

Tips for Publishing in Health Care Journals with the Medical Expenditure Panel Survey (MEPS) Data Using SAS®

Seungyoung Hwang, Johns Hopkins University Bloomberg School of Public Health

ABSTRACT

Recent decades have witnessed a large increase in academic interest in health services research. The Medical Expenditure Panel Survey (MEPS) data released from the Agency for Healthcare Research and Quality (AHRQ) have been important tools for healthcare services research and policy in an era of health care reform.

In this paper, the analysis proceeds in two phases. First, we outline eight recently published articles on health services use and expenditures associated with specific medical conditions using the MEPS data. Second, we conduct and extend their statistical analyses, considering depression with heart disease as the key variable of interest. Examples of using the user-friendly survey sampling procedures are drawn from the 2012 MEPS data that were the latest available datasets at the time of study.

Providing step-by-step instructions to perform each technique in SAS along with clear explanations of associated SAS code, it allows readers to understand how to create and manage key variables of interest, prepare SAS programs for the appropriate statistical analyses, and interpret the outputs. This paper is ideally suited to students who are beginning their study of health sciences, social sciences, or public health and to professors and research professionals who are researching in the fields in epidemiology, clinical psychology, or health services research. They find it a useful reference for their own research using MEPS data.

PHASE 1: PUBLISHED ARTICLES USING MEPS DATA

Table 1. Published articles on health services use and expenditures using the MEPS data.

Study	Journal	Age	Diagnosis
Egede et al. 2002	Diabetes Care	adults	Depression
Luo et al. 2004	The Spine Journal	≥18	Back pain
Balu et al. 2006	American Journal of Hypertension	≥18	Hypertension
Kamble et al. 2009	Journal of Asthma	<18, ≥18	Asthma
Bhattacharyya. 2011	The Laryngoscope	adults	Allergic rhinitis
Kawatkar et al. 2012	Arthritis Care & Research	≥18	Rheumatoid arthritis
Raval et al. 2012	Journal of Thyroid Research	≥21	Thyroid disorders
Shirneshan et al. 2013	Journal of Anxiety Disorders	≥18	Anxiety disorders

Table 1 presents a summary of studies on the medical expenditures of specific medical conditions published in healthcare and medical journals. They found adults with hypertension (1), asthma (2), allergic rhinitis (3), rheumatoid arthritis (4), or anxiety disorders among diabetics (5) had significantly greater mean total healthcare expenditures than their counterparts (adjusted incremental costs ranging from \$1,131 to \$2,085). They also found adults with comorbid depression among diabetics (6), back pain (7), or thyroid disorders (8) (adjusted cost ratios ranging from 1.2 to 4.5). Bhattacharyya (3) additionally investigated incremental direct healthcare utilization associated with allergic rhinitis using 2007 MEPS data. He found adult patients with allergic rhinitis had nearly 3 additional office visits and a significant 9 additional prescribed medicines compared to those without allergic rhinitis.

Sociodemographic characteristics and various types of healthcare services use and expenditures for their study samples observed from the Full Year Consolidated Data Files. The diagnoses of specific medical conditions were assessed via International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code or Clinical Classification Code (CCC) in the MEPS Medical Conditions Files.

They assessed univariate associations between potential explanatory variables and specific medical conditions using SAS/STAT® procedures for survey sampling (e.g., PROC SURVEYFREQ procedure for categorical variables and the PROC SURVEYMEANS procedure for continuous variables) (9).

In multivariate analyses, they used: (i) generalized linear models (GLMs) with and without log-transformed data (1, 3, 6-8), (ii) GLMs with log-link and Poisson distribution (2, 5), or (iii) a GLM with log-link and Gamma distribution (4) in order to estimate the incremental medical services, total expenditures, and types of healthcare expenditures (inpatient, outpatient, office-based visits, emergency room visits, prescribed medicines, and other) associated with specific medical conditions, after controlling for influential patient characteristics.

Since certain sub-groups (e.g., blacks, Hispanics, disabled, and low-income households) were oversampled in MEPS, survey sampling strata and primary sampling units (PSUs) were incorporated into their calculation of weighted mean and standard errors. In addition, all individual statistics were computed based on person weight so that their study samples represented the entire U.S. population. Analyses were not limited to a subgroup of the population (e.g., age ≥ 18). Rather, all analyses were stratified by age (e.g., 18+ or <18) using the DOMAIN statement in SAS.

