

```

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: CHARLS
12 baseline: wave 1 (2011) follow-up waves 2-4 (2013-2018)
13
14
15 TIMELINE
16
17 LATENT CLASSES OF CARDIOMETABOLIC MARKERS: WV1 (BASELINE)
18 DEMENTIA INCIDENCE: W2 - WV4 (3 TIME POINTS)
19 COVARIATES ADJUSTMENT FOR HR MODELS: WV1
20
21 */
22
23
24
25
26 * KEEP NECESSARY VARIABLES
27
28 keep ID id_12char bloodweight ///
29 C_sex C_age C_education C_educ_new C_maritalstatus_8cat C_maritalstatus_3cat C_maritalstatus_4cat
30 Cwv1_netwealth_quintiles ///
31 Cwv1_smoking_2cat Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_freq Cwv1_alcohol_status ///
32 C_cvd_comorbidity Cwv1_antidepressant Cwv1_psycholog_treat Cwv1_anytreat_psych ///
33 Cwv1_memory_wordrecall Cwv1_cognition ///
34 Cwv1_cesd_score Cwv1_depressive_symptoms ///
35 Cwv2_cesd_score Cwv2_depressive_symptoms ///
36 Cwv3_cesd_score Cwv3_depressive_symptoms ///
37 Cwv4_cesd_sumscore Cwv4_depressive_symptoms ///
38 Cwv1_crp_level Cwv1_crp Cwv1_hdl_level Cwv1_male_hdl Cwv1_female_hdl ///
39 Cwv1_meds_dyslipid Cwv1_anymeds_dyslipid Cwv1_dyslipid_evr ///
40 Cwv1_dyslipid_report Cwv1_dyslipid_report_sum Cwv1_dyslipid_report Cwv1_hdl_sum
41 Cwv1_hdl_cholesterol ///
42 Cwv1_waist Cwv1_malewaist_ao Cwv1_femalewaist_ao Cwv1_obesity_waist_sum Cwv1_obesity_waist ///
43 Cwv1_bmi_score Cwv1_obesity_bmi Cwv1_waist_bmi_sum Cwv1_obesity ///
44 Cwv1_tg_level Cwv1_tg Cwv1_triglyc_sum Cwv1_triglyc ///
45 Cwv1_systolic_mean Cwv1_diastolic_mean Cwv1_systolic_bp Cwv1_diastolic_bp ///
46 Cwv1_meds_bp Cwv1_anymeds_bp Cwv1_bp_evr Cwv1_bp_report Cwv1_bp_report_sum Cwv1_bp ///
47 Cwv1_glucose_level Cwv1_glucose Cwv1_HbA1c_level Cwv1_HbA1c ///
48 Cwv1_diabetes_evr Cwv1_diabetes_report Cwv1_diabetes_report_sum Cwv1_diabetes_report Cwv1_meds_diabetes Cwv1_anymeds_diabetes ///
49 Cwv1_diabetes_report Cwv1_meds_diabetes Cwv1_anymeds_diabetes ///
50 Cwv1_diabetes_report Cwv1_meds_diabetes Cwv1_anymeds_diabetes ///
51 Cwv1_interview_date Cwv2_interview_date Cwv3_interview_date Cwv4_interview_date ///
52 Cwv2to4_newdementia_or_lastinter Cwv2to4_dementia_free_date C_time_dementia_months ///
53 Cwv2to4_dementia_sum Cwv2to4_dementia_event ///
54 C_time_dementia_midpoint C_time_dementia_midpoint_final C_time_of_event_dementia
55
56
57
58
59
60
61
62
63 /* Latent class analysis - LCA of cardiometabolic risk factors for dementia
64
65 Useful links:

```

```

66  https://www.stata.com/meeting/uk18/slides/uk18_MacDonald.pdf
67
68  https://www.stata.com/meeting/mexico18/slides/5_Mexico18_Canette.pdf
69
70  https://www.bgsu.edu/content/dam/BGSU/college-of-arts-and-sciences/center-for-family-and-demographi
c-research/documents/Workshops/2020-latent-class-analysis.pdf
71
72  https://www.stata.com/features/overview/latent-class-analysis/
73
74  https://www.stata.com/manuals/semexample50g.pdf
75
76  https://www.stata.com/manuals/semexample51g.pdf
77
78  https://www.stata.com/manuals/semexample52g.pdf
79
80  https://www.ucl.ac.uk/population-health-sciences/sites/population_health_sciences/files/lca.pdf
81
82  https://www.stata.com/manuals/semgsemlclassoptions.pdf
83
84  https://www.stata.com/meeting/nordic-and-baltic17/slides/nordic-and-baltic17_Pitblado.pdf
85
86  https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00920/full
87
88  https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774074
89
90  https://www.statalist.org/forums/forum/general-stata-discussion/general/1412686-calculating-entropy
-for-lca-latent-class-analysis-in-stata-15
91
92  https://www.statalist.org/forums/forum/general-stata-discussion/general/1590174-how-to-calculate-en
tropy-for-lca-with-stata
93
94  https://www.statalist.org/forums/forum/general-stata-discussion/general/1390895-combine-marginsplot
-problem-with-plot-options
95
96
97
98
99  * gsem command to fit a latent class model
100
101  * gsem (var1 var2 var3 <-), logit lclass(C 3)
102
103
104
105  OR TRY
106
107  gsem (var1 var2 var3 <-), logit lclass(C 3) ///
108  startvalues(randompr, draws(20) seed(15) difficult) ///
109  emopts(iterate(30) difficult)
110
111
112
113  Binary variables of cardiometabolic markers measured at wave 2
114
115  CRP: Cwv1_crp
116
117  HDL cholesterol: Cwv1_hdl_cholesterol
118
119  Obesity by waist cir: Cwv1_obesity_waist
120
121  systolic Blood pressure: Cwv1_systolic_bp
122
123  diastolic Blood pressure: Cwv1_diastolic_bp
124
125  Diabetes: Cwv1_diabetes_report
126
127  HbA1c: Cwv1_HbA1c
128
129

```

```

130
131  */
132
133
134
135
136
137
138  *** Descriptive stats of cardiometabolic markers
139
140
141
142  tabulate Cwv1_crp
143  summarize Cwv1_crp
144
145  misstable summarize Cwv1_crp
146  misstable patterns Cwv1_crp
147
148  tabulate Cwv1_hdl_cholesterol
149  summarize Cwv1_hdl_cholesterol
150
151  misstable summarize Cwv1_hdl_cholesterol
152  misstable patterns Cwv1_hdl_cholesterol
153
154  tabulate Cwv1_obesity_waist
155  summarize Cwv1_obesity_waist
156
157  misstable summarize Cwv1_obesity_waist
158  misstable patterns Cwv1_obesity_waist
159
160  tabulate Cwv1_systolic_bp
161  summarize Cwv1_systolic_bp
162
163  misstable summarize Cwv1_systolic_bp
164  misstable patterns Cwv1_systolic_bp
165
166
167  tabulate Cwv1_diastolic_bp
168  summarize Cwv1_diastolic_bp
169
170  misstable summarize Cwv1_diastolic_bp
171  misstable patterns Cwv1_diastolic_bp
172
173
174  tabulate Cwv1_diabetes_report
175  summarize Cwv1_diabetes_report
176
177  misstable summarize Cwv1_diabetes_report
178  misstable patterns Cwv1_diabetes_report
179
180
181  tabulate Cwv1_HbA1c
182  summarize Cwv1_HbA1c
183
184  misstable summarize Cwv1_HbA1c
185  misstable patterns Cwv1_HbA1c
186
187
188  tabulate Cwv1_dementia_report
189  summarize Cwv1_dementia_report
190
191  misstable summarize Cwv1_dementia_report
192  misstable patterns Cwv1_dementia_report
193
194
195
196
197

```

```

198
199
200 *** CLEANING DATA
201
202
203 * 1. drop dementia cases at baseline
204
205 * drop dementia wave 2 missing data
206 drop if Cwv1_dementia_report==1
207 * (267 observations deleted)
208 drop if Cwv1_dementia_report== .
209 * (88 observations deleted)
210
211
212 * 2. drop missing values of cardiometabolic markers
213
214 drop if Cwv1_crp== .
215 * (180 observations deleted)
216
217 drop if Cwv1_hdl_cholesterol== .
218 * (2 observations deleted)
219
220 drop if Cwv1_obesity_waist== .
221 * (1688 observations deleted)
222
223 drop if Cwv1_systolic_bp== .
224 * (83 observations deleted)
225
226 drop if Cwv1_diastolic_bp== .
227 * (13 observations deleted)
228
229 drop if Cwv1_diabetes_report== .
230 * (91 observations deleted)
231
232 drop if Cwv1_HbA1c== .
233 * (71 observations deleted)
234
235
236
237 * 3. drop obs with no records on dementia at any wave from 2-4 follow-ups
238
239
240 search mdesc
241 search rmiss2
242 search mvpatterns
243
244 * see number of missing values vs non-missing in each variable
245 mdesc Cwv2_dementia_report Cwv3_dementia_report Cwv4_self_info_dementia
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_wv2to4=rmiss2(Cwv2_dementia_report ///
253 Cwv3_dementia_report Cwv4_self_info_dementia)
254
255 tab nmisfollowup_dementia_wv2to4
256
257 * drop observations "nmisfollowup_dementia_wv2to4" > 2 (those with 3 missing data = no records at
any wave)
258 drop if nmisfollowup_dementia_wv2to4>2
259 *(342 observations deleted)
260
261
262 * FINAL SAMPLE -> 9022
263
264

