```
capture log close
log using 01_Barger_Epidemiology_2017, replace
version 14.2 // Stata version for analyses

*******************************************************************************
*  This is Stata analytic code for analyses reported in                       *
*                                                                             *
*    Barger, SD, Cribbet, MR, & Muldoon, MF (2017).                           *
*     Leukocyte telomere length and cardiovascular risk scores for            *
*     prediction of cardiovascular mortality. Epidemiology,Vol. 28            *
*        epub 2016                                                            *
*                                                                             *
*   Users of this syntax are kindly requested to credit/cite this             *
*   publication.                                                              *
*                                                                             *
*  This syntax presumes the user has downloaded the NHANES 1999-2000          *
*  and 2001-2002 data files, merged within cycles and appended between        *
*  cycles. Required files include demographics, lab values for cholesterol,   *
*  glucose, smoking status and MEC blood pressure measurements.               *
*                                                                             *
*       1999-2000                       2001-2002                             *
*       DEMO.XPT                        DEMO_B.XPT  [demographics]            *
*       DIQ.XPT                         DIQ_B.XPT   [diabetes]                *
*       BPQ.XPT                         BPQ_B.XPT   [blood pressure tx]       *
*       MCQ.XPT                         MCQ_B.XPT   [medical history]         *
*       SMQ.XPT                         SMQ_B.XPT   [smoking]                 *
*       BPX.XPT                         BPX_B.XPT   [blood pressure]          *
*       LAB18.XPT                       L40_B.XPT   [biochemistry/glucose]    *
*       LAB13.XPT                       L13_B.XPT   [cholesterol]             *
*       TELO_A.XPT                      TELO_B.XPT  [telomere length]         *
*                                                                             *
*  Users should also include in their analytic file the mortality follow up   *
*  (through December 31, 2011) for these four survey years. These files are   *
*  available on the CDC web site and are titled:                              *
*                                                                             *
*   NHANES_1999_2000_MORT_2013_PUBLIC.dat                                     *
*   NHANES_2001_2002_MORT_2013_PUBLIC.dat                                     *
*                                                                             *
*   All source variables are in lower case  (via the _renvars_ command)       *
*     To find this package type : net search renvars                          *
*                                                                             *
*******************************************************************************

/*

Created by Steven D. Barger, PhD.
No warranties are expressed or implied.

*/

clear all
macro drop all
* insert your NHANES file name below
use "yourfile.dta" /*INSERT your file name  here*/
*
*net search renvars [install this program to run command line below]
*renvars, lower
codebook sddsrvyr riagendr ridageyr ridreth1 wtmec4yr, compact

/*
This is the output from the above command for comparison
Variable     Obs Unique      Mean  Min       Max  Label
-------------------------------------------------------------------------------
sddsrvyr   21004      2  1.525567    1         2  Data Release Number
riagendr   21004      2  1.513712    1         2  Gender
```

```
ridageyr   21004     86  29.47929    0        85  Age at Screening Adjudicated - Recode
ridreth1   21004      5    2.6743    1         5  Race/Ethnicity - Recode
wtmec4yr   21004  13632  13266.63    0  103831.2  Full Sample 4 Year MEC Exam Weight
-------------------------------------------------------------------------------
*/

* The files below are included with this document
* the user can copy that code and create these .do files
do 01a_Barger_Epidemiology_variable_coding.do // calls the variable scoring file

do 01b_Barger_Epidemiology_risk_scoring.do // scores ASCVD risk

do 01c_Barger_Epidemiology_survival_setup.do // calculates survival time

*declare survey design
***design characteristics wtmec4yr sdmvstra sdmvpsu
svyset sdmvpsu [pweight=wtmec4yr], strata(sdmvstra) vce(linearized) ///
 singleunit(missing)

*declare survival time setup
stset follow_up_time [pweight=wtmec4yr] ///
 if (wtmec4yr!=0) /*excludes those not examined in MEC*/, id(seqn) ///
 failure(mortstat==1) origin(time origin)

*Below delineates the progression from the NHANES sample to the analytic sample

* 1. Examined (in Mobile Exam Center), people aged 40-79 years
 tab mortstat if !missing(wtmec4yr) & wtmec4yr!=0 & ///
  (age >=40 & age<=79) // N=5329

* 2. Examined people 40-79 years old with ASCVD risk scores
 tab male if  !missing(wtmec4yr, ascvd_all) & wtmec4yr!=0 & ///
  (age >=40 & age<=79) // N=4780

* 3. Examined people 40-79 with ASCVD risk scores & telomere data
 tab male if  !missing(wtmec4yr, ascvd_all, telomean) & wtmec4yr!=0 & ///
  (age >=40 & age<=79) // N=4262

* 4. Examined 40-79, ASCVD risk scores, telomere data, & CVD death or alive
tab mortstat if cvd_death==1 & !missing(wtmec4yr, ascvd_all, telomean) ///
 & wtmec4yr!=0 &  (age >=40 & age<=79) // N=3652; 198 events

bysort cvd_death: tab mortstat if  !missing(wtmec4yr, ascvd_all, telomean) ///
 & wtmec4yr!=0 & (age >=40 & age<=79) // shows the 610 non-CVD deaths

tab caused if cvd_death==1 & !missing(wtmec4yr, ascvd_all, telomean) & ///
 wtmec4yr!=0 & (age >=40 & age<=79) // CVD subtypes: 148 CHD; 50 cerebrovascular

*standardized telomere length FOR THOSE 40-79