PHASE 2: POTENTIAL PAPER IDEA, DEPRESSION WITH HEART DISEASE

There is a lack of information on healthcare costs and sources of care for depressed adults with heart disease in the United States. In this manner, the 2012 MEPS data (Medical Conditions File (H154) and Full Year Consolidated Data File (H155)) were used to test differences in health services use and expenditures between depressed (599 samples; 6.4 million population) and nondepressed (1,717 samples; 17.4 million population) adults with heart disease in the United States.

The SAS code provided in the Attachment section will create the final data set, called **finaldata**. The variables are as follows:

DUPERSID	is the person id.
PERWT12F	is the person-level weight.
VARPSU	is the primary sampling unit required by the variance estimation programs.
VARSTR	is the sampling strata required by the variance estimation programs.
heart	1 = Heart disease; 2 = No heart disease.
depr	1 = Depression; 2 = No depression.
Adult	1 = 18 or older; 2 = under 18.
AGE12X	1 = 17 or younger; 2 = Between 18 and 59; 3 = 60 or older.
SEX	1 = Male; 2 = Female.
RACEV1X	1 = White; 2 = Nonwhite.
MARRY12X	1 = Married; 2 = Previously married; 3 = Never married.
EDRECODE	1 = No degree; 2 = GED/High school grad; 3 = Associate/Bachelor; 4 = Master/Doctorate.
POVCAT12	1 = Poor; 2 = Never poor; 3 = Low income; 4 = Middle income; 5 = High income.

REGION12	1 = Northeast; 2 = Midwest; 3 = South; 4 = West.
MSA12	1 = Metropolitan Statistical Area (MSA); 2 = No MSA.
UNINS12	1 = Not insured; 2 = Insured.
CCI	is the 2012 Charlson comorbidity index excluding heart disease related diagnosis.
ipvisit	is the number of hospital discharges in 2012.
opvisit	is the number of outpatient department provider visits in 2012.
officevisit	is the number of office-based provider visits in 2012.
ervisit	is the number of emergency room (ER) visits in 2012.
medicine	is the number of prescribed medicines including refills in 2012.
ipcost	is the inpatient expenditures in 2012.
opcost	is the outpatient expenditures in 2012.
officecost	is the office-based expenditures in 2012.
ercost	is the ER-based expenditures in 2012.
rxcost	is the prescribed medicine expenditures in 2012.
othercost	is the other medical expenditures in 2012.
totalcost	is the total healthcare expenditures in 2012.

Among adults with heart disease, we compared demographic, socioeconomic, and clinical characteristics of persons with and without depression in our study sample. Like the eight published articles, univariate associations with depression status were assessed using the PROC SURVEYFREQ procedure for categorical explanatory variables and the PROC SURVEYMEANS procedure for continuous variables (9). To do this, we run the following program:

```
proc surveyfreq data=finaldata;
    weight PERWT12F;
    cluster VARPSU;
    strata VARSTR;
    tables heart*adult*depr / row CL nofreq nostd;
run;
* Depression with Heart Disease      : 27.0%,  6.4 million adults) *;
* No Depression with Heart Disease   : 73.0%, 17.4 million adults) *;

proc surveyfreq data=finaldata;
    weight PERWT12F;
    cluster VARPSU;
    strata VARSTR;
    tables heart*adult*depr*AGE12X / row CL nofreq nostd chisq;
run;
/*
AGE12X
SEX
RACEV1X
MARRY12X
EDRECODE
POVCAT12
REGION12
MSA12
```

```

UNINS12
*/

proc surveymeans data=finaldata;
  weight PERWT12F;
  cluster VARPSU;
  strata VARSTR;
  var CCI;
  domain heart*adult*depr;
run;

proc surveyreg data=finaldata;
  weight PERWT12F;
  cluster VARPSU;
  strata VARSTR;
  class depr;
  domain heart*adult;
  model CCI = depr / noint solution vadjust=none;
  lsmeans depr /diff CL;
run;

proc surveymeans data=finaldata mean clm sum;
  weight PERWT12F;
  cluster VARPSU;
  strata VARSTR;
  var totalcost;
  domain heart*adult*depr;
run;
* Depression with Heart Disease: $110.16 billion in U.S. dollars *;

```

The WEIGHT statement specifies the variable that contains the sampling weights. The CLUSTER and STRATA statements specify the variables that identify clusters and strata in a complex survey sample design, respectively (9). Subgroup analyses (e.g., depressed adults with heart disease, nondepressed adults with heart disease) with the DOMAIN statement were used in order to preserve the entire survey design structure by reading in the entire person-level file.