```

```

265
266
267
268
269 /* Latent Class analysis - gsem
270
271 * 7 variables: Cwv1_crp Cwv1_hdl_cholesterol Cwv1_obesity_waist
272   Cwv1_systolic_bp Cwv1_diastolic_bp Cwv1_diabetes_report Cwv1_HbA1c
273
274 */
275
276
277 * change names to start with lowercase (STATA assumes variables starting with a capital letter
   are cont latent variables!)
278
279 rename Cwv1_crp crp_lca
280 rename Cwv1_hdl_cholesterol hdl_lca
281 rename Cwv1_obesity_waist obesity_lca
282 rename Cwv1_systolic_bp systolic_lca
283 rename Cwv1_diastolic_bp diastolic_lca
284 rename Cwv1_diabetes_report diabetes_lca
285 rename Cwv1_HbA1c hba1c_lca
286
287
288
289 * Correlation matrix of the CM variables
290
291 corr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca
292
293 pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
294
295
296 * to create quality table in word - asdoc -
297 * https://www.youtube.com/watch?v=XHBl6PHf0zs&ab\_channel=StataProfessor
298
299 help asdoc
300
301 asdoc pwcorr crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca hba1c_lca, sig
302
303
304 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)
305
306
307
308 * LCA models
309
310
311 * one-class model
312
313 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
314   hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 1)
315
316 estimates store oneclass_cm
317
318 * two-class model
319
320 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
321   hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 2)
322
323
324 estimates store twoclass_cm
325
326
327 * three-class model
328
329 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///

```

```

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

```

```

397  * for a well-fitted model p value should be non-sig. A significant p-value indicates lack of
398  model fit in absolute terms.
399  * 3-class model:  $p < 0.001$ 
400
401  * Entropy
402
403  quietly predict classpost*, classposteriorpr
404  gen sum_p_lnp = 0
405  forvalues k = 1/2 {
406      replace sum_p_lnp = sum_p_lnp + classpost`k'*ln(classpost`k')
407  }
408  summ sum_p_lnp, meanonly
409  scalar E = 1+`r(sum)'/ (e(N)*ln(2))
410  drop classpost? sum_p_lnp
411  di E
412
413
414
415
416
417
418
419
420  /* We can use the predictions of the posterior probability of class membership to evaluate an
421  individual's probability of being in each class.
422
423  */
424
425  predict classpost1*, classposteriorpr
426  list in 1, abbrev(10)
427
428  /* We can determine the expected class for each individual based on whether the posterior
429  probability
430  is greater than 0.5
431  */
432  generate expclass1 = 1 + (classpost11>0.5)
433  tabulate expclass
434
435
436  generate expclass2 = 1 + (classpost12>0.5)
437  tabulate expclass2
438
439
440  generate expclass3 = 1 + (classpost13>0.5)
441  tabulate expclass3
442
443
444
445  /* We can determine expected classification for each individual in the dataset based on the
446  predicted
447  posterior class probabilities.
448  */
449  predict cpost*, classposteriorpr
450  egen max = rowmax(cpost*)
451
452
453  * generate classes var
454
455  generate predclass = 1 if cpost1==max
456
457  replace predclass = 2 if cpost2==max
458
459  replace predclass = 3 if cpost3==max
460
461  tabulate predclass

```

```

462
463
464
465 *****function to print out summary stats
466 program summary_table_proclCla
467     preserve
468     *look at the average posterior probability
469     gen Mp = 0
470     foreach i of varlist cpost* {
471         replace Mp = `i' if `i' > Mp
472     }
473     sort predclass
474     *and the odds of correct classification
475     by predclass: gen countG = _N
476     by predclass: egen groupAPP = mean(Mp)
477     by predclass: gen counter = _n
478     gen n = groupAPP/(1 - groupAPP)
479     gen p = countG/_N
480     gen d = p/(1-p)
481     gen occ = n/d
482     *Estimated proportion for each group
483     scalar c = 0
484     gen TotProb = 0
485     foreach i of varlist cpost* {
486         scalar c = c + 1
487         quietly summarize `i'
488         replace TotProb = r(sum)/ _N if predclass == c
489     }
490     gen d_pp = TotProb/(1 - TotProb)
491     gen occ_pp = n/d_pp
492     *This displays the group number [_traj_~p],
493     *the count per group (based on the max post prob), [countG]
494     *the average posterior probability for each group, [groupAPP]
495     *the odds of correct classification (based on the max post prob group assignment), [occ]
496     *the odds of correct classification (based on the weighted post. prob), [occ_pp]
497     *and the observed probability of groups versus the probability [p]
498     *based on the posterior probabilities [TotProb]
499     list predclass countG groupAPP occ occ_pp p TotProb if counter == 1
500     restore
501 end
502
503 summary_table_proclCla
504
505
506
507
508 /*
509 Class variable manipulation
510 */
511
512 * 3-class model: rename predclass to C_lca_group3
513
514 rename predclass C_lca_group3
515
516 * labelling variable of C_lca_group3
517
518 label var C_lca_group3 "Latent classes 3 groups of cardiometabolic markers"
519
520 * labelling values
521 lab def lca_group3 1 "Relatively healthy" 2 "Obesity and Hypertension" 3 "Complex cardiometabolic disorders"
522
523 * attach category labels to the variable through label value
524
525 lab val C_lca_group3 lca_group3
526
527 ta C_lca_group3
528

```



```

529
530
531 * margins and marginsplot
532
533 * use margins to calculate marginal predictions
534 * use marginsplot to graph marginal predictions
535
536
537 margins, predict(classpr class(1)) ///
538             predict(classpr class(2)) ///
539             predict(classpr class(3))
540
541 marginsplot, xtitle ("") ytitle ("") ///
542             xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
543             title ("Predicted Latent Class Probabilities with 95% CI")
544
545
546 margins, predict(classpr class(1)) ///
547             predict(classpr class(2)) ///
548             predict(classpr class(3))
549 marginsplot, recast(bar) xtitle("") ytitle("") ///
550             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
551             title("Predicted Latent Class Probabilities with 95% CI")
552
553 margins, predict(outcome(hba1c_lca) class(1)) ///
554             predict(outcome(hba1c_lca) class(2)) ///
555             predict(outcome(hba1c_lca) class(3))
556 marginsplot, recast(bar) xtitle("") ytitle("") ///
557             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
558             title("Predicted Pr(HbA1c=1) with 95% CI")
559
560 * repeat with all CM var
561
562
563
564 * class 1
565
566 margins, predict(outcome(crp_lca) class(1)) ///
567             predict(outcome(hdl_lca) class(1)) ///
568             predict(outcome(obesity_lca) class(1)) ///
569             predict(outcome(systolic_lca) class(1)) ///
570             predict(outcome(diastolic_lca) class(1)) ///
571             predict(outcome(diabetes_lca) class(1)) ///
572             predict(outcome(hba1c_lca) class(1)) ///
573
574
575
576 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
577             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
578             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
579             ytitle("Predicted mean") ylabel(0(.20)1) name (class1)
580
581
582 * class 2
583
584 margins, predict(outcome(crp_lca) class(2)) ///
585             predict(outcome(hdl_lca) class(2)) ///
586             predict(outcome(obesity_lca) class(2)) ///
587             predict(outcome(systolic_lca) class(2)) ///
588             predict(outcome(diastolic_lca) class(2)) ///
589             predict(outcome(diabetes_lca) class(2)) ///
590             predict(outcome(hba1c_lca) class(2)) ///
591
592
593
594 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
595             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
596             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///

```

```

597             ytitle("Predicted mean") ylabel(0(.20)1) name (class2)
598
599
600
601
602 * class 3
603
604 margins, predict(outcome(crp_lca) class(3)) ///
605             predict(outcome(hdl_lca) class(3)) ///
606             predict(outcome(obesity_lca) class(3)) ///
607             predict(outcome(systolic_lca) class(3)) ///
608             predict(outcome(diastolic_lca) class(3)) ///
609             predict(outcome(diabetes_lca) class(3)) ///
610             predict(outcome(hba1c_lca) class(3)) ///
611
612
613
614 marginsplot, recast(bar) title ("Class 3") xtitle("") ///
615             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
616                 5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
617             ytitle("Predicted mean") ylabel(0(.20)1) name (class3)
618
619
620
621 graph combine class1 class2 class3, cols(3)
622
623
624
625
626
627 * ----- *
628
629
630
631
632 /* MULTIPLE IMPUTATION (MI)
633
634 To handle with missing values of baseline and time 3 covariates
635
636
637 useful sources for MI and MICE:
638
639 https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
640 https://www.stata.com/manuals/mi.pdf - see page 139
641 https://www.stata.com/meeting/switzerland16/slides/medeiros-switzerland16.pdf
642 https://www.youtube.com/watch?v=i6S0lq0mjuc&ab\_channel=StataCorpLLC
643 https://dss.princeton.edu/training/MIStata.pdf
644
645
646
647 Preparing to conduct MI
648 1. examine the number and proportion of missing values among the variables of interest
649     use the mdisc command
650 2. examine missing data patterns
651     use commands mi set and mi misstable patterns
652 3. identify potential auxiliary variables
653
654
655 Run MI using chained equations (MICE)
656 using the commands
657 1. how (in what style) to store the imputations
658     mi set wide
659 2. which variables will be imputed
660     mi register imputed
661 3. optionally, which variables will not be imputed
662     mi register regular
663 4. what imputation method is implemented to impute each of var - MICE
664     mi impute chained

