

```

1  /*
2
3  PHD PROJECT: The role of depressive symptoms and cardiometabolic risk factors in the prediction
4  of dementia: a cross-country comparison in England, the United States and China
5
6  STUDY 4: Clustering of cardiometabolic risk factors and dementia incidence
7
8  Method of analysis:
9  Latent Class Analysis (LCA)
10
11 DATASET: HRS
12 baseline: wave 8 (2004) follow-up waves 9-14 (2008-2018)
13
14 TIMELINE
15
16 LATENT CLASSES OF CARDIOMETABOLIC MARKERS: WV8 (BASELINE)
17 DEMENTIA INCIDENCE: W9 - WV14 (6 TIME POINTS)
18 COVARIATES ADJUSTMENT FOR HR MODELS: WV8
19
20 */
21
22
23 * KEEP NECESSARY VARIABLES
24
25
26 keep HHID PN RAHHIDPN HHIDPN ///
27 H_sex H_age H_education_yrs H_education H_maritalstatus_4cat H_wealthquintiles ///
28 H_ethnicity H_hispanic_ethnicity ///
29 H_smoking_2cat H_smoking_3cat H_physicalactivity H_alcohol_freq H_alcohol_status ///
30 H_cvd_comorbidity Hwv8_cognition Hwv8_memory Hwv8_loneliness_quintiles ///
31 Hwv8_cesd_sumscore Hwv8_depressive_symptoms ///
32 Hwv9_cesd_sumscore Hwv9_depressive_symptoms ///
33 Hwv10_cesd_sumscore Hwv10_depressive_symptoms ///
34 Hwv11_cesd_sumscore Hwv11_depressive_symptoms ///
35 Hwv12_cesd_sumscore Hwv12_depressive_symptoms ///
36 Hwv13_cesd_sumscore Hwv13_depressive_symptoms ///
37 Hwv14_cesd_sumscore Hwv14_depressive_symptoms ///
38 Hwv8_crp_level Hwv8_crp Hwv8_hdl_level Hwv8_male_hdl Hwv8_female_hdl ///
39 Hwv8_meds_hdl Hwv8_hdl_sum Hwv8_hdl ///
40 Hwv8_waist Hwv8_malewaist_ao Hwv8_femalewaist_ao Hwv8_obesity_waist_sum Hwv8_obesity_waist ///
41 Hwv8_bmi_score Hwv8_obesity_bmi Hwv8_waist_bmi_sum Hwv8_obesity ///
42 Hwv8_systolic_mean Hwv8_diastolic_mean Hwv8_systolic_bp Hwv8_diastolic_bp ///
43 Hwv8_meds_bp Hwv8_bp_before Hwv8_bp_report Hwv8_bpevr ///
44 Hwv8_bp_reportevr_sum Hwv8_bp_reportevr Hwv8_bp_sum Hwv8_bp ///
45 Hwv8_diabetes_before Hwv8_diabetes_report Hwv8_diabetes_sevr ///
46 Hwv8_diabetes_reportevr_sum Hwv8_diabetes_reportevr ///
47 Hwv8_meds_diabetes Hwv8_insulin_diabetes Hwv8_diabetes_anymeds_sum Hwv8_diabetes_anymeds ///
48 Hwv8_HbA1c_level Hwv8_HbA1c Hwv8_diabetes_HbA1c_sum Hwv8_glycemia ///
49 Hwv8_memory_report Hwv9_memory_report Hwv10_anydementia_report ///
50 Hwv11_anydementia_report Hwv12_anydementia_report Hwv13_anydementia_report
51 Hwv14_anydementia_report ///
52 Hwv8_interview_date Hwv9_interview_date Hwv10_interview_date ///
53 Hwv11_interview_date Hwv12_interview_date Hwv13_interview_date Hwv14_interview_date ///
54 Hwv9to14_dementia_sum Hwv9to14_dementia_event ///
55 Hwv9to14_newdementia_or_lastinte Hwv9to14_dementia_free_date H_time_dementia_months ///
56 H_time_dementia_midpoint H_time_dementia_midpoint_final H_time_of_event_dementia
57
58
59
60
61
62
63
64
65
66 /* Latent class analysis - LCA of cardiometabolic risk factors for dementia

```

```

67
68 Useful links:
69 https://www.stata.com/meeting/uk18/slides/uk18\_MacDonald.pdf
70
71 https://www.stata.com/meeting/mexico18/slides/5\_Mexico18\_Canette.pdf
72
73 https://www.bgsu.edu/content/dam/BGSU/college-of-arts-and-sciences/center-for-family-and-demographi
74 c-research/documents/Workshops/2020-latent-class-analysis.pdf
75
76 https://www.stata.com/features/overview/latent-class-analysis/
77
78 https://www.stata.com/manuals/semexample50g.pdf
79
80 https://www.stata.com/manuals/semexample51g.pdf
81
82 https://www.stata.com/manuals/semexample52g.pdf
83
84 https://www.ucl.ac.uk/population-health-sciences/sites/population\_health\_sciences/files/lca.pdf
85
86 https://www.stata.com/manuals/semgsemlclassoptions.pdf
87
88 https://www.stata.com/meeting/nordic-and-baltic17/slides/nordic-and-baltic17\_Pitblado.pdf
89
90 https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00920/full
91
92 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774074
93
94 https://www.statalist.org/forums/forum/general-stata-discussion/general/1412686-calculating-entropy
95 -for-lca-latent-class-analysis-in-stata-15
96
97 https://www.statalist.org/forums/forum/general-stata-discussion/general/1590174-how-to-calculate-en
98 tropy-for-lca-with-stata
99
100 https://www.statalist.org/forums/forum/general-stata-discussion/general/1390895-combine-marginsplot
101 -problem-with-plot-options
102
103
104 * gsem command to fit a latent class model
105
106 * gsem (var1 var2 var3 <-), logit lclass(C 3)
107
108 OR TRY
109
110 gsem (var1 var2 var3 <-), logit lclass(C 3) ///
111 startvalues(randompr, draws(20) seed(15) difficult) ///
112 emopts(iterate(30) difficult)
113
114
115 Binary variables of cardiometabolic markers measured at wave 2
116
117 CRP: Hwv8_crp
118
119 HDL cholesterol: Hwv8_hdl
120
121 Obesity by waist cir: Hwv8_obesity_waist
122
123 systolic Blood pressure: Hwv8_systolic_bp
124
125 diastolic Blood pressure: Hwv8_diastolic_bp
126
127 Diabetes: Hwv8_diabetes_reportevr
128
129 HbA1c: Hwv8_HbA1c
130

```

```

131
132  */
133
134
135
136
137
138
139  *** Descriptive stats of cardiometabolic markers
140
141  tabulate Hwv8_crp
142  summarize Hwv8_crp
143
144  misstable summarize Hwv8_crp
145  misstable patterns Hwv8_crp
146
147  tabulate Hwv8_hdl
148  summarize Hwv8_hdl
149
150  misstable summarize Hwv8_hdl
151  misstable patterns Hwv8_hdl
152
153  tabulate Hwv8_obesity_waist
154  summarize Hwv8_obesity_waist
155
156  misstable summarize Hwv8_obesity_waist
157  misstable patterns Hwv8_obesity_waist
158
159  tabulate Hwv8_systolic_bp
160  summarize Hwv8_systolic_bp
161
162  misstable summarize Hwv8_systolic_bp
163  misstable patterns Hwv8_systolic_bp
164
165
166  tabulate Hwv8_diastolic_bp
167  summarize Hwv8_diastolic_bp
168
169  misstable summarize Hwv8_diastolic_bp
170  misstable patterns Hwv8_diastolic_bp
171
172
173  tabulate Hwv8_diabetes_reportevr
174  summarize Hwv8_diabetes_reportevr
175
176  misstable summarize Hwv8_diabetes_reportevr
177  misstable patterns Hwv8_diabetes_reportevr
178
179
180  tabulate Hwv8_HbA1c
181  summarize Hwv8_HbA1c
182
183  misstable summarize Hwv8_HbA1c
184  misstable patterns Hwv8_HbA1c
185
186
187  tabulate Hwv8_memory_report
188  summarize Hwv8_memory_report
189
190  misstable summarize Hwv8_memory_report
191  misstable patterns Hwv8_memory_report
192
193
194
195
196
197
198

```

```

199 *** CLEANING DATA
200
201 * 1. drop dementia cases at baseline
202
203 * drop dementia wave 8 missing data
204 drop if Hwv8_memory_report==1
205 * (226 observations deleted)
206
207 drop if Hwv8_memory_report== .
208 * (0 observations deleted)
209
210
211 * 2. drop missing values of cardiometabolic markers
212
213 drop if Hwv8_crp== .
214 * (509 observations deleted)
215
216 drop if Hwv8_hdl== .
217 * (227 observations deleted)
218
219 drop if Hwv8_obesity_waist== .
220 * (193 observations deleted)
221
222 drop if Hwv8_systolic_bp== .
223 * (102 observations deleted)
224
225 drop if Hwv8_diastolic_bp== .
226 * (29 observations deleted)
227
228 drop if Hwv8_diabetes_reportevr== .
229 * (4 observations deleted)
230
231 drop if Hwv8_HbA1c== .
232 * (76 observations deleted)
233
234
235
236 * 3. drop obs with no records on dementia at any wave from 9-14 follow-ups
237
238
239 search mdesc
240 search rmiss2
241 search mvpatterns
242
243 * see number of missing values vs non-missing in each variable
244 mdesc Hwv9_memory_report Hwv10_anydementia_report Hwv11_anydementia_report ///
245 Hwv12_anydementia_report Hwv13_anydementia_report Hwv14_anydementia_report
246
247
248
249 /* number of missing values per observation
250 * the code below creates a variable called nmisfollowup that gives the number of missing values
251 for each observation in the variables of interest */
252 egen nmisfollowup_dementia_wv9to14=rmiss2(Hwv9_memory_report ///
253 Hwv10_anydementia_report Hwv11_anydementia_report ///
254 Hwv12_anydementia_report Hwv13_anydementia_report Hwv14_anydementia_report)
255
256 tab nmisfollowup_dementia_wv9to14
257
258 * drop observations "nmisfollowup_dementia_wv9to14" > 5 (those with 6 missing data = no records
at any wave)
259 drop if nmisfollowup_dementia_wv9to14>5
260 *(257 observations deleted)
261
262
263 * FINAL SAMPLE -> 5112
264
265