Results are shown in **Table 2**. Among adults with heart disease, the prevalence of depression in 2012 was estimated at 27% (6.4 million persons) and their total medical expenditures were estimated at \$110 billion in 2012 U.S. dollars.

Compared to nondepressed adults with heart disease, those who have been diagnosed with depression were more likely to be in the age group of 18 to 59 (44.2% vs. 29.5%; $p < 0.001$), female (55.8% vs. 44.6%; $p < 0.001$), white (88.8% vs. 82.9%; $p < 0.001$), and poor (18.3% vs. 12.4%; $p = 0.02$). Alternatively, they were less likely to be married (46.0% vs. 56.4%; $p = 0.001$). Also, the mean \pm SE of CCI score was significantly higher at 1.71 ± 0.09 , for sufferers from depression, compared to nondepressed adults at 1.35 ± 0.05 ($p < 0.001$). Education, region, Metropolitan Statistical Area (MSA), and insured status were not significantly associated with depression and were not further considered.

Table 2. Baseline characteristics of depressed and nondepressed individuals with heart disease (U.S. 2012).

Characteristics	Adults with heart disease (weighted n=23,879,008) ^a				p-value
	Depression (weighted n=6,440,639)		No Depression (weighted n=17,438,368)		
	Weighted %	95% CI	Weighted %	95% CI	
Age in yrs %					<0.001
18 - 59	44.2	38.8 - 49.5	29.5	26.4 - 32.7	
60 and order	55.8	50.5 - 61.2	70.5	67.3 - 73.6	
Gender %					<0.001
Male	44.2	39.9 - 48.5	55.4	52.5 - 58.2	
Female	55.8	51.5 - 60.1	44.6	41.8 - 47.5	
Race/ethnicity %					<0.001
White	88.8	86.3 - 91.4	82.9	80.5 - 85.3	
Nonwhite	11.2	8.6 - 13.7	17.1	14.7 - 19.5	
Marital status %					0.001
Married	46.0	41.0 - 51.0	56.4	53.0 - 59.7	
Previously married ^b	37.5	32.7 - 42.3	33.1	30.0 - 36.3	
Never married	16.5	12.0 - 21.0	10.5	8.8 - 12.2	
Poverty category %					0.02
Poor	18.3	14.5 - 22.2	12.4	10.4 - 14.4	
Near poor	7.6	5.2 - 10.0	7.2	5.5 - 8.9	
Low income	16.7	12.8 - 20.6	17.0	14.3 - 19.7	
Middle income	30.4	25.3 - 35.6	29.2	26.0 - 32.4	
High income	27.0	22.1 - 31.8	34.2	30.7 - 37.6	
CCI score, mean	1.71	1.53 - 1.89	1.35	1.25 - 1.45	<0.001

Notes: GED: Graduate Equivalency Degree; HS: High School; CCI: Charlson Comorbidity Index.

^a Weighted numbers represent projected number of individuals (i.e., national-level estimates).

^b Previously married refers to divorced, separated, or widowed elders.

