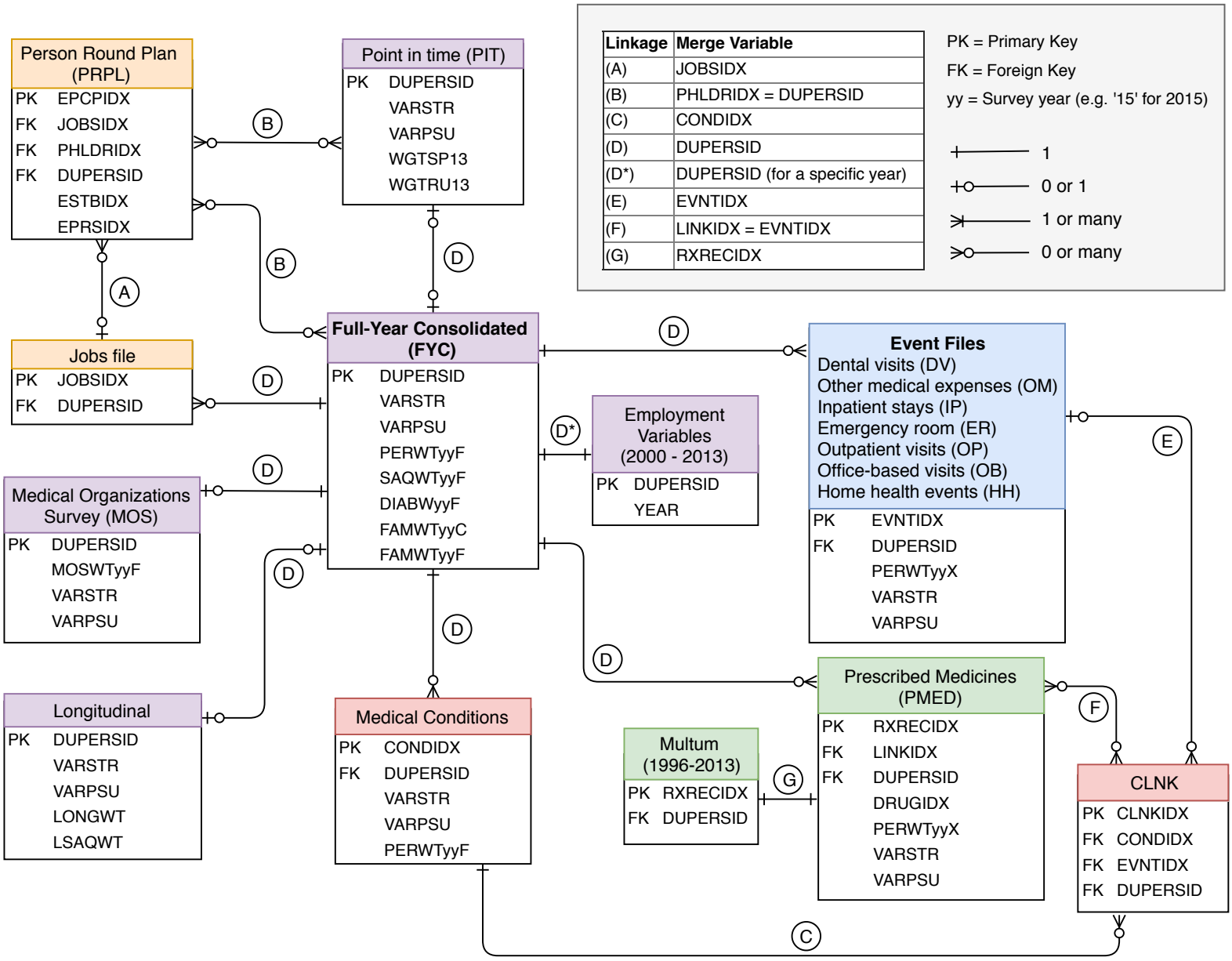
MEPS Public Use Files (PUFs)

Entity Relationship Diagram (ERD) with survey and linkage variables



Linkage	Merge Variable
(A)	JOBSIDX
(B)	PHLDRIDX = DUPERSID
(C)	CONDIDX
(D)	DUPERSID
(D*)	DUPERSID (for a specific year)
(E)	EVNTIDX
(F)	LINKIDX = EVNTIDX
(G)	RXRECIDX