```

```

266
267
268
269 /* Latent Class analysis - gsem
270
271 7 variables: Hwv8_crp Hwv8_hdl Hwv8_obesity_waist
272   Hwv8_systolic_bp Hwv8_diastolic_bp Hwv8_diabetes_reportevr Hwv8_HbA1c
273
274 */
275
276
277
278 * change names to start with lowercase (STATA assumes variables starting with a capital letter
are cont latent variables)
279
280 rename Hwv8_crp crp_lca
281 rename Hwv8_hdl hdl_lca
282 rename Hwv8_obesity_waist obesity_lca
283 rename Hwv8_systolic_bp systolic_lca
284 rename Hwv8_diastolic_bp diastolic_lca
285 rename Hwv8_diabetes_reportevr diabetes_lca
286 rename Hwv8_HbA1c hba1c_lca
287
288
289
290 * Correlation matrix of the CM variables
291
292 corr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
293
294 pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
295
296
297 * to create quality table in word - asdoc -
298 * https://www.youtube.com/watch?v=XHBL6PHf0zs&ab\_channel=StataProfessor
299
300 help asdoc
301
302 asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
303
304
305 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)
306
307
308
309 * LCA models
310
311
312 * one-class model
313
314 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
315 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 1)
316
317 estimates store oneclass_cm
318
319 * two-class model
320
321 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
322 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 2)
323
324
325 estimates store twoclass_cm
326
327
328 * three-class model
329
330 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///

```

```

331 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
332
333 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
334 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
335 startvalues(randompr, draws(20) seed(15) difficult) ///
336 emopts(iterate(30) difficult)
337
338
339 estimates store threeclass_cm
340
341
342 * four-class model
343
344 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
345 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 4)
346
347
348 estimates store fourclass_cm
349
350
351
352 * five-class model
353
354 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
355 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 5)
356
357
358
359
360
361 /*
362
363 ** Evaluating Fit to choose the number of classes **
364
365 1. a priori theory
366 2. Information Statistics
367     AIC, BIC, adjusted BIC
368 3. Chi-Square goodness of fit
369 4. Entropy
370
371 Others but not used here:
372 Lo-Mendell-Rubin (LMR)
373     Not recommended (designed for normal Y)
374 Bootstrapped Likelihood Ratio Test
375
376 */
377
378
379 * AIC and BIC to determine which of these models fits best
380
381 estimates stats oneclass_cm twoclass_cm threeclass_cm fourclass_cm
382
383
384
385 * LCA postestimation
386 * Latent class marginal probabilities - lcprob -
387 * Latent class marginal means - lcmean -
388
389
390
391 estat lcprob
392
393 estat lcmean
394
395
396 * likelihood -ratio test (G2) to evaluate whether our model fits as well as the saturated model
397
398 estat lcgof

```

```

399
400 * if p value is sig means that we reject the null hypothesis and the model doesn't fit well
401 * 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.
402 * for a well-fitted model p value should be non-sig. A significant p-value indicates lack of
  model fit in absolute terms.
403
404 * 3-class model:  $p < 0.001$ 
405
406
407
408 * Entropy
409
410 quietly predict classpost*, classposteriorpr
411 gen sum_p_lnp = 0
412 forvalues k = 1/2 {
413     replace sum_p_lnp = sum_p_lnp + classpost`k'*ln(classpost`k')
414 }
415 summ sum_p_lnp, meanonly
416 scalar E = 1+`r(sum)'/(e(N)*ln(2))
417 drop classpost? sum_p_lnp
418 di E
419
420
421
422
423
424
425
426
427
428 /* We can use the predictions of the posterior probability of class membership to evaluate an
429 individual's probability of being in each class.
430
431 */
432
433 predict classpost1*, classposteriorpr
434 list in 1, abbrev(10)
435
436 /* We can determine the expected class for each individual based on whether the posterior
  probability
437 is greater than 0.5
438 */
439
440 generate expclass1 = 1 + (classpost11>0.5)
441 tabulate expclass1
442
443
444 generate expclass2 = 1 + (classpost12>0.5)
445 tabulate expclass2
446
447
448 generate expclass3 = 1 + (classpost13>0.5)
449 tabulate expclass3
450
451
452
453 /* We can determine expected classification for each individual in the dataset based on the
  predicted
454 posterior class probabilities.
455 */
456
457 predict cpost*, classposteriorpr
458 egen max = rowmax(cpost*)
459
460
461 * generate classes var
462

```

```

463 generate predclass = 1 if cpost1==max
464
465 replace predclass = 2 if cpost2==max
466
467 replace predclass = 3 if cpost3==max
468
469 tabulate predclass
470
471
472
473 *****function to print out summary stats
474 program summary_table_proclCla
475     preserve
476     *look at the average posterior probability
477     gen Mp = 0
478     foreach i of varlist cpost* {
479         replace Mp = `i' if `i' > Mp
480     }
481     sort predclass
482     *and the odds of correct classification
483     by predclass: gen countG = _N
484     by predclass: egen groupAPP = mean(Mp)
485     by predclass: gen counter = _n
486     gen n = groupAPP/(1 - groupAPP)
487     gen p = countG/ _N
488     gen d = p/(1-p)
489     gen occ = n/d
490     *Estimated proportion for each group
491     scalar c = 0
492     gen TotProb = 0
493     foreach i of varlist cpost* {
494         scalar c = c + 1
495         quietly summarize `i'
496         replace TotProb = r(sum)/ _N if predclass == c
497     }
498     gen d_pp = TotProb/(1 - TotProb)
499     gen occ_pp = n/d_pp
500     *This displays the group number [_traj_~p],
501     *the count per group (based on the max post prob), [countG]
502     *the average posterior probability for each group, [groupAPP]
503     *the odds of correct classification (based on the max post prob group assignment), [occ]
504     *the odds of correct classification (based on the weighted post. prob), [occ_pp]
505     *and the observed probability of groups versus the probability [p]
506     *based on the posterior probabilities [TotProb]
507     list predclass countG groupAPP occ occ_pp p TotProb if counter == 1
508     restore
509 end
510
511 summary_table_proclCla
512
513
514
515 /*
516 Class variable manipulation
517 */
518
519 * 3-class model: rename predclass to H_lca_group3
520
521 rename predclass H_lca_group3
522
523 * labelling variable of H_lca_group3
524
525 label var H_lca_group3 "Latent classes 3 groups of cardiometabolic markers"
526
527 * labelling values
528 lab def lca_group3 1 "Healthy with obesity" 2 "Obesity and Hypertension" 3 "Complex
cardiometabolic disorders"
529

```



```

530 * attach category labels to the variable through label value
531
532 lab val H_lca_group3 lca_group3
533
534 ta H_lca_group3
535
536
537
538 * margins and marginsplot
539
540 * use margins to calculate marginal predictions
541 * use marginsplot to graph marginal predictions.
542
543
544
545 margins, predict(classpr class(1)) ///
546             predict(classpr class(2)) ///
547             predict(classpr class(3))
548
549 marginsplot, xtitle ("") ytitle ("")
550             xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3")
551             title ("Predicted Latent Class Probabilities with 95% CI")
552
553
554 margins, predict(classpr class(1)) ///
555             predict(classpr class(2)) ///
556             predict(classpr class(3))
557 marginsplot, recast(bar) xtitle("") ytitle("") ///
558             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
559             title("Predicted Latent Class Probabilities with 95% CI")
560
561 margins, predict(outcome(hba1c_lca) class(1)) ///
562             predict(outcome(hba1c_lca) class(2)) ///
563             predict(outcome(hba1c_lca) class(3))
564 marginsplot, recast(bar) xtitle("") ytitle("") ///
565             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
566             title("Predicted Pr(HbA1c=1) with 95% CI")
567
568 * repeat with all CM var
569
570
571 * class 1
572
573 margins, predict(outcome(crp_lca) class(1)) ///
574             predict(outcome(hdl_lca) class(1)) ///
575             predict(outcome(obesity_lca) class(1)) ///
576             predict(outcome(systolic_lca) class(1)) ///
577             predict(outcome(diastolic_lca) class(1)) ///
578             predict(outcome(diabetes_lca) class(1)) ///
579             predict(outcome(hba1c_lca) class(1)) ///
580
581
582
583 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
584             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
585             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
586             ytitle("Predicted mean") ylabel(0(.20)1) name (class1)
587
588
589 * class 2
590
591 margins, predict(outcome(crp_lca) class(2)) ///
592             predict(outcome(hdl_lca) class(2)) ///
593             predict(outcome(obesity_lca) class(2)) ///
594             predict(outcome(systolic_lca) class(2)) ///
595             predict(outcome(diastolic_lca) class(2)) ///
596             predict(outcome(diabetes_lca) class(2)) ///
597             predict(outcome(hba1c_lca) class(2)) ///

