```
/*

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 I: Independent and combined effects of depressive symptoms and cardiometabolic risk factors on dementia incidence


DATASET: CHARLS
baseline: wave 1 (2011) follow-up waves 2-4 (2013-2018)

4 time points in total: baseline and 3 follow-ups


TIMELINE

DEPRESSIVE SYMPTOMS AND CARDIOMETABOLIC RISK FACTORS: WV1 (BASELINE)
DEMENTIA INCIDENCE: W2 - WV4 (3 TIME POINTS)
COVARIATES ADJUSTMENT FOR HR MODELS: WV1

*/




* KEEP NECESSARY VARIABLES

keep ID id_12char bloodweight ///
C_sex C_age C_eduaction C_maritalstatus_8cat ///
C_maritalstatus_3cat C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
Cwv1_smoking_2cat Cwv1_smoking_3cat ///
Cwv1_physicalactivity Cwv1_alcohol_freq Cwv1_alcohol_status ///
C_cvd_comorbidity Cwv1_antidepressant Cwv1_psycholog_treat Cwv1_anytreat_psyche ///
Cwv1_memory_wordrecall Cwv1_cognition Cwv4_memory_wordrecall Cwv4_cognition ///
Cwv1_cesd_score Cwv1_depressive_symptoms ///
Cwv2_cesd_score Cwv2_depressive_symptoms ///
Cwv3_cesd_score Cwv3_depressive_symptoms ///
Cwv4_cesd_sumscore Cwv4_depressive_symptoms ///
Cwv1_crp_level Cwv1_crp Cwv1_hdl_level Cwv1_male_hdl Cwv1_female_hdl ///
Cwv1_meds_dyslipid Cwv1_anymeds_dyslipid Cwv1_dyslipid_evr ///
Cwv1_dyslipid_diagnosed Cwv1_dyslipid_report_sum ///
Cwv1_dyslipid_report Cwv1_hdl_sum Cwv1_hdl_cholesterol ///
Cwv1_waist Cwv1_malewaist_ao ///
Cwv1_femalewaist_ao Cwv1_obesity_waist_sum Cwv1_obesity_waist ///
Cwv1_bmi_score Cwv1_obesity_bmi Cwv1_waist_bmi_sum Cwv1_obesity ///
Cwv1_tg_level Cwv1_tg Cwv1_triglyc_sum Cwv1_triglyc ///
Cwv1_systolic_mean Cwv1_diastolic_mean Cwv1_systolic_bp Cwv1_diastolic_bp ///
Cwv1_meds_bp Cwv1_anymeds_bp Cwv1_bp_evr Cwv1_bp_diagnosed ///
Cwv1_bp_report_sum Cwv1_bp_report Cwv1_bp_sum Cwv1_bp ///
Cwv1_glucose_level Cwv1_glucose Cwv1_HbA1c_level Cwv1_HbA1c ///
Cwv1_diabetes_evr Cwv1_diabetes_diagnosed Cwv1_diabetes_report_sum ///
Cwv1_diabetes_report Cwv1_meds_diabetes Cwv1_anymeds_diabetes ///
Cwv1_glucose_diabetes_sum Cwv1_glycemia ///
Cwv1_ao_depress_sum Cwv1_Nao_Ndepress Cwv1_Nao_Ydepress ///
Cwv1_Yao_Ndepress Cwv1_Yao_Ydepress Cwv1_ao_depress_group ///
Cwv1_waist_depress_sum Cwv1_Nwaist_Ndepress Cwv1_Nwaist_Ydepress ///
Cwv1_Ywaist_Ndepress Cwv1_Ywaist_Ydepress Cwv1_waist_depress_group ///
Cwv1_glycemia_depress_sum Cwv1_Nglycemia_Ndepress ///
Cwv1_Nglycemia_Ydepress Cwv1_Yglycemia_Ndepress ///
Cwv1_Yglycemia_Ydepress Cwv1_glycemia_depress_group ///
Cwv1_diabet_depress_sum Cwv1_Ndiabet_Ndepress ///
Cwv1_Ndiabet_Ydepress Cwv1_Ydiabet_Ndepress ///
Cwv1_Ydiabet_Ydepress Cwv1_diabet_depress_group ///
Cwv1_hba1c_depress_sum Cwv1_Nhba1c_Ndepress Cwv1_Nhba1c_Ydepress ///
Cwv1_Yhba1c_Ndepress Cwv1_Yhba1c_Ydepress Cwv1_hba1c_depress_group ///
Cwv1_hdl_depress_sum Cwv1_Nhdl_Ndepress Cwv1_Nhdl_Ydepress ///
Cwv1_Yhdl_Ndepress Cwv1_Yhdl_Ydepress Cwv1_hdl_depress_group ///
```

```
Cwv1_bp_depress_sum Cwv1_Nbp_Ndepress Cwv1_Nbp_Ydepress ///
Cwv1_Ybp_Ndepress Cwv1_Ybp_Ydepress Cwv1_bp_depress_group ///
Cwv1_sbp_depress_sum Cwv1_Nsbp_Ndepress Cwv1_Nsbp_Ydepress ///
Cwv1_Ysbp_Ndepress Cwv1_Ysbp_Ydepress Cwv1_sbp_depress_group ///
Cwv1_dbp_depress_sum Cwv1_Ndbp_Ndepress Cwv1_Ndbp_Ydepress ///
Cwv1_Ydbp_Ndepress Cwv1_Ydbp_Ydepress Cwv1_dbp_depress_group ///
Cwv1_crp_depress_sum Cwv1_Ncrp_Ndepress Cwv1_Ncrp_Ydepress ///
Cwv1_Ycrp_Ndepress Cwv1_Ycrp_Ydepress Cwv1_crp_depress_group ///
Cwv1_tg_depress_sum Cwv1_Ntg_Ndepress Cwv1_Ntg_Ydepress ///
Cwv1_Ytg_Ndepress Cwv1_Ytg_Ydepress Cwv1_tg_depress_group ///
Cwv1_cardio_biomarkers_sum Cwv1_cardio_abnormality ///
Cwv1_ca_depress_sum Cwv1_Nca_Ndepress Cwv1_Nca_Ydepress ///
Cwv1_Yca_Ndepress Cwv1_Yca_Ydepress Cwv1_ca_depress_group ///
Cwv1_cardio3_sum Cwv1_cardio3 Cwv1_ca3_depress_sum ///
Cwv1_Nca3_Ndepress Cwv1_Nca3_Ydepress Cwv1_Yca3_Ndepress ///
Cwv1_Yca3_Ydepress Cwv1_ca3_depress_group Cwv1_cardio4_sum ///
Cwv1_cardio4 Cwv1_ca4_depress_sum Cwv1_Nca4_Ndepress ///
Cwv1_Nca4_Ydepress Cwv1_Yca4_Ndepress Cwv1_Yca4_Ydepress ///
Cwv1_ca4_depress_group Cwv1_cardio_number_sum Cwv1_cardio_number ///
Cwv1_cardio2_sum Cwv1_cardio2 Cwv1_ca2_depress_sum ///
Cwv1_Nca2_Ndepress Cwv1_Nca2_Ydepress ///
Cwv1_Yca2_Ndepress Cwv1_Yca2_Ydepress Cwv1_ca2_depress_group ///
Cwv1_dementia_report Cwv2_dementia_report ///
Cwv3_dementia_report Cwv4_self_info_dementia ///
Cwv1_interview_date Cwv2_interview_date Cwv3_interview_date Cwv4_interview_date ///
Cwv2to4_newdementia_or_lastinter Cwv2to4_dementia_free_date C_time_dementia_months ///
Cwv2to4_dementia_sum Cwv2to4_dementia_event ///
C_time_dementia_midpoint C_time_dementia_midpoint_final C_time_of_event_dementia



/* ---- MERGE DATA ----

Process to merge

Open master dataset and run merge two datasets

After merging all data both from master and using will be added
Need to keep if _merge==3
1 means cases from master data
2 means cases from using data
3 means cases from both master and using data

Drop _merge var.
*/


help merge

* Menu > Data > Combine datasets > Merge two datasets
* Choose One to many (key variable)


merge 1:m id_12char using
"S:\Research\pkstudies\Study1_biopsych_risk\CHARLS\charls_tomerge_educ.dta"

keep if _merge==3

drop _merge



/*

EXPOSURE VARIABLES