```

```

665
666 */
667
668
669
670
671
672 /*
673
674 1. examining missing values
675     install packages:
676     * install mdesc
677     * install tabmiss
678     * insatll dm31
679     * insall mvpatterna
680
681 */
682
683 search mdesc
684 search rmiss2
685 search mvpatterns
686
687
688
689
690 * examining number of missing values vs non-missing in each variable
691
692 mdesc C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
693 Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
694 C_cvd_comorbidity Cwv1_memory_wordrecall Cwv1_depressive_symptoms
695
696
697
698 * examining missing data patterns
699
700 mi set wide
701
702 mi misstable summarize C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
703 Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
704 C_cvd_comorbidity ///
705 Cwv1_memory_wordrecall Cwv1_depressive_symptoms
706
707
708
709 mi misstable patterns C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
710 Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status ///
711 C_cvd_comorbidity ///
712 Cwv1_memory_wordrecall Cwv1_depressive_symptoms
713
714
715 /*
716     identifying potential auxiliary var
717     * Auxiliary variables are either correlated with a missing variable(s)
718     (the recommendation is  $r > 0.4$ ) or are believed to be associated with missingness
719     - a priori knowledge of var that would make good auxiliary var
720     - identify potential candidates by examining associations between missing var and other var in
       the dataset
721         running correlation using the command: pwcorr v1 v2 v3, obs
722         the recommnedation for good correlation is  $r > 0.4$ 
723
724
725 Missing var to be imputed:
726
727     Cwv1_netwealth_quintiles
728     Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity
729     Cwv1_memory_wordrecall Cwv1_depressive_symptoms
730
731

```

```

732
733 Potential auxiliary var:
734 DV: Cwv2to4_dementia_event
735 IV: crp_lca hdl_lca obesity_lca Cwv1_tg systolic_lca diastolic_lca hba1c_lca diabetes_lca
736 other var: C_age C_sex C_education C_maritalstatus_4cat
737
738 */
739
740
741 * correlation
742
743 pwcorr Cwv1_netwealth_quintiles ///
744         Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
745         Cwv1_memory_wordrecall Cwv1_depressive_symptoms ///
746         Cwv2to4_dementia_event crp_lca hdl_lca obesity_lca Cwv1_tg systolic_lca diastolic_lca
747         hba1c_lca diabetes_lca ///
748         C_age C_sex C_education C_maritalstatus_4cat, obs
749
750 /* The correlation showed that all the following var are good auxiliary:
751 Cwv2to4_dementia_event obesity_lca diabetes_lca C_age C_sex C_education C_maritalstatus_4cat
752 * A good auxiliary does not have to be correlated with every variable to be useful
753 * And it's not problematic if it has missing info of it's own
754 */
755
756
757
758 /*
759 MI by chained equations (MICE)
760     see: https://stats.idre.ucla.edu/stata/seminars/mi\_in\_stata\_pt1\_new/
761
762 MICE is known as the fully conditional specification or sequential generalized regression
763 does not assume a joint MVN distribution
764 but instead uses a separate conditional distribution for each imputed variable.
765
766 The multivariate normal (MVN) model - mi imputed mvn -
767 assumes multivariate normality of all var
768
769 The multivariate imputation by chained equations (MICE) - mi imputed chained -
770 offers flexibility in how each var is modeled
771
772 mi impute chained allows to specify models for a
773 variety of variable types, including
774 continuous, binary, ordinal, nominal, truncated, and count variables
775
776
777 The MICE distributions available in Stata are:
778 binary, ordered and multinomial logistic regression for categorical variables,
779 linear regression and predictive mean matching (PMM)* for continuous variables,
780 and Poisson and negative binomial regression for count variables.
781
782
783
784 IMPUTATION PHASES
785
786 1. mi set wide
787     style to store imputations
788
789 2. mi register imputed
790     identifies which variables in the imputation model have missing information.
791
792 3. mi register regular (! optional)
793     which variables will not be imputed
794
795 4. mi impute chained
796     where the user specifies the imputation model to be used
797     and the number of imputed datasets to be created.
798     Example:

```

```

799         mi impute chained (regress) bmi age (logit) female ///
800         (mlogit) race = bpdiastr i.region, add(20)
801
802 5. mi estimate
803     is used as a prefix to the standard regress command.
804     This executes the specified estimation model within each of the 20 imputed datasets
805     to obtain 20 sets of coefficients and standard errors.
806     Stata then combines these estimates to obtain one set of inferential statistics.
807     In the output from mi estimate you will see some metrics: Imputation Diagnostics
808     information for RVI (Relative Increase in Variance),
809     FMI (Fraction of Missing Information),
810     DF (Degrees of Freedom) ,
811     RE (Relative Efficiency),
812     and the between imputation and the within imputation variance estimates
813     to examine how the standard errors (SEs) are calculated.
814
815
816 -----
817
818
819 SELECTING MY IMPUTATION MODEL
820
821 - MICE -> mi impute chained
822
823 - var to be imputed:
824
825     linear regression for continuous var (regress) ->
826     C_age Cwv1_memory_wordrecall
827
828     logistic for the binary var (logit) ->
829     C_cvd_comorbidity Cwv1_depressive_symptoms
830
831     multinomial logistic for our nominal categorical var (mlogit) ->
832     Cwv1_netwealth_quintiles
833     Cwv1_smoking_3cat Cwv1_alcohol_status
834
835
836
837 - auxiliary var:
838
839     DV -> Cwv2to4_dementia_event
840     IV -> obesity_lca diabetes_lca
841     other covariates -> C_age C_sex C_education C_maritalstatus_4cat
842
843
844
845 - imputation numbers (m) -> 20
846
847     White et al. (2010) recommendation: use the rule that m should equal the percentage of
848     incomplete cases
849
850
851 - rseed (53421) for reproducibility reasons
852
853 - (! OPTIONAL) advance impute options -> force
854
855     proceed with imputation, even when missing imputed values (e.g. auxiliary have missing data)
856     are encountered
857
858 - impute options -> savetrace (trace1)
859
860     specifies Stata to save the means and standard deviations of imputed values from each
861     iteration to a Stata dataset named "trace1
862 */
863

```

```

864  mi set wide
865
866
867  mi register imputed Cwv1_netwealth_quintiles ///
868      Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
869      Cwv1_memory_wordrecall Cwv1_depressive_symptoms
870
871
872
873  mi impute chained (logit) C_cvd_comorbidity Cwv1_depressive_symptoms ///
874      (mlogit) Cwv1_netwealth_quintiles Cwv1_smoking_3cat Cwv1_alcohol_status ///
875      (regress) Cwv1_memory_wordrecall = Cwv2to4_dementia_event obesity_lca diabetes_lca ///
876      C_age C_sex C_eduaction C_maritalstatus_4cat, add(20) rseed(53421) savetrace(trace1)
877
878
879  * save imputed data
880
881  * plot imputations
882
883
884  *it will open a file named trace1
885  use trace1, clear
886
887  describe
888
889
890  reshape wide *mean *sd, i(iter) j(m)
891
892  tsset iter
893
894
895
896
897  /*
898  The trace plot below graphs the predicted means value produced during the first imputation chain.
899  As before, the expectations is that the values would vary randomly to incorporate variation into
900  the predicted values for read.
901  */
902
903  tsline Cwv1_netwealth_quintiles_mean1, name(mice1,replace)legend(off) ytitle("Mean of wealth")
904  tsline Cwv1_smoking_3cat_mean1, name(mice1,replace)legend(off) ytitle("Mean of smoking")
905  tsline Cwv1_alcohol_status_mean1, name(mice1,replace)legend(off) ytitle("Mean of alcohol")
906  tsline C_cvd_comorbidity_mean1, name(mice1,replace)legend(off) ytitle("Mean of cvd")
907  tsline Cwv1_depressive_symptoms_mean1, name(mice1,replace)legend(off) ytitle("Mean of depression")
908  tsline Cwv1_memory_wordrecall_mean1, name(mice1,replace)legend(off) ytitle("Mean of memory")
909
910  /*
911
912  All 20 imputation chains can also be graphed simultaneously to make sure that nothing unexpected
913  occurred in a single chain.
914  Every chain is obtained using a different set of initial values and this should be unique.
915  Each colored line represents a different imputation.
916  So all 10 imputation chains are overlaid on top of one another.
917  */
918
919
920  tsline Cwv1_memory_wordrecall_mean*, name(mice1,replace)legend(off) ytitle("Mean of memory")
921  tsline Cwv1_memory_wordrecall_sd*, name(mice2, replace) legend(off) ytitle("SD of memory")
922  graph combine mice1 mice2, xcommon cols(1) title(Trace plots of summaries of imputed values)
923
924  * repeat for each imputed var
925
926
927
928
929

