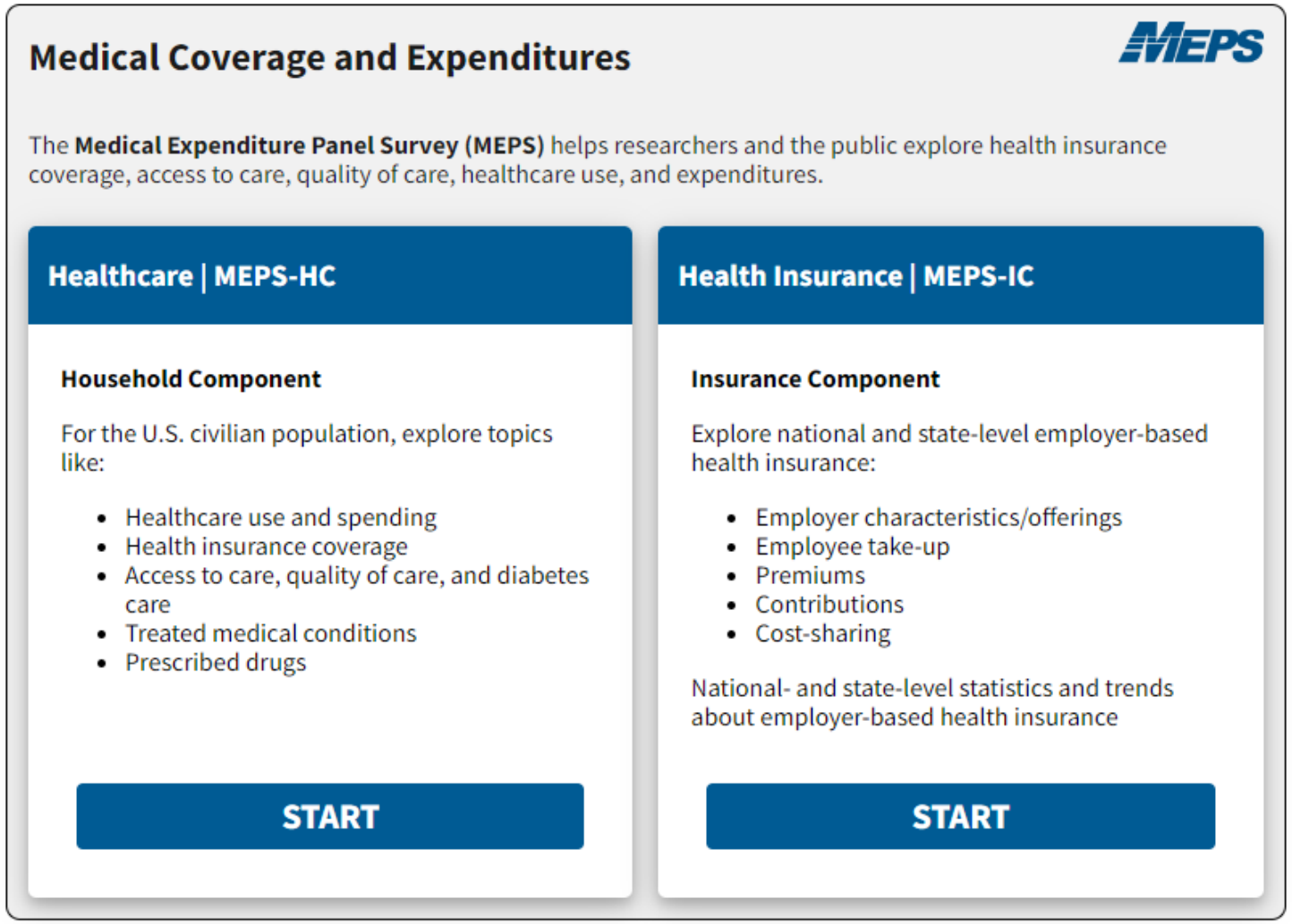
- Access to Health Care
- Children's Health
- Children's Insurance Coverage
- Elderly Health Care
- Health Care Costs/Expenditures
- Health Care Disparities
- Health Insurance
- Medical Conditions
- Medicare/Medicaid/SCHIP
- Men's Health
- Mental Health
- Obesity
- Prescription Drugs
- Projected Data/Expenditures
- Quality of Health Care
- State and Metro Area Estimates
- The Uninsured
- Women's Health

Click here for full topic list ...

MEPS Data Tools (cont.)

https://datatools.ahrq.gov





MEPS

Search Across Data Tools

Search

Medical Expenditure Panel Survey (MEPS) Household Component (HC)

Information on the health status of Americans, health insurance coverage, and access, use, and cost of health services.

For more information about MEPS, visit meps.ahrq.gov

- AHRQ Data Tools +
- MEPS Home
- Data Files +
- Educational Links +
- MEPS GitHub Repository
- Publications
- Workshops

Explore the MEPS-HC Data Tools

The MEPS Household Component collects data on all members of sample households from selected communities across the United States. These data can be used to produce nationally representative estimates of medical conditions, health status, use of medical care services, charges and payments, access to care, experience with care, health insurance coverage, income, and employment.

The summary tables provide frequently used summary estimates for the U.S. civilian non-institutionalized population.

This tool is provided as a convenience. It is the responsibility of the user to review the results for statistical significance and overall reasonableness.

Use, Expenditures, and Population

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

START

Health Insurance

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

START

Accessibility and Quality of Care

Information on access to care, preventive care, diabetes care, and patient-reported quality of doctor's visits.

START

Medical Conditions

Utilization, spending, and number of people with care for medical conditions by demographic attributes.

START

Prescribed Drugs

Purchases and spending by prescribed drug or therapeutic class.

START

MEPS

Search Across Data Tools

Search

Medical Expenditure Panel Survey (MEPS) Household Component (HC)

Information on the health status of Americans, health insurance coverage, and access, use, and cost of health services.

For more information about MEPS, visit meps.ahrq.gov

AHRQ Data Tools +

MEPS Home

Data Files +

Educational Links +

MEPS GitHub Repository

Publications

Workshops

Explore the MEPS-HC Data Tools

The MEPS Household Component collects data on all members of sample households from selected communities across the United States. These data can be used to produce nationally representative estimates of medical conditions, health status, use of medical care services, charges and payments, access to care, experience with care, health insurance coverage, income, and employment.

