```
/*

PHD PROJECT: The role of depressive symptoms and cardiometabolic risk factors in the prediction of dementia: a cross-country comparison in England, the United States and China

STUDY 4: Clustering of cardiometabolic risk factors and dementia incidence

Method of analysis:
Latent Class Analysis (LCA)

DATASET: ELSA
baseline: wave 2 (2004) follow-up waves 3-9 (2006-2018)


TIMELINE

LATENT CLASSES OF CARDIOMETABOLIC MARKERS: WV2 (BASELINE)
DEMENTIA INCIDENCE: W3 - WV9 (7 TIME POINTS)
COVARIATES ADJUSTMENT FOR HR MODELS: WV2


*/




* KEEP NECESSARY VARIABLES

keep idauniq w2wtnur w2wtbld ///
E_sex E_age E_eduaction_yrs E_eduaction E_maritalstatus_3cat E_maritalstatus_4cat ///
E_wealthquintiles E_smoking_3cat E_physicalactivity E_alcohol_freq E_alcohol_status ///
E_cvd_comorbidity E_cognitive_index E_memory_wordrecall Ewv2_loneliness_quintiles ///
Ewv2_cesd_score Ewv2_depressive_symptoms ///
Ewv3_cesd_sumscore_rand Ewv3_depressive_symptoms ///
Ewv4_cesd_sumscore_rand Ewv4_depressive_symptoms ///
Ewv5_cesd_sumscore_rand Ewv5_depressive_symptoms ///
Ewv6_cesd_sumscore_rand Ewv6_depressive_symptoms ///
Ewv7_cesd_sumscore_rand Ewv7_depressive_symptoms ///
Ewv8_cesd_sumscore Ewv8_depressive_symptoms ///
Ewv9_cesd_sumscore Ewv9_depressive_symptoms ///
Ewv2_crp_level Ewv2_crp Ewv2_fibrinogen_level Ewv2_fibrinogen ///
Ewv2_hdl_level Ewv2_male_hdl Ewv2_female_hdl ///
Ewv2_meds_hdl Ewv2_cholesterol_evr Ewv2_hdl_sum Ewv2_hdl_cholesterol ///
Ewv2_waist Ewv2_malewaist_ao Ewv2_femalewaist_ao Ewv2_obesity_waist_sum Ewv2_obesity_waist ///
Ewv2_bmi_score Ewv2_obesity_bmi Ewv2_waist_bmi_sum Ewv2_obesity ///
Ewv2_tg_level Ewv2_tg ///
Ewv2_systolic_mean Ewv2_diastolic_mean Ewv2_systolic_bp Ewv2_diastolic_bp ///
Ewv2_meds_bp Ewv2_bp_reportevr Ewv2_bp_before Ewv2_bp_diagnosed_sum Ewv2_bp_diagnosed Ewv2_bp_sum Ewv2_bp ///
Ewv2_diabetes_evr Ewv2_diabetes_before Ewv2_diabetes_diagnosed_sum Ewv2_diabetes_diagnosed ///
Ewv2_glucose_level Ewv2_glucose Ewv2_HbA1c_level Ewv2_HbA1c ///
Ewv2_meds1_diabetes Ewv2_meds2_diabetes Ewv2_insulin_diabetes Ewv2_diabetes_anymeds_sum Ewv2_diabetes_anymeds ///
Ewv2_diabetes_glucose_sum Ewv2_glycemia ///
Ewv2_anydementia_iqcode_report Ewv3_anydementia_iqcode_report ///
Ewv4_anydementia_iqcode_report Ewv6to9_dementia_event ///
Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report ///
Ewv2_interview_date Ewv3_interview_date Ewv4_interview_date ///
Ewv5_interview_date Ewv6_interview_date Ewv7_interview_date ///
Ewv8_interview_date Ewv9_interview_date ///
Ewv3to9_dementia_sum Ewv3to9_dementia_sum_no_iqcode ///
Ewv3to9_dementia_event Ewv3to9_dementia_event_no_iqcode ///
Ewv3to9_dementia_report_or_lasti Ewv3to9_dementia_report_free_dat ///
Ewv3to9_newdementia_or_lastinter Ewv3to9_dementia_free_date E_time_dementia_months ///
E_time_dementia_report_months_no E_time_dementia_midpoint ///
E_time_dementia_midpoint_final E_time_event_dementia E_time_dementia_report_midpoint_ ///
E_time_dementia_midpoint_no_iqco E_time_event_dementia_report_no_
```

```
/* Latent class analysis - LCA  of cardiomeatbolic risk factors for dementia

Useful links:
https://www.stata.com/meeting/uk18/slides/uk18_MacDonald.pdf

https://www.stata.com/meeting/mexico18/slides/5_Mexico18_Canette.pdf

https://www.bgsu.edu/content/dam/BGSU/college-of-arts-and-sciences/center-for-family-and-demographic-research/documents/Workshops/2020-latent-class-analysis.pdf

https://www.stata.com/features/overview/latent-class-analysis/

https://www.stata.com/manuals/semexample50g.pdf

https://www.stata.com/manuals/semexample51g.pdf

https://www.stata.com/manuals/semexample52g.pdf

https://www.ucl.ac.uk/population-health-sciences/sites/population_health_sciences/files/lca.pdf

https://www.stata.com/manuals/semgsemlclassoptions.pdf

https://www.stata.com/meeting/nordic-and-baltic17/slides/nordic-and-baltic17_Pitblado.pdf

https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00920/full

https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774074

https://www.statalist.org/forums/forum/general-stata-discussion/general/1412686-calculating-entropy-for-lca-latent-class-analysis-in-stata-15

https://www.statalist.org/forums/forum/general-stata-discussion/general/1590174-how-to-calculate-entropy-for-lca-with-stata

https://www.statalist.org/forums/forum/general-stata-discussion/general/1390895-combine-marginsplot-problem-with-plot-options




 * gsem command to fit a latent class model

gsem (var1 var2 var3 <-), logit lclass(C 3)

OR TRY

gsem (var1 var2 var3 <-), logit lclass(C 3) ///
startvalues(randompr, draws(20) seed(15) difficult) ///
emopts(iterate(30) difficult)




Binary variables of cardiometabolic markers measured at wave 2

CRP: Ewv2_crp

HDL cholesterol: Ewv2_hdl_cholesterol

Obesity by waist cir: Ewv2_obesity_waist

systolic Blood pressure: Ewv2_systolic_bp

```

```
diastolic Blood pressure: Ewv2_diastolic_bp

Diabetes: Ewv2_diabetes_diagnosed

HbA1c: Ewv2_HbA1c


*/






*** Descriptive stats of cardiometabolic markers

tabulate Ewv2_crp
summarize Ewv2_crp

misstable summarize Ewv2_crp
misstable patterns Ewv2_crp

tabulate Ewv2_hdl_cholesterol
summarize Ewv2_hdl_cholesterol

misstable summarize Ewv2_hdl_cholesterol
misstable patterns Ewv2_hdl_cholesterol

tabulate Ewv2_obesity_waist
summarize Ewv2_obesity_waist

misstable summarize Ewv2_obesity_waist
misstable patterns Ewv2_obesity_waist

tabulate Ewv2_systolic_bp
summarize Ewv2_systolic_bp

misstable summarize Ewv2_systolic_bp
misstable patterns Ewv2_systolic_bp


tabulate Ewv2_diastolic_bp
summarize Ewv2_diastolic_bp

misstable summarize Ewv2_diastolic_bp
misstable patterns Ewv2_diastolic_bp


tabulate Ewv2_diabetes_diagnosed
summarize Ewv2_diabetes_diagnosed

misstable summarize Ewv2_diabetes_diagnosed
misstable patterns Ewv2_diabetes_diagnosed


tabulate Ewv2_HbA1c
summarize Ewv2_HbA1c

misstable summarize Ewv2_HbA1c
misstable patterns Ewv2_HbA1c


tabulate Ewv2_anydementia_iqcode_report
summarize Ewv2_anydementia_iqcode_report

misstable summarize Ewv2_anydementia_iqcode_report
misstable patterns Ewv2_anydementia_iqcode_report

```

```
*** CLEANING DATA


* 1. drop dementia cases and missing data at baseline

drop if Ewv2_anydementia_iqcode_report==1
* (50 observations deleted)

drop if Ewv2_anydementia_iqcode_report== .