We then carried out multivariate regression analyses with log-transformed data using PROC SURVEYREG procedure in SAS in order to estimate the incremental health services and direct medical costs (inpatient, outpatient, office-based visits, emergency room visits, prescribed medicines, and other) attributable to depression, after controlling for significant patient characteristics. Adjusted models include terms for characteristics identified by their association ($p < 0.05$) with depression status: age, gender, race/ethnicity, marital status, poverty category, and Charlson comorbidity index (10). To do this, we run the following program:

```
* Unadjusted Healthcare Services & Expenditures Associated with Depression *;
proc surveyreg data=finaldata;
  weight PERWT12F;
  cluster VARPSU;
  strata VARSTR;
  class depr;
  domain heart*adult;
  model ipvisit = depr / s;
  lsmeans depr / diff CL;
```

```

run;
/*
ipvisit
opvisit
officevisit
ervisit
medicine

ipcost
opcost
officecost
ercost
rxcost
othercost
totalcost
*/

* Adjusted Healthcare Services & Expenditures Associated with Depression *;
proc surveyreg data=finaldata;
    weight PERWT12F;
    cluster VARPSU;
    strata VARSTR;
    class depr AGE12X SEX RACEV1X MARRY12X POVCAT12;
    domain heart*adult;
    model log_ipvisit = depr AGE12X SEX RACEV1X MARRY12X POVCAT12 CCI / s;
    lsmeans depr / diff CL;
    ods output DiffFs=temp1;
run;
data temp2;
    set temp1;
    HR = exp(Estimate);
    lower_HR = exp(Lower);
    upper_HR = exp(Upper);
run;
proc print data=temp2 noobs;
    var HR lower_HR upper_HR;
run;
/*
log_ipvisit
log_opvisit
log_officevisit
log_ervisit
log_medicine

log_ipcost
log_opcost
log_officecost
log_ercost
log_rxcost
log_othercost
log_totalcost
*/

```

Results are shown in **Table 3-4**.

Table 3. Comparison of health services use among depressed and nondepressed adults with heart disease (U.S. 2012).

Utilization categories	Adults with heart disease (weighted n=23,879,008)	
	Depression (weighted n=6,440,639)	No Depression (weighted n=17,438,368)
A. Unadjusted	Mean health care use (95% CI)	
Hospital discharges	0.43 (0.36 to 0.51)***	0.29 (0.26 to 0.32)
Outpatient visits	1.26 (0.93 to 1.59)	1.03 (0.84 to 1.22)
Office-based visits	14.35 (12.83 to 15.86)**	11.84 (10.52 to 13.16)
Emergency room visits	0.67 (0.53 to 0.80)***	0.41 (0.36 to 0.46)
Prescribed medicines	55.63 (49.59 to 61.67)***	31.04 (29.35 to 32.73)
B. Adjusted	Relative health care use ratio (95% CI)	
Hospital discharges	1.06 (1.02 to 1.10)**	1.00
Outpatient visits	1.03 (0.95 to 1.12)	1.00
Office-based visits	1.28 (1.15 to 1.41)***	1.00
Emergency room visits	1.08 (1.03 to 1.15)**	1.00
Prescribed medicines	1.89 (1.69 to 2.11)***	1.00

* P<0.05; ** P<0.01; *** P<0.001.

Table 4. Comparison of health care expenditures among depressed and nondepressed adults with heart disease (U.S. 2012).

Cost categories	Adults with heart disease (weighted n=23,879,008)	
	Depression (weighted n=6,440,639)	No Depression (weighted n=17,438,368)
A. Unadjusted	Mean health care expenditures (95% CI), \$	
Inpatient	5,161 (3,965 to 6,358)	4,685 (3,764 to 5,605)
Outpatient	858 (574 to 1,143)	668 (532 to 804)
Office-based	2,852 (2,391 to 3,312)	2,726 (2,333 to 3,118)
Emergency room visits	617 (369 to 866)	437 (343 to 531)
Prescription drugs	5,245 (4,400 to 6,090)***	2,484 (2,247 to 2,721)
Other medical expenses	2,370 (1,533 to 3,207)*	1,312 (914 to 1,709)
Overall expenditure	17,104 (15,060 to 19,149)***	12,311 (11,020 to 13,602)
B. Adjusted	Relative health care expenditures ratio (95% CI), \$	
Inpatient	1.40 (0.96 to 2.06)	1.00
Outpatient	0.95 (0.65 to 1.38)	1.00
Office-based	1.45 (1.18 to 1.77)***	1.00
Emergency room visits	1.46 (1.00 to 2.13)	1.00
Prescription drugs	2.61 (2.14 to 3.19)***	1.00
Other medical expenses	1.88 (1.25 to 2.82)**	1.00
Overall expenditure	1.70 (1.44 to 2.00)***	1.00

* P<0.05; ** P<0.01; *** P<0.001.

A comparison of mean healthcare services use across depression status among persons with heart disease was conducted. Adults with heart disease and depression had more hospital discharges (0.43 vs. 0.29, $p < 0.001$), office-based visits (14.35 vs. 11.84, $p = 0.002$), emergency room visits (0.67 vs. 0.41, $p < 0.001$), and prescriptions (55.63 vs. 31.04, $p < 0.001$) than their counterparts without depression. There were no statistically significant differences in the mean number of outpatient visits.