```

```

930
931  /*
932  ---- DESCRIPTIVE STATISTICS (baseline and time 3 covariates) ----
933
934  General characteristics of participants
935
936  General characteristics of participnats stratified for study inclusion
937
938  General characteristics of participants stratified for dementia occurence
939
940  Participant characteristics by CM 3-class groups
941
942  1. CHI-SQUARE (chi2) for categorical var (crosstabulation)
943      Frequency tables -> two-way tables
944          using the command tabulate, chi2
945          reporting observations, column percentage (N, %) and p-value of Pearson's r
946
947
948  2. one-way ANOVA for continuous var
949      check box plot
950      using the command oneway
951      reporting mean, sd (summary tables) and p-value of F
952  */
953
954
955
956
957  * General characteristics of CHARLS participants at baseline
958
959  * Socio-demographics
960  sum C_age
961  ta C_sex
962  ta C_educ_new
963  ta C_maritalstatus_4cat
964  ta Cwv1_netwealth_quintiles
965  * Cardiometabolic disorders
966  ta crp_lca
967  ta hdl_lca
968  ta obesity_lca
969  ta systolic_lca
970  ta diastolic_lca
971  ta diabetes_lca
972  ta hba1c_lca
973  * Cardiovascular health factors
974  ta Cwv1_smoking_3cat
975  ta Cwv1_alcohol_status
976  ta Cwv1_physicalactivity
977  ta C_cvd_comorbidity
978  * Depressive symptoms (cont and categ)
979  sum Cwv1_cesd_score
980  ta Cwv1_depressive_symptoms
981  * Memory score
982  sum Cwv1_memory_wordrecall
983
984
985
986  * General baseline characteristics of CHARLS participants by dementia status
987
988  * Socio-demographics
989  ttest C_age, by(Cwv2to4_dementia_event)
990  ta C_sex Cwv2to4_dementia_event, chi2 column row
991  ta C_educ_new Cwv2to4_dementia_event, chi2 column row
992  ta C_maritalstatus_4cat Cwv2to4_dementia_event, chi2 column row
993  ta Cwv1_netwealth_quintiles Cwv2to4_dementia_event, chi2 column row
994  * Cardiometabolic disorders
995  ta crp_lca Cwv2to4_dementia_event, chi2 column row
996  ta hdl_lca Cwv2to4_dementia_event, chi2 column row
997  ta obesity_lca Cwv2to4_dementia_event, chi2 column row

```



```

998 ta systolic_lca Cwv2to4_dementia_event, chi2 column row
999 ta diastolic_lca Cwv2to4_dementia_event, chi2 column row
1000 ta diabetes_lca Cwv2to4_dementia_event, chi2 column row
1001 ta hba1c_lca Cwv2to4_dementia_event, chi2 column row
1002 * Cardiovascular health factors
1003 ta Cwv1_smoking_3cat Cwv2to4_dementia_event, chi2 column row
1004 ta Cwv1_alcohol_status Cwv2to4_dementia_event, chi2 column row
1005 ta Cwv1_physicalactivity Cwv2to4_dementia_event, chi2 column row
1006 ta C_cvd_comorbidity Cwv2to4_dementia_event, chi2 column row
1007 * Depressive symptoms (cont and categ)
1008 ttest Cwv1_cesd_score, by(Cwv2to4_dementia_event)
1009 ta Cwv1_depressive_symptoms Cwv2to4_dementia_event, chi2 column row
1010 * Memory score
1011 ttest Cwv1_memory_wordrecall, by(Cwv2to4_dementia_event)
1012
1013
1014
1015
1016 * Sample characteristics by CM 3-class groups
1017 * crosstabs categ var (frequencies and chi2) !report column percentage!
1018 * oneway ANOVA cont var (mean, sd)
1019
1020
1021 * Socio-demographics
1022 oneway C_age C_lca_group3, tabulate
1023 ta C_sex C_lca_group3, chi2 column row
1024 ta C_educ_new C_lca_group3, chi2 column row
1025 ta C_maritalstatus_4cat C_lca_group3, chi2 column row
1026 ta Cwv1_netwealth_quintiles C_lca_group3, chi2 column row
1027 * Cardiovascular health factors
1028 ta Cwv1_smoking_3cat C_lca_group3, chi2 column row
1029 ta Cwv1_alcohol_status C_lca_group3, chi2 column row
1030 ta Cwv1_physicalactivity C_lca_group3, chi2 column row
1031 ta C_cvd_comorbidity C_lca_group3, chi2 column row
1032 * Depressive symptoms (cont and categ)
1033 oneway Cwv1_cesd_score C_lca_group3, tabulate
1034 ta Cwv1_depressive_symptoms C_lca_group3, chi2 column row
1035 * Memory score
1036 oneway Cwv1_memory_wordrecall C_lca_group3, tabulate
1037
1038
1039
1040
1041
1042
1043
1044
1045 /*
1046 ---- SURVIVAL ANALYSIS AT COMPLETE DATA ----
1047
1048 Tests of proportional-hazards assumption
1049 Kaplan Meier survival curves
1050 Person-time
1051 Cox proportional regression - Hazard ratios - stcox
1052 Postestimation tools for stcox
1053 Test of Goodness of Fit
1054
1055 *** Cox regression in full data, complete data (listwise deletion of missing data) and imputed data
1056 Cox PH regression in complete data
1057 Cox PH regression model in imputed dataset - mi estimate
1058
1059
1060 */
1061
1062
1063
1064 * check dataset variables of interest only
1065

```



```

1066 codebook C_time_of_event_dementia Cwv2to4_dementia_event C_lca_group3 ///
1067 C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
1068 Cwv1_smoking_3cat Cwv1_physicalactivity Cwv1_alcohol_status C_cvd_comorbidity ///
1069 Cwv1_depressive_symptoms Cwv1_memory_wordrecall,compact
1070
1071
1072
1073
1074 * Declare Data to be Survival Data
1075 * Time to event: C_time_of_event_dementia (months)
1076 * Censoring: Cwv2to4_dementia_event (1=dementia, 0=censored)
1077 * Command is stset TIMETOEVENT, failure(CENSORVARIABLE)
1078
1079
1080 stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
1081
1082
1083
1084 *describe survival data using commnad stsum
1085
1086 stsum
1087
1088 stsum, by(C_lca_group3)
1089
1090
1091
1092
1093 * Kaplan Meier Curve estimation
1094
1095 sts list
1096
1097 sts list, by(C_lca_group3)
1098
1099
1100
1101 * Kaplan Meier Curve Plot
1102
1103 * no frills plot
1104
1105 sts graph
1106
1107 * with frills
1108
1109 sts graph, xtitle("Time in Months") ytitle("Survival Prob") ///
1110 title("Kaplan Meier Curve") subtitle("n=9022, # events=470") ///
1111 caption("graph02.png", size(vsmall))
1112
1113
1114 * With Greenwood CI limits
1115
1116 sts graph, gwood legend(off) xtitle("Time in Months") ytitle("Survival Prob") ///
1117 title("Kaplan Meier Curve") subtitle("n=9022, # events=470") caption("graph03.png", size(vsmall))
1118
1119
1120
1121
1122 * Group Kaplan-Meier Curve Estimation
1123 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1124
1125 sort C_lca_group3
1126
1127 sts list, by(C_lca_group3)
1128
1129 * graph with frills
1130
1131 sts graph, by(C_lca_group3) xlabel(0(20)100) ylabel(0.80(.05)1) xtitle("Time in Months") ///
1132 ytitle("Survival Prob") title("Kaplan Meier Curve") subtitle("n=9022, # events=470") ///
1133 caption("graph04.png", size(vsmall))

```

```

1134
1135
1136
1137
1138 * calculate person-time and incidence rates using command stptime
1139
1140 stptime, title(Person-years)
1141
1142 stptime, title(Person-years) per(1000)
1143
1144 stptime, title(Person-years) per(10000)
1145
1146
1147 * calculate person-time by category of C_lca_group3
1148
1149 stptime, by(C_lca_group3)
1150
1151 stptime, by(C_lca_group3) per(1000)
1152
1153
1154
1155 * mean and median of follow-up
1156 sum C_time_of_event_dementia
1157 sum C_time_of_event_dementia, detail
1158
1159
1160
1161 /* Log Rank Test of equality of survival distributions
1162 (NULL: equality of survival distributions among C_lca_group3 groups)
1163 We will consider including the predictor if the test has a p-value of 0.2 - 0.25 or less.
1164 If the predictor has a p-value greater than 0.25 in a univariate analysis
1165 it is highly unlikely that it will contribute anything to a model which includes other
predictors.
1166 Command is sts test GROUPVAR
1167 */
1168
1169
1170
1171 sts test C_lca_group3, logrank
1172
1173 sts test C_age, logrank
1174
1175 sts test C_sex, logrank
1176
1177 sts test C_eduaction, logrank
1178
1179 sts test C_maritalstatus_4cat, logrank
1180
1181 sts test Cwv1_netwealth_quintiles, logrank
1182
1183 sts test Cwv1_smoking_3cat, logrank
1184
1185 sts test Cwv1_physicalactivity, logrank
1186
1187 sts test Cwv1_alcohol_status, logrank
1188
1189 sts test C_cvd_comorbidity, logrank
1190
1191 sts test Cwv1_depressive_symptoms, logrank
1192
1193 sts test Cwv1_memory_wordrecall, logrank
1194
1195
1196
1197
1198
1199
1200