* (0 observations deleted)


* 2. drop missing values of cardiometabolic markers

drop if Ewv2_crp== .
* (1,753 observations deleted)

drop if Ewv2_hdl_cholesterol== .
* (6 observations deleted)

drop if Ewv2_obesity_waist== .
* (133 observations deleted)

drop if Ewv2_systolic_bp== .
* (660 observations deleted)

drop if Ewv2_diastolic_bp== .
* (0 observations deleted)

drop if Ewv2_diabetes_diagnosed== .
* (0 observations deleted)

drop if Ewv2_HbA1c== .
* (102 observations deleted)



* 3. drop obs with no records on dementia at any wave from 3-9 follow-ups


search mdesc
search rmiss2
search mvpatterns

* see number of missing values vs non-missing in each variable
mdesc Ewv3_anydementia_iqcode_report Ewv4_anydementia_iqcode_report ///
Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report



/* number of missing values per observation
* the code below creates a variable called nmisfollowup that gives the number of missing values
for each observation in the variables of interest */
egen nmisfollowup_dementia_wv3to9=rmiss2(Ewv3_anydementia_iqcode_report ///
Ewv4_anydementia_iqcode_report Ewv5_anydementia_iqcode_report ///
Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report)

tab nmisfollowup_dementia_wv3to9

* drop observations "nmisfollowup_dementia_wv3to9" > 6 (those with 7 missing data = no records at any wave)
drop if nmisfollowup_dementia_wv3to9>6
```

```
*(451 observations deleted)


* FINAL SAMPLE -> 4511






/* Latent Class analysis - gsem

 7 variables: Ewv2_crp Ewv2_hdl_cholesterol Ewv2_obesity_waist
 Ewv2_systolic_bp Ewv2_diastolic_bp Ewv2_diabetes_diagnosed Ewv2_HbA1c

*/


* change names to start with lowercase (STATA assumes variables starting with a capital letter are cont latent variables)

rename Ewv2_crp crp_lca
rename Ewv2_hdl_cholesterol hdl_lca
rename Ewv2_obesity_waist obesity_lca
rename Ewv2_systolic_bp systolic_lca
rename Ewv2_diastolic_bp diastolic_lca
rename Ewv2_diabetes_diagnosed diabetes_lca
rename Ewv2_HbA1c hba1c_lca



* Corrrelation matrix of the CM variables

corr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca

pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig


* to create quality table in word - asdoc -
* https://www.youtube.com/watch?v=XHBl6PHfOzs&ab_channel=StataProfessor

help asdoc

asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig


asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, nonum replace cor crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca label replace star(.05) dec(2)



* LCA models


* one-class model

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 1)

estimates store oneclass_cm

* two-class model

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 2)


estimates store twoclass_cm
```

```
* three-class model

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)


gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
startvalues(randompr, draws(20) seed(15) difficult) ///
emopts(iterate(30) difficult)

estimates store threeclass_cm

* four-class model

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 4)

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 4) ///
startvalues(randompr, draws(20) seed(15) difficult) ///
emopts(iterate(30) difficult)




/*

** Evaluating Fit to choose the number of classes **

1. a priori theory
2. Information Statistics
        AIC, BIC, adjusted BIC
3. Chi-Square goodness of fit
4. Entropy

Others but not used here:
Lo-Mendell-Rubin (LMR)
        Not recommended (designed for normal Y)
Bootstrapped Likelihood Ratio Test

*/

* AIC and BIC to determine which of these models fits best

estimates stats oneclass_cm twoclass_cm threeclass_cm



* LCA postestimation
* Latent class marginal probabilities - lcprob -
* Latent class marginal means - lcmean -


estat lcprob

estat lcmean



* likelihood -ratio test (G2) to evaluate whether our model fits as well as the saturated model

estat lcgof

* if p value is sig means that we reject the null hypothesis and and the model doesn't fit well
* if p value is not sig means that we fail to reject the null hypothesis that our model fits as
```

```
well as the saturated model.
* for a well-fitted model p value should be non-sig. A significant p-value indicates lack of
model fit in absolute terms.

* 3-class model: p < 0.001



* Entropy

quietly predict classpost*, classposteriorpr
gen sum_p_lnp = 0
forvalues k = 1/2 {
        replace sum_p_lnp = sum_p_lnp + classpost`k'*ln(classpost`k')
}
summ sum_p_lnp, meanonly
scalar E = 1+`r(sum)'/(e(N)*ln(2))
drop classpost? sum_p_lnp
di E









/* We can use the predictions of the posterior probability of class membership to evaluate an
individual's probability of being in each class.

*/

predict classpost1*, classposteriorpr
list in 1, abbrev(10)

/* We can determine the expected class for each individual based on whether the posterior
probability
is greater than 0.5
*/

generate expclass1 = 1 + (classpost11>0.5)
tabulate expclass1


generate expclass2 = 1 + (classpost12>0.5)
tabulate expclass2


generate expclass3 = 1 + (classpost13>0.5)
tabulate expclass3



/* We can determine expected classification for each individual in the dataset based on the
predicted
posterior class probabilities.