Binary variables of depressive symptoms and cardiometabolic markers measured at wave 1
```

```
134
135   Depression: Cwv1_depressive_symptoms
136
137   CRP: Cwv1_crp
138
139   HDL cholesterol: Cwv1_hdl_cholesterol
140
141   Obesity by waist cir: Cwv1_obesity_waist
142
143   systolic Blood pressure: Cwv1_systolic_bp
144
145   diastolic Blood pressure: Cwv1_diastolic_bp
146
147   Diabetes: Cwv1_diabetes_report
148
149   HbA1c: Cwv1_HbA1c
150
151
152   CA number (categ 0,1,2,3,4+): Cwv1_cardio_number
153
154   CA mutlimorbidity >= 2 CA conditions: Cwv1_cardio2
155
156   Grouping of Dep-CA: Cwv1_waist_depress_group Cwv1_diabet_depress_group Cwv1_hba1c_depress_group
      Cwv1_hdl_depress_group Cwv1_sbp_depress_group Cwv1_dbp_depress_group Cwv1_crp_depress_group
      Cwv1_ca3_depress_group Cwv1_ca4_depress_group Cwv1_ca2_depress_group
157
158   OUTCOME VARIABLES
159
160   Dementia event: Cwv2to4_dementia_event
161
162   Time-to-event: C_time_of_event_dementia
163
164   */
165
166
167
168
169
170
171   *** Descriptive stats of var of interest
172
173
174   tabulate Cwv1_depressive_symptoms
175   summarize Cwv1_depressive_symptoms
176
177   misstable summarize Cwv1_depressive_symptoms
178   misstable patterns Cwv1_depressive_symptoms
179
180
181   tabulate Cwv1_crp
182   summarize Cwv1_crp
183
184   misstable summarize Cwv1_crp
185   misstable patterns Cwv1_crp
186
187   tabulate Cwv1_hdl_cholesterol
188   summarize Cwv1_hdl_cholesterol
189
190   misstable summarize Cwv1_hdl_cholesterol
191   misstable patterns Cwv1_hdl_cholesterol
192
193   tabulate Cwv1_obesity_waist
194   summarize Cwv1_obesity_waist
195
196   misstable summarize Cwv1_obesity_waist
197   misstable patterns Cwv1_obesity_waist
198
199   tabulate Cwv1_systolic_bp
```

```
summarize Cwv1_systolic_bp

misstable summarize Cwv1_systolic_bp
misstable patterns Cwv1_systolic_bp


tabulate Cwv1_diastolic_bp
summarize Cwv1_diastolic_bp

misstable summarize Cwv1_diastolic_bp
misstable patterns Cwv1_diastolic_bp


tabulate Cwv1_diabetes_report
summarize Cwv1_diabetes_report

misstable summarize Cwv1_diabetes_report
misstable patterns Cwv1_diabetes_report


tabulate Cwv1_HbA1c
summarize Cwv1_HbA1c

misstable summarize Cwv1_HbA1c
misstable patterns Cwv1_HbA1c


tabulate Cwv1_dementia_report
summarize Cwv1_dementia_report

misstable summarize Cwv1_dementia_report
misstable patterns Cwv1_dementia_report









*** CLEANING DATA


* 1. drop dementia cases and missing data at baseline

* drop dementia wave 2 missing data
drop if Cwv1_dementia_report==1
* (267 observations deleted)
drop if Cwv1_dementia_report== .
* (88 observations deleted)


* 2. drop missing values and invalid data of depresssive sym and cardiometabolic markers


drop if Cwv1_depressive_symptoms== .
* (542 observations deleted)

drop if Cwv1_crp== .
* (175 observations deleted)

* drop CRP > 100

drop if Cwv1_crp_level > 100 & Cwv1_crp_level < 300
* (13 observations deleted)

drop if Cwv1_hdl_cholesterol== .
* (2 observations deleted)

drop if Cwv1_obesity_waist== .
```

```
* (1250 observations deleted)

drop if Cwv1_systolic_bp== .
* (81 observations deleted)

* drop sbp > 900

drop if Cwv1_systolic_mean > 900 & Cwv1_systolic_mean < 999
* (14 observations deleted)

drop if Cwv1_diastolic_bp== .
* (13 observations deleted)

drop if Cwv1_diabetes_report== .
* (90 observations deleted)

drop if Cwv1_HbA1c== .
* (70 observations deleted)



* 3. drop obs with no records on dementia at any wave from 2-4 follow-ups


search mdesc
search rmiss2
search mvpatterns

* see number of missing values vs non-missing in each variable
mdesc Cwv2_dementia_report Cwv3_dementia_report Cwv4_self_info_dementia



/* 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_wv2to4=rmiss2(Cwv2_dementia_report ///
Cwv3_dementia_report Cwv4_self_info_dementia)

tab nmisfollowup_dementia_wv2to4

* drop observations "nmisfollowup_dementia_wv2to4" > 2 (those with 3 missing data = no records at
any wave)
drop if nmisfollowup_dementia_wv2to4>2
*(331 observations deleted)


* ANALYTIC SAMPLE -> 8925







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

General characteristics of participnats stratified for study inclusion

General characteristics of participants stratified for dementia occurence

1. 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

2. one-way ANOVA for continuous var
    check box plot
```

```
335       using the command oneway
336       reporting mean, sd (summary tables) and p-value of F
337   */
338
339
340
341
342   * General characteristics of CHARLS participants at baseline
343
344   * Socio-demographics
345   sum C_age
346   ta C_sex
347   ta C_educ_new
348   ta C_maritalstatus_4cat
349   ta Cwv1_netwealth_quintiles
350   * Cardiometabolic risk factors
351   ta Cwv1_crp
352   ta Cwv1_hdl_cholesterol
353   ta Cwv1_obesity_waist
354   ta Cwv1_systolic_bp
355   ta Cwv1_diastolic_bp
356   ta Cwv1_diabetes_report
357   ta Cwv1_HbA1c
358   ta Cwv1_cardio2
359   * Lifestyle and health indicators
360   ta Cwv1_smoking_3cat
361   ta Cwv1_alcohol_status
362   ta Cwv1_physicalactivity
363   ta C_cvd_comorbidity
364   * Depressive symptoms
365   ta Cwv1_depressive_symptoms
366   * Memory score
367   sum Cwv1_memory_wordrecall