```

```

598
599
600
601 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
602     xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
603     5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
604     ytitle("Predicted mean") ylabel(0(.20)1) name (class2)
605
606
607
608
609 * class 3
610
611 margins, predict(outcome(crp_lca) class(3)) ///
612     predict(outcome(hdl_lca) class(3)) ///
613     predict(outcome(obesity_lca) class(3)) ///
614     predict(outcome(systolic_lca) class(3)) ///
615     predict(outcome(diastolic_lca) class(3)) ///
616     predict(outcome(diabetes_lca) class(3)) ///
617     predict(outcome(hba1c_lca) class(3)) ///
618
619
620
621 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
622     xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
623     5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
624     ytitle("Predicted mean") ylabel(0(.20)1) name (class3)
625
626
627
628
629 graph combine class1 class2 class3, cols(3)
630
631
632
633
634
635 * ----- *
636
637
638
639
640
641
642
643
644 /* MULTIPLE IMPUTATION (MI)
645
646 To handle with missing values of covariates
647
648
649 useful sources for MI and MICE:
650
651 https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
652 https://www.stata.com/manuals/mi.pdf - see page 139
653 https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
654 https://www.youtube.com/watch?v=i6S0lq0mjuc&ab\_channel=StataCorpLLC
655 https://dss.princeton.edu/training/MIStata.pdf
656
657
658
659 Preparing to conduct MI
660 1. examine the number and proportion of missing values among the variables of interest
661     use the mdesc command
662 2. examine missing data patterns
663     use commands mi set and mi misstable patterns
664 3. identify potential auxiliary variables
665

```

```

666
667 Run MI using chained equations (MICE)
668 using the commands
669 1. how (in what style) to store the imputations
670 mi set wide
671 2. which variables will be imputed
672 mi register imputed
673 3. optionally, which variables will not be imputed
674 mi register regular
675 4. what imputation method is implemented to impute each of var - MICE
676 mi impute chained
677
678 */
679
680
681
682
683
684 /*
685
686 1. examining missing values
687 install packages:
688 * install mdesc
689 * install tabmiss
690 * insatll dm31
691 * insall mvpatterna
692
693 */
694 search mdesc
695 search rmiss2
696 search mvpatterns
697
698
699
700
701
702
703
704 * examining number of missing values vs non-missing in each variable
705
706 mdesc H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
707 H_smoking_3cat H_physicalactivity H_alcohol_status H_cvd_comorbidity ///
708 Hwv8_memory Hwv8_depressive_symptoms
709
710
711
712
713
714 * examining missing data patterns
715
716 mi set wide
717
718 mi misstable summarize H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
719 H_smoking_3cat H_physicalactivity H_alcohol_status H_cvd_comorbidity ///
720 Hwv8_memory Hwv8_depressive_symptoms
721
722
723 mi misstable patterns H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
724 H_smoking_3cat H_physicalactivity H_alcohol_status H_cvd_comorbidity ///
725 Hwv8_memory Hwv8_depressive_symptoms
726
727
728
729 /*
730 identifying potential auxiliary var
731 * Auxiliary variables are either correlated with a missing variable(s)
732 (the recommendation is  $r > 0.4$ ) or are believed to be associated with missingness
733 - a priori knowledge of var that would make good auxiliary var

```

```

734 - identify potential candidates by examining associations between missing var and other var in
the dataset
735     running correlation using the command: pwcorr v1 v2 v3, obs
736     the recommendation for good correlation is  $r > 0.4$ 
737
738
739 Missing var to be imputed:
740
741     H_smoking_3cat H_physicalactivity H_alcohol_status Hwv8_depressive_symptoms
742
743
744 Potential auxiliary var:
745 DV:  Hwv9to14_dementia_event
746 IV:  crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
747 other var:
748     H_age H_sex H_eduaction H_maritalstatus_4cat H_wealthquintiles H_cvd_comorbidity
749
750 */
751
752
753 * correlation
754
755 pwcorr H_smoking_3cat H_physicalactivity H_alcohol_status ///
756     Hwv8_depressive_symptoms ///
757     Hwv9to14_dementia_event ///
758     crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca ///
759     H_age H_sex H_eduaction H_maritalstatus_4cat H_wealthquintiles H_cvd_comorbidity, obs
760
761
762 /* The correlation showed that all the following var are good auxiliary:
763 Hwv9to14_dementia_event obesity_lca diabetes_lca H_age H_sex H_eduaction H_maritalstatus_4cat
H_cvd_comorbidity
764 A good auxiliary does not have to be correlated with every variable to be useful
765 And it's not problematic if it has missing info of it's own
766 */
767
768
769
770 /*
771 MI by chained equations (MICE)
772     see: https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
773
774 MICE is known as the fully conditional specification or sequential generalized regression
775 does not assume a joint MVN distribution
776 but instead uses a separate conditional distribution for each imputed variable.
777
778 The multivariate normal (MVN) model - mi imputed mvn -
779 assumes multivariate normality of all var
780
781 The multivariate imputation by chained equations (MICE) - mi imputed chained -
782 offers flexibility in how each var is modeled
783
784 mi impute chained allows to specify models for a
785 variety of variable types, including
786 continuous, binary, ordinal, nominal, truncated, and count variables
787
788
789 The MICE distributions available in Stata are:
790 binary, ordered and multinomial logistic regression for categorical variables,
791 linear regression and predictive mean matching (PMM)* for continuous variables,
792 and Poisson and negative binomial regression for count variables.
793
794
795
796 IMPUTATION PHASES
797
798 1. mi set wide
799     style to store imputations

```

```

800
801 2. mi register imputed
802     identifies which variables in the imputation model have missing information.
803
804 3. mi register regular (! optional)
805     which variables will not be imputed
806
807 4. mi impute chained
808     where the user specifies the imputation model to be used
809     and the number of imputed datasets to be created.
810     Example:
811         mi impute chained (regress) bmi age (logit) female ///
812         (mlogit) race = bpdia i.region, add(20)
813
814 5. mi estimate
815     is used as a prefix to the standard regress command.
816     This executes the specified estimation model within each of the 20 imputed datasets
817     to obtain 20 sets of coefficients and standard errors.
818     Stata then combines these estimates to obtain one set of inferential statistics.
819     In the output from mi estimate you will see some metrics: Imputation Diagnostics
820     information for RVI (Relative Increase in Variance),
821     FMI (Fraction of Missing Information),
822     DF (Degrees of Freedom) ,
823     RE (Relative Efficiency),
824     and the between imputation and the within imputation variance estimates
825     to examine how the standard errors (SEs) are calculated.
826
827 -----
828
829
830 SELECTING MY IMPUTATION MODEL
831
832 - MICE -> mi impute chained
833
834 - var to be imputed:
835
836     linear regression for continuous var (regress) -> none
837
838
839     logistic for the binary var (logit) ->
840     Hwv8_depressive_symptoms
841
842
843     multinomial logistic for our nominal categorical var (mlogit) ->
844     H_smoking_3cat H_physicalactivity H_alcohol_status
845
846
847
848 - auxiliary var:
849
850     DV -> Hwv9to14_dementia_event
851     IV -> obesity_lca diabetes_lca
852     other covariates -> H_age H_sex H_education H_maritalstatus_4cat H_cvd_comorbidity
853
854
855
856 - imputation numbers (m) -> 10
857
858     ELSA data were imputed 10 numbers
859
860     White et al. (2010) recommendation: use the rule that m should equal the percentage of
861     incomplete cases
862
863 - rseed (53421) for reproducibility reasons
864
865
866 - (! OPTIONAL) advance impute options -> force

```

```

867
868     proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
      are encountered
869
870 - impute options -> savetrace (trace1)
871
872     specifies Stata to save the means and standard deviations of imputed values from each
      iteration to a Stata dataset named "trace1
873 */
874
875
876 mi set wide
877
878
879 mi register imputed H_smoking_3cat H_physicalactivity H_alcohol_status ///
880     Hwv8_loneliness_quintiles Hwv8_depressive_symptoms
881
882
883
884 mi impute chained (logit) Hwv8_depressive_symptoms ///
885 (mlogit) H_smoking_3cat H_physicalactivity H_alcohol_status = Hwv9to14_dementia_event obesity_lca
      diabetes_lca ///
886 H_age H_sex H_education H_maritalstatus_4cat H_cvd_comorbidity, add(10) rseed(53421) savetrace(
      trace1)
887
888
889 * save imputed data
890
891
892 * plot imputations
893
894 *it will open a file named trace1
895 use trace1,clear
896 describe
897
898
899 reshape wide *mean *sd, i(iter) j(m)
900 tsset iter
901
902
903
904 /*
905 The trace plot below graphs the predicted means value produced during the first imputation chain.
906 As before, the expectations is that the values would vary randomly to incorporate variation into
      the predicted values for read.
907 */
908
909 tsline H_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
910 tsline H_physicalactivity_mean1, name(mice1,replace)legend(off) ytitle("Mean of physical activity")
911 tsline H_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol status")
912 tsline Hwv8_depressive_symptoms_mean1, name(mice1,replace)legend(off) ytitle("Mean of depression")
913
914
915
916 /*
917
918 All imputation chains can also be graphed simultaneously to make sure that nothing unexpected
      occurred in a single chain.
919 Every chain is obtained using a different set of initial values and this should be unique.
920 Each colored line represents a different imputation.
921 So all 10 imputation chains are overlaid on top of one another.
922
923 */
924
925
926 tsline Hwv8_depressive_symptoms_mean*, name(mice1,replace)legend(off) ytitle("Mean of depressive
      symptoms")
927 tsline Hwv8_depressive_symptoms_sd*, name(mice2, replace) legend(off) ytitle("SD of depressive

