



AGENCY FOR HEALTHCARE RESEARCH AND QUALITY



MEPS Data Tools and Programming Overview

Emily M. Mitchell, PhD

Table of Contents



Data Tools

Public Use Files (PUFs)

Programming Example (SAS, Stata, R)

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>



AHRQ Data Tools

Not sure where to start?

Click the dropdown for Topic Areas to help get you to the data you are looking for.

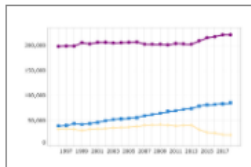
Topic Area

Select

Go

Search Across Data Tools

Featured Dashboard



Health Insurance Coverage
US, 1996 - 2018

AHRQ Priority Topic



OPI
Data and Analysis

Explore the AHRQ Data Tools

Medical Coverage and Expenditures



The **Medical Expenditure Panel Survey (MEPS)** helps researchers and the public explore health insurance coverage, access to care, quality of care, healthcare use, and expenditures.

Healthcare | MEPS-HC

Household Component

For the U.S. civilian population, explore topics like:

- Healthcare use and spending
- Health insurance coverage
- Access to care, quality of care, and diabetes care
- Treated medical conditions
- Prescribed drugs

Health Insurance | MEPS-IC

Insurance Component

Explore national and state-level employer-based health insurance:

- Employer characteristics/offerings
- Employee take-up
- Premiums
- Contributions
- Cost-sharing

National- and state-level statistics and trends about employer-based health insurance

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



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.

Health Insurance

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

Accessibility and Quality of Care

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

Medical Conditions

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

Prescribed Drugs

Purchases and spending by prescribed drug or therapeutic class.



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



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:

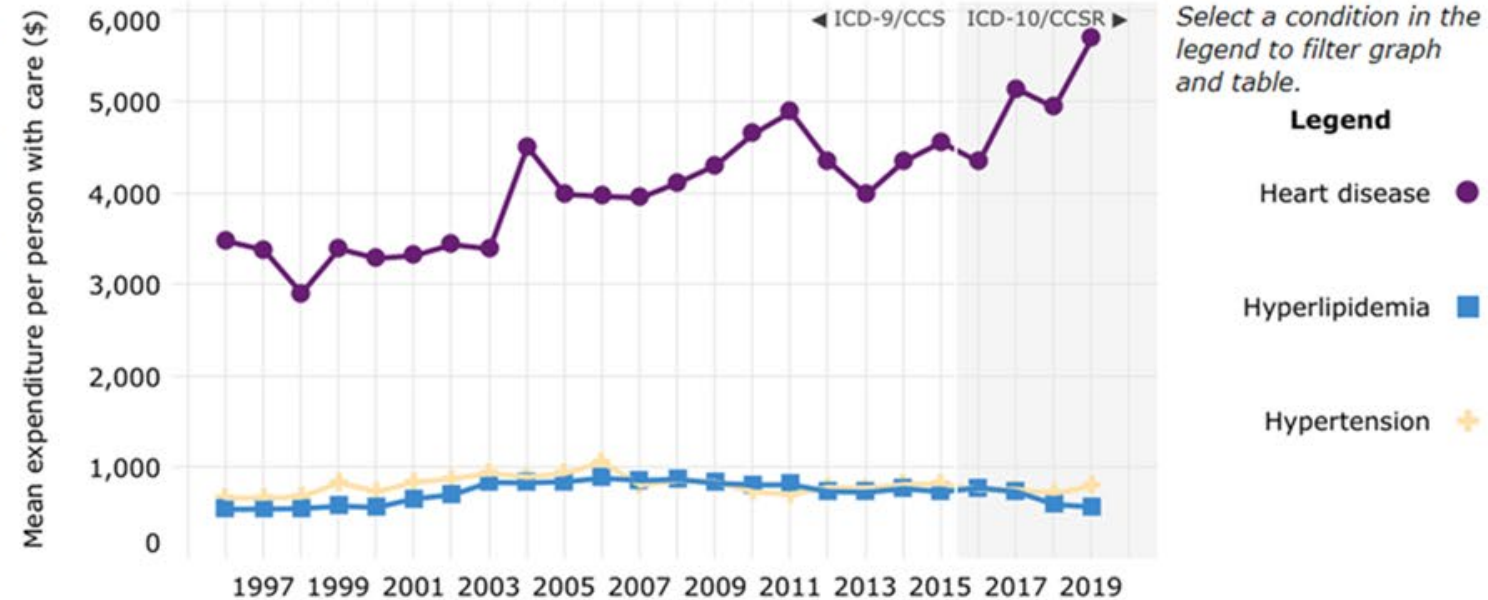
Mean expenditure per person...