368
369
370
371   * General characteristics of CHARLS participants stratified for dementia occurence
372
373   * Socio-demographics
374   ttest C_age, by(Cwv2to4_dementia_event)
375   ta C_sex Cwv2to4_dementia_event, chi2 column row
376   ta C_educ_new Cwv2to4_dementia_event, chi2 column row
377   ta C_maritalstatus_4cat Cwv2to4_dementia_event, chi2 column row
378   ta Cwv1_netwealth_quintiles Cwv2to4_dementia_event, chi2 column row
379   * Cardiometabolic risk factors
380   ta Cwv1_crp Cwv2to4_dementia_event, chi2 column row
381   ta Cwv1_hdl_cholesterol Cwv2to4_dementia_event, chi2 column row
382   ta Cwv1_obesity_waist Cwv2to4_dementia_event, chi2 column row
383   ta Cwv1_systolic_bp Cwv2to4_dementia_event, chi2 column row
384   ta Cwv1_diastolic_bp Cwv2to4_dementia_event, chi2 column row
385   ta Cwv1_diabetes_report Cwv2to4_dementia_event, chi2 column row
386   ta Cwv1_HbA1c Cwv2to4_dementia_event, chi2 column row
387   ta Cwv1_cardio2 Cwv2to4_dementia_event, chi2 column row
388   * Lifestyle and health indicators
389   ta Cwv1_smoking_3cat Cwv2to4_dementia_event, chi2 column row
390   ta Cwv1_alcohol_status Cwv2to4_dementia_event, chi2 column row
391   ta Cwv1_physicalactivity Cwv2to4_dementia_event, chi2 column row
392   ta C_cvd_comorbidity Cwv2to4_dementia_event, chi2 column row
393   * Depressive symptoms
394   ta Cwv1_depressive_symptoms Cwv2to4_dementia_event, chi2 column row
395   * Memory score
396   ttest Cwv1_memory_wordrecall, by(Cwv2to4_dementia_event)
397   ta C_age_group Cwv2to4_dementia_event, chi2 column row
398
399
400
401
402
```

```
403
404
405   /*
406   ---- SURVIVAL ANALYSIS AT COMPLETE DATA ----
407
408   Tests of proportional-hazards assumption
409   Kaplan Meier survival curves
410   Person-time
411   Cox proportional regression - Hazard ratios - stcox
412   Postestimation tools for stcox
413   Test of Goodness of Fit
414
415   *** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
416   Cox PH regression in complete data
417   Cox PH regression model in imputed dataset - mi estimate
418
419
420   */
421
422
423
424   * check dataset variables of interest only
425
426   codebook C_time_of_event_dementia Cwv2to4_dementia_event ///
427   Cwv1_depressive_symptoms Cwv1_crp Cwv1_hdl_cholesterol Cwv1_obesity_waist Cwv1_systolic_bp ///
428   Cwv1_diastolic_bp Cwv1_diabetes_report Cwv1_HbA1c ///
429   Cwv1_cardio3 Cwv1_cardio4 ///
430   Cwv1_cardio_number_sum Cwv1_cardio_number ///
431   Cwv1_waist_depress_group Cwv1_diabet_depress_group ///
432   Cwv1_hba1c_depress_group Cwv1_hdl_depress_group ///
433   Cwv1_sbp_depress_group Cwv1_dbp_depress_group ///
434   Cwv1_crp_depress_group ///
435   Cwv1_ca3_depress_group Cwv1_ca4_depress_group ///
436   C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
437   Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status C_cvd_comorbidity,compact
438
439
440
441
442   * Declare Data to be Survival Data
443   * Time to event: C_time_of_event_dementia (months)
444   * Censoring: Cwv2to4_dementia_event (1=dementia, 0=censored)
445   * Command is stset TIMETOEVENT, failure(CENSORVARIABLE)
446
447
448   stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
449
450
451
452   *describe survival data using commnad stsum
453
454   stsum
455
456   stsum, by(Cwv1_ca2_depress_group)
457
458
459
460
461   * Kaplan Meier Curve estimation
462
463   sts list
464
465   sts list, by(Cwv1_ca2_depress_group)
466
467
468
469   * Kaplan Meier Curve Plot
470
```

```
471  * no frills plot
472
473  sts graph
474
475  * with frills
476
477  sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
478  title("Kaplan Meier Curve")
479
480
481  * With Greenwood CI limits
482
483  sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
484  title("Kaplan Meier Curve")
485
486
487
488  * Group Kaplan-Meier Curve Estimation
489  * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
490
491  sort Cwv1_ca2_depress_group
492
493  sts list, by(Cwv1_ca2_depress_group)
494
495  * graph with frills
496
497  sts graph, by(Cwv1_ca2_depress_group) xlabel(0(20)100) ylabel(0.80(.05)1) xtitle("Time in Months")
      ///
498  ytitle("Survival Prob") title("Kaplan Meier Curve")
499
500
501
502  * calculate person-time and incidence rates using command ststime
503
504  stptime,title(Person-years)
505
506  stptime, title(Person-years) per(1000)
507
508
509
510
511  /*
512
513  Repeat to find incident case per category
514
515  Cwv1_depressive_symptoms
516  Cwv1_crp
517  Cwv1_hdl_cholesterol
518  Cwv1_obesity_waist
519  Cwv1_systolic_bp
520  Cwv1_diastolic_bp
521  Cwv1_diabetes_report
522  Cwv1_HbA1c
523  Cwv1_cardio_number
524  Cwv1_cardio2
525
526  Cwv1_crp_depress_group
527  Cwv1_hdl_depress_group
528  Cwv1_waist_depress_group
529  Cwv1_sbp_depress_group
530  Cwv1_dbp_depress_group
531  Cwv1_diabet_depress_group
532  Cwv1_hba1c_depress_group
533  Cwv1_ca2_depress_group
534
535  */
536
537
```

```
538  ta Cwv1_ca2_depress_group
539
540
541  * calculate person-time  by category
542
543  stptime, by(Cwv1_ca2_depress_group)
544
545  stptime, by(Cwv1_ca2_depress_group) per(1000)
546
547
548
549  * mean and median of follow-up
550  sum C_time_of_event_dementia