egen z_ascv_telomean=std(telomean) if ascvd_all<. & cvd_death==1 //  3652 cases

label var z_ascv_telomean "z telomere ascvd subsample only"

* create an indicator variable for the analytic sample
gen ascvdinsample=!missing(z_ascv_telomean, ascvd_all) & cvd_death==1
summ age if ascvdinsample, detail
label var ascvdinsample "analytic sample ascvd telomere CVD mortality"
notes ascvdinsample: generated by:gen ascvdinsample=!missing(z_ascv_telomean, ///
 ascvd_all) & cvd_death==1; for the CVD mortality study of telomere length vs ///
 ASCVD risk
tab ascvdinsample , miss

* 5. healthy baseline: CVD deaths among examined 40-79;
tab mortstat if cvd_death==1 & !missing(wtmec4yr, ascvd_healthy, telomean) ///
```

```
 & wtmec4yr!=0 & (age >=40 & age<=79) // N=3274; 118 events

*descriptives for follow up time; median=10.7 years; IQR 9.8-11.7
gen fu_years=permth_int/12
tabstat  fu_years if ascvdinsample, statistics(q)

*variable capturing baseline CVD:
gen existing_cvd=0 if ascvdinsample==1
replace existing_cvd=1 if (mi ==1 | chd ==1 | chf ==1 | stroke ==1)
bysort mortstat: tab existing if  ascvdinsample==1

*** Focal analyses here ****

 local cvdsubpop "if ascvdinsample==1"
 local strata ", strata(agecohort)"
 local survey "vce(cluster sdmvpsu) strata(agecohort)"
 local varlist "ascvd_all z_ascv_telomean"

/* Analysis #1;
    predictor: standardized leukocyte telomere length [LTL];
    criterion; CVD survival time
    covariates; sex, race/ethnicity
    stratified by: 5-year age cohorts
    */
svy linearized, subpop(`cvdsubpop')   : stcox male black hispanic ///
 othernonHrace z_ascv_telomean  `strata' // LTL Hazard Ratio=0.93

/* Analysis #2;
    predictor: standardized leukocyte telomere length [LTL];
    criterion; CVD survival time
    covariate; ascvd risk
    stratified by: 5-year age cohorts
    */
svy linearized, subpop(`cvdsubpop')   : stcox  ///
  z_ascv_telomean ascvd_all `strata' // LTL Hazard Ratio=0.92

/* Analysis #3;
    predictor: standardized leukocyte telomere length [LTL];
    criterion; CVD survival time
    covariate; ascvd risk-calculated only for initially healthy
    stratified by: 5-year age cohorts
    RESTRICTED TO PARTICIPANTS FREE OF DIAGNOSED CVD AT BASELINE
    */
svy linearized, subpop(`cvdsubpop')   : stcox  ///
  z_ascv_telomean ascvd_healthybsln `strata' // LTL Hazard Ratio=0.92

tab mortstat `cvdsubpop' & !missing(ascvd_healthybsln) // shows 118 CVD events

*** this section produces the descriptives presented in the Table

estimates clear
forvalues i = 0(1)1 {
display "mortality category = `i'"
 svy linearized, subpop(if ascvdinsample==1 & mortstat==`i') : mean age ///
  telomean ascvd_all
mat list e(_N)
}

estimates clear
forvalues i = 0(1)1 {
display "mortality category = `i'"
svy linearized, subpop(if ascvdinsample==1 & mortstat==`i') : proportion male ///
 raceethnic  educcat existing_cvd,  missing
mat list e(_N)
}
```

```
-
195
196 log close
197 exit
198
199
200
201
202
```

```
1
2  *this file is called by 01_Barger_Epidemilogy_2017.do
3
4  note: The four-year sample weights should be used (WTINT4YR, WTMEC4YR) for ///
5  combined analyses of NHANES 1999-2000 & NHANES 2001-2002 data.
6
7  label define binarylbl 0 "0 absent/no" 1 "1 present/yes"
8
9  *Part 1-demographics
10 clonevar age=ridageyr
11 note age: source variable ridageyr demographic file
12 recode age (17/24=1) (25/34=2) (35/44=3) (45/54=4) (55/64=5) (65/74=6) (75/90=7) ///
13  if age>=17 & age<=. /*17 yo INCLUDED*/, gen(agecat)
14 label define agecat  1 "1 17-24" 2 "2 25-34" 3 "3 35-44" 4 "4 45-54" ///
15  5 "5 55-64" 6 "6 65-74" 7 "7 75-90"
16 label values agecat agecat
17 label var agecat "7 level age categories"
18 note agecat: source variable age [from ridageyr] 12,060 valid adult age values
19 bysort agecat: summ ridageyr, detail // this age recode is correct
20
21 clonevar agemonths= ridagemn
22 recode agemonths (9999=.)
23 label var agemonths "age in months"
24 note agemonths: source variable ridagemn demographic file; 409 missing; ///
25  204-215 = 17 years old; >=216 months=18 years old
26 bysort agecat: summ agemonths if age!=17, detail
27 bysort agecat: summ agemonths , detail
28 disp 2752-2133
29 tab agemonths if age==17 //  619 participants aged 17 at baseline interview
30
31 *Create an age variable w 5-year age intervals for stratification in Cox models
32 gen agecohort=recode(age,22,27,32,37,42,47,52,57,62,67,72,77,82,90) // combined 83-87 &
   88-90 groups
33 label var agecohort "5 year age cohort stratification"
34 tab agecohort // this includes all below the age of 17
35
36 tab riagendr, miss
37 clonevar male=riagendr
38 recode male (2=0)
39 label define male 0 "0 female" 1 "1 male"
40 label values male male
41 tab male riagendr, miss
42 notes male: source variable riagendr
43 bysort year: tab male riagendr, miss // these match documentation
44
45 clonevar raceethnic= ridreth1
46 tab raceethnic ridreth1, miss
47 tab raceethnic
48
49 *ethnic group recodes
50 gen white = raceethnic
51 recode white  (1 2 4 5=0) (3=1)
52 label var white "White non-Hisp"
53 label define white 0 "0 non White" 1 "1 White"
54 label values white white
55 tab raceeth white, miss // good
56
57 gen black = raceeth
58 recode black (4=1) (1 2 3 5=0)
59 label var black "Black non-Hisp"
60 label define black 0 "0 not Black" 1 "1 Black"
61 label values black black
62 tab raceeth black, miss
63
64 gen hispanic = raceeth
```

```
gen hispanic = raceeth
recode hispanic (1 2=1) (3 4 5=0)
label var hispanic "hispanic"
label define hisp 0 "0 not Hisp" 1 "1 Hispanic", modify
label values hispanic hisp
tab raceeth hispanic, miss

gen othernonHrace = raceeth
recode othernonHrace  (5=1) (1/4=0)
label var othernonHrace "other non hispanic race"
label define other 0 "0 not Other" 1 "1 Other race non Hisp"
label values othernonHrace other
tab raceeth othernonHrace, miss

*education
gen educcat=dmdeduc2
label var educcat "education level"
note educcat: source variable dmdeduc2; 20 and older eligible for this question
recode educcat (7/9=.)