Start Year End Year

1996 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 +

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

Estimates:

Mean expenditure per person...

Group by:

Age groups

Group Levels:

(All)

Years:

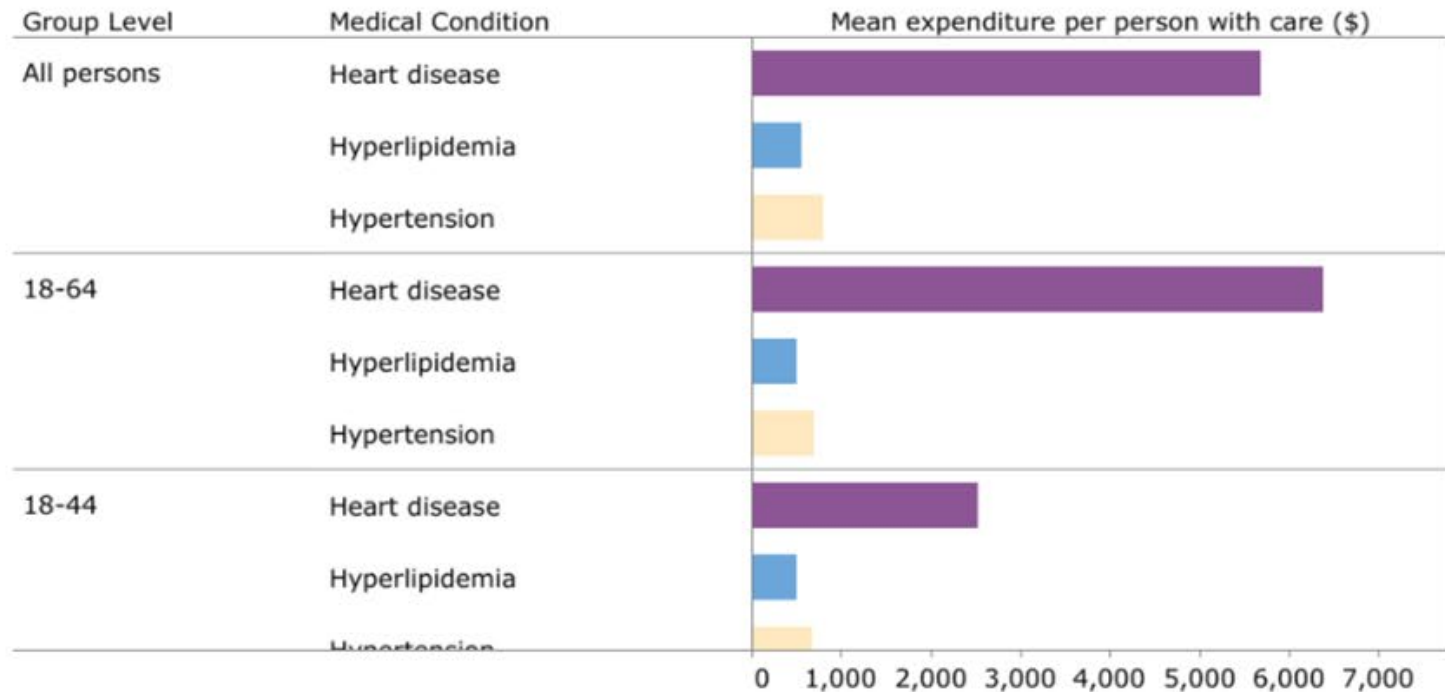
2019

☐ Show SE/95% CI

Search by
medical condition:
(press enter to search)

Clear Selected

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



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

☐ **Pooled Linkage Variable Structure** - Standardized variance strata and FSO 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	2015
EDRECODE	FYC	EDUCATION RECODE (EDITED)					EDRECODE
EDRECODE	PIT	EDUCATION RECODE (EDITED)					EDRECODE
EDUCYEAR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS					
EDUCYR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	EDUCYR
EDUCYR	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS	EDUCYR	EDUCYR	EDUCYR	EDUCYR	EDUCYR
EDUCYR1	FYC	COMPLETED YEARS OF EDUCATION-RD1					
EDUCYR1	PIT	COMPLETED YEARS OF EDUCATION					
EDUCYR13	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS					
EDUCYR2	FYC	COMPLETED YEARS OF EDUCATION-RD2					