```

```

1201
1202  /* Cox PH regression model
1203
1204  using the command stcox
1205
1206  --- Building the model ---
1207
1208  Model 1: unadjusted - single predictor of CM classes
1209  Model 2: model 1 + sociodemographics: age sex education marital status and wealth
1210  Model 3: model 2 + cvd health: smoking, alcohol consumption, cvd comorbidity
1211  Model 4: model 3 + mental health: depressive symptoms
1212  */
1213
1214
1215  * Unadjusted model - model 1 - single predictor
1216
1217  stcox C_lca_group3
1218
1219  * define design var by using i.(3 classes)
1220
1221  stcox i.C_lca_group3
1222
1223
1224  * Adjusted models - multivariable Cox model
1225  * controlling for covariates
1226
1227  * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1228
1229  stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1230
1231  * model 3: model 2 + adjust for cvd health
1232
1233  stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1234  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1235
1236
1237  * model 4: model 3 + adjust for depression
1238
1239  stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1240  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1241  i.Cwv1_depressive_symptoms
1242
1243
1244
1245
1246
1247  * Coefficients instead of hazard ratios by specifying the option nohr
1248
1249  stcox i.C_lca_group3, nohr
1250
1251
1252  stcox i.C_lca_group3 C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1253  ///
1254  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1255  i.Cwv1_depressive_symptoms, nohr
1256
1257
1258
1259  * Multivariable model development
1260  * Likelihood-ratio tests
1261
1262
1263
1264  *install eststo
1265  findit eststo
1266
1267

```

```

1268 * ---- rx controlling for age and sex -----*
1269 quietly: stcox C_age i.C_sex
1270 eststo modelagesex
1271
1272 quietly: stcox C_age i.C_sex i.C_lca_group3
1273 eststo modelagesex_3group
1274
1275 lrtest modelagesex modelagesex_3group
1276
1277
1278
1279 * ---- rx controlling for sociodemographics -----*
1280 quietly: stcox C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1281 eststo modelsociodemo
1282
1283 quietly: stcox C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles i.
1284 C_lca_group3
1285 eststo modelsociodemo_3group
1286
1287 lrtest modelsociodemo modelsociodemo_3group
1288
1289 * ---- rx controlling for cardiovascular health -----*
1290 quietly: stcox i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1291 eststo modelcardiovascular
1292
1293 quietly: stcox i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity i.C_lca_group3
1294 eststo modelcardiovascular_3group
1295
1296 lrtest modelcardiovascular modelcardiovascular_3group
1297
1298
1299 * ---- rx controlling for mental health-----*
1300 quietly: stcox i.Cwv1_depressive_symptoms Cwv1_memory_wordrecall
1301 eststo modelmentalcogn
1302
1303 quietly: stcox i.Cwv1_depressive_symptoms i.C_lca_group3
1304 eststo modelmentalcogn_3group
1305
1306 lrtest modelmentalcogn modelmentalcogn_3group
1307
1308
1309
1310
1311 * side-by-side comparison of models
1312
1313
1314 quietly: stcox i.C_lca_group3
1315 eststo model1
1316
1317 quietly: stcox C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles i.
1318 C_lca_group3
1319 eststo model2
1320
1321 quietly: stcox C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1322 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity i.C_lca_group3
1323 eststo model3
1324
1325 quietly: stcox C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1326 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1327 i.Cwv1_depressive_symptoms i.C_lca_group3
1328 eststo model4
1329
1330
1331 * Display Betas and Summary Statistics
1332 estout model1 model2 model3 model4, stats(n chi2 bic, star(chi2)) prehead("Betas")
1333

```

```

1334  /* Key Interpretation
1335  Chi2 = Value of LR test comparing the model fit ("full") to intercept only ("reduced")
1336  bic = Schwarz' Bayesian Information Criterion = It is a function of the log-likelihood.
1337  Smaller values indicate a better fit.
1338  */
1339
1340  * Display Hazard Ratios and Model Fit Statistics. Option eform produces hazard ratios
1341  estout model1 model2 model3 model4, eform stats(n chi2 bic, star(chi2)) prehead("Hazard Ratios")
1342
1343
1344
1345
1346  * Postestimation tools for stcox
1347
1348  * Test of proportional hazards
1349
1350  estat phtest, detail
1351
1352
1353  /* Proportionality Assumption - method 1
1354  We will check proportionality by including time-dependent covariates in the model
1355  by using the tvc and the texp options in the stcox command.
1356  Time dependent covariates are interactions of the predictors and time.
1357  In this analysis we choose to use the interactions with log(time)
1358  because this is the most common function of time used in time-dependent covariates
1359  but any function of time could be used.
1360  If a time-dependent covariate is significant this indicates
1361  a violation of the proportionality assumption for that specific predictor.
1362  The conclusion is that all of the time-dependent variables are not significant
1363  either collectively or individually thus supporting the assumption of proportional hazard.
1364  */
1365
1366
1367
1368  stcox i.C_lca_group3 C_age i.C_sex i.C_education i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1369  ///
1370  i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1371  i.Cwv1_depressive_symptoms, nohr ///
1372  tvc(C_lca_group3 C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
1373  Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
1374  Cwv1_depressive_symptoms) texp(ln(C_time_of_event_dementia))
1375
1376
1377  /* Proportionality Assumption - method 2
1378  by using the Schoenfeld and scaled Schoenfeld residuals
1379  In the stphtest command we test the proportionality of the model as a whole
1380  and by using the detail option we get a test of proportionality for each predictor.
1381  By using the plot option we can also obtain a graph of the scaled Schoenfeld assumption.
1382  If the tests in the table are not significance (p-values over 0.05)
1383  then we can not reject proportionality and we assume
1384  that we do not have a violation of the proportional assumption.
1385  The stphplot command uses log-log plots to test proportionality
1386  and if the lines in these plots are parallel then we have further indication
1387  that the predictors do not violate the proportionality assumption.
1388  */
1389
1390  quietly stcox C_lca_group3 C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles
1391  ///
1392  Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
1393  Cwv1_depressive_symptoms Cwv1_memory_wordrecall, schoenfeld(sch*) scaledsch(sca*)
1394  stphtest, detail
1395  stphtest, plot(C_lca_group3) msym(oh)
1396  stphtest, plot(C_age) msym(oh)
1397  stphtest, plot(C_sex) msym(oh)
1398  stphtest, plot(C_education) msym(oh)
1399  stphtest, plot(C_maritalstatus_4cat) msym(oh)
1400  stphtest, plot(Cwv1_netwealth_quintiles) msym(oh)

```

```

1400 stptest, plot(C_cvd_comorbidity) msym(oh)
1401 stptest, plot(Cwv1_smoking_3cat) msym(oh)
1402 stptest, plot(Cwv1_alcohol_status) msym(oh)
1403 stptest, plot(Cwv1_depressive_symptoms) msym(oh)
1404
1405
1406
1407
1408
1409 stphplot, by(C_lca_group3) plot1(msym(oh)) plot2(msym(th))
1410 stphplot, by(C_age) plot1(msym(oh)) plot2(msym(th))
1411 stphplot, by(C_sex) plot1(msym(oh)) plot2(msym(th))
1412 stphplot, by(C_education) plot1(msym(oh)) plot2(msym(th))
1413 stphplot, by(C_maritalstatus_4cat) plot1(msym(oh)) plot2(msym(th))
1414 stphplot, by(Cwv1_netwealth_quintiles) plot1(msym(oh)) plot2(msym(th))
1415 stphplot, by(C_cvd_comorbidity) plot1(msym(oh)) plot2(msym(th))
1416 stphplot, by(Cwv1_smoking_3cat) plot1(msym(oh)) plot2(msym(th))
1417 stphplot, by(Cwv1_alcohol_status) plot1(msym(oh)) plot2(msym(th))
1418 stphplot, by(Cwv1_depressive_symptoms) plot1(msym(oh)) plot2(msym(th))
1419
1420
1421
1422 * Assessment of PH Assumption: adjust for age and sex
1423 stphplot, by(C_lca_group3) adjust(C_age C_sex) nolntime plotlopts(symbol(none) color(black)
lpattern(dash)) ///
1424 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1425 title("Assessment of PH Assumption") subtitle(" Predictor is C_lca_group3") xtitle("months")
1426
1427
1428
1429 * Assessment of PH Assumption: adjust for model 2
1430 stphplot, by(C_lca_group3) adjust(C_age C_sex C_education C_maritalstatus_4cat
Cwv1_netwealth_quintiles) ///
1431 nolntime plotlopts(symbol(none) color(black) lpattern(dash)) ///
1432 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1433 title("Assessment of PH Assumption") subtitle(" Predictor is C_lca_group3") xtitle("months")
1434
1435
1436
1437 * Assessment of PH Assumption: adjust for model 3
1438 stphplot, by(C_lca_group3) adjust(C_age C_sex C_education C_maritalstatus_4cat
Cwv1_netwealth_quintiles
Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity) ///
1439 nolntime plotlopts(symbol(none) color(black) lpattern(dash)) ///
1440 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1441 title("Assessment of PH Assumption") subtitle(" Predictor is C_lca_group3") xtitle("months")
1442
1443
1444
1445
1446 * Assessment of PH Assumption: adjust for model 4
1447 stphplot, by(C_lca_group3) adjust(C_age C_sex C_education C_maritalstatus_4cat
Cwv1_netwealth_quintiles
Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity
Cwv1_depressive_symptoms) ///
1448 nolntime plotlopts(symbol(none) color(black) lpattern(dash)) ///
1449 plot2opts(symbol(none) color(green)) plot3opts(symbol(none) color(red)) ///
1450 title("Assessment of PH Assumption") subtitle(" Predictor is C_lca_group3") xtitle("months")
1451
1452
1453
1454
1455
1456
1457
1458 /* Test of overall goodness of fit
1459 Goodness of fit of the final model
1460 2 methods:
1461 - by using the commnad stcoxgof (good fit = non sig p-value)
1462 - by using the Cox-Snell residuals
1463 to create the Nelson-Aalen cumulative hazard function