PK = Primary Key

FK = Foreign Key

yy = Survey year (e.g. '15' for 2015)

1

0 or 1

1 or many

0 or many

Type	Variable	Label
ID	CLNKIDX	CLNK ID: CONDIDX + EVNTIDX
	CONDIDX	CONDITION ID
	DRUGIDX	DRUG ID (DUPERSID + COUNTER)
	DUPERSID	PERSON ID
	EPCPIDX	UNIQUE RECORD ID (EPRSIDX + DUPERSID)
	EPRSIDX	ESTBIDX + POLICYHOLDER ID + ROUND NUMBER
	ESTBIDX	ESTABLISHMENT ID
	EVNTIDX	EVENT ID
	JOBSIDX	JOB ID NUMBER
	LINKIDX	ID FOR LINKAGE TO COND/OTHER EVENT FILES
	PHLDRIDX	POLICYHOLDER'S DUPERSID
	RXRECIDX	UNIQUE RX/PRESCRIBED MEDICINE IDENTIFIER

Type	Variable	Label
Survey	VARPSU	VARIANCE ESTIMATION PSU
	VARSTR	VARIANCE ESTIMATION STRATUM
Weight	DIABWyyF	FINAL DIABETES CARE SUPPLEMENT WEIGHT
	FAMWTyyC	POVERTY ADJUSTED FAMILY WEIGHT (CPS FAMILY)
	FAMWTyyF	FINAL FAMILY WEIGHT
	LONGWT	LONGITUDINAL WEIGHT
	LSAQWT	LONGITUDINAL SAQ WEIGHT
	MOSWTyyF	MOS FILE FINAL PERSON WEIGHT
	PERWTyyF	FINAL PERSON WEIGHT
	SAQWTyyF	FINAL SAQ PERSON WEIGHT
	WGTRU13	FAMILY WEIGHT, ROUND 1/3
	WGTSP13	PERSON WEIGHT, ROUND 1/3

This graphic contains information for the most common linkages among the most frequently used datasets. For complete descriptions of all MEPS Public Use Files, visit: meps.ahrq.gov

AHRQ Agency for Healthcare Research and Quality

MEPS Medical Expenditure Panel Survey

Person-level files

Point-in-time (PIT): Nationally representative data for the MEPS population (civilian, non-institutionalized U.S. population) as of the first half of the year. Variables include survey administration, demographics, employment, health status, and health insurance.

Full-Year Consolidated (FYC): Nationally representative data for the MEPS population for the calendar year. Variables include survey administration, language of interview, demographics, parent identifiers, health status, disability days, access to care, employment, quality of care, patient satisfaction, health insurance, income, and healthcare utilization and expenditure variables.

Longitudinal: A two-year longitudinal file of persons in the MEPS population for all or part of the panel's two-year period. Contains all variables from the FYC data files.

Medical Organizations Survey (MOS): Usual source of care provider characteristics for persons whose providers were respondents to the MEPS MOS. Includes the following variables: survey administration, organizational characteristics, case management and use of clinical quality data, and financial arrangements.

Employment Variables (2000-2013): A supplemental release containing selected employment variables for the years 2000 through 2013. With a few exceptions, variables in this file are fully-imputed versions of variables from the FYC files. For data years 2014 and beyond, these variables are included on the FYC file.

Jobs/Insurance files

Jobs file: Jobs-level data for persons age 16 and older. Each record represents one job reported by a person in a round. Variables include wages, work hours, industry, and occupation.

Person Round Plan (PRPL): Contains records for persons insured through private establishments providing hospital/physician, medigap, dental, vision or prescription medication coverage and includes variables pertaining to managed care and out-of-pocket premiums. The PRPL file is not a person-level file, and linking it to a MEPS full year person-level file requires users making analytic decisions based on understanding the complexity of the PRPL file.

Medical Conditions

Medical Conditions: Provides information on household-reported medical conditions. Variables include ICD and Clinical Classification Codes.

CLNK file: File used for linking the MEPS Medical Conditions file with the event files for the same year.

Event files

Dental visits (DV): Variables include date of the dental visit, type of provider seen, reason for the dental visit, and expenditure data.

Other medical expenses (OM): Contains information on expenditures for medical equipment, supplies, glasses and other medical items. From 1996-2017, each record for visual aids represents total expenses during the reference period for a round, while each record for other items represents the entire year. From 2018 onward, records for all items represent expenses during a round.