551  sum C_time_of_event_dementia, detail
552
553
554
555
556  /* Log Rank Test of equality of survival distributions
557   (NULL: equality of survival distributions among groups)
558   We will consider including the predictor if the test has a p-value of 0.2 – 0.25 or less.
559   If the predictor has a p-value greater than 0.25 in a univariate analysis
560   it is highly unlikely that it will contribute anything to a model which includes other
     predictors.
561   Command is sts test GROUPVAR
562  */
563
564
565  sts test Cwv1_cardio2, logrank
566
567  sts test Cwv1_ca2_depress_group, logrank
568
569  sts test C_age, logrank
570
571  sts test C_sex, logrank
572
573  sts test C_eduaction, logrank
574
575  sts test C_maritalstatus_4cat, logrank
576
577  sts test Cwv1_netwealth_quintiles, logrank
578
579  sts test Cwv1_smoking_3cat, logrank
580
581  sts test Cwv1_physicalactivity, logrank
582
583  sts test Cwv1_alcohol_status, logrank
584
585  sts test C_cvd_comorbidity, logrank
586
587
588
589
590
591
592
593
594
595  /* Cox PH regression model
596
597  using the command stcox
598
599  --- Building the model ---
600
601  Model 1: unadjusted - single predictor of group
602  Model 2: model 1 + sociodemographics: age sex education marital status and wealth
603  Model 3: model 2 + lifestyle/health indicators: smoking, alcohol consumption, cvd comorbidity
604
```

```
*/


* Unadjusted model - model 1 - single predictor

stcox Cwv1_ca2_depress_group

* define design var by using i.(by group)

stcox i.Cwv1_ca2_depress_group


* Adjusted models - multivariable Cox model
* controlling for covariates

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

stcox i.Cwv1_ca2_depress_group C_age C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles

* model 3: model 2 + adjust for lifestyle/health indicators

stcox i.Cwv1_ca2_depress_group C_age C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity




* Coefficients instead of hazard ratios by specifing the option nohr

stcox i.Cwv1_ca2_depress_group, nohr


stcox i.Cwv1_ca2_depress_group C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity, nohr




* Multivariable model development
* Likelihood-ratio tests



*install eststo
findit eststo


* ---- rx controlling for age and sex -----*
quietly: stcox C_age i.C_sex
eststo modelagesex

quietly: stcox C_age i.C_sex i.Cwv1_ca2_depress_group
eststo modelagesex_4group

lrtest modelagesex modelagesex_4group



* ---- rx controlling for sociodemographics -----*
quietly: stcox C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
eststo modelsociodemo

quietly: stcox C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles i.Cwv1_ca2_depress_group
eststo modelsociodemo_4group
```

```
669
670  lrtest modelsociodemo modelsociodemo_4group
671
672
673  * ---- rx controlling for lifestyle/health indicators -----*
674  quietly: stcox i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
675  eststo modelcardiovascular
676
677  quietly: stcox i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity i.
     Cwv1_ca2_depress_group
678  eststo modelcardiovascular_4group
679
680  lrtest modelcardiovascular modelcardiovascular_4group
681
682
683  * side-by-side comparison of models
684
685
686  quietly: stcox i.Cwv1_ca2_depress_group
687  eststo model1
688
689  quietly: stcox  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles i.
     Cwv1_ca2_depress_group
690  eststo model2
691
692  quietly: stcox C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
693  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity i.Cwv1_ca2_depress_group
694  eststo model3
695
696
697
698
699
700  * Display Betas and Summary Statistics
701  estout model1 model2 model3, stats(n chi2 bic, star(chi2)) prehead("Betas")
702
703  /* Key Interpretattion
704  Chi2 = Value of LR test comparing the model fit (“full”) to intercept only (“reduced”)
705  bic = Schwarz’ Bayesian Information Criterion = It is a function of the log-likelihood.
706  Smaller values indicate a better fit.
707  */
708
709  * Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
710  estout model1 model2 model3, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")
711
712
713
714
715  * Postestimation tools for stcox
716
717  * Test of proportional hazards
718
719  estat phtest, detail
720
721
722  /* Proportionality Assumption - method 1
723  We will check proportionality by including time-dependent covariates in the model
724  by using the tvc and the texp options in the stcox command.
725  Time dependent covariates are interactions of the predictors and time.
726  In this analysis we choose to use the interactions with log(time)
727  because this is the most common function of time used in time-dependent covariates
728  but any function of time could be used.
729  If a time-dependent covariate is significant this indicates
730  a violation of the proportionality assumption for that specific predictor.
731  The conclusion is that all of the time-dependent variables are not significant
732  either collectively or individually thus supporting the assumption of proportional hazard.
733  */
734
```

```
735
736
737   stcox i.Cwv1_ca2_depress_group C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.