The summary tables provide frequently used summary estimates for the U.S. civilian non-institutionalized population.

This tool is provided as a convenience. It is the responsibility of the user to review the results for statistical significance and overall reasonableness.

Use, Expenditures, and Population | Health Insurance | Accessibility and Quality of Care | **Medical Conditions** | Prescribed Drugs

Statistics on the number of people with care for **medical conditions**, health care utilization, total expenditures, and mean expenditures per person by medical condition. Data can be viewed over time or for a single year by event type (such as prescription medicines or outpatient events), source of payment (such as Medicare or Medicaid), or demographic characteristics (such as age, race, or sex).

Select the **Download Data** button for an accessible MS Excel version of the data visualization. The file size will depend on parameters selected.

Medical Conditions

Trends

Mean expenditure per person with care ($) by condition, United States, 1996 to 2019

Medical Conditions

Trends

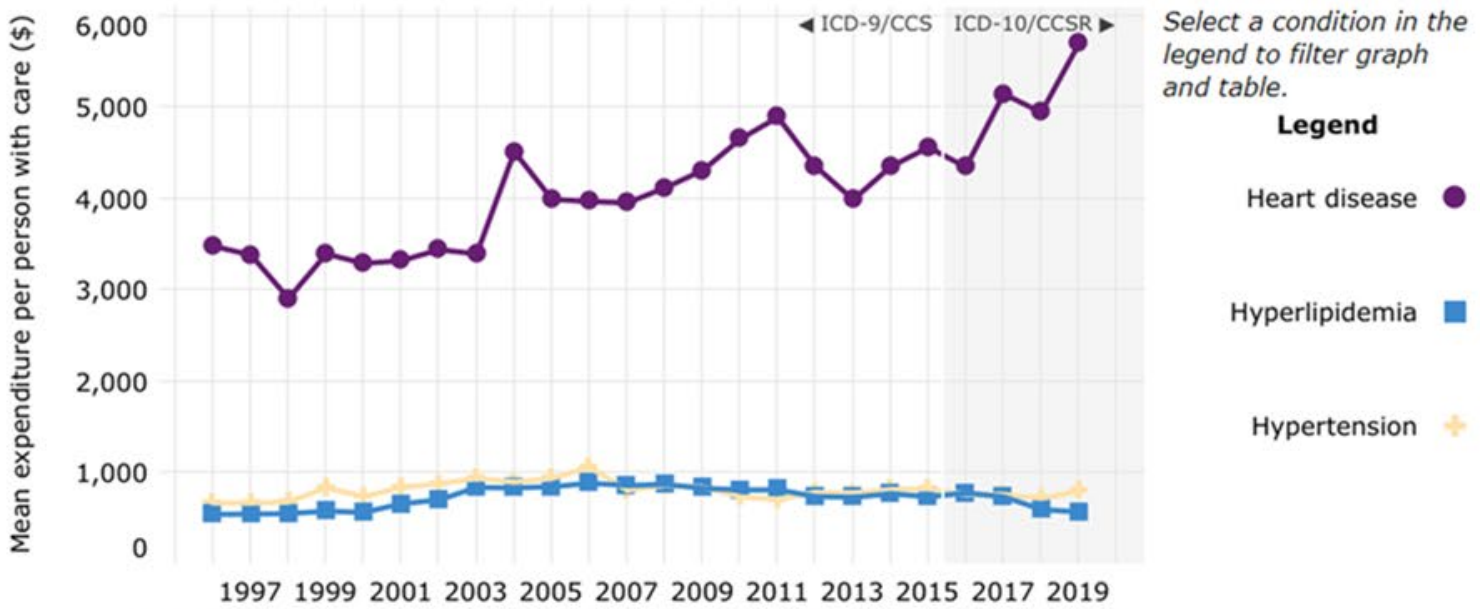
Cross-sectional

Estimates:
Mean expenditure per person...

Start Year 1996 **End Year** 2019

☐ **Show SE/95% CI**

Mean expenditure per person with care ($) by condition, United States, 1996 to 2019



Select a condition in the legend to filter graph and table.

Legend

- Heart disease
- Hyperlipidemia
- Hypertension

Starting in 2016, coding of household-reported medical conditions transitioned from ICD-9/CCS to ICD-10/CCSR codes. Take caution when comparing data before and after this transition. Refer to the Notes section for details.

Search by medical condition:
(press enter to search)

Clear Selected

Select from rows in the table below to graph data.

	1996	1997	1998	1999	2000	2001	2002	2003	2004
Acute Bronchitis and URI	160	152	166	216	165	229	266	239	223
Allergic reactions	163	140	278	244	257	256	276	270	280
Anemia and other deficiencies	1,890*	--	743*	1,235*	519	1,241	1,131	1,138	1,306
Back problems	934	982	1,096	1,208	1,235	1,183	1,256	1,318	1,488
Cancer	4,067	5,214	3,952	3,525	4,195	4,376	4,462	4,406	5,727
Cataract	1,378	1,265	1,277	1,490	1,233	1,078	1,509	1,183	1,333

Medical Conditions

Trends

Cross-sectional

Mean expenditure per person with care ($) by condition and age groups, United States, 2019