tab educcat dmdeduc2 , miss
label define educ  1 "1 <9th grade" 2 "2 9-11 no diploma" 3 "3 HS dipl or GED" ///
 4 "4 some college" 5 "5 >=college grad"
label values educcat educ

*Part 2
****LAB AND EXAM VARIABLES*******

***BLOOD PRESSURE*** These check out
*BP has up to 4 measurements at the exam
*rename source variables first

/*variable names will be
sbpavg-average systolic BP
dbpavg-average diasystolic BP
*/
local sbp "bpxsy1 bpxsy2 bpxsy3 bpxsy4"
foreach var of varlist `sbp' {
 gen sbp`var'=`var'
 }

 local dbp "bpxdi1 bpxdi2 bpxdi3 bpxdi4"
foreach var of varlist `dbp' {
 gen dbp`var'=`var'
 }

*now rename them into something sensible-sbp first
local counter  = 0
foreach var of varlist sbpbpxsy1 sbpbpxsy2 sbpbpxsy3 sbpbpxsy4 {
 local ++counter
 display "==> varname is: >`var'<"
 rename `var' sbp`counter'
 label var sbp`counter' "systolic BP `counter'"
}
*diastolic now
local bp "dbp"
local counter  = 0
foreach var of varlist `bp'bpxdi1 `bp'bpxdi2 `bp'bpxdi3 `bp'bpxdi4 {
 local ++counter
 display "==> varname is: >`var'<"
 rename `var' dbp`counter'
 label var dbp`counter' "diastolic BP `counter'"
}
note: BP was from exam sessions

egen sbpnomiss=rownonmiss(sbp1 sbp2 sbp3 sbp4)
```

```
tab sbpnomiss
egen sbpavg=rowmean(sbp1 sbp2 sbp3 sbp4)
summ sbpavg, detail
label var sbpavg "SBP average"
egen dbpnomiss=rownonmiss(dbp1 dbp2 dbp3 dbp4)
tab dbpnomiss
egen dbpavg=rowmean(dbp1 dbp2 dbp3 dbp4)
summ dbpavg, detail
label var dbpavg "DBP average"
inspect sbpavg dbpavg

***LIPIDS*** (Lab files)
*total cholesterol= lbxtc:   hdl cholesterol= lbdhdl

clonevar totchol=lbxtc
label var totchol "total cholesterol (mg/dL)"
bysort year: inspect totchol
note totchol: source variable lbxtc ///
 from documentation: In general, for most analyses, the appropriate ///
 [cholesterol] variable to use is LBXTC
summ totchol, detail

clonevar hdl=lbdhdl
label var hdl "hdl cholesterol (mg/dL)"
note hdl: source variable lbdhdl; PM exams or nonfasting were not calculated
summ hdl, detail
bysort year: inspect hdl

clonevar ldl=lbdldl
label var ldl "ldl cholesterol (mg/dL)"
note ldl: source variable lbdldl; AM fasting exams ONLY-different weights too
summ ldl, detail
bysort year: inspect ldl


******OPERATIONAL DEFINITION OF DIABETES******
/* from Pencina et al. Circulation 2009;119;3078-3084 :
"Diabetes mellitus was defined as fasting glucose >=126 mg/dL or use of insulin
 or oral hypoglycemic medications." p. 3079*/
*taking insulin;   _diq050_ 1=yes

**GLUCOSE***
clonevar glucose=lbxsgl
label var glucose "plasma glucose (mg/dL)"
note glucose: source variable lbxsgl

gen havediabetes = 0
recode havediabetes (0=1) if glucose >= 126 & !missing(glucose)
recode havediabetes (0=1) if diq050 == 1  // captures those taking insulin
label var havediabetes "diabetic status"
label values havediabetes binarylbl
tab havediabetes, miss
note havediabetes: glucose >= 126 OR taking insulin diq050 == 1 ///
 [Pencina et al Circulation 2009;119;3078-3084]

*Part 3-reported risk factor status, prior diagnosed chronic disease

gen bpmeds=bpq040a
label var bpmeds "currently taking hytns meds?"
recode bpmeds (7 9=.) (2=0)
label values bpmeds binarylbl
bysort year: tab bpmeds, miss // 1225 & 1339
note bpmeds: source variable bpq040a

gen hibpmd= bpq020
```

```
label var hibpmd "MD told you hypertension?"
recode hibpmd (7 /9=.) (2=0)
note hibpmd: source variable bpq020
label values hibpmd binarylbl
tab hibpmd bpq020, miss
bysort year:inspect hibpmd

*blood pressure
local dbp "bpxdar"
local sbp "bpxsar"
gen hyptens=.
replace hyptens = 0 if  ((`dbp' <=89.49999 & `dbp'!=0) | (`sbp'>=139.49999 & `sbp'!=0))
replace hyptens = 1 if  ((`dbp' >=89.5 & `dbp'<.) | (`sbp'>=139.50000 & `sbp'<.))
label var hyptens "binary hypertension; bpx_ar source"
note hyptens: binary < 140sbp or 90 dbp; 1= >= 140 or >=90 dbp; source ///
 variables bpxdar bpxsar
local dbp "bpxdar"
local sbp "bpxsar"
bysort hyptens: summ `sbp' `dbp', detail // looks good

gen unlabeled = .