```

```

1464     If the hazard function follows the 45 degree line then we know that it approximately
1465     has an exponential distribution with a hazard rate of one and that the model fits the data
1466     well.
1467     If the model fits the data, the plot of the cumulative hazard versus cs
1468     should approximate a straight line with slope 1.
1469 */
1470
1471 * by using the commnad stcoxgof
1472
1473 * install stcoxgof
1474 findit stcoxgof
1475
1476
1477 stcox C_lca_group3 C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles ///
1478 Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
1479 Cwv1_depressive_symptoms, mgale(mgale)
1480
1481
1482 stcoxgof
1483
1484 * Good. Do not reject. We do not have statistically significant evidence of a poor fit (p-value =
1485 .38).
1486
1487
1488
1489 * by using the Cox-Snell residuals
1490
1491 quietly stcox C_lca_group3 C_age C_sex C_education C_maritalstatus_4cat Cwv1_netwealth_quintiles
1492 ///
1493 Cwv1_smoking_3cat Cwv1_alcohol_status C_cvd_comorbidity ///
1494 Cwv1_depressive_symptoms
1495 predict cs, csnell
1496
1497 * or
1498
1499 quietly stcox E_traj_group4
1500 predict cs, csnell
1501
1502
1503 stset cs, failure(Cwv2to4_dementia_event)
1504 sts generate km = s
1505 generate H = -ln(km)
1506 line H cs cs, sort ytitle("") clstyle(. refline)
1507
1508
1509
1510
1511 * ----- COX PH REGRESSION MODEL IN IMPUTED DATASET ----- *
1512
1513
1514 * Declare Data to be Survival Data by using mi
1515
1516 mi stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
1517
1518
1519 * Run Cox regression analysis in imputed dataset by using "mi estimate:"
1520 * Building the Model: Model 1 (unadjusted), Model 2, Model 3, Model 4
1521
1522
1523
1524 * Unadjusted model - model 1 - single predictor
1525
1526 * Model 1 (default coefficients)
1527 mi estimate: stcox C_lca_group3
1528

```



```

1529 * Model 1: define design var by using i.
1530 mi estimate: stcox i.C_lca_group3
1531
1532
1533 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
1534
1535 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3
1536
1537
1538 * Adjusted models - multivariable Cox model
1539 * controlling for covariates
1540
1541 * Model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
1542
1543 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
1544 C_age i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles
1545
1546 * Model 3: model 2 + adjust for cvd health
1547
1548 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
1549 C_age i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1550 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity
1551
1552
1553 * Model 4: model 3 + adjust for depression
1554
1555 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
1556 C_age i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
1557 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
1558 i.Cwv1_depressive_symptoms
1559
1560
1561
1562
1563
1564
1565
1566
1567 /*
1568
1569
1570 *** SENSITIVITY ANALYSES ***
1571
1572 1) multigroup latent class model by sex
1573
1574 2) interactions with age and gender
1575 survival analysis stratified by age
1576 two age groups: <70 and >=70
1577
1578 3) exclude participants with cvd
1579
1580 4) Complete data
1581 Cox regression analysis on complete data (without imputed covariates)
1582 (see above)
1583 */
1584
1585
1586
1587 /*
1588 1) Multigroup latent class model by sex
1589
1590
1591 TWO STEP PROCESS
1592
1593 1) LCA by group (to build the model and get lcprob and lcmean and to get the marginplots for males)
1594
1595 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1596 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///

```



```

1597 group(C_sex) ginvariant(coef)
1598
1599 estat lcprob
1600 estat lcmean
1601 estat lcgof
1602
1603 2) LCA sort sex (to get the marginplots for females)
1604
1605 sort C_sex
1606
1607 by C_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1608 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
1609
1610
1611 */
1612
1613
1614 * LCA by group
1615 * three-class model
1616
1617
1618 gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1619 hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3) ///
1620 group(C_sex) ginvariant(coef) ///
1621 startvalues(randompr, draws(20) seed(15) difficult) ///
1622 emopts(iterate(30) difficult)
1623
1624 estimates store threeclass_cm
1625
1626
1627
1628 * LCA postestimation
1629 * Latent class marginal probabilities - lcprob -
1630 * Latent class marginal means - lcmean -
1631
1632
1633 estat lcprob
1634
1635 estat lcmean
1636
1637 estat lcgof
1638
1639
1640 /* We can use the predictions of the posterior probability of class membership to evaluate an
1641 individual's probability of being in each class.
1642
1643 */
1644
1645 predict m_classpost1*, classposteriorpr
1646 list in 1, abbrev(10)
1647
1648 /* We can determine the expected class for each individual based on whether the posterior
1649 probability
1650 is greater than 0.5
1651 */
1652 generate m_expclass1 = 1 + (m_classpost11>0.5)
1653 tabulate m_expclass1
1654
1655
1656 generate m_expclass2 = 1 + (m_classpost12>0.5)
1657 tabulate m_expclass2
1658
1659
1660 generate m_expclass3 = 1 + (m_classpost13>0.5)
1661 tabulate m_expclass3
1662
1663

```

```

1664
1665  /* We can determine expected classification for each individual in the dataset based on the
1666  predicted
1667  posterior class probabilities.
1668  */
1669  predict m_cpost*, classposteriorpr
1670  egen m_max = rowmax(m_cpost*)
1671
1672
1673  * generate classes var
1674
1675  generate m_predclass = 1 if m_cpost1==m_max
1676
1677  replace m_predclass = 2 if m_cpost2==m_max
1678
1679  replace m_predclass = 3 if m_cpost3==m_max
1680
1681  tabulate m_predclass
1682
1683
1684
1685  * margins and marginsplot for MALES
1686
1687  * use margins to calculate marginal predictions
1688  * use marginsplot to graph marginal predictions
1689
1690
1691  *Install/update combomarginsplot ado.
1692
1693  *https://www.statalist.org/forums/forum/general-stata-discussion/general/1425209-is-it-possible-to-do-multilevel-latent-class-analysis-with-stata-15-ic
1694
1695  ssc install combomarginsplot, replace
1696
1697
1698
1699  margins, predict(classpr class(1)) ///
1700          predict(classpr class(2)) ///
1701          predict(classpr class(3)) subpop(if C_sex==0) saving(margin_male, replace)
1702  marginsplot, xtitle ("") ytitle ("") ///
1703          xlabel (1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1704          title ("Predicted Latent Class Probabilities with 95% CI") ///
1705          name(margin_male, replace)
1706
1707
1708  margins, predict(classpr class(1)) ///
1709          predict(classpr class(2)) ///
1710          predict(classpr class(3)) subpop(if C_sex==0) saving(margin_male, replace)
1711  marginsplot, recast(bar) xtitle("") ytitle("") ///
1712          xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1713          title("Predicted Latent Class Probabilities with 95% CI") ///
1714          name(margin_male, replace)
1715
1716
1717
1718  * class 1
1719
1720  margins, predict(outcome(crp_lca) class(1)) ///
1721          predict(outcome(hdl_lca) class(1)) ///
1722          predict(outcome(obesity_lca) class(1)) ///
1723          predict(outcome(systolic_lca) class(1)) ///
1724          predict(outcome(diastolic_lca) class(1)) ///
1725          predict(outcome(diabetes_lca) class(1)) ///
1726          predict(outcome(hba1c_lca) class(1)) subpop(if C_sex==0) ///
1727          saving(class1_male, replace) ///
1728
1729  marginsplot, recast(bar) title ("Class 1") xtitle("") ///

```

```

1730      xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1731      5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1732      ytitle("Predicted mean") ylabel(0(.20)1) name (class1_male, replace)
1733
1734
1735  * class 2
1736
1737  margins, predict(outcome(crp_lca) class(2)) ///
1738      predict(outcome(hdl_lca) class(2)) ///
1739      predict(outcome(obesity_lca) class(2)) ///
1740      predict(outcome(systolic_lca) class(2)) ///
1741      predict(outcome(diastolic_lca) class(2)) ///
1742      predict(outcome(diabetes_lca) class(2)) ///
1743      predict(outcome(hba1c_lca) class(2)) subpop(if C_sex==0) ///
1744      saving(class2_male, replace) ///
1745
1746  marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1747      xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1748      5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1749      ytitle("Predicted mean") ylabel(0(.20)1) name (class2_male, replace)
1750
1751
1752
1753
1754  * class 3
1755
1756  margins, predict(outcome(crp_lca) class(3)) ///
1757      predict(outcome(hdl_lca) class(3)) ///
1758      predict(outcome(obesity_lca) class(3)) ///
1759      predict(outcome(systolic_lca) class(3)) ///
1760      predict(outcome(diastolic_lca) class(3)) ///
1761      predict(outcome(diabetes_lca) class(3)) ///
1762      predict(outcome(hba1c_lca) class(3)) subpop(if C_sex==0) ///
1763      saving(class3_male, replace) ///
1764
1765  marginsplot, recast(bar) title ("Class 3") xtitle("") ///
1766      xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1767      5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1768      ytitle("Predicted mean") ylabel(0(.20)1) name (class3_male, replace)
1769
1770
1771
1772  graph combine class1_male class2_male class3_male, cols(3)
1773
1774
1775
1776
1777  * LCA sort by sex
1778  * three-class model
1779
1780  sort C_sex
1781
1782  by C_sex: gsem (crp_lca hdl_lca obesity_lca systolic_lca diastolic_lca diabetes_lca ///
1783  hba1c_lca <- _cons), family(bernoulli) link(logit) lclass(C 3)
1784
1785
1786  estat lcprob
1787
1788  estat lcmean
1789
1790  estat lcgof
1791
1792
1793
1794  /* We can use the predictions of the posterior probability of class membership to evaluate an
1795  individual's probability of being in each class.
1796
1797  */