Pooled Linkage Variable Structure - Standardized variance strata and FSO variables for a pooled analysis that includes
☐ Preventive Care SA

Go **Reset**

MEPS-HC Variable Explorer

Quick Search

Search Table

educ

Advanced Search

Variable

Selecting a variable under the years will navigate

Variable	Data	Description
EDRECODE	FYC	EDUCATION RECORD
EDRECODE	PIT	EDUCATION RECORD
EDUCYEAR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS
EDUCYR	FYC	YEARS OF EDUC WHEN FIRST ENTERED MEPS
EDUCYR	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS
EDUCYR1	FYC	COMPLETED YEARS OF EDUCATION-RD1
EDUCYR1	PIT	COMPLETED YEARS OF EDUCATION
EDUCYR13	PIT	YEARS OF EDUC WHEN FIRST ENTERED MEPS
EDUCYR2	FYC	COMPLETED YEARS OF EDUCATION-RD2

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

EDUCYR



Table of Contents



Data Tools

Public Use Files (PUFs)

Programming Example (SAS, Stata, R)

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](#)

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

[DVTTCH19](#)

[DVTTRI19](#)

[DVTVA19](#)

[DVTWCP19](#)

[EDUCYR](#)

[ELGRND19](#)

[ELGRND31](#)

[ELGRND42](#)

[ELGRND53](#)

[EMPHAGED](#)

[EMPHDX](#)

[EMPST31](#)

[EMPST31H](#)

[EMPST42](#)

[EMPST42H](#)

[EMPST53](#)

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

1088

1090

EMPLOYMENT STATUS RD 3/1

1393

1394

EMPLOYMENT STATUS RD 3/1 (IMP)

1091

1093

EMPLOYMENT STATUS RD 4/2

1395

1396

EMPLOYMENT STATUS RD 4/2 (IMP)

1094

1096

EMPLOYMENT STATUS RD 5/3

⚡ 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





[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)


  







 **HHS-AHRQ / MEPS** Public

[Unpin](#) [Unwatch](#) 26 [Fork](#) 56 [Starred](#) 101

[Code](#) [Pull requests](#) [Actions](#) [Security](#) [Insights](#) [Settings](#)

[master](#) 1 branch 0 tags [Go to file](#) [Add file](#) [Code](#)

 **e-mitchell** Fixing typo in README links 4d4349b on Jan 27 322 commits


 Quick_Reference_Guides	Updating ICD10 labels with 2019 data	5 months ago
 R	Fixing typo in README links	2 months ago
 SAS	Fixing typo in README links	2 months ago
 Stata	Fixing typo in README links	2 months ago
 _images	Updating SOP categories with note about 'other public' and 'other pri...	6 months ago
 README.md	Updating READMEs after new file types added to 2017 files.	3 months ago

About

This repository provides example code for loading and analyzing data from AHRQ's Medical Expenditure Panel Survey (MEPS). More information about the survey and access to public use data files is available on our website


meps.ahrq.gov

[r](#) [stata](#) [sas](#) [survey-data](#) [meps](#) [ahrq](#)




 [Readme](#)


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





[Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)


  







 **HHS-AHRQ / MEPS** Public

[Unpin](#) [Unwatch](#) 26 [Fork](#) 56 [Starred](#) 101

[Code](#) [Pull requests](#) [Actions](#) [Security](#) [Insights](#) [Settings](#)

[master](#) [1 branch](#) [0 tags](#) [Go to file](#) [Add file](#) [Code](#)

 **e-mitchell** Fixing typo in README links 4d4349b on Jan 27 322 commits


	Quick_Reference_Guides	Updating ICD10 labels with 2019 data	5 months ago
	R	Fixing typo in README links	2 months ago
	SAS	Fixing typo in README links	2 months ago
	Stata	Fixing typo in README links	2 months ago
	_images	Updating SOP categories with note about 'other public' and 'other pri...	6 months ago
	README.md	Updating READMEs after new file types added to 2017 files.	3 months ago