*/

predict cpost*, classposteriorpr
egen max = rowmax(cpost*)


* generate classes var

generate predclass = 1 if cpost1==max

replace predclass = 2 if cpost2==max

```

```
replace predclass = 3 if cpost3==max

tabulate predclass



**********function to print out summary stats
program summary_table_procLCla
    preserve
    *look at the average posterior probability
    gen Mp = 0
    foreach i of varlist cpost* {
        replace Mp = `i' if `i' > Mp
    }
    sort predclass
    *and the odds of correct classification
    by predclass: gen countG = _N
    by predclass: egen groupAPP = mean(Mp)
    by predclass: gen counter = _n
    gen n = groupAPP/(1 - groupAPP)
    gen p = countG/ _N
    gen d = p/(1-p)
    gen occ = n/d
    *Estimated proportion for each group
    scalar c = 0
    gen TotProb = 0
    foreach i of varlist cpost* {
       scalar c = c + 1
       quietly summarize `i'
       replace TotProb = r(sum)/ _N if predclass == c
    }
    gen d_pp = TotProb/(1 - TotProb)
    gen occ_pp = n/d_pp
    *This displays the group number [_traj_~p],
    *the count per group (based on the max post prob), [countG]
    *the average posterior probability for each group, [groupAPP]
    *the odds of correct classification (based on the max post prob group assignment), [occ]
    *the odds of correct classification (based on the weighted post. prob), [occ_pp]
    *and the observed probability of groups versus the probability [p]
    *based on the posterior probabilities [TotProb]
    list predclass countG groupAPP occ occ_pp p TotProb if counter == 1
    restore
end

summary_table_procLCla




/*
Class variable manipulation
*/

* 3-class model: rename  predclass to E_lca_group3

rename  predclass E_lca_group3

* labelling variable of E_lca_group3

label var E_lca_group3 "Latent classes 3 groups of cardiometabolic markers"

* labelling values
lab def lca_group3 1 "Healthy with obesity" 2 "Obesity and Hypertension" 3 "Complex
cardiometabolic disorders"

* attach category labels to the variable through label value

lab val E_lca_group3 lca_group3
```

```
ta E_lca_group3



* margins and marginsplot

* use margins to calculate marginal predictions
* use marginsplot to graph marginal predictions


margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3))

marginsplot, xtitle ("") ytitle ("") ///
             xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
             title ("Predicted Latent Class Probabilities with 95% CI")


margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3))
marginsplot, recast(bar) xtitle("") ytitle("") ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
             title("Predicted Latent Class Probabilities with 95% CI")


margins, predict(outcome(hba1c_lca) class(1)) ///
         predict(outcome(hba1c_lca) class(2)) ///
         predict(outcome(hba1c_lca) class(3))
marginsplot, recast(bar) xtitle("") ytitle("") ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
             title("Predicted Pr(HbA1c=1) with 95% CI")

* repeat with all CM var


* class 1

margins, predict(outcome(crp_lca) class(1)) ///
         predict(outcome(hdl_lca) class(1)) ///
         predict(outcome(obesity_lca) class(1)) ///
         predict(outcome(systolic_lca) class(1)) ///
         predict(outcome(diastolic_lca) class(1)) ///
         predict(outcome(diabetes_lca) class(1)) ///
         predict(outcome(hba1c_lca) class(1)) ///



marginsplot, recast(bar) title ("Class 1") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1) name (class1)


* class 2

margins, predict(outcome(crp_lca) class(2)) ///
         predict(outcome(hdl_lca) class(2)) ///
         predict(outcome(obesity_lca) class(2)) ///
         predict(outcome(systolic_lca) class(2)) ///
         predict(outcome(diastolic_lca) class(2)) ///
         predict(outcome(diabetes_lca) class(2)) ///
         predict(outcome(hba1c_lca) class(2)) ///
```

```
marginsplot, recast(bar) title ("Class 2") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1)name (class2)




* class 3

margins, predict(outcome(crp_lca) class(3)) ///
         predict(outcome(hdl_lca) class(3)) ///
         predict(outcome(obesity_lca) class(3)) ///
         predict(outcome(systolic_lca) class(3)) ///
         predict(outcome(diastolic_lca) class(3)) ///
         predict(outcome(diabetes_lca) class(3)) ///
         predict(outcome(hba1c_lca) class(3)) ///



marginsplot, recast(bar) title ("Class 3") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1) name (class3)



graph combine class1 class2 class3, cols(3)









* ------------------------------- *






/* MULTIPLE IMPUTATION (MI)

To handle with missing values of covariates


useful sources for MI and MICE:

https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/
https://www.stata.com/manuals/mi.pdf - see page 139
https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
https://www.youtube.com/watch?v=i6SOlq0mjuc&ab_channel=StataCorpLLC
https://dss.princeton.edu/training/MIStata.pdf



Preparing to conduct MI
1. examine the number and proportion of missing values among the variables of interest
    use the mdesc command
2. examine missing data patterns
    use commands mi set and mi misstable patterns
3. identify potential auxiliary variables


Run MI using chained equations (MICE)
using the commands
```

```
1. how (in what style) to store the imputations
 mi set wide
2. which variables will be imputed
 mi register imputed
3. optionally, which variables will not be imputed
 mi register regular
4. what imputation method is implemented to impute each of var - MICE
 mi impute chained

*/





/*

1. examining missing values
    install packages:
    * install mdesc
    * install tabmiss
    * insatll dm31
    * insall mvpatterna

*/

search mdesc
search rmiss2
search mvpatterns


* examining number of missing values vs non-missing in each variable

mdesc E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
E_memory_wordrecall Ewv2_depressive_symptoms




* examining missing data patterns

mi set wide

mi misstable summarize E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
E_memory_wordrecall Ewv2_depressive_symptoms


mi misstable patterns E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
E_memory_wordrecall Ewv2_depressive_symptoms


/*
 identifying potential auxiliary var
* Auxiliary variables are either correlated with a missing variable(s)
(the recommendation is r > 0.4) or are believed to be associated with missingness
- a priori knowledge of var that would make good auxiliary var
- identify potential candidates by examining associations between missing var and other var in the dataset
    running correlation using the command: pwcorr v1 v2 v3, obs
    the recommnedation for good correlation is r > 0.4


Missing var to be imputed:

    E_eduaction E_wealthquintiles
```

```
    E_smoking_3cat E_physicalactivity E_alcohol_status
    E_memory_wordrecall Ewv2_depressive_symptoms



Potential auxiliary var:
DV:  Ewv3to9_dementia_event
IV:  crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
other var:
    E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity

*/


* correlation

pwcorr E_eduaction E_wealthquintiles ///
    E_smoking_3cat E_physicalactivity E_alcohol_status ///
    E_memory_wordrecall Ewv2_depressive_symptoms ///
    Ewv3to9_dementia_event crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca ///
    diabetes_lca hba1c_lca ///
    E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity, obs


/* The correlation showed that all the following var are good auxiliary:
Ewv3to9_dementia_event obesity_lca diabetes_lca E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity
A good auxiliary does not have to be correlated with every variable to be useful
And it's not problematic if it has missing info of it's own
*/



/*
MI by chained equations (MICE)
    see: https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/

MICE is known as the fully conditional specification or sequential generalized regression
does not assume a joint MVN distribution
but instead uses a separate conditional distribution for each imputed variable.

The multivariate normal (MVN) model - mi imputed mvn -
assumes multivariate normality of all var

The multivariate imputation by chained equations (MICE) - mi imputed chained -
offers flexibility in how each var is modeled

mi impute chained allows to specify models for a
variety of variable types, including
continuous, binary, ordinal, nominal, truncated, and count variables


The MICE distributions available in Stata are:
binary, ordered and multinomial logistic regression for categorical variables,
linear regression and predictive mean matching (PMM)* for continuous variables,
and Poisson and negative binomial regression for count variables.



IMPUTATION PHASES

1. mi set wide
    style to store imputations

2. mi register imputed
    identifies which variables in the imputation model have missing information.

3. mi register regular (! optional)
    which variables will not be imputed
```

```

4. mi impute chained
    where the user specifies the imputation model to be used
    and the number of imputed datasets to be created.
    Example:
        mi impute chained (regress) bmi age (logit) female ///
        (mlogit) race = bpdiast i.region, add(20)

5. mi estimate
    is used as a prefix to the standard regress command.
    This executes the specified estimation model within each of the 20 imputed datasets
    to obtain 20 sets of coefficients and standard errors.
    Stata then combines these estimates to obtain one set of inferential statistics.
    In the output from mi estimate you will see some metrics: Imputation Diagnostics
    information for RVI (Relative Increase in Variance),
    FMI (Fraction of Missing Information),
    DF (Degrees of Freedom) ,
    RE (Relative Efficiency),
    and the between imputation and the within imputation variance estimates
    to examine how the standard errors (SEs) are calculated.


