



U.S. Department of Health and Human Services



Agency for Healthcare Research and Quality

Advancing Excellence in Health Care • www.ahrq.gov

MEPS Programming Overview

Emily Mitchell, PhD



MEPS Public Use Files

www.meps.ahrq.gov



MEPS Public Use Files

Data files and related documentation are available for downloading from the MEPS Web site: www.meps.ahrq.gov/mepsweb

- ▶ Data files in SAS transport and ASCII formats
- ▶ Documentation files
- ▶ Codebooks
- ▶ SAS, SPSS, and Stata users notes
 - Sample code to read SAS transport and ASCII files, including variable positions
 - Variable labels
 - Variable formats
- ▶ Questionnaires



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

EXAMPLES

-1 Inapplicable

FYC file: [Pregnant](#)

-7 Refused

Event file: [Phone Calls](#)
and [Expenditures](#)

-8 Don't Know

Jobs file: [Hourly Wage](#)

-9 Not Ascertained

-10 Top-Coded ← **Jobs file**



Variable Naming Conventions

Edited Variables end in an “X”

RACEX

Year-specific variables use last two digits of year

TOTEXP14
PERWT14F

Round-specific variables, use two-digit round

- ▶ Certain questions or instrument sections are only asked in certain rounds, e.g. the Self-Administered Questionnaire in rounds 2 and 4

AGE31X
AGE42X
AGE53X



Record-Level and Identifiers

Person-level

- ▶ Full-year consolidated file
- ▶ Longitudinal files

Event-level

- ▶ Event files: RX, 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	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



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



Estimation Variables

Weight Variables – [which one?](#)

- ▶ Person-level (e.g. PERWT13F, DIABW13F, SAQWT13F)
- ▶ Family-level (e.g. FAMWT13F, FAMWT13C)
- ▶ Longitudinal (e.g. LONGWT)

Variance-Estimation Variables:

- ▶ After 2002 FY data: VARSTR, VARPSU
- ▶ 1996-2001 FY data: VARSTRyy, VARPSUyy
 - When calculating variances with **pooled data**, use STRA9614, PSU9614 in HC-036



Common Mistakes

Mistakes

Not reading documentation or checking results

Ignoring reserve codes: -1, -9, -7, -8, etc.

Sub-setting to specific populations

Using default analysis methods
(e.g. proc means)

Using incorrect weights
(e.g. using PERWT for family-level analysis)

Incorrect merging
- can end up double-counting expenditures

Solutions

Read the documentation!
Check your results!

Check summary statistics of all variables

Use survey-specific subsetting procedures
(e.g. 'domain' analysis)

Using complex survey methods
(e.g. proc surveymeans)

Verify weights correspond to analysis

Check dataset at every step



Common Mistakes

Not thinking about what statistic is desired, e.g.

- ▶ Expenditures for persons who reported being diabetic
Conditions file + FYC file

- ▶ Expenditures for persons who were treated with at least one diabetes-associated event
Event file + FYC file

- ▶ Expenditures for events associated with diabetes
Event file







Programming Example


How do 2014 medical expenses for the elderly (age 65 and over) compare to those for persons under 65?*

* Not including people that have \$0 in expenses



MEPS GitHub page





 This organization Search Pull requests Issues Gist   




Agency for Healthcare Research and Quality (AHRQ)

The Agency for Healthcare Research and Quality (AHRQ) is the lead Federal agency charged with improving the safety and quality of America's health care system.

Rockville, MD <http://www.ahrq.gov/>


Repositories  People 5  Teams 0  Projects 0  Settings

Type: All Language: All Customize pinned repositories 


MEPS

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

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


 SAS Updated 6 hours ago

Top languages

 SAS

People

5 >



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



Process

1. Download / load datasets
2. Create Analytic File
3. Quality checks on new variables
4. Run a survey procedure
5. Examine results

1. Download / load datasets

2. Create Analytic File

3. Quality checks on new variables

4. Run a survey procedure

5. Examine results



Download datasets

FYC 2014 (h171)