These associations were unchanged on adjustment for age, gender, race/ethnicity, marital status, poverty level, and Charlson comorbidity index. Specifically, adults with heart disease and depression had more hospital discharges (relative ratio (RR) = 1.06, 95% confidence interval (CI) [1.02 to 1.10]), office-based visits (RR = 1.28, 95% CI [1.15 to 1.41]), emergency room visits (RR = 1.08, 95% CI [1.03 to 1.15]), and prescribed medicines (RR = 1.89, 95% CI [1.69, 2.11]) than their counterparts without depression.

Among adults with heart disease, comparisons of average total, inpatient, outpatient, office-based, emergency room, prescription drug, and other medical expenditures for persons with and without depression were conducted. Adults with heart disease and depression had significantly greater average annual total medical expenditures (\$17,104 vs. \$12,311) than those without depression. In other words, the annual overall incremental medical expenditure associated with depression for persons with heart disease was estimated at \$4,293 per person (SE: \$1,234; $p < 0.001$). Prescription drugs, estimated at \$2,761 (SE: \$419; $p < 0.001$) accounted for the largest proportion of the overall incremental medical expenditures, followed by other medical expenses at \$1,058 (SE: \$468; $p = 0.02$). That is, prescription medications and other medical expenses together accounted for approximately 89% of the total incremental cost. There were no statistically significant cost differences in inpatient, outpatient, office-based, and emergency room visits.

After adjustment, adults with heart disease and depression spent 70% more annual total health care expenditures than those without depression (RR = 1.70, 95% CI [1.44, 2.00]). Likewise, cost of office-based (RR = 1.45, 95% CI [1.18, 1.77]), prescription drugs (RR = 2.61, 95% CI [2.14, 3.19]), and other medical expenses (RR = 1.88, 95% CI [1.25, 2.82]) were significantly associated with depression for adults with heart disease.

In conclusion, depression in adults with heart disease is associated with increased health care use and expenditures, even after adjusting for differences in age, gender, race/ethnicity, marital status, poverty level, and medical comorbidity.

DISCUSSION

In this paper, we provided easy-to-understand explanation with the example SAS code on examining healthcare resource utilization and expenditures using the MEPS data with similar analytic strategies over previously published work. This paper gently guides all SAS users—even those with limited training in statistics or who have never used SAS—through a step-by-step approach to analyze MEPS data and interpret the results. With the provided example SAS code, students and researchers will find all the information they need in order to perform analyses with user-friendly survey sampling procedures, master solutions for problems often encountered in their real-world research, and to publish in healthcare journals with the MEPS data using SAS.

REFERENCE

1. Balu S, Thomas J, 3rd: Incremental expenditure of treating hypertension in the United States. *American journal of hypertension* 2006; 19:810-816; discussion 817
2. Kamble S, Bharmal M: Incremental direct expenditure of treating asthma in the United States. *The Journal of asthma : official journal of the Association for the Care of Asthma* 2009; 46:73-80
3. Bhattacharyya N: Incremental healthcare utilization and expenditures for allergic rhinitis in the United States. *The Laryngoscope* 2011; 121:1830-1833
4. Kawatkar AA, Jacobsen SJ, Levy GD, et al: Direct medical expenditure associated with

- rheumatoid arthritis in a nationally representative sample from the medical expenditure panel survey. *Arthritis care & research* 2012; 64:1649-1656
5. Shirneshan E, Bailey J, Relyea G, et al: Incremental direct medical expenditures associated with anxiety disorders for the U.S. adult population: evidence from the Medical Expenditure Panel Survey. *Journal of anxiety disorders* 2013; 27:720-727
 6. Egede LE, Zheng D, Simpson K: Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. *Diabetes care* 2002; 25:464-470
 7. Luo X, Pietrobon R, Sun SX, et al: Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. *Spine* 2004; 29:79-86
 8. Raval AD, Sambamoorthi U: Incremental Healthcare Expenditures Associated with Thyroid Disorders among Individuals with Diabetes. *Journal of thyroid research* 2012; 2012:418345
 9. SAS/STAT 9.3 User's Guide, Cary, NC, SAS Institute Inc., 2011
 10. D'Hoore W, Bouckaert A, Tilquin C: Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of clinical epidemiology* 1996; 49:1429-1433

