Using MEPS Data Files

Programming Overview

Public Use Files (PUF)

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 ...

What's New Highlights

Public Use Files (PUF)

- ☐ **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**
- ☐ **Longitudinal Data files**
- ☐ **Preventive Care Self-Administered Questionnaire file (2014)**
- ☐ **Supplemental Variables files (1996-2000)**
- ☐ **Health Insurance Plan Abstraction file (1996)**
- ☐ **Long Term Care file (1998)**
- ☐ **Household Component Event files** ⓘ

  Data for the calendar year on unique household-reported medical events.
  - ☐ **Prescribed Medicines files**
  - ☐ **Dental Visits files**
  - ☐ **Other Medical Expenses files**
  - ☐ **Hospital Inpatient Stays files**
  - ☐ **Emergency Room Visits files**
  - ☐ **Outpatient Visits files**
  - ☐ **Office-Based Medical Provider Visits files**
  - ☐ **Home Health files**
  - ☐ **Appendix to MEPS Event files**
- ☐ **Household Component Point-in-Time files** ⓘ

Public Use Files (PUF)

Documentation	File type
Documentation	PDF (1.0 MB) / HTML
Codebook	PDF (1.1 MB) / HTML*
SAS Programming Statements	TXT (497 KB)
SPSS Programming Statements	TXT (339 KB)
STATA Programming Statements	TXT (413 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 (9.5 MB) / EXE (6.7 MB)
Data File, SAS transport format	ZIP (13 MB) / EXE (7.5 MB)

Public Use Files (PUF)

Data files in SAS transport format

```
PROC XCOPY IN=IN1 OUT=PUFLIB IMPORT; /*convert transport to SAS data*/


import sasxport "C:\MEPS\data\h181.ssp"    /*to Stata data*/
save "C:\MEPS\data\h181.dta", replace


h163 = read.xport("C:\MEPS\data\h181.ssp")  /*to R data (need 'foreign' package) */


GET SAS Data='C:\MEPS\data\h181.ssp'  /* Use SPSS to read transport file*/
```

Public Use Files (PUF)

Data files in ASCII format

o SAS, SPSS, and Stata user statements
o Variable labels
o Variable formats

```
INPUT @1      DUID      5.0         LABEL DUID     ='DWELLING UNIT ID'
      @6      PID       3.0               PID      ='PERSON NUMBER'
      @9      DUPERSID $8.0               DUPERSID ='PERSON ID (DUID + PID)'
      @17     PANEL     2.0               PANEL    ='PANEL NUMBER'
      @19     FAMID31  $2.0

VALUE DSA1C53F
  -9 = '-9 NOT ASCERTAINED'
  -8 = '-8 DK'
  -7 = '-7 REFUSED'
  -1 = '-1 INAPPLICABLE'
  0 = '0 NUMBER OF TIMES'
  1 - 94 = '1 - 94 NUMBER OF TIMES'
  95 = '95 95 TIMES OR MORE'
  96 = '96 DID NOT HAVE A1C BLOOD TEST'
  ;
```

Public Use Files (PUF)

Documentation Files

o Contain general information about MEPS
o Contain survey information specific to each file
o Contain variable-source crosswalk to link back to questionnaire items
o Discuss all variables on the file. For example,

For user convenience this file contains a constructed variable INSCOV15 that summarizes health insurance coverage for the person in 2015, with the following three values:

1 = ANY PRIVATE (Person had any private insurance coverage [including TRICARE/CHAMPVA] any time during 2015)

2 = PUBLIC ONLY (Person had only public insurance coverage during 2015)

3 = UNINSURED (Person was uninsured during all of 2015)

Public Use Files (PUF)

Codebooks

○ Contain formatted frequencies for all variables on the file
○ Good source of overview information
○ List both weighted and un-weighted estimates

NAME	DESCRIPTION	FORMAT	TYPE	START	END
HRWG31X	HOURLY WAGE RD 3/1 CMJ (IMP)	6.2	NUM	1038	1043

VALUE	UNWEIGHTED	WEIGHTED BY PERWT15F
-10 HOURLY WAGE >= $80.00	162	2,195,356
-9 NOT ASCERTAINED	206	1,352,724
-1 INAPPLICABLE	21,531	182,248,861
.28 - 79.55 HOURLY WAGE	13,528	135,626,309
TOTAL	35,427	321,423,251

MEPS Reserved 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 Hourly wage >=$80** – Hourly wage was top-coded for confidentiality

Variable Naming Conventions

- Edited Variables end in an "X"
  - ► For example: RACEX, RACEV1X
- Names of year specific variables use last two digits of year
  - ► For example: TOTEXP15, PERWT15F, AGE15X
- For round specific variables, round designation is indicated at the end of the variable or immediately before the "X" in the case of edited variables
  - ► For example: AGE31X, AGE42X, AGE53X
  - ► Certain questions or instrument sections are only asked in certain rounds, e.g. the Self-Administered Questionnaire in rounds 2 and 4

Record Level and Identifiers

- Person-level files have DUPERSID as the identifier
- Event-level files have EVNTIDX (DUPERSID+Event Number)
- Condition-level files have CONDIDX (DUPERSID+CONDN)
- Job-level files have JOBSIDX (DUPERSID+RN {Round Identifier}+JOBSN {Job Number})

MEPS File Structure

- Person-level file (DUPERSID)
  - ► DUPERSID = DUID + PID
  - ► For example: 30011 (DUID) + 102 (PID) = 30011102 (DUPERSID)

- Person-condition-level file (CONDIDX)
  - ► CONDIDX = DUPERSID + CONDN
  - ► For example: 30011102 (DUPERSID) + 0021 (CONDN) = 300111020021 (CONDIDX)

- Person-event-level file, except prescribed medicines (EVNTIDX)
  - ► EVNTIDX = DUPERSID + event number
  - ► For example: 30011102 (DUPERSID) + 0281 (event number) = 300111020281 (EVNTIDX)

- NOTE: CONDN and event numbers cannot be used for counts

Estimation Variables

- **Weight Variables**
  - ► Person-level (e.g. PERWT15F, DIABW15F, SAQWT15F)
  - ► Family-level (e.g. FAMWT15F, FAMWT15C)

  Some of the distinctions between CPS-and MEPS-defined families are that MEPS families include and CPS families do not include: non-married partners, foster children, and in-laws. These persons are considered as members of separate families for CPS-like families.