Documentation	File type
Documentation	PDF (982 KB) / HTML
Codebook	PDF (1.6 MB) / HTML *
SAS Programming Statements	TXT (500 KB)
SPSS Programming Statements	TXT (342 KB)
STATA Programming Statements	TXT (416 KB)
2007 Industry Codes File	PDF (229 KB) / HTML
2010 Occupation Codes File	PDF (225 KB) / HTML
Data	File type**
Data File, ASCII format	ZIP (9.3 MB) / EXE (6.3 MB)
Data File, SAS transport format	ZIP (12 MB) / EXE (7.5 MB)



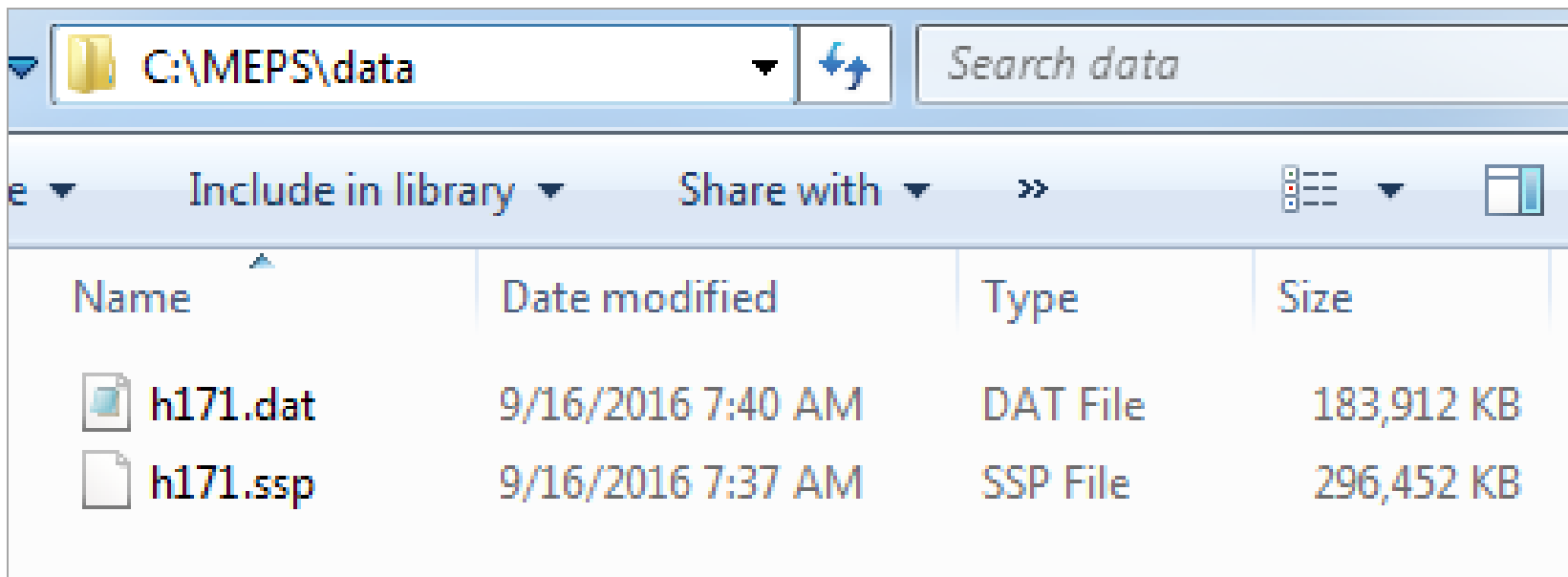
Download datasets



FYC 2014 (h171)

Documentation	File type
Documentation	PDF (982 KB) / HTML
Codebook	PDF (1.6 MB) / HTML *
SAS Programming Statements	TXT (500 KB)
SPSS Programming Statements	TXT (342 KB)
STATA Programming Statements	TXT (416 KB)
2007 Industry Codes File	PDF (229 KB) / HTML
2010 Occupation Codes File	PDF (225 KB) / HTML
Data	File type**
Data File, ASCII format	ZIP (9.3 MB) / EXE (6.3 MB)
Data File, SAS transport format	ZIP (12 MB) / EXE (7.5 MB)

Download datasets

Store .dat and .ssp files in a local directory:



Name	Date modified	Type	Size
 h171.dat	9/16/2016 7:40 AM	DAT File	183,912 KB
 h171.ssp	9/16/2016 7:37 AM	SSP File	296,452 KB