```

```

symptoms")
928 graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)
929
930 * repeat for each imputed var
931
932
933
934
935
936
937 /*
938 ---- DESCRIPTIVE STATISTICS ----
939
940 General characteristics of participants
941
942 General characteristics of participnats stratified for study inclusion
943
944 General characteristics of participants stratified for dementia occurence
945
946 Participant characteristics by CM 3-class groups
947
948
949 1. CHI-SQUARE (chi2) for categorical var (crosstabulation)
950     Frequency tables -> two-way tables
951         using the command tabulate, chi2
952         reporting observations, column percentage (N, %) and p-value of Pearson's r
953
954
955 2. one-way ANOVA for continuous var
956     check box plot
957     using the command oneway
958     reporting mean, sd (summary tables) and p-value of F
959 */
960
961
962
963
964
965 * General characteristics of HRS participants at baseline
966
967
968 * Socio-demographics
969 sum H_age
970 ta H_sex
971 ta H_education
972 ta H_maritalstatus_4cat
973 ta H_wealthquintiles
974 * Cardiometabolic disorders
975 ta crp_lca
976 ta hdl_lca
977 ta obesity_lca
978 ta systolic_lca
979 ta diastolic_lca
980 ta diabetes_lca
981 ta hba1c_lca
982 * Cardiovascular health factors
983 ta H_smoking_3cat
984 ta H_physicalactivity
985 ta H_alcohol_status
986 ta H_cvd_comorbidity
987 * Depressive symptoms (cont and categ)
988 sum Hwv8_cesd_sumscore
989 ta Hwv8_depressive_symptoms
990 * Memory score
991 sum Hwv8_memory
992
993
994

```

```

995
996 * General baseline characteristics of HRS participants by dementia status
997
998 * Socio-demographics
999 ttest H_age, by(Hwv9to14_dementia_event)
1000 ta H_sex Hwv9to14_dementia_event, chi2 column row
1001 ta H_education Hwv9to14_dementia_event, chi2 column row
1002 ta H_maritalstatus_4cat Hwv9to14_dementia_event, chi2 column row
1003 ta H_wealthquintiles Hwv9to14_dementia_event, chi2 column row
1004 * Cardiometabolic disorders
1005 ta crp_lca Hwv9to14_dementia_event, chi2 column row
1006 ta hdl_lca Hwv9to14_dementia_event, chi2 column row
1007 ta obesity_lca Hwv9to14_dementia_event, chi2 column row
1008 ta systolic_lca Hwv9to14_dementia_event, chi2 column row
1009 ta diastolic_lca Hwv9to14_dementia_event, chi2 column row
1010 ta diabetes_lca Hwv9to14_dementia_event, chi2 column row
1011 ta hba1c_lca Hwv9to14_dementia_event, chi2 column row
1012 * Cardiovascular health factors
1013 ta H_smoking_3cat Hwv9to14_dementia_event, chi2 column row
1014 ta H_physicalactivity Hwv9to14_dementia_event, chi2 column row
1015 ta H_alcohol_status Hwv9to14_dementia_event, chi2 column row
1016 ta H_cvd_comorbidity Hwv9to14_dementia_event, chi2 column row
1017 * Depressive symptoms (cont and categ)
1018 ttest Hwv8_cesd_sumscore, by(Hwv9to14_dementia_event)
1019 ta Hwv8_depressive_symptoms Hwv9to14_dementia_event, chi2 column row
1020 * Memory score
1021 ttest Hwv8_memory, by(Hwv9to14_dementia_event)
1022
1023
1024
1025
1026
1027
1028
1029 * Sample characteristics by CM 3-class groups
1030 * crosstabs categ var (frequencies and chi2) !report column percentage!
1031 * oneway ANOVA cont var (mean, sd)
1032
1033
1034 * Socio-demographics
1035 oneway H_age H_lca_group3, tabulate
1036 ta H_sex H_lca_group3, chi2 column row
1037 ta H_education H_lca_group3, chi2 column row
1038 ta H_maritalstatus_4cat H_lca_group3, chi2 column row
1039 ta H_wealthquintiles H_lca_group3, chi2 column row
1040 * Cardiovascular health factors
1041 ta H_smoking_3cat H_lca_group3, chi2 column row
1042 ta H_physicalactivity H_lca_group3, chi2 column row
1043 ta H_alcohol_status H_lca_group3, chi2 column row
1044 ta H_cvd_comorbidity H_lca_group3, chi2 column row
1045 * Depressive symptoms (cont and categ)
1046 oneway Hwv8_cesd_sumscore H_lca_group3, tabulate
1047 ta Hwv8_depressive_symptoms H_lca_group3, chi2 column row
1048 * Memory score
1049 oneway Hwv8_memory H_lca_group3, tabulate
1050
1051
1052
1053
1054
1055
1056
1057 /*
1058 ---- SURVIVAL ANALYSIS IN COMPLETE DATA ----
1059
1060 Tests of proportional-hazards assumption
1061 Kaplan Meier survival curves
1062 Person-time

```



```

1063 Cox proportional regression - Hazard ratios - stcox
1064 Postestimation tools for stcox
1065 Test of Goodness of Fit
1066
1067 *** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
1068 Cox PH regression in complete data
1069 Cox PH regression model in imputed dataset - mi estimate
1070
1071
1072 */
1073
1074
1075
1076 * check dataset variables of interest only
1077
1078 codebook H_time_of_event_dementia Hwv9to14_dementia_event H_lca_group3 ///
1079 H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
1080 H_smoking_3cat H_physicalactivity H_alcohol_status H_cvd_comorbidity ///
1081 Hwv8_depressive_symptoms Hwv8_memory,compact
1082
1083
1084
1085 * Declare Data to be Survival Data
1086 * Time to event: H_time_of_event_dementia (months)
1087 * Censoring: Hwv9to14_dementia_event (1=dementia, 0=censored)
1088 * Command is stset TIMETOEVENT, failure(CENSORVARIABLE)
1089
1090
1091 stset H_time_of_event_dementia, failure (Hwv9to14_dementia_event==1) id(RAHHIDPN)
1092
1093
1094
1095 *describe survival data using commnad stsum
1096
1097 stsum
1098
1099 stsum, by(H_lca_group3)
1100
1101
1102
1103
1104 * Kaplan Meier Curve estimation
1105
1106 sts list
1107
1108 sts list, by(H_lca_group3)
1109
1110
1111
1112 * Kaplan Meier Curve Plot
1113
1114 * no frills plot
1115
1116 sts graph
1117
1118 * with frills
1119
1120 sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
1121 title("Kaplan Meier Curve") subtitle("n=5112, # events=476") ///
1122 caption("graph02.png", size(vsmall))
1123
1124
1125 * With Greenwood CI limits
1126
1127 sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
1128 title("Kaplan Meier Curve") subtitle("n=n=5112, # events=476") caption("graph03.png", size(vsmall))
1129
1130

```

```

1131
1132
1133 * Group Kaplan-Meier Curve Estimation
1134 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1135
1136 sort H_lca_group3
1137
1138 sts list, by(H_lca_group3)
1139
1140 * graph with frills
1141
1142 sts graph, by(H_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in Months") ///
1143 ytitle("Survival Prob") title("Kaplan Meier Curve") subtitle("n=5112, # events=476") ///
1144 caption("graph04.png", size(vsmall))
1145
1146
1147
1148
1149 * calculate person-time and incidence rates using command stptime
1150
1151 stptime, title(Person-years)
1152
1153 stptime, title(Person-years) per(1000)
1154
1155 stptime, title(Person-years) per(10000)
1156
1157
1158 * calculate person-time by category of H_lca_group3
1159
1160 stptime, by(H_lca_group3)
1161
1162 stptime, by(H_lca_group3) per(1000)
1163
1164
1165
1166
1167 * mean and median of follow-up
1168 sum H_time_of_event_dementia
1169 sum H_time_of_event_dementia, detail
1170
1171
1172
1173 /* Log Rank Test of equality of survival distributions
1174 (NULL: equality of survival distributions among H_lca_group3)
1175 We will consider including the predictor if the test has a p-value of 0.2 - 0.25 or less.
1176 If the predictor has a p-value greater than 0.25 in a univariate analysis
1177 it is highly unlikely that it will contribute anything to a model which includes other
predictors.
1178 Command is sts test GROUPVAR
1179 */
1180
1181
1182 sts test H_lca_group3, logrank
1183
1184 sts test H_age, logrank
1185
1186 sts test H_sex, logrank
1187
1188 sts test H_eduaction, logrank
1189
1190 sts test H_maritalstatus_4cat, logrank
1191
1192 sts test H_wealthquintiles, logrank
1193
1194 sts test H_smoking_3cat, logrank
1195
1196 sts test H_physicalactivity, logrank
1197