----------------

SELECTING MY IMPUTATION MODEL

- MICE -> mi impute chained

- var to be imputed:

    linear regression for continuous var (regress) ->
    E_memory_wordrecall

    logistic for the binary var (logit) ->
    Ewv2_depressive_symptoms


    multinomial logistic for our nominal categorical var (mlogit) ->
    E_eduaction E_wealthquintiles
    E_smoking_3cat E_physicalactivity E_alcohol_status



- auxiliary var:

    DV -> Ewv3to9_dementia_event
    IV -> obesity_lca diabetes_lca
    other covariates -> E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity



- imputation numbers (m) -> 10

    White et al. (2010) recommendation: use the rule that m should equal the percentage of incomplete cases


- rseed (53421) for reproducability reasons


- (! OPTIONAL) advance impute options -> force

    proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data) are encountered

- impute options -> savetrace (trace1)

    specifies Stata to save the means and standard deviations of imputed values from each
```

```
iteration to a Stata dataset named “trace1
*/


mi set wide


mi register imputed E_eduaction E_wealthquintiles ///
    E_smoking_3cat E_physicalactivity E_alcohol_status ///
    E_memory_wordrecall Ewv2_depressive_symptoms



mi impute chained (logit) Ewv2_depressive_symptoms ///
(mlogit) E_eduaction E_wealthquintiles ///
    E_smoking_3cat E_physicalactivity E_alcohol_status ///
(regress) E_memory_wordrecall = Ewv3to9_dementia_event obesity_lca diabetes_lca ///
    E_age E_sex E_maritalstatus_4cat E_cvd_comorbidity, add(10) rseed(53421) savetrace(trace1)




* save imputed data

* plot imputations


*it will open a file named trace1
use trace1, clear

describe


reshape wide *mean *sd, i(iter) j(m)

tsset iter




/*
The trace plot below graphs the predicted means value produced during the first imputation chain.
As before, the expectations is that the values would vary randomly to incorporate variation into the predicted values for read.
*/

tsline E_eduaction_mean1, name(mice1,replace)legend(off) ytitle("Mean of education")
tsline E_wealthquintiles_mean1, name(mice1,replace)legend(off) ytitle("Mean of wealth")
tsline E_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
tsline E_physicalactivity_mean1, name(mice1,replace)legend(off) ytitle("Mean of physical activity")
tsline E_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol status")
tsline E_memory_wordrecall_mean1, name(mice1,replace)legend(off) ytitle("Mean of memory")


/*

All 10 imputation chains can also be graphed simultaneously to make sure that nothing unexpected occurred in a single chain.
Every chain is obtained using a different set of initial values and this should be unique.
Each colored line represents a different imputation.
So all 10 imputation chains are overlaid on top of one another.

*/


tsline E_memory_wordrecall_mean*, name(mice1,replace)legend(off) ytitle("Mean of memory")
tsline E_memory_wordrecall_sd*, name(mice2, replace) legend(off) ytitle("SD of memory")
graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)
```

```
* repeat for each imputed var






/*
---- DESCRIPTIVE STATISTICS ----

General characteristics of participants

General characteristics of participnats stratified for study inclusion

General characteristics of participants stratified for dementia occurence

Participant characteristics by CM 3-class groups

CHI-SQUARE (chi2) for categorical var (crosstabulation)
    Frequency tables -> two-way tables
        using the command tabulate, chi2
        reporting observations, column percentage (N, %) and p-value of Pearson's r

one-way ANOVA for continuous var
    check box plot
    using the command oneway
    reporting mean, sd (summary tables) and p-value of F
*/


* General characteristics of ELSA participants at baseline

* Socio-demographics
sum E_age
ta E_sex
ta E_eduaction
ta E_maritalstatus_4cat
ta E_wealthquintiles
* Cardiometabolic disorders
ta crp_lca
ta hdl_lca
ta obesity_lca
ta systolic_lca
ta diastolic_lca
ta diabetes_lca
ta hba1c_lca
* Cardiovascular health factors
ta E_smoking_3cat
ta E_physicalactivity
ta E_alcohol_status
ta E_cvd_comorbidity
* Depressive symptoms (cont and categ)
sum Ewv2_cesd_score
ta Ewv2_depressive_symptoms
* Memory score
sum E_memory_wordrecall



* General baseline characteristics of ELSA participants by dementia status

* Socio-demographics
ttest E_age, by(Ewv3to9_dementia_event)
ta E_sex Ewv3to9_dementia_event, chi2 column row
ta E_eduaction Ewv3to9_dementia_event, chi2 column row
ta E_maritalstatus_4cat Ewv3to9_dementia_event, chi2 column row
ta E_wealthquintiles Ewv3to9_dementia_event, chi2 column row
```

```
* Cardiometabolic disorders
ta crp_lca Ewv3to9_dementia_event, chi2 column row
ta hdl_lca Ewv3to9_dementia_event, chi2 column row
ta obesity_lca Ewv3to9_dementia_event, chi2 column row
ta systolic_lca Ewv3to9_dementia_event, chi2 column row
ta diastolic_lca Ewv3to9_dementia_event, chi2 column row
ta diabetes_lca Ewv3to9_dementia_event, chi2 column row
ta hba1c_lca Ewv3to9_dementia_event, chi2 column row
* Cardiovascular health factors
ta E_smoking_3cat Ewv3to9_dementia_event, chi2 column row
ta E_physicalactivity Ewv3to9_dementia_event, chi2 column row
ta E_alcohol_status Ewv3to9_dementia_event, chi2 column row
ta E_cvd_comorbidity Ewv3to9_dementia_event, chi2 column row
* Depressive symptoms (cont and categ)
ttest Ewv2_cesd_score, by(Ewv3to9_dementia_event)
ta Ewv2_depressive_symptoms Ewv3to9_dementia_event, chi2 column row
* Memory score
ttest E_memory_wordrecall, by(Ewv3to9_dementia_event)



* Sample characteristics by CM 3-class groups
* crosstabs categ var (frequencies and chi2) !report column percentage!