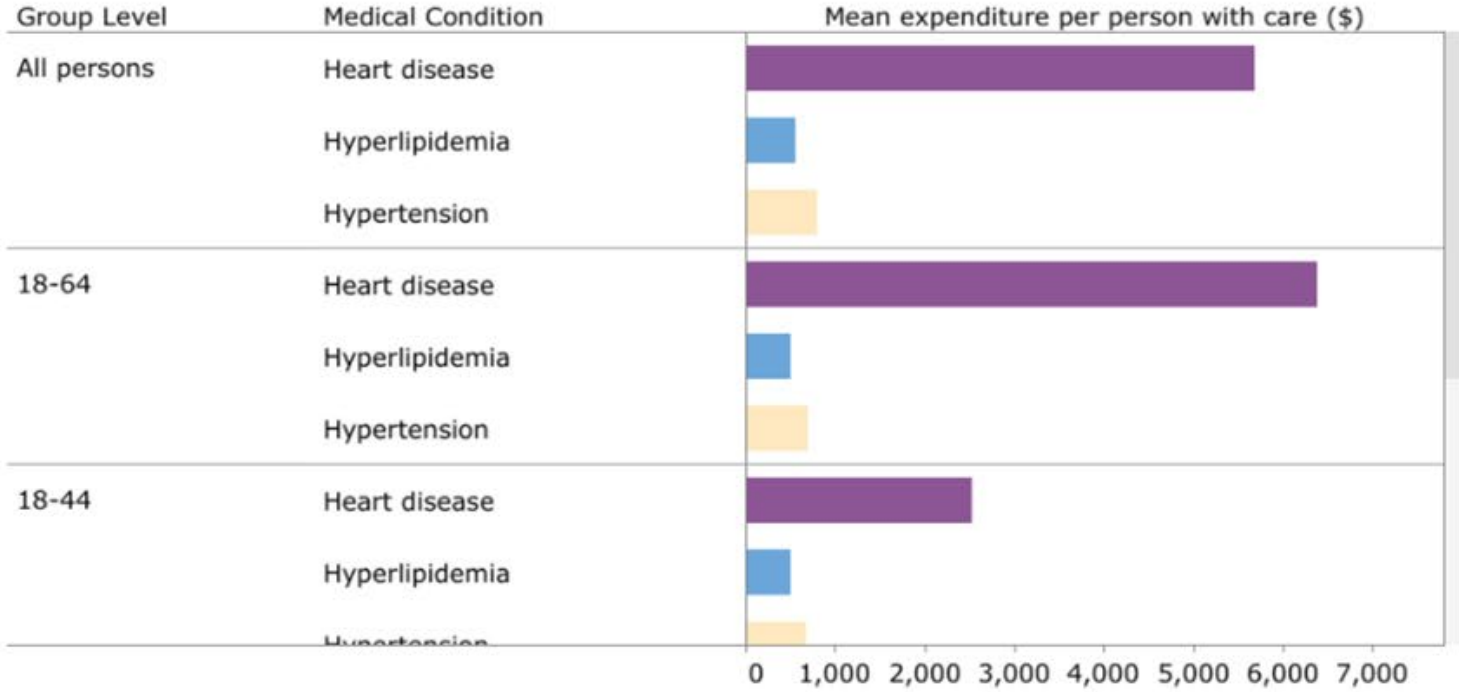
Estimates:
Mean expenditure per person...

Group by:
Age groups

Group Levels:
(All)

Years:
2019

☐ Show SE/95% CI



Search by medical condition:
(press enter to search)

Clear Selected

Select from rows in the table below to graph data.

	All persons	Under 18	Under 5	5-17	18-64	18-44
Acute Bronchitis and URI	418	365	513	--	429	434
Allergic reactions	591	565	446	--	629	595
Anemia and other deficiencies	1,914	--	--	--	1,436	1,331*
Back problems	3,169	--	--	--	2,693	2,300
Cancer	7,769	--	--	--	11,162	9,301

- **Food Security Files (2016-2017)** - Data pertaining to food security
- **Longitudinal Data Files** - A two-year longitudinal file representing each Panel in the MEPS survey
- **Medical Organization Survey (2015-2016)** - Characteristics for usual source of care providers
- **Pooled Linkage Variance Structure** - Standardized variance strata and PSU variables for a pooled analysis that includes data from the years 1996-2001
- **Preventive Care SAQ (2014)** - Contains various person-level preventive health care data for adults

Go Reset

MEPS-HC Variable Explorer Tool: Annual/Main Public Use Files (PUFs) 1996 - 2019

Quick Search

Search Table: educ

Advanced Search

Variable | Data File: (All) | Description | Years (All)

Selecting a variable under the years will navigate to the codebook on the AHRQ Medical Expenditure Panel Survey website.

Variable	Data	Description	2019	2018	2017	2016	
EDRECODE	FYC	EDUCATION RECODE (EDITED)					ED
EDRECODE	PIT	EDUCATION RECODE (EDITED)					ED
EDUCYEAR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS					
EDUCYR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	E
EDUCYR	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	E
EDUCYR1	FYC	COMPLETED YEARS OF EDUCATION-RD1					
EDUCYR1	PIT	COMPLETED YEARS OF EDUCATION					

Food Security Files (2016-2017) - Data pertaining to food security

Longitudinal Data Files - A two-year longitudina

Medical Organization Survey (2015-2016) - Cha

Pooled Linkage Variance Structure - Standardiz

Preventive Care SAQ (2014) - Contains various

Go Reset

MEPS-HC Variable Explor

Quick Search

Search Table

educ

Advanced Searc

Variable

Selecting a variable under the years will navigate

Variable	Data	Description					
EDRECODE	FYC	EDUCATION RECOD					
EDRECODE	PIT	EDUCATION RECOD					
EDUCYEAR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS					
EDUCYR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	E
EDUCYR	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	E
EDUCYR1	FYC	COMPLETED YEARS OF EDUCATION-RD1					
EDUCYR1	PIT	COMPLETED YEARS OF EDUCATION					

VALUE	UNWEIGHTED	WEIGHTED
-15 CANNOT BE COMPUTED	1	11,121
-8 DK	160	1,545,846
-7 REFUSED	24	197,843
-1 INAPPLICABLE	2,041	24,727,352
0 NO SCHOOL/KINDERGARTEN ONLY	945	10,020,656
1 - 8 ELEMENTARY GRADES 1 - 8	4,197	42,854,987
9 - 11 HIGH SCHOOL GRADES 9 - 11	2,872	29,435,742
12 GRADE 12	6,663	68,954,120
13 1 YEAR COLLEGE	1,452	18,228,523
14 2 YEARS COLLEGE	2,866	37,565,268
15 3 YEARS COLLEGE	722	9,617,079
16 4 YEARS COLLEGE	3,937	50,987,116
17 5+ YEARS COLLEGE	2,632	33,251,040
TOTAL	28,512	327,396,693

Table of Contents

MEPS Public Use Data Files

https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp

- :: Data Files
- :: Data Centers

Communication

- :: What's New
- :: Mailing List
- :: Discussion Forum

- :: Participants' Corner

Select by year and/or data file type

Year: All available years

Data file types to include in search (check all that apply). Click information icon for file details. Click link for full list of file types in category.