```

```

1198 sts test H_alcohol_status, logrank
1199
1200 sts test H_cvd_comorbidity, logrank
1201
1202 sts test Hwv8_depressive_symptoms, logrank
1203
1204 sts test Hwv8_memory, logrank
1205
1206
1207
1208
1209
1210 /* Cox PH regression model
1211
1212 using the command stcox
1213
1214 --- Building the model ---
1215
1216 Model 1: unadjusted - single predictor of CM classes
1217 Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1218 Model 3: model 2 + cvd health: smoking, alcohol consumption, cvd comorbidity
1219 Model 4: model 3 + mental health: depressive symptoms
1220
1221 !! I didn't adjust for physical activity because this variable can't be used in CHARLS (missing
1222 values)
1223 */
1224
1225
1226 * Unadjusted model - model 1 - single predictor
1227
1228 stcox H_lca_group3
1229
1230 * define design var by using i.(3 classes)
1231
1232 stcox i.H_lca_group3
1233
1234
1235 * Adjusted models - multivariable Cox model
1236 * controlling for covariates
1237
1238 * model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
1239
1240 stcox i.H_lca_group3 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles
1241
1242
1243 * model 3: model 2 + adjust for cvd health
1244
1245 stcox i.H_lca_group3 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
1246 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity
1247
1248 * model 4: model 3 + adjust for mental health
1249
1250 stcox i.H_lca_group3 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
1251 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
1252 i.Hwv8_depressive_symptoms
1253
1254
1255
1256
1257
1258
1259
1260
1261 * Multivariable model development
1262 * Likelihood-ratio tests
1263
1264

```

```

1265
1266 *install eststo
1267 findit eststo
1268
1269
1270 * ---- rx controlling for age and sex -----*
1271 quietly: stcox H_age i.H_sex
1272 eststo modelagesex
1273
1274 quietly: stcox H_age i.H_sex i.H_lca_group3
1275 eststo modelagesex_3group
1276
1277 lrtest modelagesex modelagesex_3group
1278
1279
1280
1281 * ---- rx controlling for sociodemographics -----*
1282 quietly: stcox H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles
1283 eststo modelsociodemo
1284
1285 quietly: stcox H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles i.
H_lca_group3
1286 eststo modelsociodemo_3group
1287
1288 lrtest modelsociodemo modelsociodemo_3group
1289
1290
1291 * ---- rx controlling for cardiovascular health -----*
1292 quietly: stcox i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity
1293 eststo modelcardiovascular
1294
1295 quietly: stcox i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity i.H_lca_group3
1296 eststo modelcardiovascular_3group
1297
1298 lrtest modelcardiovascular modelcardiovascular_3group
1299
1300
1301
1302 * ---- rx controlling for mental health-----*
1303 quietly: stcox i.Hwv8_depressive_symptoms Hwv8_memory
1304 eststo modelmental
1305
1306 quietly: stcox i.Hwv8_depressive_symptoms i.H_lca_group3
1307 eststo modelmental_3group
1308
1309 lrtest modelmental modelmental_3group
1310
1311
1312
1313
1314 * side-by-side comparison of models
1315
1316
1317 quietly: stcox i.H_lca_group3
1318 eststo model1
1319
1320
1321 quietly: stcox H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles i.
H_lca_group3
1322 eststo model2
1323
1324
1325 quietly: stcox H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
1326 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity i.H_lca_group3
1327 eststo model3
1328
1329
1330 quietly: stcox H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
```

```

1331 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity i.Hwv8_depressive_symptoms i.H_lca_group3
1332 eststo model4
1333
1334
1335
1336
1337 * Display Betas and Summary Statistics
1338 estout model1 model2 model3 model4, stats(n chi2 bic, star(chi2)) prehead("Betas")
1339
1340 /* Key Interpretation
1341 Chi2 = Value of LR test comparing the model fit ("full") to intercept only ("reduced")
1342 bic = Schwarz' Bayesian Information Criterion = It is a function of the log-likelihood.
1343 Smaller values indicate a better fit.
1344 */
1345
1346 * Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
1347 estout model1 model2 model3 model4, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")
1348
1349
1350
1351
1352 * Postestimation tools for stcox
1353
1354 /* Test of proportional hazards
1355
1356 If the tests in the table are not significance (p-values over 0.05)
1357 then we can not reject proportionality and we assume
1358 that we do not have a violation of the proportional assumption.
1359 */
1360
1361 estat phtest, detail
1362
1363
1364 /* Proportionality Assumption - method 1
1365 We will check proportionality by including time-dependent covariates in the model
1366 by using the tvc and the texp options in the stcox command.
1367 Time dependent covariates are interactions of the predictors and time.
1368 In this analysis we choose to use the interactions with log(time)
1369 because this is the most common function of time used in time-dependent covariates
1370 but any function of time could be used.
1371 If a time-dependent covariate is significant this indicates
1372 a violation of the proportionality assumption for that specific predictor.
1373 The conclusion is that all of the time-dependent variables are not significant
1374 either collectively or individually thus supporting the assumption of proportional hazard.
1375 */
1376
1377
1378
1379 stcox i.H_lca_group3 H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
1380 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
1381 i.Hwv8_depressive_symptoms, nohr ///
1382 tvc(H_lca_group3 H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
1383 H_smoking_3cat H_alcohol_status H_cvd_comorbidity Hwv8_depressive_symptoms) texp(ln(
1384 H_time_of_event_dementia))
1385
1386
1387 /* Proportionality Assumption - method 2
1388 by using the Schoenfeld and scaled Schoenfeld residuals
1389 In the stphtest command we test the proportionality of the model as a whole
1390 and by using the detail option we get a test of proportionality for each predictor.
1391 By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
1392 If the tests in the table are not significance (p-values over 0.05)
1393 then we can not reject proportionality and we assume
1394 that we do not have a violation of the proportional assumption.
1395 The stphtest command uses log-log plots to test proportionality
1396 and if the lines in these plots are parallel then we have further indication
1397 that the predictors do not violate the proportionality assumption.

```

```

1398 */
1399
1400 quietly stcox H_lca_group3 H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
1401 H_smoking_3cat H_alcohol_status H_cvd_comorbidity Hwv8_depressive_symptoms Hwv8_memory, schoenfeld
(sch*) scaledsch(sca*)
1402 stphtest, detail
1403 stphtest, plot(H_lca_group3) msym(oh)
1404 stphtest, plot(H_age) msym(oh)
1405 stphtest, plot(H_sex) msym(oh)
1406 stphtest, plot(H_education) msym(oh)
1407 stphtest, plot(H_maritalstatus_4cat) msym(oh)
1408 stphtest, plot(H_wealthquintiles) msym(oh)
1409 stphtest, plot(H_smoking_3cat) msym(oh)
1410 stphtest, plot(H_alcohol_status) msym(oh)
1411 stphtest, plot(H_cvd_comorbidity) msym(oh)
1412 stphtest, plot(Hwv8_depressive_symptoms) msym(oh)
1413
1414
1415
1416
1417
1418
1419 stphplot, by(H_lca_group3) plot1(msym(oh)) plot2(msym(th))
1420 stphplot, by(H_age) plot1(msym(oh)) plot2(msym(th))
1421 stphplot, by(H_sex) plot1(msym(oh)) plot2(msym(th))
1422 stphplot, by(H_education) plot1(msym(oh)) plot2(msym(th))
1423 stphplot, by(H_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
1424 stphplot, by(H_wealthquintiles) plot1(msym(oh)) plot2(msym(th))
1425 stphplot, by(H_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
1426 stphplot, by(H_alcohol_status) plot1(msym(oh)) plot2(msym(th))
1427 stphplot, by(H_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
1428 stphplot, by(Hwv8_depressive_symptoms) plot1(msym(oh)) plot2(msym(th))
1429
1430
1431
1432
1433
1434 * Assessment of PH Assumption: adjust for age and sex
1435 stphplot, by(H_lca_group3) adjust(H_age H_sex) nolntime plot1opts(symbol(none) color(black)
lpattern(dash)) ///
1436 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1437 title("Assessment of PH Assumption") subtitle(" Predictor is H_lca_group3") xtitle("months")
1438
1439
1440
1441 * Assessment of PH Assumption: adjust for model 2
1442 stphplot, by(H_lca_group3) adjust(H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles)
///
1443 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1444 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1445 title("Assessment of PH Assumption") subtitle(" Predictor is H_lca_group3") xtitle("months")
1446
1447
1448
1449 * Assessment of PH Assumption: adjust for model 3
1450 stphplot, by(H_lca_group3) adjust(H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles
///
1451 H_smoking_3cat H_alcohol_status H_cvd_comorbidity) ///
1452 nolntime plot1opts(symbol(none) color(black) lpattern(dash)) ///
1453 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1454 title("Assessment of PH Assumption") subtitle(" Predictor is H_lca_group3") xtitle("months")
1455
1456
1457
1458 * Assessment of PH Assumption: adjust for model 4
1459 stphplot, by(H_lca_group3) adjust(H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles
///
1460 H_smoking_3cat H_alcohol_status H_cvd_comorbidity) ///