ACKNOWLEDGMENTS

This work is dedicated to my parents, Jungja Han and Okgil Hwang. All I have and will accomplish are only possible due to their love and sacrifices. As always, most special thanks to Yun Kyoung Ryu for her advice and encouragement in getting me to finish this paper. She has been incredibly generous with her time in reviewing the draft and making many helpful suggestions. All remaining errors, omissions, and weaknesses are my sole responsibility.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Seungyoung Hwang, MS, MSE
Department of Mental Health,
Bloomberg School of Public Health
Johns Hopkins University
624 North Broadway
Baltimore, MD 21205
Email: shwang25@jhu.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

ATTACHMENT

```
***** BEGINNING OF PROGRAM *****;

* set library *;
options ps=66 ls=94 center pageno=1 mprint nodate nofmterr nolabel;
libname MyData "C:\MEPS 2012";

* copy to work directory *;
proc datasets;
    copy in=MyData out=work;
    select h154
           h155;
quit;

* ===== MEPS 2012 Full Year Consolidated Data File (H155) ===== *;
proc sql;
    create table consoll as
    select DUPERID,

           PERWT12F,
           VARPSU,
           VARSTR,

           AGE12X,
           SEX,
           RACEV1X,
           MARRY12X,
           EDRECODE,
           POVCAT12,
           REGION12,
           MSA12,
           UNINS12,

           IPDIS12      as ipvisit,
           OPTOTV12     as opvisit,
           OBTOTV12     as officevisit,
           ERTOT12      as ervisit,
           RXTOT12      as medicine,

           IPTEXP12     as ipcost,
           OPTEXP12     as opcost,
           OBVEXP12     as officecost,
           ERTEXP12     as ercost,
           RXEXP12      as rxcost,
           TOTEXP12 - sum(IPTEXP12, OPTEXP12, OBVEXP12, ERTEXP12, RXEXP12)
              as othercost,
           TOTEXP12     as totalcost
    from h155
    where PERWT12F > 0      /* positive person weights */
    order by DUPERID;
quit;
* N = 37,182 *;

* ===== MEPS 2012 Medical Conditions File (H154) ===== *;
proc sql;
```

```

create table condition1 as
select DUPERSID,
       PERWT12F,
       VARPSU,
       VARSTR,

       ICD9CODX,
       CCCODEX
from h154
where PERWT12F > 0 and CCCODEX not in ("-1", "-9")
/* positive person weights with nonmissing clinical classification code */
order by DUPERSID;
quit;
* Obs = 114,011 *;

data condition2;
set condition1;

if "100" <= CCCODEX <= "108" or CCCODEX in ("96", "97") then heart = 1;
else heart = 0;
format heart 1.;

if ICD9CODX in ("300", "301", "309", "311") then depr = 1;
else depr = 0;
format depr 1.;

* Diagnoses for Charlson comorbidity index*;
if "440" <= ICD9CODX <= "447" then peri = 1;
else peri = 0;
format peri 1.;

if ICD9CODX in ("290", "291", "294") then deme = 1;
else deme = 0;
format deme 1.;

if ICD9CODX in ("430", "431", "432", "433", "435") then cere = 1;
else cere = 0;
format cere 1.;

if "491" <= ICD9CODX <= "493" then chro = 1;
else chro = 0;
format chro 1.;

if ICD9CODX in ("710", "714", "725") then conn = 1;
else conn = 0;
format conn 1.;

if "531" <= ICD9CODX <= "534" then ulce = 1;
else ulce = 0;
format ulce 1.;

if ICD9CODX in ("571", "573") then mild = 1;
else mild = 0;
format mild 1.;

if ICD9CODX in ("342", "434", "436", "437") then hemi = 1;
else hemi = 0;