About

This repository provides example code for loading and analyzing data from AHRQ's Medical Expenditure Panel Survey (MEPS). More information about the survey and access to public use data files is available on our website

meps.ahrq.gov

[r](#) [stata](#) [sas](#) [survey-data](#) [meps](#) [ahrq](#)

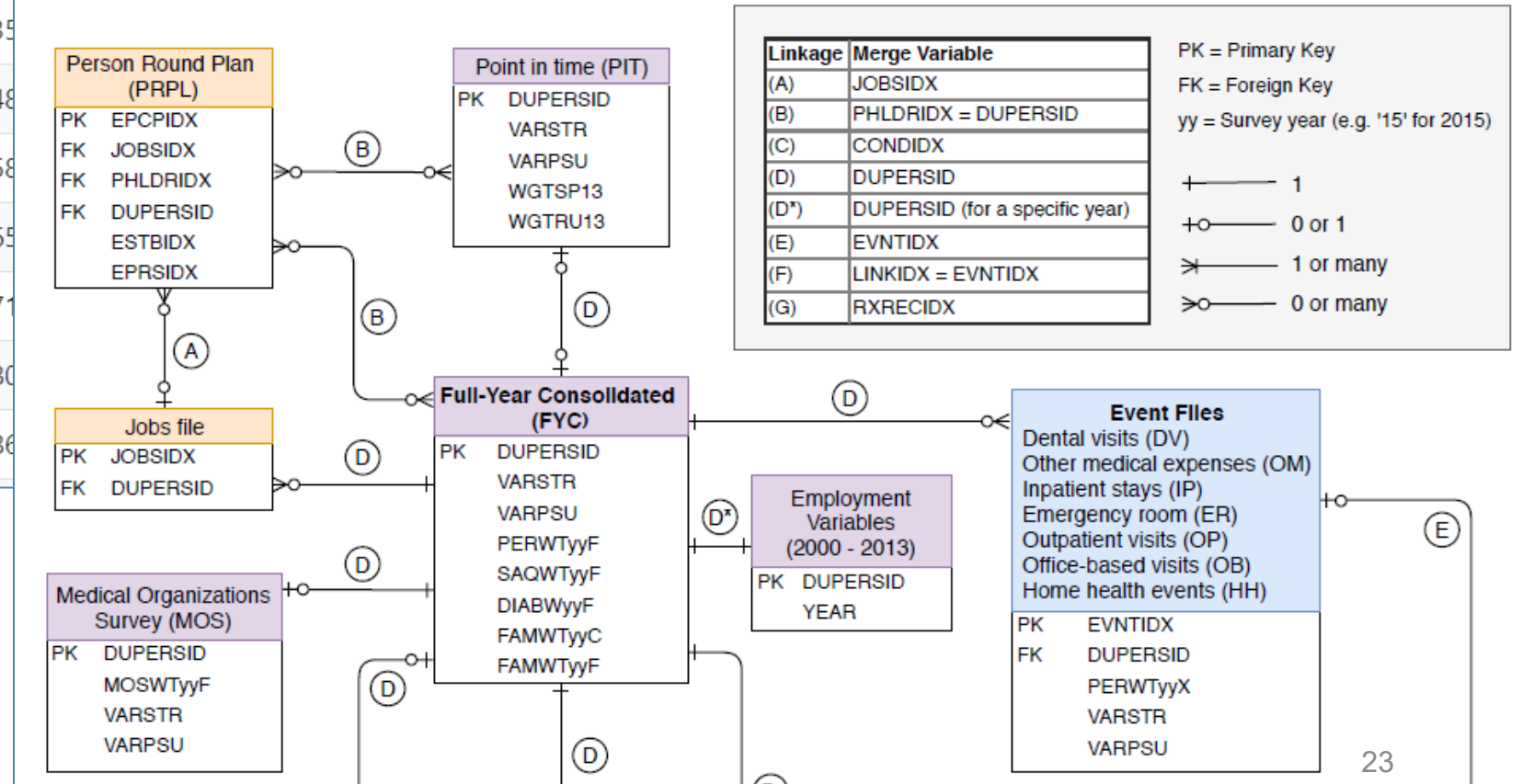
 [Readme](#)

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

MEPS Public Use Files (PUFs)