      Cwv1_netwealth_quintiles ///
738   i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity, nohr ///
739   tvc(Cwv1_ca2_depress_group C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles
      ///
740   Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity) texp(ln(C_time_of_event_dementia))
741
742
743
744   /* Proportionality Assumption - method 2
745   by using the Schoenfeld and scaled Schoenfeld residuals
746   In the stphtest command we test the proportionality of the model as a whole
747   and by using the detail option we get a test of proportionality for each predictor.
748   By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
749   If the tests in the table are not significance (p-values over 0.05)
750   then we can not reject proportionality and we assume
751   that we do not have a violation of the proportional assumption.
752   The stphplot command uses log-log plots to test proportionality
753   and if the lines  in these plots are parallel then we have further indication
754   that the predictors do not violate the proportionality assumption.
755   */
756
757   quietly stcox Cwv1_ca2_depress_group C_age C_sex C_eduaction C_maritalstatus_4cat
      Cwv1_netwealth_quintiles ///
758   Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity, schoenfeld(sch*) scaledsch(sca*)
759   stphtest, detail
760   stphtest, plot(Cwv1_ca2_depress_group) msym(oh)
761   stphtest, plot(C_age) msym(oh)
762   stphtest, plot(C_sex) msym(oh)
763   stphtest, plot(C_eduaction) msym(oh)
764   stphtest, plot(C_maritalstatus_4cat) msym(oh)
765   stphtest, plot(Cwv1_netwealth_quintiles) msym(oh)
766   stphtest, plot(C_cvd_comorbidity) msym(oh)
767   stphtest, plot(Cwv1_smoking_3cat) msym(oh)
768   stphtest, plot(Cwv1_alcohol_status) msym(oh)
769
770
771
772
773
774
775   stphplot, by(Cwv1_ca2_depress_group) plot1(msym(oh)) plot2(msym(th))
776   stphplot, by(C_age) plot1(msym(oh)) plot2(msym(th))
777   stphplot, by(C_sex) plot1(msym(oh)) plot2(msym(th))
778   stphplot, by(C_eduaction) plot1(msym(oh)) plot2(msym(th))
779   stphplot, by(C_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
780   stphplot, by(Cwv1_netwealth_quintiles) plot1(msym(oh)) plot2(msym(th))
781   stphplot, by(C_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
782   stphplot, by(Cwv1_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
783   stphplot, by(Cwv1_alcohol_status) plot1(msym(oh)) plot2(msym(th))
784
785
786
787   * Assessment of PH Assumption: adjust for age and sex
788   stphplot, by(Cwv1_ca2_depress_group) adjust(C_age C_sex) nolntime plot1opts(symbol(none) color(
      black) lpattern(dash)) ///
789   plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
790   title("Assessment of PH Assumption") subtitle(" Predictor is Cwv1_ca4_depress_group") xtitle(
      "months")
791
792
793
794   * Assessment of PH Assumption: adjust for model 2
795   stphplot, by(Cwv1_ca2_depress_group) adjust(C_age C_sex C_eduaction C_maritalstatus_4cat
      Cwv1_netwealth_quintiles) ///
796   nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
```

```
797   plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
798   title("Assessment of PH Assumption") subtitle(" Predictor is Cwv1_ca4_depress_group") xtitle(
      "months")
799
800
801
802   * Assessment of PH Assumption: adjust for model 3
803   stphplot, by(Cwv1_ca2_depress_group) adjust(C_age C_sex C_eduaction C_maritalstatus_4cat
      Cwv1_netwealth_quintiles ///
804   Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity) ///
805   nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
806   plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
807   title("Assessment of PH Assumption") subtitle(" Predictor is Cwv1_ca4_depress_group") xtitle(
      "months")
808
809
810
811
812
813   /* Test of overall goodness of fit
814   Goodness of fit of the final model
815   2 methods:
816    - by using the commnad stcoxgof (good fit = non sig p-value)
817    - by using the Cox-Snell residuals
818       to create the Nelson-Aalen cumulative hazard function
819       If the hazard function follows the 45 degree line then we know that it approximately
820       has an exponential distribution with a hazard rate of one and that the model fits the data
      well.
821       If the model fits the data, the plot of the cumulative hazard versus cs
822       should approximate a straight line with slope 1.
823   */
824
825
826   * by using the commnad stcoxgof
827
828   * install stcoxgof
829   findit stcoxgof
830
831
832   stcox Cwv1_ca2_depress_group C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles
       ///
833   Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity, mgale(mgale)
834
835
836   stcoxgof
837
838
839
840   * by using the Cox-Snell residuals
841
842   quietly stcox Cwv1_ca2_depress_group C_age C_sex C_eduaction C_maritalstatus_4cat
      Cwv1_netwealth_quintiles ///
843   Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity
844   predict cs, csnell
845
846   * or
847
848   quietly stcox Cwv1_ca2_depress_group
849   predict cs, csnell
850
851
852   stset cs, failure(Cwv2to4_dementia_event)
853   sts generate km = s