```

```

format hemi 1.;

if ICD9CODX in ("403","404") or "580"<=ICD9CODX<="586" then rena = 1;
else rena = 0;
format rena 1.;

if ICD9CODX in ("250") then diab = 1;
else diab = 0;
format diab 1.;

if "140" <= ICD9CODX <= "195" then tumo = 1;
else tumo = 0;
format tumo 1.;

if "204" <= ICD9CODX <= "208" then leuk = 1;
else leuk = 0;
format leuk 1.;

if ICD9CODX in ("200", "202", "203") then lymp = 1;
else lymp = 0;
format lymp 1.;

if ICD9CODX in ("070", "570", "572") then live = 1;
else live = 0;
format live 1.;

if "196" <= ICD9CODX <= "199" then meta = 1;
else meta = 0;
format meta 1.;

run;
* Obs = 114,011 *;

proc sql;
create table condition3 as
select DUPERID,

max(heart) as heart format=1.,
max(depr) as depr format=1.,

max(peri) as peri format=1.,
max(deme) as deme format=1.,
max(cere) as cere format=1.,
max(chro) as chro format=1.,
max(conn) as conn format=1.,
max(ulce) as ulce format=1.,
max(mild) as mild format=1.,
max(hemi) as hemi format=1.,
max(rena) as rena format=1.,
max(diab) as diab format=1.,
max(tumo) as tumo format=1.,
max(leuk) as leuk format=1.,
max(lymp) as lymp format=1.,
max(live) as live format=1.,
max(meta) as meta format=1.
from condition2
group by DUPERID;

```

```

quit;
* N = 28,342 *;

data condition4 (keep=DUPERSID heart depr CCI);
  set condition3;
  CCI = 1*(peri + deme + cere + chro + conn + ulce + mild) +
        2*(hemi + rena + diab + tumo + leuk + lymp) +
        3*(live) +
        6*(meta);
  format CCI 2.;      /* CCI without heart disease */
run;
* N = 28,342 *;

* ===== Merging ===== *;
proc sql;
  create table merged1 as
  select A.DUPERSID,

         A.PERWT12F,
         A.VARPSU,
         A.VARSTR,

         B.heart,
         B.depr,

         A.AGE12X,
         A.SEX,
         A.RACEV1X,
         A.MARRY12X,
         A.EDRECODE,
         A.POVCAT12,
         A.REGION12,
         A.MSA12,
         A.UNINS12,
         B.CCI,

         A.ipvisit,
         A.opvisit,
         A.officevisit,
         A.ervisit,
         A.medicine,

         A.ipcost,
         A.opcost,
         A.officecost,
         A.ercost,
         A.rxcost,
         A.othercost,
         A.totalcost
  from consoll as A left join condition4 as B on A.DUPERSID = B.DUPERSID
  order by A.DUPERSID;
quit;
* N = 37,182 *;

data merged2;
  set merged1;

```

```

if heart in (.,0) then heart = 2;

if depr in (.,0) then depr = 2;

if AGE12X >= 18 then adult = 1;
else adult = 2;

if AGE12X < 0 then AGE12X = .;
else if AGE12X < 18 then AGE12X = 1;
else if AGE12X < 60 then AGE12X = 2;
else AGE12X = 3;

if RACEV1X = 1 then RACEV1X = 1;
else RACEV1X = 2;

if MARRY12X = 1 then MARRY12X = 1;
else if MARRY12X in (2,3,4) then MARRY12X = 2;
else if MARRY12X = 5 then MARRY12X = 3;
else MARRY12X = .;

if EDRECODE < 0 then EDRECODE = .;
else if EDRECODE <= 12 then EDRECODE = 1;
else if EDRECODE <= 13 then EDRECODE = 2;
else if EDRECODE <= 15 then EDRECODE = 3;
else EDRECODE = 4;

if REGION12 = -1 then REGION12 = .;

if MSA12 = -1 then MSA12 = .;
if MSA12 = 0 then MSA12 = 2;

if CCI = . then CCI = 0;

run;
* N = 37,182 *;

data finaldata;
set merged2;

log_ipvisit = log(1 + ipvisit);
log_opvisit = log(1 + opvisit);
log_officevisit = log(1 + officevisit);
log_ervisit = log(1 + ervisit);
log_medicine = log(1 + medicine);

log_ipcost = log(1 + ipcost);
log_opcost = log(1 + opcost);
log_officecost = log(1 + officecost);
log_ercost = log(1 + ercost);
log_rxcost = log(1 + rxcost);
log_othercost = log(1 + othercost);
log_totalcost = log(1 + totalcost);

run;

***** END OF PROGRAM *****;

```