- Search all data files
- Household Component Full-Year files

  Expenditure and utilization data for the calendar year from several rounds of data collection.
  - Full-Year Consolidated Data files
  - Full-Year Population Characteristics files
  - Full-Year Medical Organizations Survey Final file
  - Full-Year Medical Organizations Survey Preliminary file
  - Medical Conditions files
  - Risk Adjustment Scores files
  - Employment Variables file
  - Jobs files
  - Person Round Plan files

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

Update notes

Documentation	File type
Documentation	PDF (540 KB) / HTML
Codebook	PDF (212 KB) / HTML*
SAS Programming Statements	TXT (74 KB)
SPSS Programming Statements	TXT (6.2 KB)
STATA Programming Statements	TXT (8.4 KB)
R Programming Statements	TXT (5.3 KB)

Data	File type**
Data File, ASCII format	ZIP (1.3 MB) / EXE (1.8 MB)
Data File, SAS transport format	ZIP (1.5 MB) / EXE (2.0 MB)
Data File, SAS V9 format	ZIP (1.8 MB)
Data File, Stata format	ZIP (1.8 MB)
Data File, XLSX format	ZIP (6.9 MB)

Questionnaires — see Survey Questionnaires

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

Update notes

Documentation	File type
Documentation	PDF (540 KB) / HTML
Codebook	PDF (212 KB) / HTML*
SAS Programming Statements	TXT (74 KB)
SPSS Programming Statements	TXT (6.2 KB)
STATA Programming Statements	TXT (8.4 KB)
R Programming Statements	TXT (5.3 KB)

Data	File type**
Data File, ASCII format	ZIP (1.3 MB) / EXE (1.8 MB)
Data File, SAS transport format	ZIP (1.5 MB) / EXE (2.0 MB)
Data File, SAS V9 format	ZIP (1.8 MB)
Data File, Stata format	ZIP (1.8 MB)
Data File, XLSX format	ZIP (6.9 MB)

Questionnaires — see **Survey Questionnaires**

DVTTCH19	3733	3737	TOTAL DENTAL CARE VISIT CHARGES 19
DVTTRI19	3768	3771	ALL DENTAL CARE - TRICARE AMT 19
DVTVA19	3763	3767	ALL DENTAL CARE - VA/CHAMPVA AMT 19
DVTWCP19	3780	3782	ALL DENTAL CARE - WORKERS COMP AMT 19
EDUCYR	234	236	YEARS OF EDUC WHEN FIRST ENTERED MEPS
ELGRND19	163	163	ELIGIBILITY STATUS AS OF 12/31/19
ELGRND31	160	160	ELIGIBILITY - R3/1
ELGRND42	161	161	ELIGIBILITY - R4/2
ELGRND53	162	162	ELIGIBILITY - R5/3
EMPHAGED	341	342	AGE OF DIAGNOSIS-EMPHYSEMA
EMPHDX	338	340	EMPHYSEMA DIAGNOSIS (>17)
EMPST31	1088	1090	EMPLOYMENT STATUS RD 3/1
EMPST31H	1393	1394	EMPLOYMENT STATUS RD 3/1 (IMP)
EMPST42	1091	1093	EMPLOYMENT STATUS RD 4/2
EMPST42H	1395	1396	EMPLOYMENT STATUS RD 4/2 (IMP)
EMPST53	1094	1096	EMPLOYMENT STATUS RD 5/3

DVTTCH19	3723	3727	TOTAL DENTAL CARE VISIT CHARGES 19
DVTTRI19			
DVTVA19			
DVTWCP19			
EDUCYR			
ELGRND19			
ELGRND31			
ELGRND42			
ELGRND53			
EMPHAGED			
EMPHDX			
EMPST31	1088	1090	EMPLOYMENT STATUS RD 3/1
EMPST31H	1393	1394	EMPLOYMENT STATUS RD 3/1 (IMP)
EMPST42	1091	1093	EMPLOYMENT STATUS RD 4/2
EMPST42H	1395	1396	EMPLOYMENT STATUS RD 4/2 (IMP)
EMPST53	1094	1096	EMPLOYMENT STATUS RD 5/3

VALUE	UNWEIGHTED	WEIGHTED
-15 CANNOT BE COMPUTED	1	11,121
-8 DK	160	1,545,846
-7 REFUSED	24	197,843
-1 INAPPLICABLE	2,041	24,727,352
0 NO SCHOOL/KINDERGARTEN ONLY	945	10,020,656
1 - 8 ELEMENTARY GRADES 1 - 8	4,197	42,854,987
9 - 11 HIGH SCHOOL GRADES 9 - 11	2,872	29,435,742
12 GRADE 12	6,663	68,954,120
13 1 YEAR COLLEGE	1,452	18,228,523
14 2 YEARS COLLEGE	2,866	37,565,268
15 3 YEARS COLLEGE	722	9,617,079
16 4 YEARS COLLEGE	3,937	50,987,116
17 5+ YEARS COLLEGE	2,632	33,251,040
TOTAL	28,512	327,396,693

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

Update notes

Documentation	File type
Documentation	PDF (540 KB) / HTML
Codebook	PDF (212 KB) / HTML*
SAS Programming Statements	TXT (74 KB)
SPSS Programming Statements	TXT (6.2 KB)
STATA Programming Statements	TXT (8.4 KB)
R Programming Statements	TXT (5.3 KB)

Data	File type**
Data File, ASCII format	ZIP (1.3 MB) / EXE (1.8 MB)
Data File, SAS transport format	ZIP (1.5 MB) / EXE (2.0 MB)
Data File, SAS V9 format	ZIP (1.8 MB)
Data File, Stata format	ZIP (1.8 MB)
Data File, XLSX format	ZIP (6.9 MB)