```

```

1461 Hwv8_depressive_symptoms) ///
1462 nolntime plotlopts(symbol(none) color(black) lpattern(dash)) ///
1463 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1464 title("Assessment of PH Assumption") subtitle(" Predictor is H_lca_group3") xtitle("months")
1465
1466
1467
1468
1469
1470 /* Test of overall goodness of fit
1471 Goodness of fit of the final model
1472 2 methods:
1473 - by using the commnad stcoxgof (good fit = non sig p-value)
1474 - by using the Cox-Snell residuals
1475     to create the Nelson-Aalen cumulative hazard function
1476     If the hazard function follows the 45 degree line then we know that it approximately
1477     has an exponential distribution with a hazard rate of one and that the model fits the data
1478     well.
1479     If the model fits the data, the plot of the cumulative hazard versus cs
1480     should approximate a straight line with slope 1.
1481 */
1482
1483 * by using the commnad stcoxgof
1484
1485 * install stcoxgof
1486 findit stcoxgof
1487
1488
1489 stcox H_lca_group3 H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
1490 H_smoking_3cat H_alcohol_status H_cvd_comorbidity ///
1491 Hwv8_depressive_symptoms, mgale(mgale)
1492
1493
1494 stcoxgof
1495
1496
1497
1498
1499 * by using the Cox-Snell residuals
1500
1501 quietly stcox H_lca_group3 H_age H_sex H_education H_maritalstatus_4cat H_wealthquintiles ///
1502 H_smoking_3cat H_alcohol_status H_cvd_comorbidity ///
1503 Hwv8_depressive_symptoms
1504 predict cs, csnell
1505
1506 * or
1507
1508 quietly stcox H_lca_group3
1509 predict cs, csnell
1510
1511
1512 stset cs, failure(Hwv9to14_dementia_event)
1513 sts generate km = s
1514 generate H = -ln(km)
1515 line H cs cs, sort ytitle("") clstyle(. refline)
1516
1517
1518
1519
1520
1521
1522 * ----- COX PH REGRESSION MODEL IN IMPUTED DATASET ----- *
1523
1524
1525 * Declare Data to be Survival Data by using mi
1526
1527 mi stset H_time_of_event_dementia, failure (Hwv9to14_dementia_event==1) id(RAHHIDPN)

```



```

1528
1529
1530 * Run Cox regression analysis in imputed dataset by using "mi estimate:"
1531 * Building the Model: Model 1 (unadjusted), Model 2, Model 3, Model 4
1532
1533
1534 * Model 1 (default coefficients)
1535 mi estimate: stcox H_lca_group3
1536
1537 * Model 1: define design var by using i.
1538 mi estimate: stcox i.H_lca_group3
1539
1540
1541 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1542
1543 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3
1544
1545
1546 * Adjusted models - multivariable Cox model
1547 * controlling for covariates
1548
1549 * Model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
1550 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
1551   H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles
1552
1553
1554
1555 * Model 3: model 2 + adjust for cvd health
1556
1557 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
1558   H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
1559   i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity
1560
1561
1562 * Model 4: model 3 + adjust for depression
1563
1564 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
1565   H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
1566   i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
1567   i.Hwv8_depressive_symptoms
1568
1569
1570
1571
1572
1573
1574
1575
1576 /*
1577
1578
1579 *** SENSITIVITY ANALYSES ***
1580
1581 1) multigroup latent class model by sex
1582
1583 2) interactions with age and gender
1584 survival analysis stratified by age
1585 two age groups: <70 and >=70
1586
1587 3) exclude participants with cvd
1588
1589 4) survival analysis limiting to 5 year follow-up
1590
1591 5) Complete data
1592 Cox regression analysis on complete data (without imputed covariates)
1593 (see above)
1594 */
1595

```



```

1596
1597  /*
1598  1) Multigroup latent class model by sex
1599
1600
1601  TWO STEP PROCESS
1602
1603  1) LCA by group (to build the model and get lcprob and lcmean and to get the marginplots for males)
1604
1605  gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1606  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1607  group(H_sex) ginvariant(coef)
1608
1609  estat lcprob
1610  estat lcmean
1611  estat lcgof
1612
1613  2) LCA sort sex (to get the marginplots for females)
1614
1615  sort H_sex
1616
1617  by H_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1618  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
1619
1620
1621  */
1622
1623
1624
1625  * LCA by group
1626  * three-class model
1627
1628
1629  gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1630  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1631  group(H_sex) ginvariant(coef) ///
1632  startvalues(randompr, draws(20) seed(15) difficult) ///
1633  emopts(iterate(30) difficult)
1634
1635
1636  estimates store threeclass_cm
1637
1638
1639  * LCA postestimation
1640  * Latent class marginal probabilities - lcprob -
1641  * Latent class marginal means - lcmean -
1642
1643
1644  estat lcprob
1645
1646  estat lcmean
1647
1648  estat lcgof
1649
1650
1651
1652
1653  /* We can use the predictions of the posterior probability of class membership to evaluate an
1654  individual's probability of being in each class.
1655
1656  */
1657
1658  predict m_classpost1*, classposteriorpr
1659  list in 1, abbrev(10)
1660
1661  /* We can determine the expected class for each individual based on whether the posterior
1662  probability
1663  is greater than 0.5

```

```

1663 */
1664
1665 generate m_expclass1 = 1 + (m_classpost1>0.5)
1666 tabulate m_expclass1
1667
1668
1669 generate m_expclass2 = 1 + (m_classpost12>0.5)
1670 tabulate m_expclass2
1671
1672
1673 generate m_expclass3 = 1 + (m_classpost13>0.5)
1674 tabulate m_expclass3
1675
1676
1677
1678
1679 /* We can determine expected classification for each individual in the dataset based on the
1680 predicted
1681 posterior class probabilities.
1682 */
1683 predict m_cpost*, classposteriorpr
1684 egen m_max = rowmax(m_cpost*)
1685
1686
1687 * generate classes var
1688
1689 generate m_predclass = 1 if m_cpost1==m_max
1690
1691 replace m_predclass = 2 if m_cpost2==m_max
1692
1693 replace m_predclass = 3 if m_cpost3==m_max
1694
1695 tabulate m_predclass
1696
1697
1698
1699
1700
1701 * margins and marginsplot for MALES
1702
1703 * use margins to calculate marginal predictions
1704 * use marginsplot to graph marginal predictions
1705
1706
1707
1708 *Install/update combomarginsplot ado.
1709
1710 *https://www.statalist.org/forums/forum/general-stata-discussion/general/1425209-is-it-possible-to-do-multilevel-latent-class-analysis-with-stata-15-ic
1711
1712 ssc install combomarginsplot, replace
1713
1714
1715 margins, predict(classpr class(1)) ///
1716         predict(classpr class(2)) ///
1717         predict(classpr class(3)) subpop(if H_sex==0) saving(margin_male, replace)
1718 marginsplot, xtitle("") ytitle("") ///
1719         xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1720         title("Predicted Latent Class Probabilities with 95% CI") ///
1721         name(margin_male, replace)
1722
1723
1724 margins, predict(classpr class(1)) ///
1725         predict(classpr class(2)) ///
1726         predict(classpr class(3)) subpop(if H_sex==0) saving(margin_male, replace)
1727 marginsplot, recast(bar) xtitle("") ytitle("") ///
1728         xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///

```

```

1729         title("Predicted Latent Class Probabilities with 95% CI") ///
1730         name(margin_male, replace)
1731
1732
1733 * class 1
1734
1735 margins, predict(outcome(crp_lca) class(1)) ///
1736         predict(outcome(hdl_lca) class(1)) ///
1737         predict(outcome(obesity_lca) class(1)) ///
1738         predict(outcome(systolic_lca) class(1)) ///
1739         predict(outcome(diastolic_lca) class(1)) ///
1740         predict(outcome(diabetes_lca) class(1)) ///
1741         predict(outcome(hba1c_lca) class(1)) subpop(if H_sex==0) ///
1742         saving(class1_male, replace) ///
1743
1744 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
1745         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1746         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1747         ytitle("Predicted mean") ylabel(0(.20)1) name (class1_male, replace)
1748
1749
1750 * class 2
1751
1752 margins, predict(outcome(crp_lca) class(2)) ///
1753         predict(outcome(hdl_lca) class(2)) ///
1754         predict(outcome(obesity_lca) class(2)) ///
1755         predict(outcome(systolic_lca) class(2)) ///
1756         predict(outcome(diastolic_lca) class(2)) ///
1757         predict(outcome(diabetes_lca) class(2)) ///
1758         predict(outcome(hba1c_lca) class(2)) subpop(if H_sex==0) ///
1759         saving(class2_male, replace) ///
1760
1761 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1762         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1763         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1764         ytitle("Predicted mean") ylabel(0(.20)1) name (class2_male, replace)
1765
1766
1767
1768
1769 * class 3
1770
1771 margins, predict(outcome(crp_lca) class(3)) ///
1772         predict(outcome(hdl_lca) class(3)) ///
1773         predict(outcome(obesity_lca) class(3)) ///
1774         predict(outcome(systolic_lca) class(3)) ///
1775         predict(outcome(diastolic_lca) class(3)) ///
1776         predict(outcome(diabetes_lca) class(3)) ///
1777         predict(outcome(hba1c_lca) class(3)) subpop(if H_sex==0) ///
1778         saving(class3_male, replace) ///
1779
1780 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
1781         xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1782         5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1783         ytitle("Predicted mean") ylabel(0(.20)1) name (class3_male, replace)
1784
1785
1786
1787 graph combine class1_male class2_male class3_male, cols(3)
1788
1789
1790
1791
1792
1793
1794 * LCA sort by sex
1795 * three-class model
1796