replace unlabeled = 0 if (hyptens == 1 & hibpmd == 1)
replace unlabeled = 1 if (hyptens == 1 & hibpmd == 0)
label define unlabeled 0 "0 labeled hypertensive" 1 "1 UNlabeled hypertensive"
label values unlabeled unlabeled
bysort unlabeled: summ  bpxsar bpxdar, detail
tab unlabeled hibpmd, miss
tab hyptens hibpmd, miss // looks good

*prior chronic disease
*heart attack: mcq160e

local missval "(7 9=.) (2=0)"
gen mi= mcq160e
label var mi "md told you had heart attack?"
recode mi `missval'
label values mi binarylbl
bysort year: tab mi, miss // 221 and 258 yes
note mi: source variable mcq160e

*stroke
local missval "(7 9=.) (2=0)"
gen stroke= mcq160f
label var stroke "md told you had stroke?"
recode stroke (7 9=.) (2=0)
label values stroke binarylbl
bysort year: tab stroke mcq160f, miss
note stroke: source var mcq160f

*prior CHF
gen chf= mcq160b
label var chf "md told you had chf?"
recode chf (7 9=.) (2=0)
label values chf binarylbl
note chf: congestive heart failure; source variable mcq160b
bysort year: tab chf mcq160b, miss

*any chd
gen chd= mcq160c
label var chd "md told you had CHD?"
recode chd (7 9=.) (2=0)
label values chd binarylbl
note chd: congestive heart failure; source variable mcq160c
bysort year: tab chd mcq160c, miss // 203 & 240
```

```
258
259 *PART 4-smoking
260
261 *smoking smq020; 20 years and older eligible for the question
262 clonevar eversmoke=smq020
263 recode eversmok (7 9=.) (2=0)
264 label var eversmok "ever smoke 100 cigs?"
265 label values eversmok binarylbl
266 tab eversmok smq020, miss
267 note eversmoke: sourcevar smq020
268
269 clonevar currsmoker=smq040 // categories: every day, some days, not at all
270 recode currsmoker (7 9=.) (2=1) (3=0)
271 label var currsmoker "current smoker?"
272 label values currsmoker binarylbl
273 bysort year: tab currsmoker smq040, miss
274 note currsmoker: sourcevar smq040: categories: every day, some days, not at all: ///
275  recode currsmoker (7 9=.) (2=1) (3=0)
276 *these smoking variables match documentation.
277
278 gen neversmoker=1 if eversmok==0
279 gen formersmoker=1 if eversmok==1 & (currsmoker==0 | smq040==9) // this captures
280 *the 1 "former" smoker who was missing response [9] to "do you smoke now?"
281
282 tab1 neversmoker formersmoker currsmoker, miss
283
284 *one more smoking variable-a single one w 3 levels
285 clonevar smokestatus=smq020
286 recode smokestatus (7 9=.) (2=0) // codes "no" to zero
287 replace smokestatus =2 if (smq040==1 | smq040==2)
288 tab smokestatus, miss
289 foreach variable of varlist neversmoker formersmoker currsmoker {
290  tab smokestatus `variable', miss
291 }
292 label var smokestatus "3 level smok status"
293 label define smokestatus 0 "0 never" 1 "1 former" 2 "2 current"
294 label values smokestatus smokestatus
295 tab smokestatus
296
297 *binary smoking
298 clonevar smokbin=smokestatus
299 recode smokbin (2=1) (1=0)
300 label variable smokbin "binary smoking"
301 label define smokbin 0 "0 never/former" 1 "1 current"
302 label values smokbin smokbin
303 tab smokbin smokestatus, miss
304
305
306
307
308
309
310
```

```
*this file is called by 01_Barger_Epidemiology_2017.do

*01b_Barger_Epidemiology_risk_scoring.do

*the purpose of this .do file is to score 10 year CVD risk for men and
*women-black versus white (hispanic, other)

*  ************************************************************************
*  This is Stata analytic code for analyses reported in                    *
*                                                                          *
*    Barger, SD, Cribbet, MR, & Muldoon, MF (2017).                        *
*     Leukocyte telomere length and cardiovascular risk scores for         *
*     prediction of cardiovascular mortality. Epidemiology,Vol 28          *
*        epub 2016                                                         *
*                                                                          *
*   Users of this syntax are kindly requested to credit/cite this          *
*   publication.
*   CREATED BY STEVEN D. BARGER PHD                                        *
*  ************************************************************************

*American College of Cardiology/American Heart Association CVD risk score coding

*****  ACC/AHA Scoring *****
/* source document
(Circulation. 2014;129(suppl 2):S49-S73)
2013 ACC/AHA Guideline on the Assessment of
Cardiovascular Risk: A Report of the American College of Cardiology/American
Heart Association Task Force on Practice Guidelines
The algorithms are in the supplement-Tables A & B

"These documents also describe the process for the development of novel,
comprehensive multivariable risk equations for the prediction of 10-year
risk of development of ASCVD in non-Hispanic African-American and non-Hispanic
white men and women from 40 to 79 years of age. These equations were developed
from several long-standing population-based cohort studies funded by the NHLBI.