* oneway ANOVA cont var (mean, sd)


* Socio-demographics
oneway E_age E_lca_group3, tabulate
ta E_sex E_lca_group3, chi2 column row
ta E_eduaction E_lca_group3, chi2 column row
ta E_maritalstatus_4cat E_lca_group3, chi2 column row
ta E_wealthquintiles E_lca_group3, chi2 column row
* Cardiovascular health factors
ta E_smoking_3cat E_lca_group3, chi2 column row
ta E_physicalactivity E_lca_group3, chi2 column row
ta E_alcohol_status E_lca_group3, chi2 column row
ta E_cvd_comorbidity E_lca_group3, chi2 column row
* Depressive symptoms (cont and categ)
oneway Ewv2_cesd_score E_lca_group3, tabulate
ta Ewv2_depressive_symptoms E_lca_group3, chi2 column row
* Memory score
oneway E_memory_wordrecall E_lca_group3, tabulate





/*
---- SURVIVAL ANALYSIS IN COMPLETE DATA ----

Tests of proportional-hazards assumption
Kaplan Meier survival curves
Person-time
Cox proportional regression - Hazard ratios - stcox
Postestimation tools for stcox
Test of Goodness of Fit

*** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
Cox PH regression in complete data
Cox PH regression model in imputed dataset - mi estimate


*/



* check dataset variables of interest only

```

```
codebook E_time_event_dementia Ewv3to9_dementia_event E_lca_group3 ///
E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_physicalactivity E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms E_memory_wordrecall,compact



* Declare Data to be Survival Data
* Time to event: E_time_event_dementia (months)
* Censoring: Ewv3to9_dementia_event (1=dementia, 0=censored)
* Command is stset TIMETOEVENT, failure(CENSORVARIABLE)


stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)


*describe survival data using commnad stsum

stsum

stsum, by(E_lca_group3)



* Kaplan Meier Curve estimation

sts list

sts list, by(E_lca_group3)



* Kaplan Meier Curve Plot

* no frills plot

sts graph

* with frills

sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
title("Kaplan Meier Curve") subtitle("n=4511, # events=284") ///
caption("graph02.png"), size(vsmall))


* With Greenwood CI limits

sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
title("Kaplan Meier Curve") subtitle("n=4511, # events=284") caption("graph03.png", size(vsmall))




* Group Kaplan-Meier Curve Estimation
* Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first

sort E_lca_group3

sts list, by(E_lca_group3)

* graph with frills

sts graph, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in Months") ///
ytitle("Survival Prob") title("Kaplan Meier Curve") subtitle("n=4511, # events=284") ///
caption("graph04.png", size(vsmall))
```

```
* calculate person-time and incidence rates using command ststime

stptime,title(Person-years)

stptime, title(Person-years) per(1000)

stptime, title(Person-years) per(10000)


* calculate person-time  by category of E_lca_group3

stptime, by(E_lca_group3)

stptime, by(E_lca_group3) per(1000)





* mean and median of follow-up
sum E_time_event_dementia
sum E_time_event_dementia, detail



/* Log Rank Test of equality of survival distributions
 (NULL: equality of survival distributions among E_lca_group3 groups)
 We will consider including the predictor if the test has a p-value of 0.2 - 0.25 or less.
 If the predictor has a p-value greater than 0.25 in a univariate analysis
 it is highly unlikely that it will contribute anything to a model which includes other predictors.
 Command is sts test GROUPVAR
*/


sts test E_lca_group3, logrank

sts test E_age, logrank

sts test E_sex, logrank

sts test E_eduaction, logrank

sts test E_maritalstatus_4cat, logrank

sts test E_wealthquintiles, logrank

sts test E_smoking_3cat, logrank

sts test E_physicalactivity, logrank

sts test E_alcohol_status, logrank

sts test E_cvd_comorbidity, logrank

sts test Ewv2_depressive_symptoms, logrank

sts test E_memory_wordrecall, logrank





/* Cox PH regression model

using the command stcox

```

```
--- Building the model ---

Model 1: unadjusted - single predictor of CM classes
Model 2: model 1 + sociodemographics: age sex education marital status and wealth
Model 3: model 2 + cvd health: smoking, alcohol consumption, cvd comorbidity
Model 4: model 3 + mental health: depressive symptoms


!! I didn't adjust for physical activity because this variable can't be used in CHARLS (missing values)

*/


* Unadjusted model - model 1 - single predictor

stcox E_lca_group3

* define design var by using i.(3 classes)

stcox i.E_lca_group3


* Adjusted models - multivariable Cox model
* controlling for covariates

* model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth

stcox i.E_lca_group3 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles


* model 3: model 2 + adjust for cvd health

stcox i.E_lca_group3 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity


* model 4: model 3 + adjust for depression

stcox i.E_lca_group3 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms






* Coefficients instead of hazard ratios by specifing the option nohr

stcox i.E_lca_group3, nohr


stcox i.E_lca_group3 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms, nohr





* Multivariable model development
* Likelihood-ratio tests



*install eststo
findit eststo
```

```
* ---- rx controlling for age and sex -----*
quietly: stcox E_age i.E_sex
eststo modelagesex

quietly: stcox E_age i.E_sex i.E_lca_group3
eststo modelagesex_3group

lrtest modelagesex modelagesex_3group



* ---- rx controlling for sociodemographics -----*
quietly: stcox E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles
eststo modelsociodemo

quietly: stcox E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles i.
E_lca_group3
eststo modelsociodemo_3group

lrtest modelsociodemo modelsociodemo_3group


* ---- rx controlling for cardiovascular health -----*
quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity
eststo modelcardiovascular

quietly: stcox i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_lca_group3
eststo modelcardiovascular_3group

lrtest modelcardiovascular modelcardiovascular_3group



* ---- rx controlling for mental health-----*
quietly: stcox i.Ewv2_depressive_symptoms
eststo modelmentalcogn

quietly: stcox i.Ewv2_depressive_symptoms i.E_lca_group3
eststo modelmentalcogn_3group

lrtest modelmentalcogn modelmentalcogn_3group




* side-by-side comparison of models


quietly: stcox i.E_lca_group3
eststo model1


quietly: stcox E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles i.
E_lca_group3
eststo model2


quietly: stcox E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity i.E_lca_group3
eststo model3


quietly: stcox E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms i.E_lca_group3
eststo model4
```

```
* Display Betas and Summary Statistics
estout model1 model2 model3 model4, stats(n chi2 bic, star(chi2)) prehead("Betas")

/* Key Interpretattion
Chi2 = Value of LR test comparing the model fit (“full”) to intercept only (“reduced”)
bic = Schwarz’ Bayesian Information Criterion = It is a function of the log-likelihood.
Smaller values indicate a better fit.
*/

* Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
estout model1 model2 model3 model4, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")




* Postestimation tools for stcox

* Test of proportional hazards

estat phtest, detail


/* Proportionality Assumption - method 1
We will check proportionality by including time-dependent covariates in the model
by using the tvc and the texp options in the stcox command.
Time dependent covariates are interactions of the predictors and time.
In this analysis we choose to use the interactions with log(time)
because this is the most common function of time used in time-dependent covariates
but any function of time could be used.
If a time-dependent covariate is significant this indicates
a violation of the proportionality assumption for that specific predictor.
The conclusion is that all of the time-dependent variables are not significant
either collectively or individually thus supporting the assumption of proportional hazard.
*/



stcox i.E_lca_group3 E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms, nohr ///
tvc(E_lca_group3 E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms) texp(ln(E_time_event_dementia))



/* Proportionality Assumption - method 2
by using the Schoenfeld and scaled Schoenfeld residuals
In the stphtest command we test the proportionality of the model as a whole
and by using the detail option we get a test of proportionality for each predictor.
By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
If the tests in the table are not significance (p-values over 0.05)
then we can not reject proportionality and we assume
that we do not have a violation of the proportional assumption.
The stphplot command uses log-log plots to test proportionality
and if the lines  in these plots are parallel then we have further indication
that the predictors do not violate the proportionality assumption.