Load datasets

SAS 9.4

```
FILENAME in_h171 'C:\MEPS\data\h171.ssp';  
  
proc xcopy in = in_h171 out = WORK IMPORT;  
run;
```

Stata

```
import sasxport "C:\MEPS\data\h171.ssp"
```

R

```
install.packages("foreign")  
library(foreign)  
  
h171 = read.xport("C:/MEPS/data/h171.ssp")
```

1. Download / load datasets

2. Create Analytic File

3. Quality checks on new variables

4. Run a survey procedure

5. Examine results



Create Analytic File

SAS 9.4

```
data h171;  
  set h171; /*keep= is optional*/  
  
  if 0 <= AGELAST <= 64 then AGECA1=1;  
  else if AGELAST > 64 then AGECA1=2;  
  
  if TOTEXP14>0 then HAS_EXP=1;  
  else if TOTEXP14=0 then HAS_EXP=2;  
run;
```

Stata

```
gen agecat = 1  
replace agecat = 2 if agelast > 64  
  
gen has_exp = 1  
replace has_exp = 2 if (totexp14 <= 0)
```

R

```
install.packages("dplyr")  
library(dplyr)  
  
h171 = h171 %>%  
  mutate(  
    agecat = ifelse(AGELAST > 64, 2, 1),  
    has_exp = ifelse(TOTEXP14 <= 0, 2, 1)  
  )
```



(optional) Add formats

SAS 9.4

```
proc format;  
  value AGECAT  
    1 = '0-64'  
    2 = '65+';  
run;
```

Stata

```
label define agecat 1 "0-64" 2 "65+"  
label value agecat agecat
```

R

```
h171 = h171 %>%  
  mutate(  
    AGECAT = recode_factor(agecat,  
      '1'='0-64',  
      '2'='65+')  
  )
```


1. Download / load datasets
2. Create Analytic File
- 3. Quality checks on new variables**
4. Run a survey procedure
5. Examine results

QC Checks on New Variables

SAS 9.4

```
proc means data = h171;  
  format agecat agecat. ;  
  class agecat;  
  var agelast;  
run;
```

```
proc means data = h171;  
  class has_exp;  
  var totexp14;  
run;
```

Stata

```
bys agecat:  sum agelast  
bys has_exp: sum totexp14
```

R

```
h171 %>% group_by(AGECAT) %>%  
  summarise(min=min(AGELAST),  
            max=max(AGELAST))
```

```
h171 %>% group_by(has_exp) %>%  
  summarise(min=min(TOTEXP14),  
            max=max(TOTEXP14))
```

QC Checks on New Variables

SAS 9.4

Analysis Variable : AGELAST PERSON'S AGE LAST TIME ELIGIBLE						
AGECAT	N Obs	N	Mean	Std Dev	Minimum	Maximum
0-64	30770	30770	29.8894053	18.4666760	0	64.0000000
65+	4105	4105	73.8068210	6.6630803	65.0000000	85.0000000

Analysis Variable : TOTEXP14 TOTAL HEALTH CARE EXP 14						
HAS_EXP	N Obs	N	Mean	Std Dev	Minimum	Maximum
1	27650	27650	4938.68	14269.34	1.0000000	491858.00
2	7225	7225	0	0	0	0

Stata

```
. bys agecat:    sum agelast
```

```
-> agecat = 0-64
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agelast	30,770	29.88941	18.46668	0	64

```
-> agecat = 65+
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agelast	4,105	73.80682	6.66308	65	85

QC Checks on New Variables

R

AGECAT	min	max
<fctr>	<dbl>	<dbl>
0-64	0	64
65+	65	85

has_exp	min	max
<dbl>	<dbl>	<dbl>
1	1	491858
2	0	0

1. Download / load datasets
2. Create Analytic File
3. Quality checks on new variables
- 4. Run a survey procedure**
5. Examine results



Run a Survey Procedure

SAS 9.4

```
proc surveymeans data = h171 mean ;  
    stratum VARSTR;  
    cluster VARPSU;  
    weight PERWT14F;  
    var TOTEXP14;  
    domain has_exp * AGECAT;  
    format agecat agecat.;  
  
run;
```