854   generate H = -ln(km)
855   line H cs cs, sort ytitle("") clstyle(. refline)
856
857
858
```

```
859
860
861   /* Cox PH regression model for independent depressive symptoms and CA exposure variable
862
863   Cwv1_depressive_symptoms
864   Cwv1_crp
865   Cwv1_hdl_cholesterol
866   Cwv1_obesity_waist
867   Cwv1_systolic_bp
868   Cwv1_diastolic_bp
869   Cwv1_diabetes_report
870   Cwv1_HbA1c
871   Cwv1_cardio_number
872   Cwv1_cardio2
873   Cwv1_cardio3
874   Cwv1_cardio4
875
876   */
877
878
879
880   stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
881
882
883
884
885
886   * Unadjusted model 1
887
888
889   stcox i.Cwv1_depressive_symptoms
890
891
892   * Adjusted models - multivariable Cox model
893   * controlling for covariates
894
895   * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
896
897   stcox i.Cwv1_depressive_symptoms C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.
      Cwv1_netwealth_quintiles
898
899   * model 3: model 2 + adjust for lifestyle / health indicators
900
901   stcox i.Cwv1_depressive_symptoms C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.
      Cwv1_netwealth_quintiles ///
902   i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
903
904   * repeat for each independent variable from the list above
905
906
907
908
909   ** ---------------------- **
910
911
912
913
914
915   /* MULTIPLE IMPUTATION (MI)
916
917   To handle with missing values of baseline and time 3 covariates
918
919
920   useful sources for MI and MICE:
921
922   https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/
923   https://www.stata.com/manuals/mi.pdf - see page 139
924   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 C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
C_cvd_comorbidity Cwv1_memory_wordrecall



*** physical activity showed > 50% missing values and so cannot be used or imputed as covariates



* examining missing data patterns

mi set wide

mi misstable summarize C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
C_cvd_comorbidity
```

```
993
994
995
996   mi misstable patterns C_age C_sex C_eduaction C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
997   Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
998   C_cvd_comorbidity
999
1000
1001  /*
1002   identifying potential auxiliary var
1003  * Auxiliary variables are either correlated with a missing variable(s)
1004  (the recommendation is r > 0.4) or are believed to be associated with missingness
1005  - a priori knowledge of var that would make good auxiliary var
1006  - identify potential candidates by examining associations between missing var and other var in
      the dataset
1007      running correlation using the command: pwcorr v1 v2 v3, obs
1008      the recommnedation for good correlation is r > 0.4
1009
1010
1011  Missing var to be imputed:
1012
1013      Cwv1_netwealth_quintiles
1014      Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity
1015
1016
1017
1018
1019  Potential auxiliary var:
1020  DV:  Cwv2to4_dementia_event
1021  IV:  Cwv1_depressive_symptoms Cwv1_crp Cwv1_hdl_cholesterol Cwv1_obesity_waist
1022  Cwv1_systolic_bp Cwv1_diastolic_bp Cwv1_diabetes_report Cwv1_HbA1c
1023  other var: C_age C_sex C_eduaction C_maritalstatus_4cat
1024
1025  */
1026
1027
1028  * correlation
1029
1030  pwcorr Cwv1_netwealth_quintiles ///
1031      Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
1032      Cwv2to4_dementia_event Cwv1_depressive_symptoms Cwv1_crp ///
1033      Cwv1_hdl_cholesterol Cwv1_obesity_waist ///
1034      Cwv1_systolic_bp Cwv1_diastolic_bp Cwv1_diabetes_report Cwv1_HbA1c ///
1035      C_age C_sex C_eduaction C_maritalstatus_4cat, obs
1036
1037
1038  /* The correlation showed that all the following var are good auxiliary:
1039  Cwv2to4_dementia_event Cwv1_depressive_symptoms Cwv1_hdl_cholesterol Cwv1_obesity_waist
1040  Cwv1_diabetes_report C_age C_sex C_eduaction
1041  * A good auxiliary does not have to be correlated with every variable to be useful
1042  * And it's not problematic if it has missing info of it's own
1043  */
1044
1045
1046
1047  /*
1048  MI by chained equations (MICE)
1049      see: https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/
1050
1051  MICE is known as the fully conditional specification or sequential generalized regression
1052  does not assume a joint MVN distribution
1053  but instead uses a separate conditional distribution for each imputed variable.
1054
1055  The multivariate normal (MVN) model - mi imputed mvn -
1056  assumes multivariate normality of all var
1057
1058  The multivariate imputation by chained equations (MICE) - mi imputed chained -
1059  offers flexibility in how each var is modeled
```

```
1060
1061   mi impute chained allows to specify models for a
1062   variety of variable types, including
1063   continuous, binary, ordinal, nominal, truncated, and count variables
1064
1065
1066   The MICE distributions available in Stata are:
1067   binary, ordered and multinomial logistic regression for categorical variables,
1068   linear regression and predictive mean matching (PMM)* for continuous variables,
1069   and Poisson and negative binomial regression for count variables.
1070
1071
1072