```

```

1797  sort H_sex
1798
1799  by H_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1800  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1801  startvalues(randompr, draws(20) seed(15) difficult) ///
1802  emopts(iterate(30) difficult)
1803
1804  estat lcprob
1805
1806  estat lcmean
1807
1808  estat lcgof
1809
1810
1811
1812
1813  /* We can use the predictions of the posterior probability of class membership to evaluate an
1814  individual's probability of being in each class.
1815
1816  */
1817
1818  predict f_classpost1*, classposteriorpr
1819  list in 1, abbrev(10)
1820
1821  /* We can determine the expected class for each individual based on whether the posterior
1822  probability
1823  is greater than 0.5
1824  */
1825  generate f_expclass1 = 1 + (f_classpost1>0.5)
1826  tabulate f_expclass1
1827
1828
1829  generate f_expclass2 = 1 + (f_classpost12>0.5)
1830  tabulate f_expclass2
1831
1832
1833  generate f_expclass3 = 1 + (f_classpost13>0.5)
1834  tabulate f_expclass3
1835
1836
1837
1838  /* We can determine expected classification for each individual in the dataset based on the
1839  predicted
1840  posterior class probabilities.
1841  */
1842  predict f_cpost*, classposteriorpr
1843  egen f_max = rowmax(f_cpost*)
1844
1845
1846  * generate classes var
1847
1848  generate f_predclass = 1 if f_cpost1==f_max
1849
1850  replace f_predclass = 2 if f_cpost2==f_max
1851
1852  replace f_predclass = 3 if f_cpost3==f_max
1853
1854  tabulate f_predclass
1855
1856
1857
1858
1859  * margins and marginsplot for FEMALES
1860
1861  * use margins to calculate marginal predictions
1862  * use marginsplot to graph marginal predictions

```

```

1863
1864
1865
1866 margins, predict(classpr class(1)) ///
1867       predict(classpr class(2)) ///
1868       predict(classpr class(3)) subpop(if H_sex==1) saving(margin_female, replace)
1869 marginsplot, xtitle("") ytitle("") ///
1870       xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1871       title("Predicted Latent Class Probabilities with 95% CI") ///
1872       name(margin_female, replace)
1873
1874
1875 margins, predict(classpr class(1)) ///
1876       predict(classpr class(2)) ///
1877       predict(classpr class(3)) subpop(if H_sex==1) saving(margin_female, replace)
1878 marginsplot, recast(bar) xtitle("") ytitle("") ///
1879       xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1880       title("Predicted Latent Class Probabilities with 95% CI") ///
1881       name(margin_female, replace)
1882
1883
1884 * class 1
1885
1886 margins, predict(outcome(crp_lca) class(1)) ///
1887       predict(outcome(hdl_lca) class(1)) ///
1888       predict(outcome(obesity_lca) class(1)) ///
1889       predict(outcome(systolic_lca) class(1)) ///
1890       predict(outcome(diastolic_lca) class(1)) ///
1891       predict(outcome(diabetes_lca) class(1)) ///
1892       predict(outcome(hba1c_lca) class(1)) subpop(if H_sex==1) ///
1893       saving(class1_female, replace) ///
1894
1895 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
1896       xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1897       5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1898       ytitle("Predicted mean") ylabel(0(.20)1) name (class1_female, replace)
1899
1900
1901 * class 2
1902
1903 margins, predict(outcome(crp_lca) class(2)) ///
1904       predict(outcome(hdl_lca) class(2)) ///
1905       predict(outcome(obesity_lca) class(2)) ///
1906       predict(outcome(systolic_lca) class(2)) ///
1907       predict(outcome(diastolic_lca) class(2)) ///
1908       predict(outcome(diabetes_lca) class(2)) ///
1909       predict(outcome(hba1c_lca) class(2)) subpop(if H_sex==1) ///
1910       saving(class2_female, replace) ///
1911
1912 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1913       xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1914       5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1915       ytitle("Predicted mean") ylabel(0(.20)1) name (class2_female, replace)
1916
1917
1918
1919
1920 * class 3
1921
1922 margins, predict(outcome(crp_lca) class(3)) ///
1923       predict(outcome(hdl_lca) class(3)) ///
1924       predict(outcome(obesity_lca) class(3)) ///
1925       predict(outcome(systolic_lca) class(3)) ///
1926       predict(outcome(diastolic_lca) class(3)) ///
1927       predict(outcome(diabetes_lca) class(3)) ///
1928       predict(outcome(hba1c_lca) class(3)) subpop(if H_sex==1) ///
1929       saving(class3_female, replace) ///
1930

```

```

1931 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
1932 xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1933 5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1934 ytitle("Predicted mean") ylabel(0(.20)1) name (class3_female, replace)
1935
1936
1937
1938 graph combine class1_female class2_female class3_female, cols(3)
1939
1940
1941
1942
1943
1944
1945 *combine margins male and female class probabilities
1946
1947 graph combine margin_male margin_female, cols(3)
1948
1949 *combine margins male and female 3 classes mean
1950
1951 graph combine class1_male class2_male class3_male class1_female class2_female class3_female, cols(
1952 3)
1953
1954
1955
1956
1957 /* 2) Interaction with age and gender
1958 Survival analysis stratified by age
1959
1960 generate age group variable
1961 Age groups: 1) young old (< 70) 2) old old (>= 70)
1962
1963 Kaplan Meier curves
1964 Cox regression models in imputed data
1965
1966 young old <70
1967 if H_age_group==1
1968
1969 old old >70
1970 if H_age_group==2
1971
1972 */
1973
1974
1975
1976 gen H_age_group=1 if H_age < 70
1977 replace H_age_group=2 if H_age >=70 & ///
1978 !missing(H_age)
1979
1980 label var H_age_group "Age groups <70 young-old / 70 old-old"
1981 lab def age_group 1 "young old <70" 2 "old old >70"
1982 lab val H_age_group age_group
1983
1984 tab H_age_group
1985
1986
1987
1988 stset H_time_of_event_dementia, failure (Hwv9to14_dementia_event==1) id(RAHHIDPN)
1989
1990
1991 * YOUNG OLD <70 Kaplan Meier
1992
1993 * Group Kaplan-Meier Curve Estimation
1994 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1995
1996 sort H_lca_group3
1997

```

```

1998 sts list if H_age_group==1, by(H_lca_group3)
1999
2000
2001 * graph with frills
2002
2003 sts graph if H_age_group==1, by(H_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in
Months") ///
2004 ytitle("Survival Prob") title("Kaplan Meier Curve <70 years") ///
2005 caption("graph04.png", size(vsmall))
2006
2007
2008 * calculate person-time and incidence rates using command ststime
2009
2010 stptime if H_age_group==1, title(Person-years)
2011
2012 stptime if H_age_group==1, title(Person-years) per(1000)
2013
2014
2015 * calculate person-time by category of H_lca_group3
2016
2017 stptime if H_age_group==1, by(H_lca_group3)
2018
2019 stptime if H_age_group==1, by(H_lca_group3) per(1000)
2020
2021
2022
2023
2024 * OLD OLD >70 Kaplan Meier
2025
2026
2027 sts list if H_age_group==2, by(H_lca_group3)
2028
2029
2030 * graph with frills
2031
2032 sts graph if H_age_group==2, by(H_lca_group3) xlabel(0(20)180) ylabel(0.80(.05)1) xtitle("Time in
Months") ///
2033 ytitle("Survival Prob") title("Kaplan Meier Curve >=70 years") ///
2034 caption("graph04.png", size(vsmall))
2035
2036
2037 * calculate person-time and incidence rates using command ststime
2038
2039 stptime if H_age_group==2, title(Person-years)
2040
2041 stptime if H_age_group==2, title(Person-years) per(1000)
2042
2043
2044 * calculate person-time by category of H_lca_group3
2045
2046 stptime if H_age_group==2, by(H_lca_group3)
2047
2048 stptime if H_age_group==2, by(H_lca_group3) per(1000)
2049
2050
2051
2052
2053 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2054
2055 * Declare Data to be Survival Data by using mi
2056
2057 mi stset H_time_of_event_dementia, failure (Hwv9to14_dementia_event==1) id(RAHHIDPN)
2058
2059
2060
2061
2062
2063

```



```

2064 *** INTERACTION gender*cardiometabolic cluster ***
2065
2066 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 i.H_sex#i.H_lca_group3
2067
2068 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2069 H_age i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2070 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
2071 i.Hwv8_depressive_symptoms i.H_sex#i.H_lca_group3
2072
2073
2074
2075 *** INTERACTION age*cardiometabolic cluster ***
2076
2077 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 c.H_age#i.H_lca_group3
2078
2079 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2080 H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2081 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
2082 i.Hwv8_depressive_symptoms c.H_age#i.H_lca_group3
2083
2084
2085
2086
2087 * YOUNG OLD <70 Cox regression models
2088
2089 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2090
2091 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 if H_age_group==1
2092
2093
2094 * Adjusted models - multivariable Cox model
2095 * controlling for covariates
2096
2097 * Model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
2098 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2099 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles if H_age_group==1
2100
2101
2102 * Model 3: model 2 + adjust for cvd health
2103
2104 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2105 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2106 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity if H_age_group==1
2107
2108
2109 * Model 4: model 3 + adjust for depression
2110
2111 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2112 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2113 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
2114 i.Hwv8_depressive_symptoms if H_age_group==1
2115
2116
2117
2118 * OLD OLD >70 Cox regression models
2119
2120 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2121
2122 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 if H_age_group==2
2123
2124
2125 * Adjusted models - multivariable Cox model
2126 * controlling for covariates
2127
2128 * Model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
2129 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2130 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles if H_age_group==2
2131