Inpatient stays (IP): Variables include date of the hospital inpatient stay, reason for the stay, types of care and services received, condition codes, whether or not medicines were prescribed, and expenditure data.

Emergency room (ER): Variables include date of the ER visit, types of care and services received, condition codes, whether or not medicines were prescribed, and expenditure data.

Outpatient visits (OP): Variables include date of the OP visit, type of provider seen, types of care and services received, condition codes, whether or not medicines were prescribed, and expenditure data.

Office-based visits (OB): Variables include date of the visit, type of provider seen, types of care and services received, condition codes, whether or not medicines were prescribed, and expenditure data.

Home health events (HH): Variables include month of the visit, type of provider, type of services received, and expenditure data. Each record represents one month of home health care.

Prescribed Medicines

Prescribed Medicines (PMED): Each record on this event file represents a unique prescribed medicine event; that is, a prescribed medicine reported as being purchased or otherwise obtained by the household respondent. Variables include an identifier for each unique prescribed medicine; detailed characteristics associated with the event; selected Multum Lexicon variables; conditions, if any, associated with the medicine; the date on which the person first used the medicine; total expenditure and sources of payments; and types of pharmacies that filled the household's prescriptions.

Multum (1996-2013): Variables include generic drug name and therapeutic classes derived from the Multum Lexicon database at Cerner Multum, Inc. For data years 2014 and beyond, these variables are included on the PMED file.

Programming Notes

When using non-person-level files for analysis (e.g. Medical Conditions file or Event files), best programming practices include merging with the FYC file to preserve survey variables for all persons in the data. Example programming codes to calculate average expenditures for an inpatient stay in 2015 are provided below.

SAS

```
libname pufs "C:\MEPS";
ods graphics off;

data h181; set pufs.h181; run;
data h178d; set pufs.h178d; run;

* Merge IP (h178d) and FYC (h181)files
* (keep survey variables from FYC);
proc sort data = h181 out = fyc_svy
 (keep = DUPERSID VARSTR VARPSU PERWT15F);
  by DUPERSID;
run;

proc sort data = h178d out = ip_svy
 (drop = VARSTR VARPSU PERWT15F);
  by DUPERSID;
run;

data ip_merged;
  merge ip_svy fyc_svy;
  by DUPERSID;
  /* Create indicator for IP expenses: */
  any_IP = (IPXP15X >= 0);
run;

* Average expense per IP stay;
proc surveymeans data = ip_merged mean;
  strata VARSTR;
  cluster VARPSU;
  weight PERWT15F;
  var IPXP15X;
  domain any_IP;
run;
```

R

```
options(digits = 10)

library(foreign)
library(dplyr)
library(survey)

h178d <- read.xport('C:/MEPS/h178d.ssp')
h181 <- read.xport('C:/MEPS/h181.ssp')

# Merge IP (h178d) and FYC (h181) files
# (keep survey variables from FYC)
fyc_svy <- h181 %>%
  select(VARSTR, VARPSU, PERWT15F, DUPERSID)

ip_svy <- h178d %>%
  select(-VARSTR, -VARPSU, -PERWT15F)

ip_merged <- full_join(ip_svy, fyc_svy,
  by = "DUPERSID")

# Create design object
MEPSdsgn <- svydesign(
  id = ~VARPSU,
  strata = ~VARSTR,
  weights = ~PERWT15F,
  data = ip_merged,
  nest = TRUE)

# Average expense per IP stay
svymean(~IPXP15X,
   design = MEPSdsgn, na.rm = T)
```

Stata

```
cd "C:\MEPS"

set more off
clear

use h178d
drop varstr varpsu perwt15f

* Merge IP (h178d) and FYC (h181) files
* (keep survey variables from FYC)
merge m:m dupersid using h181, ///
  keepusing(dupersid varstr varpsu perwt15f)

* Create indicator for IP expenses:
gen any_IP = (ipxp15x < .)

* Create design object
svyset [pweight= perwt15f], ///
  strata(varstr)  ///
  psu(varpsu)     ///
  vce(linearized) ///
  singleunit(missing)

* Average expense per IP stay
svy, subpop(any_IP): mean ipxp15x
```

For additional programming examples, visit:
github.com/HHS-AHRQ/MEPS

AHRQ Publication No. 19-M013
March 2019