Ten-year risk was defined as the risk of developing a first ASCVD event, defined
as non- fatal myocardial infarction or coronary heart disease (CHD) death or
fatal or nonfatal stroke, over a 10-year period among people free from ASCVD
at the beginning of the period." p. S53

"Because the estimated probabilities can become unstable when approaching the
limits of the sample data, the risk probabilities are truncated at 1% and 30%."
p. S56
*/

/*  SOURCE VARIABLES
GENDER                  male (0=female/1=male)
AGE                     age
RACE                    white black hispanic othernonHrace
                        - code "other" as (hispanic==1 | othernonHrace  ==1)
TOTAL CHOLESTEROL       totchol
HDL CHOLESTEROL         hdl
SYSTOLIC BP             sbpavg
HYPERTENSION TREATMENT  bpmeds (0=no/1=yes)
DIABETES                havediabetes (0=no/1=yes)
SMOKING                 smokbin (0=never/former; 1=current)

** CREATE A  SYSTOLIC BP MEASURE THAT REFLECTS TREATED OR UNTREATED STATUS
*/

gen tx_sbp=sbpavg if bpmeds==1
label var tx_sbp "systolic BP-taking meds"
notes tx_sbp: treated systolic BP for ACC/AHA calculations: ///
 gen tx_sbp=sbpavg if bpmeds==1
```

```
tab bpmeds , miss
bysort bpmeds: summ tx_sbp, detail
gen untx_sbp=sbpavg if (bpmeds==0 | hibpmd==0) // grabs those who have no MD label
bysort bpmeds: summ untx_sbp, detail
label var untx_sbp "systolic BP-NOT taking meds"
notes untx_sbp: UNtreated systolic BP for ACC/AHA calculations: ///
 gen untx_sbp=sbpavg if bpmeds==0

*Prediction formulas are conditional on BP treatment.
*Thus, TWO separate formulas for each race/sex group to account for this.

*1st create log values of age, total cholesterol, HDL, SBP (have to distinguish
* between treated and untreated bp)

foreach var of varlist age totchol hdl untx_sbp tx_sbp  {
 gen ln_`var'=ln(`var')
 }

*additional components
*age squared
gen ln_age_squared=(ln_age)^2
label var ln_age_squared "log age squared"
note ln_age_squared: used only for White females (which includes Hispanic or ///
 other race)

*age by total cholesterol
gen ln_ageXln_totchol= (ln_age*ln_totchol)
label var ln_ageXln_totchol "interaction log age log total cholesterol"
note ln_ageXln_totchol: log age times log total cholesterol; ///
 White men & women only (includes Hispanic or other race)

*age by HDL cholesterol
gen ln_ageXln_hdl= (ln_age*ln_hdl)
label var ln_ageXln_hdl "interaction term log age log HDL"
note ln_ageXln_hdl: log age times log HDL cholesterol; used for all but Black men

*log age X log treated bp; Coefficient = -6.432; Black women only
gen ln_ageXln_tx_sbp= (ln_age*ln_tx_sbp)
label var ln_ageXln_tx_sbp "log age X log treated BP"
notes ln_ageXln_tx_sbp: log age X log treated BP; Black women only

gen ln_ageXln_untx_sbp= (ln_age*ln_untx_sbp)
label var ln_ageXln_untx_sbp "log age X log UNtreated BP"
notes ln_ageXln_untx_sbp: log age X log UNtreated BP; Black women only

gen ln_ageXsmokbin= (ln_age*smokbin)
label var ln_ageXsmokbin "log age X log smoking"
notes ln_ageXsmokbin: log age X log smoking; White women & White men; ///
 no Black coefficients

tab smokbin

* have to sum the coefficients and then subtract the subgroup mean (a constant)
* listed in the Table
* white women=-29.18; black women=86.61; white men=61.18; black men=19.54
* that difference is the exponent for the risk estimate

*white women-UNTREATED BP
gen white_other_women_sum= . if male ==0

local whiteotherwomenuntreated " male==0 & (raceeth ==1 | raceeth ==2 | raceeth ==3 |
raceeth ==5) & (age >=40 & age <=79) & !missing(untx_sbp)" // all but Black

** SECTION 1A WHITE WOMEN-UNTREATED BP
```

```
** SECTION 1B WHITE WOMEN-TREATED BP

** 1A: WHITE WOMEN-UNTREATED BP
 replace white_other_women_sum=  ///
 ln_age*(-29.799) + ///
 ln_age_squared*(4.884)+ ///
 ln_totchol*(13.540)+ ///
 ln_ageXln_totchol*(-3.114)+ ///
 ln_hdl*(-13.578)+ ///
  /* ln_tx_sbp*(2.019)+ ///  this is treated */ ///
 ln_ageXln_hdl*(3.149)+ ///
 /* black women only: ln_ageXln_txsbp*(-6.432) */ ///
 ln_untx_sbp*(1.957)+ ///
 /* black women only: ln_ageXln_untx_sbp*(-6.087) */ ///
 smokbin*(7.574)+ ///
 ln_ageXsmokbin*(-1.665)+ /* white only*/ ///
 havediabetes*(0.661) if `whiteotherwomenuntreated'

local whitecoefvalue "-29.18"
gen white_other_women_diff=white_other_women_sum-`whitecoefvalue'
label var white_other_women_diff "difference for exponent"
gen ten_yr_white_woman_ASCVD_risk=1- (0.9665)^(exp(white_other_women_diff))