Questionnaires — see Survey Questionnaires

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

Update notes

Documentation	File type
Documentation	PDF (540 KB) / HTML
Codebook	PDF (212 KB) / HTML*
SAS Programming Statements	TXT (74 KB)
SPSS Programming Statements	TXT (6.2 KB)
STATA Programming Statements	TXT (8.4 KB)
R Programming Statements	TXT (5.3 KB)

For loading ASCII (.dat) fixed-width files

Data	File type**
Data File, ASCII format	ZIP (1.3 MB) / EXE (1.8 MB)
Data File, SAS transport format	ZIP (1.5 MB) / EXE (2.0 MB)
Data File, SAS V9 format	ZIP (1.8 MB)
Data File, Stata format	ZIP (1.8 MB)
Data File, XLSX format	ZIP (6.9 MB)

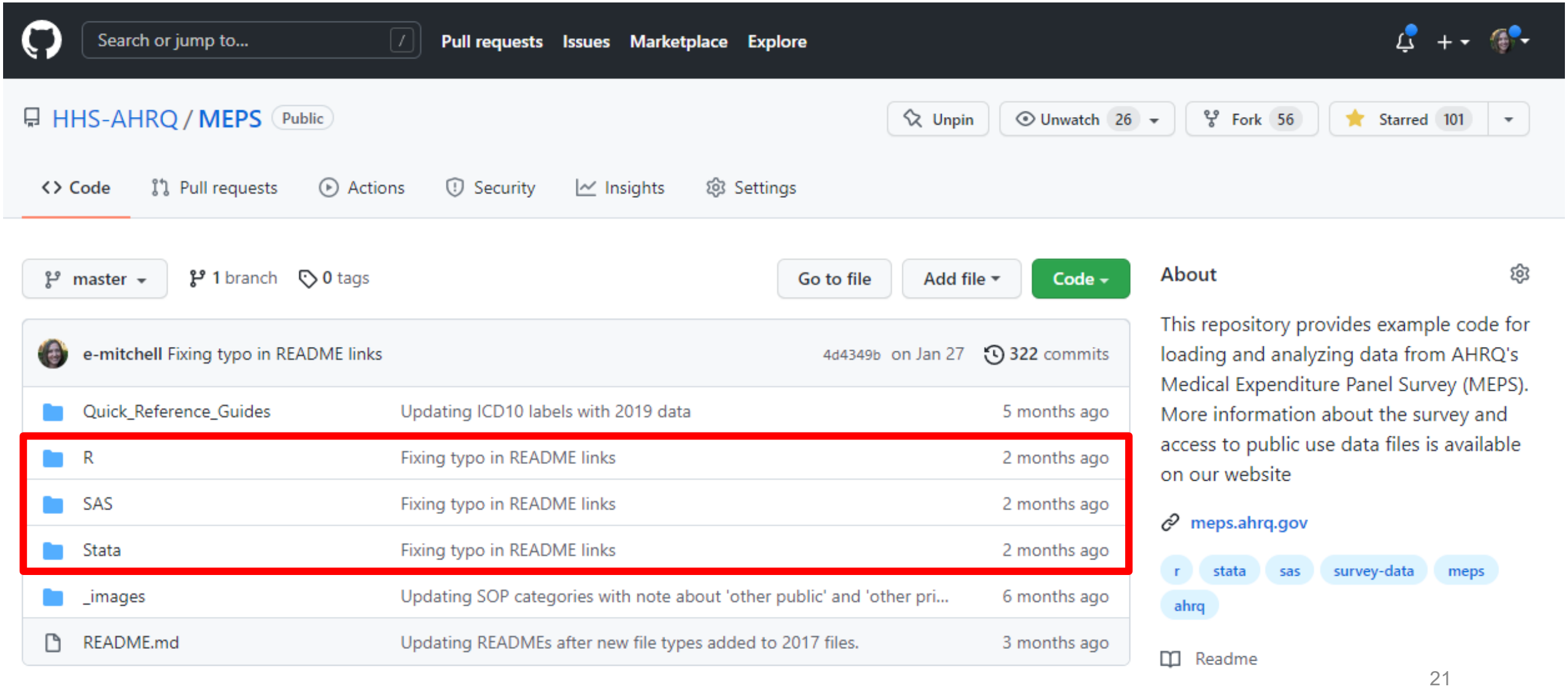
Questionnaires — see Survey Questionnaires

MEPS Public Use Data Files (cont.)

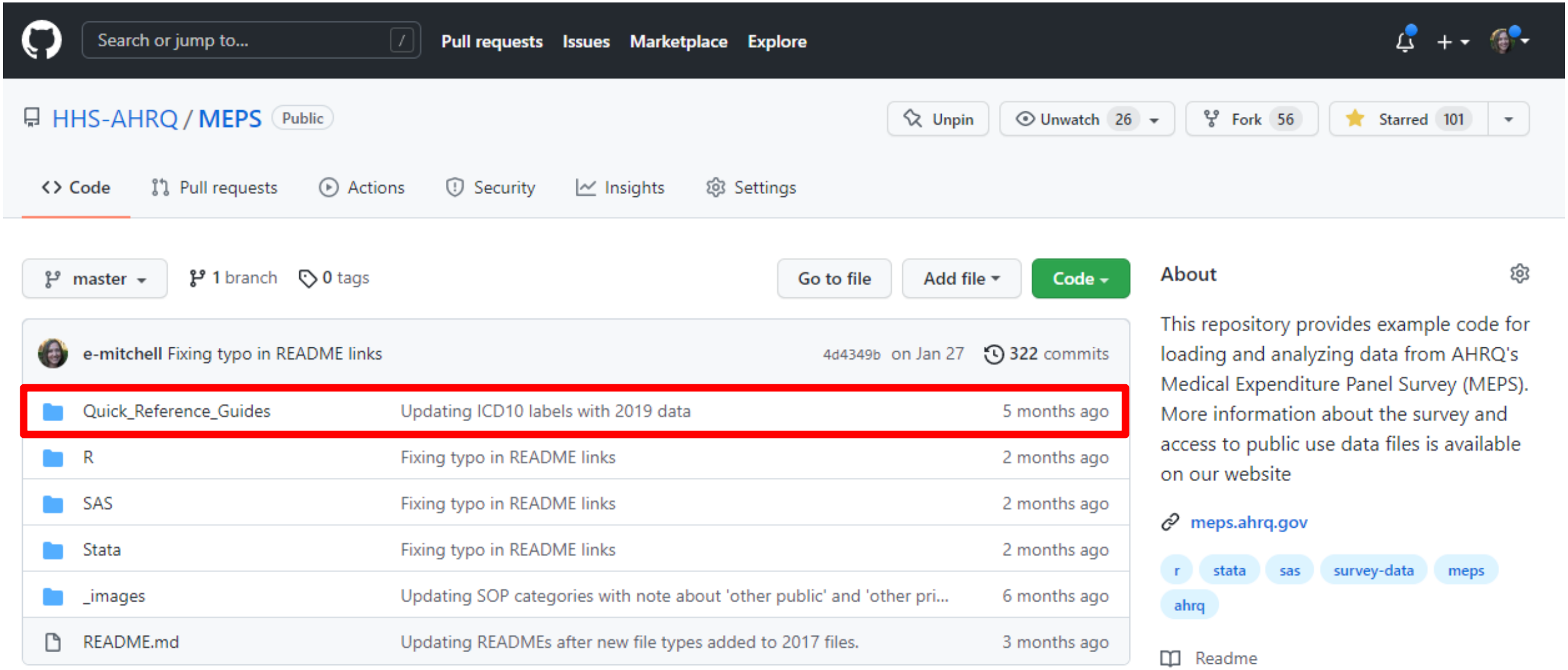
!!!