Entity Relationship Diagram (ERD) with survey and linkage variables



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
24	<u>2490001</u>	101	<u>2490001</u> 101
24	<u>2490001</u>	102	<u>2490001</u> 102
24	<u>2490002</u>	101	<u>2490002</u> 101

Event files

DUPERSID	EVNTIDX
2490001101	<u>2490001101</u> 003301
2490001101	<u>2490001101</u> 003401
2490002101	<u>2490002101</u> 002601
2490002101	<u>2490002101</u> 205301

Conditions file

DUPERSID	CONDN	CONDIDX
2490001102	3	<u>2490001102</u> 003
2490002101	2	<u>2490002101</u> 002
2490002101	8	<u>2490002101</u> 008
2490002101	11	<u>2490002101</u> 011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2490001101	3	101	<u>2490001101</u> 3101
2490001101	3	104	<u>2490001101</u> 3104
2490001101	4	104	<u>2490001101</u> 4104
2490001102	3	103	<u>2490001102</u> 3103

Record Level and Identifiers (cont.)

Person-level files

PANEL	DUID	PID	DUPERSID
24	<u>2490001</u>	101	<u>2490001</u> 1101
24	<u>2490001</u>	102	<u>2490001</u> 1102
24	<u>2490002</u>	101	<u>2490002</u> 2101

Event files

DUPERSID	EVNTIDX
2490001101	<u>2490001101</u> 1003301
2490001101	<u>2490001101</u> 1003401
2490002101	<u>2490002101</u> 1002601
2490002101	<u>2490002101</u> 1205301

Conditions file

DUPERSID	CONDN	CONDIDX
2490001102	3	<u>2490001102</u> 003
2490002101	2	<u>2490002101</u> 002
2490002101	8	<u>2490002101</u> 008
2490002101	11	<u>2490002101</u> 011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2490001101	3	101	<u>2490001101</u> 3101
2490001101	3	104	<u>2490001101</u> 3104
2490001101	4	104	<u>2490001101</u> 4104
2490001102	3	103	<u>2490001102</u> 3103

Record Level and Identifiers (cont.)

Person-level files

PANEL	DUID	PID	DUPERSID
24	<u>2490001</u>	101	<u>2490001</u> 101
24	<u>2490001</u>	102	<u>2490001</u> 102
24	<u>2490002</u>	101	<u>2490002</u> 101

Event files

DUPERSID	EVNTIDX
2490001101	<u>2490001101</u> 003301
2490001101	<u>2490001101</u> 003401
2490002101	<u>2490002101</u> 002601
2490002101	<u>2490002101</u> 205301

Conditions file

DUPERSID	CONDN	CONDIDX
2490001102	3	<u>2490001102</u> 003
2490002101	2	<u>2490002101</u> 002
2490002101	8	<u>2490002101</u> 008
2490002101	11	<u>2490002101</u> 011

Jobs file

DUPERSID	RN	JOBNUM	JOBSIDX
2490001101	3	101	<u>2490001101</u> 3101
2490001101	3	104	<u>2490001101</u> 3104
2490001101	4	104	<u>2490001101</u> 4104
2490001102	3	103	<u>2490001102</u> 3103

Variable Naming Conventions

Edited variables end in “X”

RACEX

Year-specific variables use last two digits of year

TOTEXP20
PERWT20F

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

AGE31X
AGE42X
AGE53X

2018 design changes indicated by “_M18” suffix

JTPAIN31_M18

Estimation Variables

Weight Variables

- ▶ Person-level (e.g., PERWT20F, DIABW20F, SAQWT20F)
- ▶ Family-level (e.g., FAMWT20F, FAMWT20C)
- ▶ 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: STRA9620, PSU9620 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)

Table of Contents

Data Tools

Public Use Files (PUFs)

Programming Example (SAS, Stata, R)

Programming Example

Compare average medical expenses for persons
under age 65 vs. 65 and older in 2020.*

* Not including people that have \$0 in expenses

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2020.

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 2020.

1. Load datasets.

2. Create new variables.

3. Run survey procedure.






4. Examine results.

2020 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 MB)



C:\MEPS\data			Search data
Name	Date modified	Type	
 h224.dat	9/7/2022 10:29 AM	DAT File	
 h224.dta	9/7/2022 9:37 AM	Stata Dataset	
 h224.sas7bdat	9/7/2022 9:44 AM	SAS Data Set	
 h224.ssp	9/7/2022 10:29 AM	SSP File	
 h224.xlsx	9/7/2022 10:30 AM	Microsoft Excel W...	