R

```
library(survey); options(survey.lonely.psu='adjust');  
  
mepsdsgn = svydesign(  
    id = ~VARPSU,  
    strata = ~VARSTR,  
    weights = ~PERWT14F,  
    data = h171,  
    nest = TRUE)  
  
svyby(~TOTEXP14,  
    by = ~AGECAT,  
    FUN = svymean,  
    design = subset(mepsdsgn,has_exp==1))
```

Stata

```
svyset [pweight=perwt14f], strata(varstr) psu(varpsu) vce(linearized) singleunit(missing)  
  
svy, subpop(if has_exp==1): mean totexp14, over(agecat)
```

Run a Survey Procedure

SAS

Domain Statistics in HAS_EXP*AGECAT					
HAS_EXP	AGECAT	Variable	Label	Mean	Std Error of Mean
1	0-64	TOTEXP14	TOTAL HEALTH CARE EXP 14	4429.921466	149.960906
	65+	TOTEXP14	TOTAL HEALTH CARE EXP 14	10890	302.614176
2	0-64	TOTEXP14	TOTAL HEALTH CARE EXP 14	0	0
	65+	TOTEXP14	TOTAL HEALTH CARE EXP 14	0	0

```

_subpop_1: agecat = 0-64
_subpop_2: agecat = 65+

```

Stata

Over	Linearized			
	Mean	Std. Err.	[95% Conf. Interval]	
totexp14				
_subpop_1	4429.921	149.9609	4134.223	4725.62
_subpop_2	10890	302.6142	10293.29	11486.7

Run a Survey Procedure

SAS

Domain Statistics in HAS_EXP*AGECAT					
HAS_EXP	AGECAT	Variable	Label	Mean	Std Error of Mean
1	0-64	TOTEXP14	TOTAL HEALTH CARE EXP 14	4429.921466	149.960906
	65+	TOTEXP14	TOTAL HEALTH CARE EXP 14	10890	302.614176
2	0-64	TOTEXP14	TOTAL HEALTH CARE EXP 14	0	0
	65+	TOTEXP14	TOTAL HEALTH CARE EXP 14	0	0

R

	AGECAT	TOTEXP14	se
0-64	0-64	4429.921	149.9609
65+	65+	10889.997	302.6142

1. Download / load datasets
2. Create Analytic File
3. Quality checks on new variables
4. Run a survey procedure
- 5. Examine results**



Examine the Results

Table 1: Total Health Services-Median and Mean Expenses per Person With Expense and Distribution of Expenses by Source of Payment: United States, 2014

Population Characteristic	Population (in thousands)	Percent with expense	Per person with an expense	
			Median	Mean
Total	318,440	85.1	1,373	5,531
Age in years				
Under 65	270,479	83.1	1,040	4,430
Under 5	19,912	90.7	599	2,588
5-17	53,888	86.1	632	2,294
18-44	112,965	76.4	893	3,522
45-64	83,714	88.5	2,165	7,274
65 and over	47,961	96.4	4,875	10,890

Exercises (★ difficulty level)

SAS / Stata

1. National health care expenses by type of service ★
2. Expenditures and utilization of antipsychotics (from SB #275) ★ ★
3. Constructing family-level estimates ★ ★
4. Use and expenditures for persons with diabetes ★ ★
5. Expenditures for all events associated with diabetes ★ ★ ★ ★ ★
6. Pooling multiple years of MEPS data ★ ★ ★
7. Constructing insurance status variables from monthly insurance variables ★ ★ ★ ★ ★
8. Pooling longitudinal files ★ ★ ★





Exercises (★ difficulty level)


R

1. Estimates from MEPS summary table 2013, Table 1. ★
2. National health care expenses in 2013 (tables and Figure 1 from Stat brief #491)
=> ** Includes ggplot example ** ★★ ★



MEPS GitHub page





 This organization Search Pull requests Issues Gist   




Agency for Healthcare Research and Quality (AHRQ)

The Agency for Healthcare Research and Quality (AHRQ) is the lead Federal agency charged with improving the safety and quality of America's health care system.

Rockville, MD <http://www.ahrq.gov/>


Repositories  People 5  Teams 0  Projects 0  Settings

Type: All Language: All Customize pinned repositories 


MEPS

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

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


 SAS Updated 6 hours ago

Top languages

 SAS

People

5 >



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

Questions?

emily.mitchell@ahrq.hhs.gov