  - ► Longitudinal (e.g. LONGWT)
    - o Panel-specific for analysis over a 2-year period

- **Variance-Estimation Variables**
  - ► 2002-2015 FY data: VARSTR, VARPSU
  - ► 1996-2001 FY data: VARSTRyy, VARPSUyy
  - ► When calculating variances with pooled data, use STRA9615, PSU9615 in HC-036 (pooled 1996-2015 data)

Pre-Programming Questions

- What level analysis?
  - ► Per person
  - ► Per event (e.g. office visit, inpatient stay)
  - ► Per family
- Adequate sample size?
  - ► Recommendation to have unweighted N >= 100
- Analysis variables from a supplement?
  - ► Diabetes Care Supplement (DCS)
  - ► Self-Administered Questionnaire (SAQ)

Programming Tips

- General programming steps
  - ► Create analytic files by selecting the variables you need (the full-year file has more than 1,800 variables)
  - ► Create new variables if needed
  - ► Format variables if needed
  - ► Merge analytic files if needed
    - ○ Merging files requires that they be sorted by key variable (e.g. DUPERSID)
  - ► Run survey procedures
    - ○ For example: (SAS) SURVEYFREQ, SURVEYMEANS

Create an Analytic File

```
DATA FY1;
SET CDATA.H181 (KEEP=….);
   IF 0 LE AGELAST LE 64 THEN AGECAT=1;
   ELSE IF AGELAST>64 THEN AGECAT=2;


   IF TOTEXP15>0 THEN HAS_EXP=1;
   ELSE IF TOTEXP15=0 THEN HAS_EXP=2;
RUN;
```

Format Variables

```
PROC FORMAT;
   VALUE AGELAST
       0 – 64 = '0-64'
       65 – HIGH = '65+';
   VALUE AGECAT
       1 = '1 0-64'
       2 = '2 65+';
   VALUE GTZERO
       0 = '0'
       0< – HIGH = '>0' ;
RUN;
```

QC Checks on New Variables

PROC FREQ DATA=FY1;

   TABLES AGECAT*AGELAST
          HAS_EXP*TOTEXP15
          / LIST MISSING;

   FORMAT TOTEXP15 GTZERO.
            AGELAST AGELAST.;
RUN;

QC Checks on New Variables

Supporting crosstabs for the new variables

AGECAT	AGELAST	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0-64	30946	87.35	30946	87.35
2	65+	4481	12.65	35427	100.00

HAS_EXP	TOTEXP15	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	>0	28491	80.42	28491	80.42
2	0	6936	19.58	35427	100.00

Run a Survey Procedure

PROC SURVEYMEANS DATA=FY1 MEAN STDERR;

    STRATUM VARSTR;

    CLUSTER VARPSU;

    WEIGHT PERWT15F;

    VAR TOTEXP15;

    DOMAIN HAS_EXP * AGECAT;

    FORMAT AGECAT AGECAT.;

RUN;

Run a Survey Procedure (SAS)

Domain Analysis: HAS_EXP*AGECAT

HAS_EXP	AGECAT	Variable	Mean	Std Error of Mean
1	0-64	TOTEXP15	4946.39	159.06
	65+	TOTEXP15	11446.00	342.99
2	0-64	TOTEXP15	0	0
	65+	TOTEXP15	0	0

Importance of Weights

RACETHX	UNWT %	WT %
HISPANIC	32.54	17.75
WHITE	37.11	60.36
BLACK	19.48	12.30
ASIAN	7.09	5.55
OTHER	3.79	4.04

Post-Programming Questions

- Does the output make sense?
  - ► Population estimates
  - ► Right denominator (eligible population)
- Is it consistent with other published results?
  - ► Stat Briefs or other AHRQ publications
  - ► Independent sources
- Is it the right level analysis?
- Are the estimates reliable?
  - ► Sample size OK?
  - ► Standard errors OK?

Common Mistakes

- Not looking at the documentation
- Not checking results
- Not taking into account the missing data values: -1, -9, -7, -8, etc.
- Subsetting to specific populations prior to running statistical analyses
- Using standard procedures that do not take into account the complex sampling design
- Not using the appropriate weights: SAQ weight for SAQ analytic variables, diabetes weight for diabetes analysis, family weight for family-level analysis

Common Mistakes

- Not being clear about the levels – family, person, event, condition, etc. When merging condition and multiple event files, ending up with multiple records of the same events, thus overestimating expenditures
- Not thinking about what statistic is desired, e.g.
  - All expenditures for persons who reported being diabetic
  - All expenditures for persons who were treated with at least one diabetes-associated event
  - Expenditures for events associated with diabetes

General Tips

- Read the documentation and codebooks
- Do not subset to specific populations prior to running statistical analyses
- Compare program output with codebooks
- Use the correct weight, stratum and PSU variables (including supplement weights)
- Read the documentation!