```

```

1798
1799 predict f_classpost1*, classposteriorpr
1800 list in 1, abbrev(10)
1801
1802 /* We can determine the expected class for each individual based on whether the posterior
1803 probability
1804 is greater than 0.5
1805 */
1806 generate f_expclass1 = 1 + (f_classpost11>0.5)
1807 tabulate f_expclass1
1808
1809
1810 generate f_expclass2 = 1 + (f_classpost12>0.5)
1811 tabulate f_expclass2
1812
1813
1814 generate f_expclass3 = 1 + (f_classpost13>0.5)
1815 tabulate f_expclass3
1816
1817
1818
1819 /* We can determine expected classification for each individual in the dataset based on the
1820 predicted
1821 posterior class probabilities.
1822 */
1823 predict f_cpost*, classposteriorpr
1824 egen f_max = rowmax(f_cpost*)
1825
1826
1827 * generate classes var
1828
1829 generate f_predclass = 1 if f_cpost1==f_max
1830
1831 replace f_predclass = 2 if f_cpost2==f_max
1832
1833 replace f_predclass = 3 if f_cpost3==f_max
1834
1835 tabulate f_predclass
1836
1837
1838
1839
1840 * margins and marginsplot for FEMALES
1841
1842 * use margins to calculate marginal predictions
1843 * use marginsplot to graph marginal predictions
1844
1845
1846
1847 margins, predict(classpr class(1)) ///
1848         predict(classpr class(2)) ///
1849         predict(classpr class(3)) subpop(if C_sex==1) saving(margin_female, replace)
1850 marginsplot, xtitle("") ytitle("") ///
1851         xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1852         title("Predicted Latent Class Probabilities with 95% CI") ///
1853         name(margin_female, replace)
1854
1855
1856 margins, predict(classpr class(1)) ///
1857         predict(classpr class(2)) ///
1858         predict(classpr class(3)) subpop(if C_sex==1) saving(margin_female, replace)
1859 marginsplot, recast(bar) xtitle("") ytitle("") ///
1860         xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3") ///
1861         title("Predicted Latent Class Probabilities with 95% CI") ///
1862         name(margin_female, replace)
1863

```

```

1864
1865 * class 1
1866
1867 margins, predict(outcome(crp_lca) class(1)) ///
1868             predict(outcome(hdl_lca) class(1)) ///
1869             predict(outcome(obesity_lca) class(1)) ///
1870             predict(outcome(systolic_lca) class(1)) ///
1871             predict(outcome(diastolic_lca) class(1)) ///
1872             predict(outcome(diabetes_lca) class(1)) ///
1873             predict(outcome(hba1c_lca) class(1)) subpop(if C_sex==1) ///
1874             saving(class1_female, replace) ///
1875
1876 marginsplot, recast(bar) title ("Class 1") xtitle("") ///
1877             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1878             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1879             ytitle("Predicted mean") ylabel(0(.20)1) name (class1_female, replace)
1880
1881
1882 * class 2
1883
1884 margins, predict(outcome(crp_lca) class(2)) ///
1885             predict(outcome(hdl_lca) class(2)) ///
1886             predict(outcome(obesity_lca) class(2)) ///
1887             predict(outcome(systolic_lca) class(2)) ///
1888             predict(outcome(diastolic_lca) class(2)) ///
1889             predict(outcome(diabetes_lca) class(2)) ///
1890             predict(outcome(hba1c_lca) class(2)) subpop(if C_sex==1) ///
1891             saving(class2_female, replace) ///
1892
1893 marginsplot, recast(bar) title ("Class 2") xtitle("") ///
1894             xlabel(1 "crp" 2 "hdl" 3 "obesity" 4 "systolic BP" ///
1895             5 "diastolic BP" 6 "diabetes" 7 "hba1c", angle(45)) ///
1896             ytitle("Predicted mean") ylabel(0(.20)1) name (class2_female, replace)
1897
1898
1899
1900
1901 * class 3
1902
1903 margins, predict(outcome(crp_lca) class(3)) ///
1904             predict(outcome(hdl_lca) class(3)) ///
1905             predict(outcome(obesity_lca) class(3)) ///
1906             predict(outcome(systolic_lca) class(3)) ///
1907             predict(outcome(diastolic_lca) class(3)) ///
1908             predict(outcome(diabetes_lca) class(3)) ///
1909             predict(outcome(hba1c_lca) class(3)) subpop(if C_sex==1) ///
1910             saving(class3_female, replace) ///
1911
1912 marginsplot, recast(bar) title ("Class 3") 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 (class3_female, replace)
1916
1917
1918
1919 graph combine class1_female class2_female class3_female, cols(3)
1920
1921
1922
1923
1924
1925
1926 *combine margins male and female class probabilities
1927
1928 graph combine margin_male margin_female, cols(3)
1929
1930 *combine margins male and female 3 classes mean
1931

```

```

1932 graph combine class1_male class2_male class3_male class1_female class2_female class3_female, cols(
1933 3)
1934
1935
1936
1937
1938
1939 /* 2) Interaction with age and gender
1940 Survival analysis stratified by age
1941
1942 generate age group variable
1943 Age groups: 1) young old (< 70) 2) old old (>= 70)
1944
1945 Kaplan Meier curves
1946 Cox regression models in imputed data
1947
1948 young old <70
1949 if C_age_group==1
1950
1951 old old >70
1952 if C_age_group==2
1953
1954
1955 */
1956
1957
1958 gen C_age_group=1 if C_age < 70
1959 replace C_age_group=2 if C_age >=70 & ///
1960 !missing(C_age)
1961
1962 label var C_age_group "Age groups <70 young-old / 70 old-old"
1963 lab def age_group 1 "young old <70" 2 "old old >70"
1964 lab val C_age_group age_group
1965
1966 tab C_age_group
1967
1968
1969
1970
1971 stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
1972
1973
1974 * YOUNG OLD <70 Kaplan Meier
1975
1976 * Group Kaplan-Meier Curve Estimation
1977 * Command is sts graph, by(GROUPVAR) OPTION OPTION OPTION Note: Must have sorted by GROUPVAR first
1978
1979 sort C_lca_group3
1980
1981 sts list if C_age_group==1, by(C_lca_group3)
1982
1983 * graph with frills in males
1984
1985 sts graph if C_age_group==1, by(C_lca_group3) xlabel(0(20)100) ylabel(0.80(.05)1) xtitle("Time in
1986 Months") ///
1987 ytitle("Survival Prob") title("Kaplan Meier Curve <70 years") caption("graph04.png", size(vsmall))
1988
1989
1990 * calculate person-time and incidence rates using command stptime
1991
1992 stptime if C_age_group==1, title(Person-years)
1993
1994 stptime if C_age_group==1, title(Person-years) per(1000)
1995
1996
1997 * calculate person-time by category of C_lca_group3

```

```

1998
1999     stptime if C_age_group==1, by(C_lca_group3)
2000
2001     stptime if C_age_group==1, by(C_lca_group3) per(1000)
2002
2003
2004
2005     * OLD OLD >70 Kaplan Meier
2006
2007     sts list if C_age_group==2, by(C_lca_group3)
2008
2009     * graph with frills in males
2010
2011     sts graph if C_age_group==2, by(C_lca_group3) xlabel(0(20)100) ylabel(0.80(.05)1) xtitle("Time in
Months") ///
2012     ytitle("Survival Prob") title("Kaplan Meier Curve >= 70 years") caption("graph04.png", size(vsmall
))
2013
2014
2015
2016     * calculate person-time and incidence rates using command ststime
2017
2018     stptime if C_age_group==2, title(Person-years)
2019
2020     stptime if C_age_group==2, title(Person-years) per(1000)
2021
2022
2023     * calculate person-time by category of C_lca_group3
2024
2025     stptime if C_age_group==2, by(C_lca_group3)
2026
2027     stptime if C_age_group==2, by(C_lca_group3) per(1000)
2028
2029
2030
2031
2032
2033
2034     * COX PH REGRESSION MODEL IN IMPUTED DATASET
2035
2036
2037     * Declare Data to be Survival Data by using mi
2038
2039     mi stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
2040
2041
2042
2043
2044
2045     *** INTERACTION gender*cardiometabolic cluster ***
2046
2047     mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 i.C_sex#i.C_lca_group3
2048
2049
2050
2051     mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2052     C_age i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2053     i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
2054     i.Cwv1_depressive_symptoms i.C_sex#i.C_lca_group3
2055
2056
2057
2058
2059     *** INTERACTION age*cardiometabolic cluster ***
2060
2061     mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 c.C_age#i.C_lca_group3
2062
2063