```



```

2132
2133 * Model 3: model 2 + adjust for cvd health
2134
2135 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2136 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2137 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity if H_age_group==2
2138
2139
2140 * Model 4: model 3 + adjust for depression
2141
2142 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2143 i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2144 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
2145 i.Hwv8_depressive_symptoms if H_age_group==2
2146
2147
2148
2149
2150
2151 /*
2152
2153 3) exclude participants with cvd
2154
2155 use the command if H_cvd_comorbidity==0
2156
2157 */
2158
2159
2160
2161 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2162
2163 * Declare Data to be Survival Data by using mi
2164
2165 mi stset H_time_of_event_dementia, failure (Hwv9to14_dementia_event==1) id(RAHHIDPN)
2166
2167
2168 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2169
2170 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 if H_cvd_comorbidity==0
2171
2172
2173 * Adjusted models - multivariable Cox model
2174 * controlling for covariates
2175
2176 * Model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
2177 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2178 H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles if H_cvd_comorbidity==0
2179
2180
2181 * Model 3: model 2 + adjust for cvd health
2182
2183 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2184 H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2185 i.H_smoking_3cat i.H_alcohol_status if H_cvd_comorbidity==0
2186
2187
2188 * Model 4: model 3 + adjust for depression
2189
2190 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2191 H_age i.H_sex i.H_education i.H_maritalstatus_4cat i.H_wealthquintiles ///
2192 i.H_smoking_3cat i.H_alcohol_status ///
2193 i.Hwv8_depressive_symptoms if H_cvd_comorbidity==0
2194
2195
2196
2197
2198 /*
2199 4) survival analysis limiting to 5 year follow-up period

```

```

2200
2201 hrs follow-up wave 9-12
2202
2203 */
2204
2205
2206
2207 merge 1:m RAHHIDPN using "S:\Research\pkstudies\Study3_cardio_lca\HRS\hrs_lca data sensitivity
9to12followup.dta"
2208
2209
2210
2211 * COX PH REGRESSION MODEL IN IMPUTED DATASET
2212
2213 * Declare Data to be Survival Data by using mi
2214
2215 mi stset Hwv9to12_time_of_event_dementia, failure (Hwv9to12_dementia_event==1) id(RAHHIDPN)
2216
2217
2218
2219 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2220
2221 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3
2222
2223
2224 * Adjusted models - multivariable Cox model
2225 * controlling for covariates
2226
2227 * Model 2: model 1 + adjust for socio-demographics: age sex education marital status and wealth
2228 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2229 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles
2230
2231
2232
2233 * Model 3: model 2 + adjust for cvd health
2234
2235 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2236 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
2237 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity
2238
2239
2240 * Model 4: model 3 + adjust for depression
2241
2242 mi estimate, eform("Haz. Ratio"): stcox i.H_lca_group3 ///
2243 H_age i.H_sex i.H_eduaction i.H_maritalstatus_4cat i.H_wealthquintiles ///
2244 i.H_smoking_3cat i.H_alcohol_status i.H_cvd_comorbidity ///
2245 i.Hwv8_depressive_symptoms
2246
2247
2248
2249 * 5) complete data (see above)
2250
2251
2252 * ----- *
2253
2254
2255
2256
2257 *** EXTRA SENSITIVITY ANALYSES FOR THE PAPER ***
2258
2259 /*
2260
2261 compare baseline characteristics between complete sample (before exclusion) and sample with
missing data (overall after exclusion)
2262
2263 */
2264
2265

```

```

2266 * General characteristics of ELSA participants at baseline
2267
2268
2269 * Socio-demographics
2270 sum H_age
2271 ta H_sex
2272 ta H_education
2273 ta H_maritalstatus_4cat
2274 ta H_wealthquintiles
2275 * Cardiometabolic disorders
2276 ta Hwv8_crp
2277 ta Hwv8_hdl
2278 ta Hwv8_obesity_waist
2279 ta Hwv8_systolic_bp
2280 ta Hwv8_diastolic_bp
2281 ta Hwv8_diabetes_reportevr
2282 ta Hwv8_HbA1c
2283 * Cardiovascular health factors
2284 ta H_smoking_3cat
2285 ta H_physicalactivity
2286 ta H_alcohol_status
2287 ta H_cvd_comorbidity
2288 * Depressive symptoms
2289 ta Hwv8_depressive_symptoms
2290 * Memory score
2291 sum Hwv8_memory
2292
2293
2294
2295
2296
2297 * compare health characteristics between those survived and dropped out
2298
2299
2300
2301 *** CLEANING DATA to keep those who dropped out
2302
2303
2304 * 1. drop dementia cases at baseline
2305
2306 * drop dementia wave 8 missing data
2307 drop if Hwv8_memory_report==1
2308 * (226 observations deleted)
2309
2310 drop if Hwv8_memory_report== .
2311 * (0 observations deleted)
2312
2313
2314 * 2. drop missing values of cardiometabolic markers
2315
2316 drop if Hwv8_crp== .
2317 * (509 observations deleted)
2318
2319 drop if Hwv8_hdl== .
2320 * (227 observations deleted)
2321
2322 drop if Hwv8_obesity_waist== .
2323 * (193 observations deleted)
2324
2325 drop if Hwv8_systolic_bp== .
2326 * (102 observations deleted)
2327
2328 drop if Hwv8_diastolic_bp== .
2329 * (29 observations deleted)
2330
2331 drop if Hwv8_diabetes_reportevr== .
2332 * (4 observations deleted)
2333

```

```

2334 drop if Hwv8_HbA1c== .
2335 * (76 observations deleted)
2336
2337
2338
2339 * 3. drop obs with no records on dementia at any wave from 9-14 follow-ups
2340
2341
2342 search mdesc
2343 search rmiss2
2344 search mvpatterns
2345
2346 * see number of missing values vs non-missing in each variable
2347 mdesc Hwv9_memory_report Hwv10_anydementia_report Hwv11_anydementia_report ///
2348 Hwv12_anydementia_report Hwv13_anydementia_report Hwv14_anydementia_report
2349
2350
2351
2352 /* number of missing values per observation
2353 * the code below creates a variable called nmisfollowup that gives the number of missing values
2354 for each observation in the variables of interest */
2355 egen nmisfollowup_dementia_wv9to14=rmiss2(Hwv9_memory_report ///
2356 Hwv10_anydementia_report Hwv11_anydementia_report ///
2357 Hwv12_anydementia_report Hwv13_anydementia_report Hwv14_anydementia_report)
2358
2359 tab nmisfollowup_dementia_wv9to14
2360
2361 * drop observations "nmisfollowup_dementia_wv9to14" < 6
2362 drop if nmisfollowup_dementia_wv9to14<6
2363
2364
2365
2366 * Socio-demographics
2367 sum H_age
2368 ta H_sex
2369 ta H_education
2370 ta H_maritalstatus_4cat
2371 ta H_wealthquintiles
2372 * Cardiometabolic disorders
2373 ta Hwv8_crp
2374 ta Hwv8_hdl
2375 ta Hwv8_obesity_waist
2376 ta Hwv8_systolic_bp
2377 ta Hwv8_diastolic_bp
2378 ta Hwv8_diabetes_reportevr
2379 ta Hwv8_HbA1c
2380 * Cardiovascular health factors
2381 ta H_smoking_3cat
2382 ta H_physicalactivity
2383 ta H_alcohol_status
2384 ta H_cvd_comorbidity
2385 * Depressive symptoms
2386 ta Hwv8_depressive_symptoms
2387 * Memory score
2388 sum Hwv8_memory
2389
2390
2391
2392
2393 * compare health characteristics bewteen <70 and >=70
2394
2395
2396
2397 * General baseline characteristics of HRS participants by age group
2398
2399 * Socio-demographics
2400 ttest H_age, by(H_age_group)
2401 ta H_sex H_age_group, chi2 column row

```

```
2402 ta H_education H_age_group, chi2 column row
2403 ta H_maritalstatus_4cat H_age_group, chi2 column row
2404 ta H_wealthquintiles H_age_group, chi2 column row
2405 * Cardiometabolic disorders
2406 ta crp_lca H_age_group, chi2 column row
2407 ta hdl_lca H_age_group, chi2 column row
2408 ta obesity_lca H_age_group, chi2 column row
2409 ta systolic_lca H_age_group, chi2 column row
2410 ta diastolic_lca H_age_group, chi2 column row
2411 ta diabetes_lca H_age_group, chi2 column row
2412 ta hba1c_lca H_age_group, chi2 column row
2413 * Cardiovascular health factors
2414 ta H_smoking_3cat H_age_group, chi2 column row
2415 ta H_physicalactivity H_age_group, chi2 column row
2416 ta H_alcohol_status H_age_group, chi2 column row
2417 ta H_cvd_comorbidity H_age_group, chi2 column row
2418 * Depressive symptoms
2419 ta Hwv8_depressive_symptoms H_age_group, chi2 column row
2420 * Memory score
2421 ttest Hwv8_memory, by(H_age_group)
2422 ta H_lca_group3 H_age_group, chi2 column row
2423
2424
2425
2426
2427 * ----- *
2428
2429
2430
```