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.h224;  
  set "C:\MEPS\data\h224.sas7bdat";  
run;
```

Stata

```
import sasxport5 "C:\MEPS\data\h192.ssp"  
rename *, lower
```

```
use "C:\MEPS\data\h224.dta", clear  
rename *, lower
```

R

```
install.packages("foreign")  
library(foreign)  
  
h192 = read.xport("C:/MEPS/data/h192.ssp")
```

```
install.packages("haven")  
library(haven)  
  
h224 = read_dta("C:/MEPS/data/h224.dta")
```

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2020.

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

Age groups:

AGELAST < 65

AGELAST >= 65

Any expenditures:

TOTEXP20 > 0

Create New Variables

SAS

```
data h224;  
  set h224;  
  
  if 0 <= AGELAST <= 64 then agecat = 1;  
  else if AGELAST > 64 then agecat = 2;  
  
  if TOTEXP20 > 0 then has_exp = 1;  
  else if TOTEXP20 = 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 (totexp20 <= 0)
```

R

```
install.packages("dplyr")  
library(dplyr)  
  
h224 = h224 %>% mutate(  
  agecat = ifelse(AGELAST > 64, 2, 1),  
  has_exp = ifelse(TOTEXP20 <= 0, 0, 1) )
```

Create New Variables (cont.)

Quality check on new variables

agecat	AGELAST		
	Min	Mean	Max
1 (< 65)	0	33.0	64
2 (65+)	65	73.8	85

has_exp	TOTEXP20		
	Min	Mean	Max
0	0	0	0
1	1	7,656	1,662,894

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 2020.

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

Mean TOTEXP20

- by Age groups
- if has_exp == 1

Run Survey Procedure

!!!

SAS

```
proc surveymeans data = h224 mean;  
  stratum VARSTR;  
  cluster VARPSU;  
  weight PERWT20F;  
  var TOTEXP20;  
  domain has_exp * AGECAT;  
run;
```

R

```
install.packages("survey")  
library(survey);  
options(survey.lonely.psu='adjust');  
  
mepsdsgn = svydesign(  
  id = ~VARPSU, strata = ~VARSTR, weights = ~PERWT20F,  
  data = h224, nest = TRUE)  
  
svyby(~TOTEXP20, by = ~agecat, FUN = svymean,  
  design = subset(mepsdsgn, has_exp==1))
```

Stata

```
svyset [pweight=perwt20f], strata(varstr) psu(varpsu) vce(linearized) singleunit(missing)  
  
svy, subpop(if has_exp==1): mean totexp20, over(agecat)
```


Run Survey Procedure (cont.)

has_exp	agecat	totexp20	
		Mean	Std. Err.
1	1 (< 65)	5,987	202.1
	2 (65+)	12,866	379.2

Why Survey Procedures?

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,987	202.1
	2 (65+)	12,866	379.2

Why Survey Procedures? (cont.)

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,987	202.1
	2 (65+)	12,866	379.2

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,987	190.5
	2 (65+)	12,866	430.7

Why Survey Procedures? (cont.)

Correct Analysis

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,987	202.1
	2 (65+)	12,866	379.2

Ignoring VARSTR, VARPSU

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,987	190.5
	2 (65+)	12,866	430.7

Ignoring VARSTR, VARPSU, and PERWT

has_exp	agecat	Mean	Std. Err.
1	1 (< 65)	5,991	172.2
	2 (65+)	12,735	361.2

Process

Compare average medical expenses for persons under 65 vs. 65 and older in 2020.

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



Examine Results

Does output make sense?

- ▶ Well-defined question
- ▶ Population estimates
- ▶ Inflation adjustment?

Consistent with other published results?

- ▶ Statistical Briefs
- ▶ MEPS-HC Data Tools

Are estimates reliable?

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

Programming Checks

- ☐ Checked documentation
- ☐ Reserve codes addressed (-1, -9, -15, etc.)
- ☐ Datasets merged correctly
- ☐ Survey procedures
 - ☐ VARSTR, VARPSU, weights
 - ☐ Using correct weights (PERWT / FAMWT / LONGWT)
 - ☐ Correct subset analysis
 - ☐ **SAS**: domain
 - ☐ **R/Stata**: subset after defining survey design

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