**  1B: WHITE WOMEN-TREATED BP
local whiteotherwomen_treated " male==0 & (raceeth ==1 | raceeth ==2 | raceeth ==3 |
raceeth ==5) & (age >=40 & age <=79) & !missing(tx_sbp)" // all but Black
replace white_other_women_sum=  ///
 ln_age*(-29.799) + ///
 ln_age_squared*(4.884)+ ///
 ln_totchol*(13.540)+ ///
 ln_ageXln_totchol*(-3.114)+ ///
 ln_hdl*(-13.578)+ ///
  /* ln_tx_sbp*(2.019)+ ///  this is treated */ ///
 ln_ageXln_hdl*(3.149)+ ///
 /* black women only: ln_ageXln_txsbp*(-6.432) */ ///
 ln_tx_sbp*(2.019)+  ///
 /* comment out untreated: ln_untx_sbp*(1.957)+ */  ///
 /* black women only: ln_ageXln_untx_sbp*(-6.087) */ ///
 smokbin*(7.574)+ ///
 ln_ageXsmokbin*(-1.665)+ /* white only*/ ///
 havediabetes*(0.661) if `whiteotherwomen_treated'

replace white_other_women_diff=white_other_women_sum-`whitecoefvalue' ///
 if `whiteotherwomen_treated'
*baseline survival rate; white women=0.9665; black women=0.9533;
* white men=0.9144; black men=0.8954
replace ten_yr_white_woman_ASCVD_risk=1- (0.9665)^(exp(white_other_women_diff)) ///
 if `whiteotherwomen_treated'

*Black women
** SECTION 2A BLACK WOMEN-UNTREATED BP
** SECTION 2B BLACK WOMEN-TREATED BP

* 2A: BLACK WOMEN, UNTREATED BP  [exclude treated BP term(s)]
local blackwomenuntreated " male==0 & (raceeth ==4) & (age >=40 & age <=79) &
!missing(untx_sbp)" // Black only

gen black_women_sum= . if male ==0
label var black_women_sum "individual coeff sum-black women"
replace black_women_sum=  ///
 ln_age*(17.114) + ///
 /* doesn't use this term; ln_age_squared*(4.884)+ */ ///
 ln_totchol*(0.940)+ ///
 /* doesn't use this term; ln_ageXln_totchol*(-3.114)+ */  ///
 ln hdl*(-18.920)+ ///
```

```
 ln_ageXln_hdl*(4.475)+ ///
   /* ln_tx_sbp*(2.019)+ ///  this is treated */ ///
 /* black women only: ln_ageXln_txsbp*(-6.432) */ ///
 ln_untx_sbp*(27.820)+ ///
 /* black women only:*/ ln_ageXln_untx_sbp*(-6.087) +  ///
 smokbin*(0.691)+ ///
 /* black women formula doesn't use this term; ln_ageXsmokbin*(-1.665)+ */  ///
 havediabetes*(0.874) if `blackwomenuntreated'

local blackcoefvalue "86.61"
gen black_women_diff=black_women_sum-`blackcoefvalue'
label var black_women_diff "difference for exponent-black women"

gen ten_yr_black_woman_ASCVD_risk=1- (0.9533)^(exp(black_women_diff))

** 2B: BLACK WOMEN-TREATED BP

local blackwomen_treated " male==0 & (raceeth ==4) & (age >=40 & age <=79) &
!missing(tx_sbp)" // Black only
/*gen black_women_sum= . if male ==0
label var black_women_sum "individual coeff sum-black women"
*/
replace black_women_sum=  ///
 ln_age*(17.114) + ///
 /* doesn't use this term; ln_age_squared*(4.884)+ */ ///
 ln_totchol*(0.940)+ ///
 /* doesn't use this term; ln_ageXln_totchol*(-3.114)+ */  ///
 ln_hdl*(-18.920)+ ///
 ln_ageXln_hdl*(4.475)+ ///
    ln_tx_sbp*(29.291)+ ///  this is treated */ ///
 /* black women only: */ ln_ageXln_tx_sbp*(-6.432) + ///
 /* omit-treated only here: ln_untx_sbp*(27.820)+ */  ///
 /* black women only: omit; treated only here ln_ageXln_untx_sbp*(-6.087) + */ ///
 smokbin*(0.691)+ ///
 /* black women formula doesn't use this term; ln_ageXsmokbin*(-1.665)+ */  ///
 havediabetes*(0.874) if `blackwomen_treated'

local blackcoefvalue "86.61"
replace black_women_diff=black_women_sum-`blackcoefvalue' if `blackwomen_treated'
replace ten_yr_black_woman_ASCVD_risk=1- (0.9533)^(exp(black_women_diff)) ///
 if `blackwomen_treated'

** SECTION 3A WHITE MEN-UNTREATED BP
** SECTION 3B WHITE MEN-TREATED BP

** 3A: WHITE MEN-UNTREATED BP

gen white_other_men_sum= . if male ==1

local whiteothermenuntreated " male==1 & (raceeth ==1 | raceeth ==2 | raceeth ==3 | raceeth
==5) & (age >=40 & age <=79) & !missing(untx_sbp)" // all but Black

replace white_other_men_sum=  ///
 ln_age*(12.3444) + ///
 /* not used for men ln_age_squared*(4.884)+ */ ///
 ln_totchol*(11.853)+ ///
 ln_ageXln_totchol*(-2.664)+ ///
 ln_hdl*(-7.990)+ ///
 ln_ageXln_hdl*(1.769)+ ///
  /* omit treated: ln_tx_sbp*(1.797)+ */ ///
  ln_untx_sbp*(1.764)+ /* this is untreated */ ///
 smokbin*(7.837)+ ///
 ln_ageXsmokbin*(-1.795)+ /* white only*/ ///
 havediabetes*(0.658) if `whiteothermenuntreated'
```

```
254
255 local whitecoefvalue "61.18"
256 gen white_other_men_diff=white_other_men_sum-`whitecoefvalue'
257 label var white_other_men_diff "difference for exponent"
258 gen ten_yr_white_man_ASCVD_risk=1- (0.9144)^(exp(white_other_men_diff))
259
260 **  3B: WHITE MEN-TREATED BP
261 local whiteothermen_treated " male==1 & (raceeth ==1 | raceeth ==2 | raceeth ==3 | raceeth