1996–2016	2017 and later
ASCII (.dat)	ASCII (.dat)
BEST **SAS transport (.ssp)**	~~SAS transport (.ssp)~~
	BEST **SAS V9 (.sas7bdat)**
	BEST **Stata (.dta)**
	Excel (.xlsx)

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

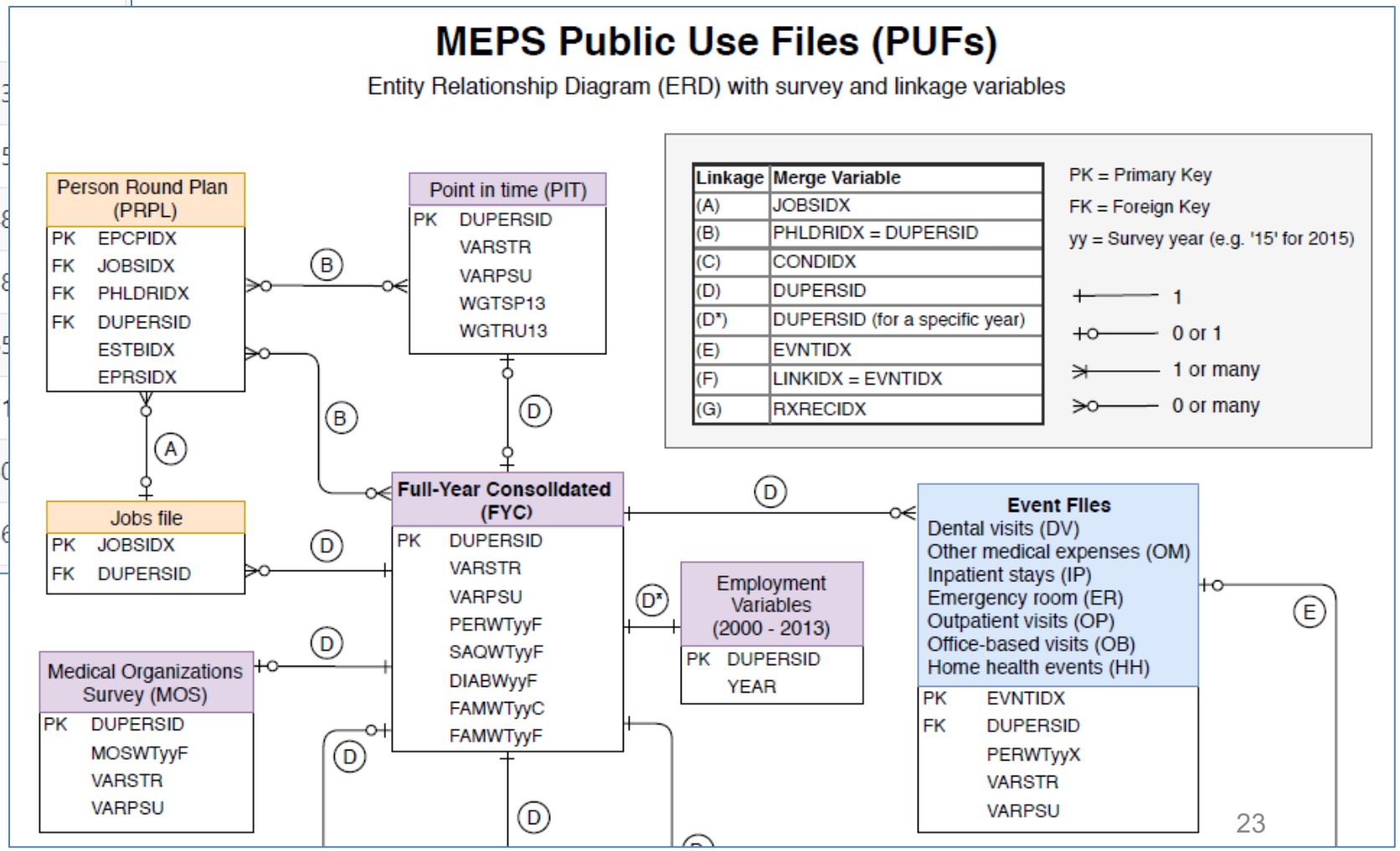


https://github.com/HHS-AHRQ/MEPS



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

- FYC file
- Longitudinal files
- Point-in-Time file
- Medical Organizations Survey

Event level

- ER visits
- Inpatient stays
- Outpatient visits
- Office-based visits
- Dental visits
- Prescribed medicines
- Other medical expenses
- Home health

Conditions level

- Medical Conditions file

Jobs/Insurance-level

- Jobs file
- Person Round Plan file

Record Level and Identifiers (cont.)