1073   IMPUTATION PHASES
1074
1075   1. mi set wide
1076       style to store imputations
1077
1078   2. mi register imputed
1079       identifies which variables in the imputation model have missing information.
1080
1081   3. mi register regular (! optional)
1082       which variables will not be imputed
1083
1084   4. mi impute chained
1085       where the user specifies the imputation model to be used
1086       and the number of imputed datasets to be created.
1087       Example:
1088           mi impute chained (regress) bmi age (logit) female ///
1089           (mlogit) race = bpdiast i.region, add(20)
1090
1091   5. mi estimate
1092       is used as a prefix to the standard regress command.
1093       This executes the specified estimation model within each of the 20 imputed datasets
1094       to obtain 20 sets of coefficients and standard errors.
1095       Stata then combines these estimates to obtain one set of inferential statistics.
1096       In the output from mi estimate you will see some metrics: Imputation Diagnostics
1097       information for RVI (Relative Increase in Variance),
1098       FMI (Fraction of Missing Information),
1099       DF (Degrees of Freedom) ,
1100       RE (Relative Efficiency),
1101       and the between imputation and the within imputation variance estimates
1102       to examine how the standard errors (SEs) are calculated.
1103
1104
1105
1106   ----------------
1107
1108   SELECTING MY IMPUTATION MODEL
1109
1110   - MICE -> mi impute chained
1111
1112   - var to be imputed:
1113
1114       linear regression for continuous var (regress) -> none
1115
1116       logistic for the binary var (logit) ->
1117       C_cvd_comorbidity
1118
1119       multinomial logistic for our nominal categorical var (mlogit) ->
1120       Cwv1_netwealth_quintiles
1121       Cwv1_smoking_3cat Cwv1_alcohol_status
1122
1123
1124
1125   - auxiliary var:
1126
1127       DV -> Cwv2to4_dementia_event
```

```
1128        IV -> Cwv1_depressive_symptoms Cwv1_hdl_cholesterol Cwv1_obesity_waist
1129    Cwv1_diabetes_report
1130        other covariates -> C_age C_sex C_eduaction
1131
1132
1133
1134    - imputation numbers (m) -> 10
1135
1136        White et al. (2010) recommendation: use the rule that m should equal the percentage of
        incomplete cases
1137
1138
1139    - rseed (53421) for reproducability reasons
1140
1141
1142    - (! OPTIONAL) advance impute options -> force
1143
1144        proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
        are encountered
1145
1146    - impute options -> savetrace (trace1)
1147
1148        specifies Stata to save the means and standard deviations of imputed values from each
        iteration to a Stata dataset named “trace1
1149    */
1150
1151
1152
1153    mi set wide
1154
1155
1156    mi register imputed Cwv1_netwealth_quintiles ///
1157        Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity
1158
1159
1160
1161    mi impute chained (logit) C_cvd_comorbidity ///
1162    (mlogit) Cwv1_netwealth_quintiles Cwv1_smoking_3cat Cwv1_alcohol_status = Cwv2to4_dementia_event
        Cwv1_depressive_symptoms ///
1163    Cwv1_hdl_cholesterol Cwv1_obesity_waist Cwv1_diabetes_report ///
1164    C_age C_sex C_eduaction, add(10) rseed(53421) savetrace(trace1)
1165
1166
1167    * save imputed data
1168
1169    * plot imputations
1170
1171
1172    *it will open a file named trace1
1173    use trace1, clear
1174
1175    describe
1176
1177
1178    reshape wide *mean *sd, i(iter) j(m)
1179
1180    tsset iter
1181
1182
1183
1184
1185    /*
1186    The trace plot below graphs the predicted means value produced during the first imputation chain.
1187    As before, the expectations is that the values would vary randomly to incorporate variation into
        the predicted values for read.
1188    */
1189
1190    tsline Cwv1_netwealth_quintiles_mean1, name(mice1,replace)legend(off) ytitle("Mean of wealth")
```

```
tsline Cwv1_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
tsline Cwv1_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol")
tsline C_cvd_comorbidity_mean1, name(mice1,replace)legend(off) ytitle("Mean of cvd")


/*

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 C_cvd_comorbidity_mean*, name(mice1,replace)legend(off) ytitle("Mean of cvd")
tsline C_cvd_comorbidity_sd*, name(mice2, replace) legend(off) ytitle("SD of cvd")
graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)

* repeat for each imputed var









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


* Declare Data to be Survival Data by using mi

mi stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)


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

/*
Independent risk factors

Cwv1_depressive_symptoms
Cwv1_crp
Cwv1_hdl_cholesterol
Cwv1_obesity_waist
Cwv1_systolic_bp
Cwv1_diastolic_bp
Cwv1_diabetes_report
Cwv1_HbA1c
Cwv1_cardio_number
Cwv1_cardio2

*/



* Depressive symptoms

* Unadjusted model - model 1 - single predictor

* Model 1 (default coefficents)
mi estimate: stcox Cwv1_depressive_symptoms

* Model 1: define design var by using i.
mi estimate: stcox i.Cwv1_depressive_symptoms


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