    ==5) & (age >=40 & age <=79) & !missing(tx_sbp)" // all but Black
262
263 replace white_other_men_sum=  ///
264  ln_age*(12.3444) + ///
265  /* not used for men ln_age_squared*(4.884)+ */ ///
266  ln_totchol*(11.853)+ ///
267  ln_ageXln_totchol*(-2.664)+ ///
268  ln_hdl*(-7.990)+ ///
269  ln_ageXln_hdl*(1.769)+ ///
270   /* include treated:*/ ln_tx_sbp*(1.797)+  ///
271  /* ln_untx_sbp*(1.764)+  exclude untreated */ ///
272  smokbin*(7.837)+ ///
273  ln_ageXsmokbin*(-1.795)+ /* white only*/ ///
274  havediabetes*(0.658) if `whiteothermen_treated'
275
276 local whitecoefvalue "61.18"
277 replace white_other_men_diff=white_other_men_sum-`whitecoefvalue' if ///
278  `whiteothermen_treated'
279 replace ten_yr_white_man_ASCVD_risk=1- (0.9144)^(exp(white_other_men_diff)) ///
280  if `whiteothermen_treated'
281
282 ** SECTION 4A BLACK MEN-UNTREATED BP
283 ** SECTION 4B BLACK MEN-TREATED BP
284
285 **  4A: BLACK MEN-UNTREATED BP
286 gen black_men_sum= . if male ==1
287 local blackmenuntreated "male==1 & (raceeth ==4) & (age >=40 & age <=79) &
    !missing(untx_sbp)"
288
289 replace black_men_sum=  ///
290  ln_age*(2.469) + ///
291  /* not used for men ln_age_squared*(4.884)+ */ ///
292  ln_totchol*(0.302)+ ///
293  /* not used black men: ln_ageXln_totchol*(-2.664)+  */ ///
294  ln_hdl*(-0.307)+ ///
295  /* not used black men: ln_ageXln_hdl*(1.769)+ */ ///
296   /* omit treated: ln_tx_sbp*(1.916)+ */ ///
297   ln_untx_sbp*(1.809)+ /* this is untreated */ ///
298  smokbin*(0.549)+ ///
299  /* not used black men: ln_ageXsmokbin*(-1.795)+ */  ///
300  havediabetes*(0.645) if `blackmenuntreated'
301
302 local blackcoefvalue "19.54"
303 gen black_men_diff=black_men_sum-`blackcoefvalue'
304 label var black_men_diff "difference for exponent"
305 gen ten_yr_black_man_ASCVD_risk=1- (0.8954)^(exp(black_men_diff))
306
307 **  4B: BLACK MEN-TREATED BP
308
309 local blackmen_treated "male==1 & (raceeth ==4) & (age >=40 & age <=79) & !missing(tx_sbp)"
    // only Black
310
311 replace black_men_sum=  ///
312  ln_age*(2.469) + ///
313  /* not used for men ln_age_squared*(4.884)+ */ ///
314  ln_totchol*(0.302)+ ///
315  /* not used black men: ln_ageXln_totchol*(-2.664)+  */ ///
```

```
 ln_hdl*(-0.307)+ ///
 /* not used black men: ln_ageXln_hdl*(1.769)+ */ ///
  /* include treated: */ ln_tx_sbp*(1.916)+  ///
  /* exclude untreated: ln_untx_sbp*(1.809)+  */ ///
 smokbin*(0.549)+ ///
 /* not used black men: ln_ageXsmokbin*(-1.795)+ */  ///
 havediabetes*(0.645) if `blackmen_treated'

local blackcoefvalue "19.54"
replace black_men_diff=black_men_sum-`blackcoefvalue' if `blackmen_treated'

replace ten_yr_black_man_ASCVD_risk=1- (0.8954)^(exp(black_men_diff)) if ///
 `blackmen_treated'

*This is a new bit-create a  risk score in a single variable
inspect ten_yr_black_man_ASCVD_risk
gen acc_aha_10yr_cvdrisk_all=.
replace acc_aha_10yr_cvdrisk=ten_yr_black_man_ASCVD_risk if ///
 !missing(ten_yr_black_man_ASCVD_risk)
inspect ten_yr_white_man_ASCVD_risk
replace acc_aha_10yr_cvdrisk=ten_yr_white_man_ASCVD_risk if ///
 !missing(ten_yr_white_man_ASCVD_risk)

inspect ten_yr_black_woman_ASCVD_risk
replace acc_aha_10yr_cvdrisk=ten_yr_black_woman_ASCVD_risk if ///
 !missing(ten_yr_black_woman_ASCVD_risk)

inspect ten_yr_white_woman_ASCVD_risk
replace acc_aha_10yr_cvdrisk=ten_yr_white_woman_ASCVD_risk if ///
 !missing(ten_yr_white_woman_ASCVD_risk)
bysort agecat: summ acc_aha_10yr_cvdrisk, detail
label var  acc_aha_10yr_cvdrisk_all "10 year CVD risk-includes bsline CVD!"
note  acc_aha_10yr_cvdrisk:  This is the ACC/AHA 10 year CVD risk score. Initial ///
 calculations done for ALL participants, including those with extant CVD

* CREATE THE ASCVD SCORE: INCLUDES ONLY THOSE AGED 40-79 WITH NO BASELINE CVD
gen ten_yr_cvd_risk_nobsln_cvd=.
replace ten_yr_cvd_risk_nobsln_cvd=acc_aha_10yr_cvdrisk if ///
 (mi ==0 & stroke==0 & chf==0 & chd==0)
replace ten_yr_cvd_risk_nobsln_cvd=acc_aha_10yr_cvdrisk if ///
 (mi !=1 & stroke!=1 & chf!=1 & chd!=1)

label var ten_yr_cvd_risk_nobsln_cvd "AHA 10yr CVD-no bslin CVD"
note ten_yr_cvd_risk_nobsln_cvd: ACC/AHA 10-year CVD risk; calculated ONLY ///
 for those 40-79 free of baseline CVD

gen ascvd_healthybsln=ten_yr_cvd_risk_nobsln_cvd
gen ascvd_all=acc_aha_10yr_cvdrisk_all

/*
** NOW TRUNCATE THOSE risk scores as per the Goff 2014 Circulation guidelines

"Because the estimated probabilities can become unstable when approaching the
limits of the sample data, the risk probabilities are truncated at 1% and 30%."