*/

quietly stcox E_lca_group3 E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms, schoenfeld(sch*) scaledsch(sca*)
stphtest, detail
```

```
stphtest, plot(E_lca_group3) msym(oh)
stphtest, plot(E_age) msym(oh)
stphtest, plot(E_sex) msym(oh)
stphtest, plot(E_eduaction) msym(oh)
stphtest, plot(E_maritalstatus_4cat) msym(oh)
stphtest, plot(E_wealthquintiles) msym(oh)
stphtest, plot(E_smoking_3cat) msym(oh)
stphtest, plot(E_alcohol_status) msym(oh)
stphtest, plot(E_cvd_comorbidity) msym(oh)
stphtest, plot(Ewv2_depressive_symptoms) msym(oh)






stphplot, by(E_lca_group3) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_age) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_sex) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_eduaction) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_wealthquintiles) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_alcohol_status) plot1(msym(oh)) plot2(msym(th))
stphplot, by(E_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
stphplot, by(Ewv2_depressive_symptoms) plot1(msym(oh)) plot2(msym(th))




* Assessment of PH Assumption: adjust for age and sex
stphplot, by(E_lca_group3) adjust(E_age E_sex) nolntime plot1opts(symbol(none) color(black)
lpattern(dash)) ///
plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")



* Assessment of PH Assumption: adjust for model 2
stphplot, by(E_lca_group3) adjust(E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles)
///
nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")



* Assessment of PH Assumption: adjust for model 3
stphplot, by(E_lca_group3) adjust(E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles
///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity) ///
nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")



* Assessment of PH Assumption: adjust for model 4
stphplot, by(E_lca_group3) adjust(E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles
///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms) ///
nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
title("Assessment of PH Assumption") subtitle(" Predictor is E_lca_group3") xtitle("months")
```

```
/* Test of overall goodness of fit
Goodness of fit of the final model
2 methods:
 - by using the commnad stcoxgof (good fit = non sig p-value)
 - by using the Cox-Snell residuals
    to create the Nelson-Aalen cumulative hazard function
    If the hazard function follows the 45 degree line then we know that it approximately
    has an exponential distribution with a hazard rate of one and that the model fits the data well.
    If the model fits the data, the plot of the cumulative hazard versus cs
    should approximate a straight line with slope 1.
*/


* by using the commnad stcoxgof

* install stcoxgof

findit stcoxgof


stcox E_lca_group3 E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms, mgale(mgale)


stcoxgof




* by using the Cox-Snell residuals

quietly stcox E_lca_group3 E_age E_sex E_eduaction E_maritalstatus_4cat E_wealthquintiles ///
E_smoking_3cat E_alcohol_status E_cvd_comorbidity ///
Ewv2_depressive_symptoms
predict cs, csnell

* or

quietly stcox E_lca_group3
predict cs, csnell


stset cs, failure(Ewv3to9_dementia_event)
sts generate km = s
generate H = -ln(km)
line H cs cs, sort ytitle("") clstyle(. refline)









* -------- COX PH REGRESSION MODEL IN IMPUTED DATASET --------- *


* Declare Data to be Survival Data by using mi

mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)


* Run Cox regression analysis in imputed dataset by using "mi estimate:"
```

```
* Building the Model: Model 1 (unadjusted), Model 2, Model 3, Model 4



* Unadjusted model - model 1 - single predictor

* Model 1 (default coefficents)
mi estimate: stcox E_lca_group3

* Model 1: define design var by using i.
mi estimate: stcox i.E_lca_group3


* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: age sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles

* Model 3: model 2 + adjust for cvd health

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity


* Model 4: model 3 + adjust for depression
mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms




/*

*** SENSITIVITY ANALYSES ***

1) multigroup latent class model by sex

2) interactions with age and gender
survival analysis stratified by age
two age groups: <70 and >=70

3) exclude participants with cvd

4) exclude dementia cases identified by IQCODE

5) survival analysis limiting to 5 year follow-up

6) Complete data
 Cox regression analysis on complete data (without imputed covariates)
 (see above)

*/


/*
1) Multigroup latent class model by sex

```

```

TWO STEP PROCESS

1) LCA by group (to build the model and get lcprob and lcmean and to get the marginplots for males)

gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
group(E_sex) ginvariant(coef)

estat lcprob
estat lcmean
estat lcgof

2) LCA sort sex (to get the marginplots for females)

sort E_sex

by E_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)


*/


* LCA by group
* three-class model


gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
group(E_sex) ginvariant(coef)


estimates store threeclass_cm




* LCA postestimation
* Latent class marginal probabilities - lcprob -
* Latent class marginal means - lcmean -


estat lcprob

estat lcmean

estat lcgof



/* We can use the predictions of the posterior probability of class membership to evaluate an
individual’s probability of being in each class.

*/

predict m_classpost1*, classposteriorpr
list in 1, abbrev(10)

/* We can determine the expected class for each individual based on whether the posterior probability
is greater than 0.5
*/

generate m_expclass1 = 1 + (m_classpost11>0.5)
tabulate m_expclass1

```

```
generate m_expclass2 = 1 + (m_classpost12>0.5)
tabulate m_expclass2


generate m_expclass3 = 1 + (m_classpost13>0.5)
tabulate m_expclass3




/* We can determine expected classification for each individual in the dataset based on the predicted
posterior class probabilities.
*/

predict m_cpost*, classposteriorpr
egen m_max = rowmax(m_cpost*)


* generate classes var

generate m_predclass = 1 if m_cpost1==m_max

replace m_predclass = 2 if m_cpost2==m_max

replace m_predclass = 3 if m_cpost3==m_max

tabulate m_predclass



* margins and marginsplot for MALES

* use margins to calculate marginal predictions
* use marginsplot to graph marginal predictions



*Install/update combomarginsplot ado.

*https://www.statalist.org/forums/forum/general-stata-discussion/general/1425209-is-it-possible-to-do-multilevel-latent-class-analysis-with-stata-15-ic

ssc install combomarginsplot, replace


margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3)) subpop(if E_sex==0) saving(margin_male, replace)
marginsplot, xtitle ("") ytitle ("") ///
             xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
             title ("Predicted Latent Class Probabilities with 95% CI") ///
             name(margin_male, replace)


margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3)) subpop(if E_sex==0) saving(margin_male, replace)
marginsplot, recast(bar) xtitle("") ytitle("") ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
             title("Predicted Latent Class Probabilities with 95% CI") ///
             name(margin_male, replace)


* class 1

margins, predict(outcome(crp_lca) class(1)) ///
         predict(outcome(hdl_lca) class(1)) ///
```

```
         predict(outcome(obesity_lca) class(1)) ///
         predict(outcome(systolic_lca) class(1)) ///
         predict(outcome(diastolic_lca) class(1)) ///
         predict(outcome(diabetes_lca) class(1)) ///
         predict(outcome(hba1c_lca) class(1)) subpop(if E_sex==0) ///
         saving(class1_male, replace) ///

marginsplot, recast(bar) title ("Class 1") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1) name (class1_male, replace)


* class 2

margins, predict(outcome(crp_lca) class(2)) ///
         predict(outcome(hdl_lca) class(2)) ///
         predict(outcome(obesity_lca) class(2)) ///
         predict(outcome(systolic_lca) class(2)) ///
         predict(outcome(diastolic_lca) class(2)) ///
         predict(outcome(diabetes_lca) class(2)) ///
         predict(outcome(hba1c_lca) class(2)) subpop(if E_sex==0) ///
          saving(class2_male, replace) ///

marginsplot, recast(bar) title ("Class 2") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1) name (class2_male, replace)




* class 3

margins, predict(outcome(crp_lca) class(3)) ///
         predict(outcome(hdl_lca) class(3)) ///
         predict(outcome(obesity_lca) class(3)) ///
         predict(outcome(systolic_lca) class(3)) ///
         predict(outcome(diastolic_lca) class(3)) ///
         predict(outcome(diabetes_lca) class(3)) ///
         predict(outcome(hba1c_lca) class(3)) subpop(if E_sex==0) ///
         saving(class3_male, replace) ///

marginsplot, recast(bar) title ("Class 3") xtitle("") ///
             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
             ytitle("Predicted mean") ylabel(0(.20)1) name (class3_male, replace)



graph combine class1_male class2_male class3_male, cols(3)








* LCA sort by sex
* three-class model

sort E_sex

by E_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
startvalues(randompr, draws(20) seed(15) difficult) ///
emopts(iterate(30) difficult)

```

```
estat lcprob

estat lcmean

estat lcgof




/* We can use the predictions of the posterior probability of class membership to evaluate an
individual’s probability of being in each class.