```



```

2064
2065 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2066 C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2067 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
2068 i.Cwv1_depressive_symptoms c.C_age#i.C_lca_group3
2069
2070
2071
2072
2073
2074 * YOUNG OLD <70 Cox regression models
2075
2076 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2077
2078 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 if C_age_group==1
2079
2080
2081 * Adjusted models - multivariable Cox model
2082 * controlling for covariates
2083
2084 * Model 2: model 1 + adjust for sociodemographics: age education marital status and wealth
2085
2086 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2087 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles if C_age_group==1
2088
2089 * Model 3: model 2 + adjust for cvd health
2090
2091 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2092 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2093 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity if C_age_group==1
2094
2095
2096 * Model 4: model 3 + adjust for depression
2097
2098 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2099 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2100 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
2101 i.Cwv1_depressive_symptoms if C_age_group==1
2102
2103
2104
2105
2106 * OLD OLD >70 Cox regression models
2107
2108 * Model 1 ask for hazard ratio by using the option eform("Haz.Ratio")
2109
2110 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 if C_age_group==2
2111
2112
2113 * Adjusted models - multivariable Cox model
2114 * controlling for covariates
2115
2116 * Model 2: model 1 + adjust for sociodemographics: sex education marital status and wealth
2117
2118 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2119 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles if C_age_group==2
2120
2121 * Model 3: model 2 + adjust for cvd health
2122
2123 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2124 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2125 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity if C_age_group==2
2126
2127
2128 * Model 4: model 3 + adjust for depression
2129
2130 mi estimate, eform("Haz. Ratio"): stcox i.C_lca_group3 ///
2131 i.C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///

```



```

2132 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status i.C_cvd_comorbidity ///
2133 i.Cwv1_depressive_symptoms if C_age_group==2
2134
2135
2136
2137
2138
2139
2140
2141 /*
2142
2143 3) exclude participants with cvd
2144
2145 use the command if C_cvd_comorbidity==0
2146
2147 */
2148
2149
2150
2151 * Declare Data to be Survival Data by using mi
2152
2153 stset C_time_of_event_dementia, failure (Cwv2to4_dementia_event==1) id(id_12char)
2154
2155
2156
2157 * define design var by using i.(3 classes)
2158
2159 stcox i.C_lca_group3 if C_cvd_comorbidity==0
2160
2161
2162 * Adjusted models - multivariable Cox model
2163 * controlling for covariates
2164
2165 * model 2: model 1 + adjust for sociodemographics: age sex education marital status and wealth
2166
2167 stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat ///
2168 i.Cwv1_netwealth_quintiles if C_cvd_comorbidity==0
2169
2170 * model 3: model 2 + adjust for cvd health
2171
2172 stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2173 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status if C_cvd_comorbidity==0
2174
2175
2176 * model 4: model 3 + adjust for depression
2177 stcox i.C_lca_group3 C_age C_sex i.C_educ_new i.C_maritalstatus_4cat i.Cwv1_netwealth_quintiles ///
2178 i.Cwv1_smoking_3cat i.Cwv1_alcohol_status ///
2179 i.Cwv1_depressive_symptoms if C_cvd_comorbidity==0
2180
2181
2182
2183
2184
2185 * 4) complete data (see above)
2186
2187
2188 * ----- *
2189
2190
2191
2192
2193
2194 *** EXTRA SENSITIVITY ANALYSES FOR THE PAPER ***
2195
2196 /*
2197
2198 compare baseline characteristics between complete sample (before exclusion) and sample with
missing data (overall after exclusion)

```

```

2199
2200 */
2201
2202
2203
2204 * General characteristics of CHARLS participants at baseline
2205
2206 * Socio-demographics
2207 sum C_age
2208 ta C_sex
2209 ta C_educ_new
2210 ta C_maritalstatus_4cat
2211 ta Cwv1_netwealth_quintiles
2212 * Cardiometabolic disorders
2213 ta Cwv1_crp
2214 ta Cwv1_hdl_cholesterol
2215 ta Cwv1_obesity_waist
2216 ta Cwv1_systolic_bp
2217 ta Cwv1_diastolic_bp
2218 ta Cwv1_diabetes_report
2219 ta Cwv1_HbA1c
2220 * Cardiovascular health factors
2221 ta Cwv1_smoking_3cat
2222 ta Cwv1_alcohol_status
2223 ta Cwv1_physicalactivity
2224 ta C_cvd_comorbidity
2225 * Depressive symptoms
2226 ta Cwv1_depressive_symptoms
2227 * Memory score
2228 sum Cwv1_memory_wordrecall
2229
2230
2231
2232
2233 * compare health characteristics between those survived and dropped out
2234
2235
2236
2237 *** CLEANING DATA to keep those who dropped out
2238
2239
2240
2241 * 1. drop dementia cases at baseline
2242
2243 * drop dementia wave 2 missing data
2244 drop if Cwv1_dementia_report==1
2245 * (267 observations deleted)
2246 drop if Cwv1_dementia_report== .
2247 * (88 observations deleted)
2248
2249
2250 * 2. drop missing values of cardiometabolic markers
2251
2252 drop if Cwv1_crp== .
2253 * (180 observations deleted)
2254
2255 drop if Cwv1_hdl_cholesterol== .
2256 * (2 observations deleted)
2257
2258 drop if Cwv1_obesity_waist== .
2259 * (1688 observations deleted)
2260
2261 drop if Cwv1_systolic_bp== .
2262 * (83 observations deleted)
2263
2264 drop if Cwv1_diastolic_bp== .
2265 * (13 observations deleted)
2266

```

```

2267 drop if Cwv1_diabetes_report== .
2268 * (91 observations deleted)
2269
2270 drop if Cwv1_HbA1c== .
2271 * (71 observations deleted)
2272
2273
2274
2275 * 3. drop obs with no records on dementia at any wave from 2-4 follow-ups
2276
2277
2278 search mdesc
2279 search rmiss2
2280 search mvpatterns
2281
2282 * see number of missing values vs non-missing in each variable
2283 mdesc Cwv2_dementia_report Cwv3_dementia_report Cwv4_self_info_dementia
2284
2285
2286
2287 /* number of missing values per observation
2288 * the code below creates a variable called nmisfollowup that gives the number of missing values
2289 for each observation in the variables of interest */
2290 egen nmisfollowup_dementia_wv2to4=rmiss2(Cwv2_dementia_report ///
2291 Cwv3_dementia_report Cwv4_self_info_dementia)
2292
2293 tab nmisfollowup_dementia_wv2to4
2294
2295 * drop observations "nmisfollowup_dementia_wv2to4" < 3
2296 drop if nmisfollowup_dementia_wv2to4<3
2297
2298
2299 * General characteristics of CHARLS participants at baseline
2300
2301 * Socio-demographics
2302 sum C_age
2303 ta C_sex
2304 ta C_educ_new
2305 ta C_maritalstatus_4cat
2306 ta Cwv1_netwealth_quintiles
2307 * Cardiometabolic disorders
2308 ta Cwv1_crp
2309 ta Cwv1_hdl_cholesterol
2310 ta Cwv1_obesity_waist
2311 ta Cwv1_systolic_bp
2312 ta Cwv1_diastolic_bp
2313 ta Cwv1_diabetes_report
2314 ta Cwv1_HbA1c
2315 * Cardiovascular health factors
2316 ta Cwv1_smoking_3cat
2317 ta Cwv1_alcohol_status
2318 ta Cwv1_physicalactivity
2319 ta C_cvd_comorbidity
2320 * Depressive symptoms
2321 ta Cwv1_depressive_symptoms
2322 * Memory score
2323 sum Cwv1_memory_wordrecall
2324
2325
2326
2327 * compare health characteristics between <70 and >=70
2328
2329
2330
2331 * General baseline characteristics of CHARLS participants by age group
2332
2333 * Socio-demographics
2334 ttest C_age, by(C_age_group)

```

```

2335 ta C_sex C_age_group, chi2 column row
2336 ta C_educ_new C_age_group, chi2 column row
2337 ta C_maritalstatus_4cat C_age_group, chi2 column row
2338 ta Cwv1_netwealth_quintiles C_age_group, chi2 column row
2339 * Cardiometabolic disorders
2340 ta crp_lca C_age_group, chi2 column row
2341 ta hdl_lca C_age_group, chi2 column row
2342 ta obesity_lca C_age_group, chi2 column row
2343 ta systolic_lca C_age_group, chi2 column row
2344 ta diastolic_lca C_age_group, chi2 column row
2345 ta diabetes_lca C_age_group, chi2 column row
2346 ta hba1c_lca C_age_group, chi2 column row
2347 * Cardiovascular health factors
2348 ta Cwv1_smoking_3cat C_age_group, chi2 column row
2349 ta Cwv1_alcohol_status C_age_group, chi2 column row
2350 ta Cwv1_physicalactivity C_age_group, chi2 column row
2351 ta C_cvd_comorbidity C_age_group, chi2 column row
2352 * Depressive symptoms
2353 ta Cwv1_depressive_symptoms C_age_group, chi2 column row
2354 * Memory score
2355 ttest Cwv1_memory_wordrecall, by(C_age_group)
2356 ta C_lca_group3 C_age_group, chi2 column row
2357
2358
2359
2360
2361 * ----- *
2362
2363
2364
2365
2366
2367

```