p. S56

*/

foreach var of varlist ascvd_healthybsln ascvd_all {
disp "working variable: `var'"
disp "this is the frequency below the lower truncated value"
summ `var' if `var' <.01, detail
display r(N)
recode `var' (0/0.0099999 = 0.01)
disp "this is the frequency after recoding the lower truncated values"
```

```
disp "this is the frequency after recoding the lower truncated values
summ `var' if `var' <.01, detail
disp "this is the frequency including the higher (> 0.30) values"
summ `var' , detail // top range
disp "this displays the higher (> 0.30) values to be truncated"
summ `var' if `var' >.30, detail
disp "this truncates the higher (> 0.30) values to be 0.30"
recode `var' (0.30/1 = 0.30)
disp "this should show no values below 0.01 or above 0.30"
summ `var' , detail
}
note ascvd_healthybsln: 10 year ASCVD risk; baseline healthy only; 40-79  ///
 years old; TRUNCATED at .01 and 0.30 as per Goff 2014 Circulation   ///
 scoring guidelines (p. S56); INCLUDES people with SBP>200mm/Hg; this results ///
 in slightly higher risk estimates for these 24 people; also includes risk ///
 estimate for 82 people with LDL>190 mg/dL

*** THIS COMPLETES THE ACC/AHA CODING FOR MEN AND WOMEN

```

```
*this file is called by 01_Barger_Epidemilogy_2017.do

*from 01c_Barger_Epidemiology_survival_setup.do

codebook eligstat mortstat mortsrce_ndi mortsrce_ssa permth_int ///
 permth_exm causeavl ucod, compact
/*

    ****This is the output from the above command***
Variable        Obs Unique      Mean  Min  Max  Label
-------------------------------------------------------------------------------------------------
-----------------
eligstat      11432      1         1    1    1  eligibility status for mortality follow-up
mortstat      11432      2  .1787964    0    1  final mortality status
mortsrce_ndi   2030      1         1    1    1  mortality source: NDI match
mortsrce_ssa   1112      1         1    1    1  mortality source: SSA information
permth_int    11432    154  119.3984    0  153
permth_exm    10559    153  120.0539    0  152
causeavl       2044      2  .9980431    0    1  cause of death data available
ucod_leading   2040     10  5.355392    1   10  underlying cause of death recode from
UCOD_113 leading causes
-------------------------------------------------------------------------------------------------
-----------------
*/

*Survival set up for NHANES 1999-2002 merged data
clonevar causedeath = ucod_ // grabs the global ICD-10 categories

gen cvd_death=0
replace cvd_death = 1 if causedeath == 1 | causedeath == 5 | mortstat==0  // All CVD codes;
label var cvd_death "binary CVD mortality code"
label define cvd_death 0 "0 non-CVD mortality" 1 "1 CVD mortality or alive"
label values cvd_death cvd_death
note cvd_death: this codes only those alive at follow up (N=9388) or who died ///
 from cardiovascular causes (N=509); total N=9897
tab cvd_death mortstat, miss
tab mortstat if cvd_death

note: 2044 deaths; 11432 eligible for NDI follow up

*** The NHANES has 2 variables with time to event: they are person/months
* since interview; person/months since MEC exam
*    permth_int  permth_exm

/**
final NDI ascertainment was  was December 31, 2011; interviews/exams were
conducted  between 1999 and 2002
*/
*Quantify the difference between MEC and home exams in months
gen fudiff =  permth_exm -  permth_int if agecat !=.
label var fudiff "months btw MEC & home assessments"
tab fudiff
tab fudiff if agecat !=. , miss
note fudiff: subtracts person-months from the interview from person-months ///
 from the exam; most values negative (interview first)
*Since the exam tended to follow the interview the best follow up time is post-MEC
*Use the follow up time from the home exam to create time variable

* KEY AGE VARIABLES:  age in months at MEC exam; ridageex:
** age in months at home exam; ridagemn[coded as agemonths]

summ ridageex if agecat!=., detail // N=10868 valid adults for the MEC exam
summ ridagemn if agecat!=., detail // N=11651 valid adults for the home interview
disp 10868/11651 // 93% of adults have examination data
```

```
* I am substituting _agemonths_ for _ridagemn_ (which had 409 missing values)

/* ***********
Below generates a time variable for survival analysis. Person-months
 of follow up is coded in the mortality data files. However, time at risk
 is age-based rather than time on study based. Thus, will add age in months
 to the number of follow up months to create the time variable.

 Age is top coded at 85 years those people will be assigned to be 85 years old
***********
*/

* drop the ineligible people (< 18 years old).
drop if mortstat==. | agecat==. //  drops those without mortality follow up (n=9572)
* and those 17 years at baseline interview
* those <18 at baseline were ineligible for NDI follow up:
gen follow_up_time= agemonths+permth_int if agecat!=.  // 409 missing values for
 * age (in months) at interview

note follow_up_time: sum of age in months @ interview plus months of follow-up; ///
  this is the time at risk variable (age until death or censoring)
tab follow_up_time
gen origin= agemonths-agemonths if agecat!=. // puts time at zero for everyone

*this bit evaluates those 409 observations missing _permth_int_ values
tab permth_exm if agemonths==., miss // these are the 409 cases: 280 have
 *follow-up time from the Mobile Exam
list permth_exm mortstat if agemonths==.
tab age if agemonths==. // all people with missing age in months are the
* top-censored people ("85" years old)
tab permth_int if agemonths==. // all 409 people have person-months of follow
 * up from the interview
tab permth_exm if agemonths==.
*age in months at exam- ridageex

* replace that time variable using the interview person-months variable
*the 409 will still be omitted from survival analyses because they have
* an invalid entry time
* I will fix that by substituting the origin with (age-age).
* This will underestimate the oldest participants' baseline age

replace origin = (age-age) if origin==. // recodes those 409

*gives 85 year olds 1,020 months of age (85*12)
replace agemonths = (1020) if agemonths==. & age==85

replace follow_up_time= (agemonths + permth_int) if follow_up_time==. // fixed 409 of 409
tab follow_up_time, miss
*this includes all 11,432 adults eligible for NDI follow up

**THIS IS THE CODING FOR THE AGE-SCALE MODEL-
*WILL GENERATE A SEPARATE FILE WITH THIS TIME SCALE***
stset follow_up_time [pw=wtint4yr /*interview weight*/], failure(mortstat==1) ///
 origin(time origin)
*11,432 observations; 2044 events
```