Person-level files

PANEL	DUID	PID	DUPERSID
22	2290001	101	2290001101
22	2290001	102	2290001102
22	2290002	101	2290002101

Event files

DUPERSID	EVNTIDX
2290001101	2290001101003301
2290001101	2290001101003401
2290002101	2290002101002601
2290002101	2290002101205301

Conditions file

DUPERSID	CONDN	CONDIDX
2290001102	3	2290001102003
2290002101	2	2290002101002
2290002101	8	2290002101008
2290002101	11	2290002101011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2290001101	3	101	2290001101 3101
2290001101	3	104	2290001101 3104
2290001101	4	104	2290001101 4104
2290001102	3	103	2290001102 3103

Record Level and Identifiers (cont.)

Person-level files

PANEL	DUID	PID	DUPERSID
22	2290001	101	2290001101
22	2290001	102	2290001102
22	2290002	101	2290002101

Event files

DUPERSID	EVNTIDX
2290001101	22900011010033 01
2290001101	22900011010034 01
2290002101	22900021010026 01
2290002101	22900021012053 01

Conditions file

DUPERSID	CONDN	CONDIDX
2290001102	3	2290001102003
2290002101	2	2290002101002
2290002101	8	2290002101008
2290002101	11	2290002101011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2290001101	3	101	2290001101 3101
2290001101	3	104	2290001101 3104
2290001101	4	104	2290001101 4104
2290001102	3	103	2290001102 3103

Record Level and Identifiers (cont.)

Person-level files

PANEL	DUID	PID	DUPERSID
22	2290001	101	2290001101
22	2290001	102	2290001102
22	2290002	101	2290002101

Event files

DUPERSID	EVNTIDX
2290001101	2290001101003301
2290001101	2290001101003401
2290002101	2290002101002601
2290002101	2290002101205301

Conditions file

DUPERSID	CONDN	CONDIDX
2290001102	3	2290001102003
2290002101	2	2290002101002
2290002101	8	2290002101008
2290002101	11	2290002101011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2290001101	3	101	2290001101 3101
2290001101	3	104	2290001101 3104
2290001101	4	104	2290001101 4104
2290001102	3	103	2290001102 3103

Record Level and Identifiers (cont.)

2017

Person-level files

PANEL	DUID	PID	DUPERSID
22	90001	101	90001101
22	90001	102	90001102
22	90002	101	90002101

Jobs file

DUPERSID	RN	JOBSN	JOBSIDX
90001101	3	1	90001101301
90001101	3	4	90001101304
90001101	4	4	90001101404
90001102	3	3	90001102303

2018

Person-level files

PANEL	DUID	PID	DUPERSID
22	2290001	101	2290001101
22	2290001	102	2290001102
22	2290002	101	2290002101

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2290001101	3	101	22900011013101
2290001101	3	104	22900011013104
2290001101	4	104	22900011014104
2290001102	3	103	22900011023103

Variable Naming Conventions

Edited variables end in “X”

RACE**X**

Year-specific variables use last two digits of year

TOTEXP**19**
PERWT**19**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

2018 design changes indicated by “_M18” suffix

JTPAIN31_**M18**

Estimation Variables

Weight Variables

- Person-level (e.g., PERWT19F, DIABW19F, SAQWT19F)
- Family-level (e.g., FAMWT19F, FAMWT19C)
- Longitudinal (e.g., LONGWT)

Variance-Estimation Variables (Stratum and PSU)

- Data after FY 2002: VARSTR, VARPSU
- FY 1996–2001 data: VARSTRyy, VARPSUyy
- Pooling data across 2002 OR 2019: STRA9619, PSU9619 in data file HC-036

Example in hands-on sessions!

MEPS Reserve Codes

-1	Inapplicable	Question was not asked due to skip pattern
-7	Refused	Question was asked and respondent refused to answer
-8	Don’t know	Question was asked and respondent did not know answer
-9	Not ascertained	Interviewer did not record the data
-15	Cannot be computed	Value cannot be derived from data
-10	Top-coded	Variable was top-coded for confidentiality (e.g., hourly wage)

New starting in Data Year 2018

Table of Contents

Programming Example

Compare average medical expenses for persons under age 65 vs. 65 and older in 2019.*

* Not including people that have $0 in expenses

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2019.

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

Process (cont.)

Compare average medical expenses for persons under 65 vs. 65 and older in 2019.

1. Load datasets.

2. Create new variables.

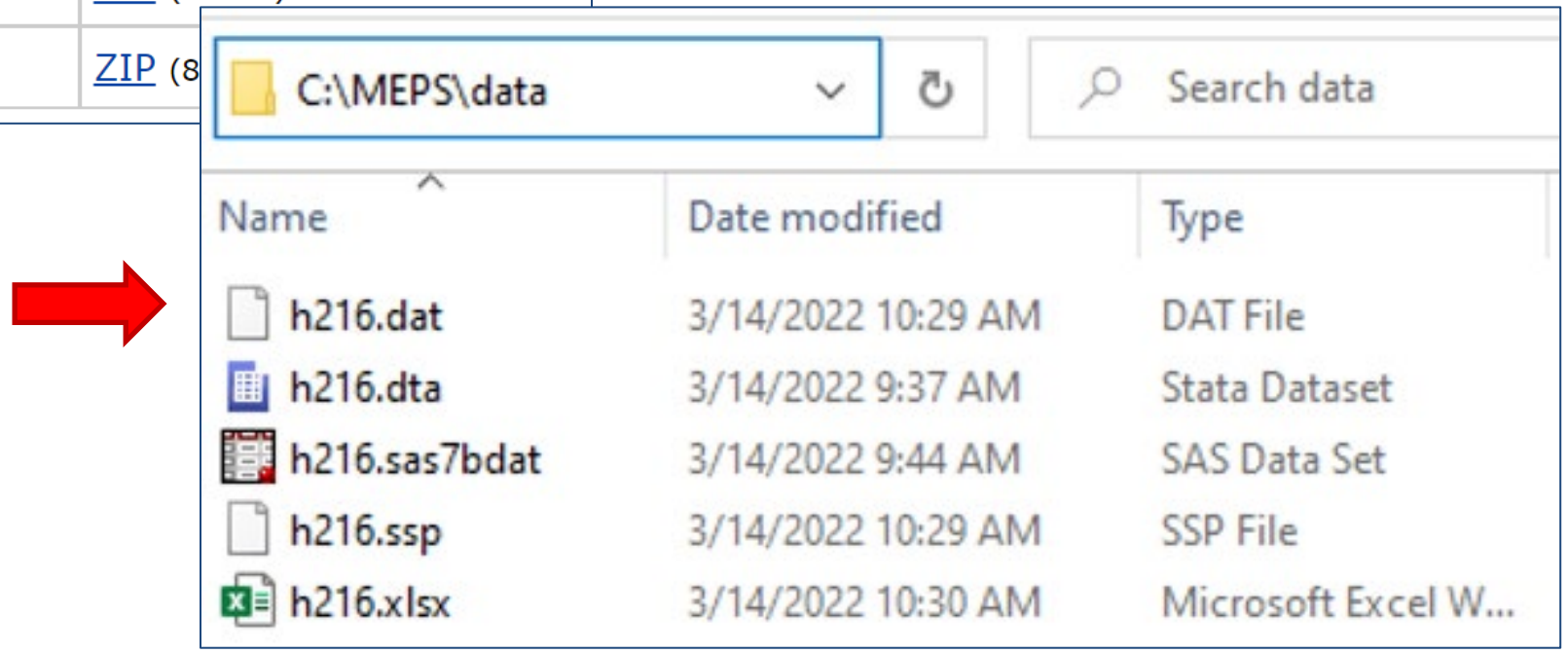
3. Run survey procedure.

