

```

/* Date 1/9/2019

    */
/* PROGRAM CREATED FOR CTS STANDARD DATA
                                                                    */
/* PROGRAM CREATED BY NADIA CHUNG

    */

/* The %INCLUDE uses PROC SQL to call the specific data needed for the
STANDARD DATA analysis.
                                                                    */
%Include 'S:\Researcher Projects\CTS Standard Data\CTS Standard Data SAS
Files\Read Only Source Code\SAS Source Part 2 MSSQL.sas';

/***** The approximate run time to call the data using this
%INCLUDE is 10 seconds *****/
/***** Actual program run times could increase based on the
amount of additional coding *****/

/*****
*****/
/* INSTRUCTIONS - PLEASE READ THIS FIRST.

    */
/* Because this %INCLUDE line calls data directly from the CTS analysis Data
Mart, you do not need to save
                                                                    */
/* or create any temporary or permanent SAS datasets in the code
below.
                                                                    */
/* Instead, perform all of your data/variable manipulations (reclassify
variables, rename variables, etc.) in
                                                                    */
/* PART 1, and then run your procedures (PROC FREQ, PROC PHREG,
etc.) in PART 2.
                                                                    */
/* Save changes to this code and use the _MMDDYYYY.sas indicator in your
SAS code.
                                                                    */
/* Every time this code, or subsequent versions based on it, is run, it will
call the specific data from the CTS
                                                                    */
/* analysis Data Mart that are needed for the analysis mentioned above.
                                                                    */
/* If you need this program to call any additional original variables,
please contact NADIA CHUNG
                                                                    */
/*****
*****/

/* PART 1: PERFORM ANALYSIS-SPECIFIC DATA/VARIABLE MANIPULATIONS HERE
                                                                    */
/* EXAMPLE BELOW SHOWS HOW TO CREATE CATEGORICAL VARIABLE FOR BMI AT CTS
BASELINE
                                                                    */

DATA temporary; SET standarddata;
    * IF YOUR ANALYSIS IS NOT BREAST CANCER RESEARCH THEN EXCLUDE THOSE THAT
CONSENTED ONLY TO BREAST CANCER RESEARCH;
    IF consent_status=0;

```

```

* APPLY CTS CUSTOMARY EXCLUSION THAT TAKES OUT PARTICIPANTS OUTSIDE CA AT
BASELINE IF ANALYZING HOSPITALIZATION
    OR CANCER DATA;
IF California_resident_at_baseline=1;

* CREATE INDICATOR VARIABLE TO IDENTIFY PARTICIPANTS WITH MISSING BMI
VALUE AT BASELINE;
bmi=0;
IF bmi_at_baseline ne . THEN bmi=1;
label bmi='Indicator: BMI=1 if known BMI; BMI=0 if BMI missing';

* CREATE CATEGORICAL VARIABLE FOR WHO BMI CATEGORIES;
* VALUES INCLUDE MISSING, UNDERWEIGHT, NORMAL WEIGHT, OVERWEIGHT,
OBESSE, AND SEVERE OBESITY;
if bmi_at_baseline>=35 then BaselineBMIWHO=6;
else if bmi_at_baseline>=30 then BaselineBMIWHO=5;
else if bmi_at_baseline>=25 then BaselineBMIWHO=4;
else if bmi_at_baseline>=18.5 then BaselineBMIWHO=3;
else if bmi_at_baseline>=0 then BaselineBMIWHO=2;
else if bmi_at_baseline=. then BaselineBMIWHO=1;
label BaselineBMIWHO='1: Missing, 2: <18.5, 3: 18.5-24; 4: 25-29; 5:
30-34; 6: 35+';

* DEFINE CANCER ENDPOINT: EXAMPLE IS 1ST INCIDENT LUNG CANCER FOR
PARTICIPANTS WHO WERE CANCER-FREE AT BASELINE;
lungcancer=0;
if cancer_survivor_at_baseline=0 then do;
    if cancer_SEER_sitel=22030 then do;
        lungcancer=1;
        cancer_dx_date=cancer_diagnosis_date1;
    end;
    if cancer_survivor_at_baseline=1 then lungcancer=.;
end;

* DEFINE FOLLOW-UP TIME FOR THIS ANALYSIS;
* STANDARD CTS ANALYSES CENSOR TIME-TO-EVENT ANALYSES AT THE EARLIEST
OF THE FOLLOWING DATES:
    DATE OF CANCER DIAGNOSIS
    DEATH
    MOVE OUT OF CALIFORNIA (& THEREFORE NO LONGER HAVE CANCER
CAPTURED VIA THE CALIF. CANCER REGISTRY)
    END OF FOLLOW-UP, I.E., 12/31/2015;

*Set cancer diagnosis date to Jan. 1, 2030 for participants who
do not have lung cancer;
if lungcancer=0 then cancer_dx_date='01jan2030'd;

*Use the 'minimum' function to define new variable 'enddate' as
the earlier of diagnosis or censoring dates;
end_date=min(cancer_dx_date,analysis_censoring_date);
format end_date mmddyy10.;

* TO SUMMARIZE THE START & STOP DATES FOR AN ANALYSIS
baseline_date = Date of CTS baseline
end_date = Date of end of follow-up for this analysis;

```

```

      * CALCULATE FOLLOW-UP TIME AT TIME IN DAYS FROM BASELINE TO END OF
FOLLOW-UP;
      followuptime=(end_date - baseline_date)/365.25;

/* PART 3: RUN UNIVARIATE AND MULTIVARIATE PROCEDURES HERE
*/
/* EXAMPLE BELOW SHOWS HOW TO GENERATE FREQUENCIES OF BMI AT BASELINE
*/

PROC FREQ DATA=TEMPORARY;

      TABLES bmi_at_baseline BaselineBMIWHO / MISSING;

/* PART 4: RUN PROPORTIONAL HAZARDS REGRESSION HERE
*/
/*
      *FIRST OPTION: TEST THE PROPORTIONAL HAZARDS ASSUMPTION USING
SCHOENFELD RESIDUALS;
      PROC PHREG DATA=TEMPORARY
      multipass
          nosummary;
      MODEL followuptime*lungcancer(0)= BaselineBMIWHO race / ties=exact
alpha=0.05 rl;
      output out=outdata ressch=schoen;

      RUN;

      ods graphics on/antialiasmax=102000 loessmaxobs=102000;
      PROC SGPLOT data=outdata;
          loess x=followuptime y=schoen /clm;
      RUN;
*/
      *SECOND OPTION: TEST THE PROPORTIONAL HAZARDS ASSUMPTION USING PROC
LIFETEST;
/*
      PROC LIFETEST DATA=TEMPORARY
          method=km
          plots=survival;
      time followuptime*lungcancer(0);
      RUN;
*/

      *ASSUMING NO VIOLATION OF PH ASSUMPTION, RUN MODEL TO GENERATE HRs;
/*
      PROC PHREG DATA=TEMPORARY
          multipass
          nosummary;
      CLASS BaselineBMIWHO (REF='2') race(REF='1');
      MODEL followuptime*lungcancer(0)= BaselineBMIWHO race / ties=exact
alpha=0.05 rl;
*/
      RUN;

```