*/

predict f_classpost1*, classposteriorpr
list in 1, abbrev(10)

/* We can determine the expected class for each individual based on whether the posterior probability
is greater than 0.5
*/

generate f_expclass1 = 1 + (f_classpost11>0.5)
tabulate f_expclass1


generate f_expclass2 = 1 + (f_classpost12>0.5)
tabulate f_expclass2


generate f_expclass3 = 1 + (f_classpost13>0.5)
tabulate f_expclass3



/* We can determine expected classification for each individual in the dataset based on the predicted
posterior class probabilities.
*/

predict f_cpost*, classposteriorpr
egen f_max = rowmax(f_cpost*)


* generate classes var

generate f_predclass = 1 if f_cpost1==f_max

replace f_predclass = 2 if f_cpost2==f_max

replace f_predclass = 3 if f_cpost3==f_max

tabulate f_predclass




* margins and marginsplot for FEMALES

* use margins to calculate marginal predictions
* use marginsplot to graph marginal predictions



margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3)) subpop(if E_sex==1) saving(margin_female, replace)
marginsplot, xtitle ("") ytitle ("") ///
```

```
                xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
                title ("Predicted Latent Class Probabilities with 95% CI") ///
                name(margin_female, replace)


margins, predict(classpr class(1)) ///
         predict(classpr class(2)) ///
         predict(classpr class(3)) subpop(if E_sex==1) saving(margin_female, replace)
marginsplot, recast(bar) xtitle("") ytitle("") ///
                xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
                title("Predicted Latent Class Probabilities with 95% CI") ///
                name(margin_female, replace)


* class 1

margins, predict(outcome(crp_lca) class(1)) ///
         predict(outcome(hdl_lca) class(1)) ///
         predict(outcome(obesity_lca) class(1)) ///
         predict(outcome(systolic_lca) class(1)) ///
         predict(outcome(diastolic_lca) class(1)) ///
         predict(outcome(diabetes_lca) class(1)) ///
         predict(outcome(hba1c_lca) class(1)) subpop(if E_sex==1) ///
         saving(class1_female, replace) ///

marginsplot, recast(bar) title ("Class 1") xtitle("") ///
                xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
                5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
                ytitle("Predicted mean") ylabel(0(.20)1) name (class1_female, replace)


* class 2

margins, predict(outcome(crp_lca) class(2)) ///
         predict(outcome(hdl_lca) class(2)) ///
         predict(outcome(obesity_lca) class(2)) ///
         predict(outcome(systolic_lca) class(2)) ///
         predict(outcome(diastolic_lca) class(2)) ///
         predict(outcome(diabetes_lca) class(2)) ///
         predict(outcome(hba1c_lca) class(2)) subpop(if E_sex==1) ///
          saving(class2_female, replace) ///

marginsplot, recast(bar) title ("Class 2") xtitle("") ///
                xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
                5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
                ytitle("Predicted mean") ylabel(0(.20)1) name (class2_female, replace)




* class 3

margins, predict(outcome(crp_lca) class(3)) ///
         predict(outcome(hdl_lca) class(3)) ///
         predict(outcome(obesity_lca) class(3)) ///
         predict(outcome(systolic_lca) class(3)) ///
         predict(outcome(diastolic_lca) class(3)) ///
         predict(outcome(diabetes_lca) class(3)) ///
         predict(outcome(hba1c_lca) class(3)) subpop(if E_sex==1)  ///
         saving(class3_female, replace) ///

marginsplot, recast(bar) title ("Class 3") xtitle("") ///
                xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
                5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
                ytitle("Predicted mean") ylabel(0(.20)1) name (class3_female, replace)
```

```
graph combine class1_female class2_female class3_female, cols(3)






*combine margins male and female class probabilities

graph combine margin_male margin_female, cols(3)

*combine margins male and female 3 classes mean

graph combine class1_male class2_male class3_male class1_female class2_female class3_female, cols(
3)







/* 2) Interaction with age and gender
Survival analysis stratified by age

generate age group variable
Age groups: 1) young old (< 70) 2) old old (>= 70)

Kaplan Meier curves
Cox regression models in imputed data

young old <70
if E_age_group==1

old old >70
if E_age_group==2


*/




gen E_age_group=1 if E_age < 70
replace E_age_group=2 if E_age >=70 & !missing(E_age)

label var E_age_group "Age groups <70 young-old / 70 old-old"
lab def age_group 1 "young old <70" 2 "old old >70"
lab val E_age_group age_group

tab E_age_group


stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)


*  <70 Kaplan Meier

* Group Kaplan-Meier Curve Estimation
* Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first

sort E_lca_group3

sts list if E_age_group==1, by(E_lca_group3)

* graph with frills

sts graph if E_age_group==1, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in
```

```
Months") ///
ytitle("Survival Prob") title("Kaplan Meier Curve <70 years")  ///
caption("graph04.png", size(vsmall))


* calculate person-time and incidence rates using command ststime

stptime if E_age_group==1,title(Person-years)

stptime if E_age_group==1, title(Person-years) per(1000)


* calculate person-time  by category of E_lca_group3

stptime if E_age_group==1, by(E_lca_group3)

stptime if E_age_group==1, by(E_lca_group3) per(1000)




*  >70 Kaplan Meier

sts list if E_age_group==2, by(E_lca_group3)

* graph with frills

sts graph if E_age_group==2, by(E_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in
Months") ///
ytitle("Survival Prob") title("Kaplan Meier Curve >=70 years")  ///
caption("graph04.png", size(vsmall))


* calculate person-time and incidence rates using command ststime

stptime if E_age_group==2,title(Person-years)

stptime if E_age_group==2, title(Person-years) per(1000)


* calculate person-time  by category of E_lca_group3

stptime if E_age_group==2, by(E_lca_group3)

stptime if E_age_group==2, by(E_lca_group3) per(1000)




* COX PH REGRESSION MODEL IN IMPUTED DATASET


* Declare Data to be Survival Data by using mi

mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)





*** INTERACTION sex*cardiometabolic cluster ***

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 c.E_sex#i.E_lca_group3


mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
```

```
i.Ewv2_depressive_symptoms c.E_sex#i.E_lca_group3




*** INTERACTION age*cardiometabolic cluster ***

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 c.E_age#i.E_lca_group3


mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms c.E_age#i.E_lca_group3




* YOUNG OLD <70 Cox regression models

* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_age_group==1


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles if E_age_group==1

* Model 3: model 2 + adjust for cvd health

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity if E_age_group==1


* Model 4: model 3 + adjust for depression