4. Examine results.

2019 Full-Year Consolidated file
Person-level

Load Datasets

Data	File type**
Data File, ASCII format	ZIP (11 MB) / EXE (12 MB)
Data File, SAS transport format	ZIP (11 MB) / EXE (12 MB)
Data File, SAS V9 format	ZIP (13 MB)
Data File, Stata format	ZIP (13 MB)
Data File, XLSX format	ZIP (8



Load Datasets (cont.)

	1996–2016	2017 and later
SAS	FILENAME in1 'C:\MEPS\data**h192.ssp**'; **proc xcopy** in = in1 out = WORK IMPORT; **run**; /* creates dataset WORK.h192 */	**data** WORK.h216; set "C:\MEPS\data**h216.sas7bdat**"; **run**;
Stata	**import** sasxport5 "C:\MEPS\data**h192.ssp**" **rename** *, lower	**use** "C:\MEPS\data**h216.dta**", clear **rename** *, lower
R	**install.packages**("foreign") **library**(foreign) **h192** = read.xport("C:/MEPS/data/**h192.ssp**")	**install.packages**("haven") **library**(haven) **h216** = read_dta("C:/MEPS/data/**h216.dta**")

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2019.

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

Age groups:
AGELAST < 65
AGELAST >= 65

Any expenditures:
TOTEXP19 > 0

Create New Variables

SAS

```
data h216;
  set h216;

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

  if TOTEXP19 > 0 then has_exp = 1;
  else if TOTEXP19 = 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 (totexp19 <= 0)
```

R

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

h216 = h216 %>% mutate(
   agecat = ifelse(AGELAST > 64, 2, 1),
   has_exp = ifelse(TOTEXP19 <= 0, 0, 1) )
```

Create New Variables (cont.)

Quality check on new variables

agecat	AGELAST		
	Min	Mean	Max
1 (< 65)	0	32.3	64
2 (65+)	65	74.0	85

has_exp	TOTEXP19		
	Min	Mean	Max
0	0	0	0
1	1	7,746	797,307

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 2019.

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

Mean TOTEXP19
- by Age groups
- if has_exp == 1

Run Survey Procedure

SAS

```
proc surveymeans data = h216 mean;
        stratum VARSTR;
        cluster VARPSU;
        weight PERWT19F;
        var TOTEXP19;
        domain has_exp * AGECAT;
run;
```

R

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

mepsdsgn = svydesign(
  id = ~VARPSU, strata = ~VARSTR, weights = ~PERWT19F,
  data = h216,  nest = TRUE)

svyby(~TOTEXP19, by = ~agecat, FUN = svymean,
     design = subset(mepsdsgn, has_exp==1))
```

Stata

```
svyset [pweight=perwt19f],  strata(varstr)   psu(varpsu)   vce(linearized)   singleunit(missing)

svy, subpop(if has_exp==1): mean totexp19, over(agecat)
```

Run Survey Procedure (cont.)

has_exp	agecat	totexp19	
		Mean	Std. Err.
1	1 (< 65)	5,847	123.3
	2 (65+)	13,464	369.1

Why Survey Procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,847	123.3
	2 (65+)	13,464	369.1

Why Survey Procedures? (cont.)

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,847	123.3
	2 (65+)	13,464	369.1

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,847	**133.0**
	2 (65+)	13,464	**338.4**

Why Survey Procedures? (cont.)

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,847	123.3
	2 (65+)	13,464	369.1

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,847	**133.0**
	2 (65+)	13,464	**338.4**

Ignoring VARSTR, VARPSU, and PERWT

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	**6,059**	**131.3**
	2 (65+)	**13,656**	**323.7**

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2019.

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?

- Statistical Briefs
- MEPS-HC Data Tools

Programming Checklist

- ❑ Well-defined question
- ❑ Checked documentation
- ❑ Reserve codes addressed (-1, -9, -15, etc.)
- ❑ Datasets merged correctly
- ❑ Adequate sample size/ precision
- ❑ Survey procedures
  - ❑ PERWT, VARSTR, VARPSU
  - ❑ Using correct weights (PERWT / FAMWT / LONGWT)
  - ❑ “Domain” analysis for subsets (SAS)
- ❑ Results make sense

Exercises (⭐ difficulty level)

SAS / Stata / R

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

1. National healthcare expenses ⭐
2. Purchases and expenses for narcotic analgesics ⭐⭐
3. Pooling multiple years of MEPS data ⭐⭐⭐⭐
4. Logistic regression ⭐⭐⭐

Thank you!

Emily.Mitchell@ahrq.hhs.gov