```

```
mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_depressive_symptoms

* Adjusted models - multivariable Cox model
* controlling for covariates

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

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_depressive_symptoms ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles

* model 3: model 2 + adjust for lifestyle/ health indicators

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_depressive_symptoms ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity



* repeat for each independent variable from the list above




/*
Combined effects Cox regression models

Cwv1_crp_depress_group
Cwv1_hdl_depress_group
Cwv1_waist_depress_group
Cwv1_sbp_depress_group
Cwv1_dbp_depress_group
Cwv1_diabet_depress_group
Cwv1_hba1c_depress_group
Cwv1_ca2_depress_group

*/



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

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group

* Adjusted models - multivariable Cox model
* controlling for covariates

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

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles

* model 3: model 2 + adjust for lifestyle/ health indicators

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity



* repeat for each variable from the list above
```

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



/*


*** SENSITIVITY ANALYSES ***

1) interarction effect of gender and age_group

2) survival analysis stratified by age
two age groups: young old <70 and old old >=70

3) depressive symptoms as continuous variable
and >= 3 and >=4 cardiometabolic multimorbidity

4) exclude participants with cvd

5) Complete data







Repeat on all independent and combined variables

Cwv1_depressive_symptoms
Cwv1_crp
Cwv1_hdl_cholesterol
Cwv1_obesity_waist
Cwv1_systolic_bp
Cwv1_diastolic_bp
Cwv1_diabetes_report
Cwv1_HbA1c
Cwv1_cardio_number
Cwv1_cardio2

Cwv1_crp_depress_group
Cwv1_hdl_depress_group
Cwv1_waist_depress_group
Cwv1_sbp_depress_group
Cwv1_dbp_depress_group
Cwv1_diabet_depress_group
Cwv1_hba1c_depress_group
Cwv1_ca2_depress_group




*/


* 1) Interaction effect

* sex*risk factor


mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio2 i.C_sex#i.Cwv1_cardio2

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio2 ///
```

```
1394  C_age i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1395  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1396   i.C_sex#i.Cwv1_cardio2
1397
1398
1399
1400  * age*risk factor
1401
1402  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio2 c.C_age#i.Cwv1_cardio2
1403
1404  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio2 ///
1405  C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1406  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1407  c.C_age#i.Cwv1_cardio2
1408
1409
1410
1411
1412  /* 2) Survival analysis stratified by age
1413
1414  generate age group variable
1415  Age groups: 1) young old (< 70) 2) old old (>= 70)
1416
1417  Kaplan Meier curves
1418  Cox regression models in imputed data
1419
1420  young old <70
1421  if C_age_group==1
1422
1423  old old >70
1424  if C_age_group==2
1425
1426
1427  */
1428
1429
1430  gen C_age_group=1 if C_age < 70
1431  replace C_age_group=2 if C_age >=70 & ///
1432  !missing(C_age)
1433
1434  label var C_age_group "Age groups <70 young-old / 70 old-old"
1435  lab def age_group 1 "young old <70" 2 "old old >70"
1436  lab val C_age_group age_group
1437
1438  tab C_age_group
1439
1440
1441
1442
1443  * COX PH REGRESSION MODEL IN IMPUTED DATASET
1444
1445
1446  * Declare Data to be Survival Data by using mi
1447
1448  mi stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
1449
1450
1451
1452  * YOUNG OLD <70 Cox regression models
1453
1454  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1455
1456  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group if C_age_group==1
1457
1458
1459  * Model 3: model 2 + adjust for lifestyle/health indicators
1460
1461  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group ///
```

```
1462  i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1463  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity if C_age_group==1
1464
1465
1466
1467  * OLD OLD >70 Cox regression models
1468
1469  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1470
1471  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group if C_age_group==2
1472
1473  * Model 3: model 2 + adjust for lifestyle/health indicators
1474
1475  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca2_depress_group ///
1476  i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1477  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity if C_age_group==2
1478
1479
1480
1481
1482
1483
1484  * 3) On depressive symptoms continuous variable
1485
1486
1487
1488  * COX PH REGRESSION MODEL IN COMPLETE DATASET
1489
1490  * Declare Data to be Survival Data by using mi
1491
1492  mi stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
1493
1494
1495
1496  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1497
1498  mi estimate, eform("Haz. Ratio"): stcox Cwv1_cesd_score
1499
1500  * Adjusted models - multivariable Cox model
1501  * controlling for covariates
1502
1503  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1504
1505  mi estimate, eform("Haz. Ratio"): stcox Cwv1_cesd_score ///
1506  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1507
1508  * model 3: model 2 + adjust for lifestyle/ health indicators
1509
1510  mi estimate, eform("Haz. Ratio"): stcox Cwv1_cesd_score ///
1511  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1512  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1513
1514
1515
1516
1517
1518
1519  * Cardiometabolic multimorbidty >= 3
1520
1521
1522  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1523
1524  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio3
1525
1526  * Adjusted models - multivariable Cox model
1527  * controlling for covariates
1528
1529  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
```

```
1530
1531  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio3 ///
1532  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1533
1534  * model 3: model 2 + adjust for lifestyle/ health indicators
1535
1536  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio3 ///
1537  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1538  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1539
1540
1541
1542
1543  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1544
1545  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca3_depress_group
1546
1547  * Adjusted models - multivariable Cox model
1548  * controlling for covariates
1549
1550  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1551
1552  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca3_depress_group ///
1553  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1554
1555  * model 3: model 2 + adjust for lifestyle/ health indicators
1556
1557  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca3_depress_group ///
1558  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1559  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1560
1561
1562
1563
1564  * Cardiometabolic multimorbidty >=4
1565
1566
1567
1568  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1569
1570  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio4
1571
1572  * Adjusted models - multivariable Cox model
1573  * controlling for covariates
1574
1575  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1576
1577  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio4 ///
1578  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1579
1580  * model 3: model 2 + adjust for lifestyle/ health indicators
1581
1582  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_cardio4 ///
1583  C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1584  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1585
1586
1587
1588
1589  * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1590
1591  mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca4_depress_group
1592
1593  * Adjusted models - multivariable Cox model
1594  * controlling for covariates
1595
1596  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1597
```

```
mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca4_depress_group ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles

* model 3: model 2 + adjust for lifestyle/ health indicators

mi estimate, eform("Haz. Ratio"): stcox i.Cwv1_ca4_depress_group ///
C_age i.C_sex i.C_eduaction i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity









/*

4) exclude participants with cvd

use the command if C_cvd_comorbidity==0

*/


* COX PH REGRESSION MODEL IN COMPLETE DATASET


* Declare Data to be Survival Data by using mi

stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)



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

stcox i.Cwv1_ca2_depress_group if C_cvd_comorbidity==0


* Adjusted models - multivariable Cox model
* controlling for covariates


* model 3: model 2 + adjust for lifestyle/health indicators

stcox i.Cwv1_ca2_depress_group C_age C_sex i.C_eduaction i.C_maritalstatus_4cat i.
Cwv1_netwealth_quintiles ///
i.Cwv1_smoking_3cat i.Cwv1_alcohol_status if C_cvd_comorbidity==0




* 4) On complete cases (see above)







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