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms if E_age_group==1




* OLD OLD >70 Cox regression models


* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_age_group==2


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles if E_age_group==2

* Model 3: model 2 + adjust for cvd health

```

```
mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity if E_age_group==2


* Model 4: model 3 + adjust for depression

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms if E_age_group==2







/*

3) exclude participants with cvd

use the command if E_cvd_comorbidity==0

*/


* COX PH REGRESSION MODEL IN IMPUTED DATASET


* Declare Data to be Survival Data by using mi

mi stset E_time_event_dementia, failure (Ewv3to9_dementia_event==1) id(idauniq)


* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 if E_cvd_comorbidity==0


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles if E_cvd_comorbidity==0

* Model 3: model 2 + adjust for cvd health

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status if E_cvd_comorbidity==0


* Model 4: model 3 + adjust for depression

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status ///
i.Ewv2_depressive_symptoms if E_cvd_comorbidity==0





/*
4) survival analysis excluding cases diagnosed with IQCODE
```

```
Variables
Ewv3to9_dementia_event_no_iqcode E_time_event_dementia_report_no_

*/




* COX PH REGRESSION MODEL IN IMPUTED DATASET



* Declare Data to be Survival Data by using mi

mi stset E_time_event_dementia_report_no_, failure (Ewv3to9_dementia_event_no_iqcode==1) id(
idauniq)


* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: age sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles

* Model 3: model 2 + adjust for cvd health

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity


* Model 4: model 3 + adjust for depression

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms









/*
5) survival analysis limiting to 5 year follow-up period

elsa follow-up wave 3-6

*/


merge 1:m idauniq using "S:\Research\pkstudies\Study3_cardio_lca\ELSA\elsa_lca data sensitivity
3to6followup.dta"


* COX PH REGRESSION MODEL IN IMPUTED DATASET


```

```
* Declare Data to be Survival Data by using mi

mi stset Ewv3to6_time_event_dementia, failure (Ewv3to6_dementia_event==1) id(idauniq)



* Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3


* Adjusted models - multivariable Cox model
* controlling for covariates

* Model 2: model 1 + adjust for demographics: age sex education marital status and wealth

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles

* Model 3: model 2 + adjust for cvd health

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity


* Model 4: model 3 + adjust for depression

mi estimate, eform("Haz. Ratio"): stcox i.E_lca_group3 ///
E_age i.E_sex i.E_eduaction i.E_maritalstatus_4cat i.E_wealthquintiles ///
i.E_smoking_3cat i.E_alcohol_status i.E_cvd_comorbidity ///
i.Ewv2_depressive_symptoms


* 6) complete data (see above)


* ------------------------------ *




*** EXTRA SENSITIVITY ANALYSES FOR PAPER ***

/*

compare baseline characteristics between complete sample (before exclusion) and sample with missing data (overall after exclusion)

*/


* General characteristics of ELSA participants at baseline

* Socio-demographics
sum E_age
ta E_sex
ta E_eduaction
ta E_maritalstatus_4cat
ta E_wealthquintiles
* Cardiometabolic disorders
ta Ewv2_crp
ta Ewv2_hdl_cholesterol
ta Ewv2_obesity_waist
ta Ewv2_systolic_bp
ta Ewv2_diastolic_bp
ta Ewv2_diabetes_diagnosed
```

```
ta Ewv2_HbA1c
* Cardiovascular health factors
ta E_smoking_3cat
ta E_physicalactivity
ta E_alcohol_status
ta E_cvd_comorbidity
* Depressive symptoms
ta Ewv2_depressive_symptoms
* Memory score
sum E_memory_wordrecall



* compare health characteristics between those survived and dropped out



*** CLEANING DATA to keep those who dropped out


* 1. drop dementia cases and missing data at baseline

drop if Ewv2_anydementia_iqcode_report==1
* (50 observations deleted)

drop if Ewv2_anydementia_iqcode_report== .
* (0 observations deleted)


* 2. drop missing values of cardiometabolic markers

drop if Ewv2_crp== .
* (1,753 observations deleted)

drop if Ewv2_hdl_cholesterol== .
* (6 observations deleted)

drop if Ewv2_obesity_waist== .
* (133 observations deleted)

drop if Ewv2_systolic_bp== .
* (660 observations deleted)

drop if Ewv2_diastolic_bp== .
* (0 observations deleted)

drop if Ewv2_diabetes_diagnosed== .
* (0 observations deleted)

drop if Ewv2_HbA1c== .
* (102 observations deleted)



* 3. drop obs with no records on dementia at any wave from 3-9 follow-ups


search mdesc
search rmiss2
search mvpatterns

* see number of missing values vs non-missing in each variable
mdesc Ewv3_anydementia_iqcode_report Ewv4_anydementia_iqcode_report ///
Ewv5_anydementia_iqcode_report Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report
```

```
/* number of missing values per observation
* the code below creates a variable called nmisfollowup that gives the number of missing values
for each observation in the variables of interest */
egen nmisfollowup_dementia_wv3to9=rmiss2(Ewv3_anydementia_iqcode_report ///
Ewv4_anydementia_iqcode_report Ewv5_anydementia_iqcode_report ///
Ewv6_anydementia_iqcode_report Ewv7_anydementia_iqcode_report ///
Ewv8_anydementia_iqcode_report Ewv9_anydementia_iqcode_report)

tab nmisfollowup_dementia_wv3to9

* drop observations "nmisfollowup_dementia_wv3to9" < 7
drop if nmisfollowup_dementia_wv3to9<7


* Socio-demographics
sum E_age
ta E_sex
ta E_eduaction
ta E_maritalstatus_4cat
ta E_wealthquintiles
* Cardiometabolic disorders
ta Ewv2_crp
ta Ewv2_hdl_cholesterol
ta Ewv2_obesity_waist
ta Ewv2_systolic_bp
ta Ewv2_diastolic_bp
ta Ewv2_diabetes_diagnosed
ta Ewv2_HbA1c
* Cardiovascular health factors
ta E_smoking_3cat
ta E_physicalactivity
ta E_alcohol_status
ta E_cvd_comorbidity
* Depressive symptoms
ta Ewv2_depressive_symptoms
* Memory score
sum E_memory_wordrecall



* compare health characteristics bewteen <70 and >=70



* General baseline characteristics of ELSA participants by age group

* Socio-demographics
ttest E_age, by(E_age_group)
ta E_sex E_age_group, chi2 column row
ta E_eduaction E_age_group, chi2 column row
ta E_maritalstatus_4cat Ewv3to9_dementia_event, chi2 column row
ta E_wealthquintiles E_age_group, chi2 column row
* Cardiometabolic disorders
ta crp_lca E_age_group, chi2 column row
ta hdl_lca E_age_group, chi2 column row
ta obesity_lca E_age_group, chi2 column row
ta systolic_lca E_age_group, chi2 column row
ta diastolic_lca E_age_group, chi2 column row
ta diabetes_lca E_age_group, chi2 column row
ta hba1c_lca E_age_group, chi2 column row
* Cardiovascular health factors
ta E_smoking_3cat E_age_group, chi2 column row
ta E_physicalactivity E_age_group, chi2 column row
ta E_alcohol_status E_age_group, chi2 column row
ta E_cvd_comorbidity E_age_group, chi2 column row
* Depressive symptoms
ta Ewv2_depressive_symptoms E_age_group, chi2 column row
* Memory score
```

```
ttest E_memory_wordrecall, by(E_age_group)
ta E_lca_group3 E_age_group, chi2 column row





* ------------------------------ *
```