```stata
1    capture log cl
2    log using $apnyls/Analysedata/Log/analysisdata.log, replace
3
4    /* Corresponds to data05_alle, but incorporates the new samples from
5    DNBC and the Birth registry.
6    Is used for Table 1b to obtain info on non-participants.
7    */
8
9    /* The four "main" datasets are merged (as in counting_fish.do by Jakob
     Grove.
10   Used for constructing the reduced dataset.
11   */
12
13   clear
14   set mo off
15   set mem 500m
16
17   set seed 50000
18
19   use "$hod\BSMB data og spørgeskema\BSMB\int1_livsstil.dta", clear
20   /* 3567 individuals*/
21   /*
22   The variable a184 is recoded in lifestyle05, but not in int1_livsstil.
23   Hence it is done here.
24   */
25   recode a184 1=0 2=1 3/4=., gen(premarital)
26   lab var premarital "Maternal prenatal marital status"
27   label define premarital 0"With partner" 1"Single", modify
28   lab val premarital premarital
29
30   /*
31   list lbgravnr if premarital==.
32        +----------+
33        | lbgravnr |
34        |----------|
35   1467. |  1474951 |
36   3188. |  1892121 |
37   3561. |  1970101 |
38        +----------+
39   */
40
41   merge lbgravnr using "$hod\Datasæt\udtraek_samlet_kategori.dta", sort
42   tab _merge
43   /* The 3 extra individuals in udtraek__samlet_kategori.dta are dropped*/
44   drop if _merge == 2
45   drop _merge
46
47   merge lobnr gravnr using "$hod/Datasæt/t_livsreg.dta", sort keep(kategori
      brev_date koen)
48   /*3489 are matched*/
49   keep if _merge == 3
50   drop _merge
51
52   /*variables are dropped due to memory constraints, when using 'update'*/
53   drop a001-a019 a025-a025b a026_4-a126a a128-a136
54
55   /*Use update to maintain the "extra" individual only found in the
     Lifestyle_dataset_05 dataset*/
56   sort lobnr gravnr
57   merge m:m lobnr gravnr using "$hod/Implementering af
     rensninger/05/Lifestyle_dataset_05", update /*sort keep(test_date)*/
58   drop _merge
```

```stata
 59    /*1933 matches, total is 3490*/
 60
 61    /*
 62    list lbgravnr if premarital==.
 63          +---------+
 64          | lbgravnr |
 65          |---------|
 66    1440. |  1474951 |
 67    3125. |  1892121 |
 68    3483. |  1970101 |
 69    3490. |  1609571 |
 70          +---------+
 71    */
 72    /*The final individual originates from lifestyle05 where a184 is coded
 73    as premarital*/
 74    replace premarital=a184 if lbgravnr==1609571
 75
 76    rename kategori category
 77    sort category
 78    gen fo_status=(category=="FO")
 79
 80    /* Categorisation (newcategory) of individuals sampled on fish oil (FO)
 81    intake */
 82    preserve
 83    keep if fo_status==1
 84    qui do "$apnyls/SamplingPlan/Program/fo_classif"
 85    tempfile fodat
 86    sa "$apnyls/SamplingPlan/Data/fodat", replace
 87    restore
 88
 89    merge m:1 lbgravnr using "$apnyls/SamplingPlan/Data/fodat"
 90    /* in total 3490 (1783 tested)*/
 91    drop _merge
 92    replace newcategory=category if newcategory==""
 93    rename newcategory kategori
 94    /* One FO-individual is not categorized and hence dropped */
 95    *list lbgravnr a145a a145 if kategori=="FO"
 96    drop if kategori=="FO"
 97
 98    gen kategori1=substr(kategori,1,1)
 99    label var kategori1 "Main category"
100    /* turns cat1 into a numeric variable*/
101    encode kategori1, gen(kat1num)
102
103    rename a004a a004amid
104    rename paritet paritetmid
105    /*merge new variable on a reduced dataset (memory concerns)*/
106    preserve
107    keep lbgravnr
108    merge 1:m lbgravnr using "$apnyls\SamplingPlan\Data\BSMB_all_cat"
109    keep if _merge==3
110    drop _merge
111    save "$apnyls\SamplingPlan\Data\BSMB_samp", replace
112    restore
113
114    /* New variables from DNBC are merged on */
115    merge 1:1 lbgravnr using "$apnyls\SamplingPlan\Data\BSMB_samp", update
116    drop _merge
117    rename paritet paritet_extra /*information on non-participants used
118    for table 1b*/
119    rename a004a a004a_extra
120    rename paritetmid paritet
```

```
121    rename a004amid a004a
122
123    /*Merge variable with sampling weights 'vagt' from hs_19022010_fr
       dataset unto current dataset */
124    rename vagt vagt_l
125    preserve
126    keep lbgravnr
127    merge 1:m lbgravnr using "$apnyls\SamplingPlan\Data\hs_19022010_fr"
128    keep if _merge==3
129    drop _merge
130    keep lbgravnr vagt
131    rename vagt vagt_extra
132    save "$apnyls\SamplingPlan\Data\fr_samp", replace
133    restore
134    rename vagt_l vagt
135
136
137
138    merge 1:m lbgravnr using "$apnyls\SamplingPlan\Data\fr_samp", update
139    recode vagt_extra 0=.
140
141
142    duplicates list lbgravnr
143    duplicates tag lbgravnr, gen(twins)
144    tab twins if test_date < .
145    tab twins if test_date == .
146
147    /*
148    Note, that 11 twin pairs appear among the 3498 individuals. Since the twin
149    to be included cannot be uniquely identified from the information in
150    our dataset and the birth registry, we randomly sample one twin from
151    each twin pair.
152    */
153
154    gen tmp=uniform()
155    bysort lbgravnr (tmp): keep if _n==1  /* Twins are sorted randomly, and
       the first is maintained. */
156
157    /* In spite of the above we decide to exclude all twins, as that
158    apparently was the original intent */
159    drop if twins
160
161
162
163    /*merge sampling fractions */
164    drop category category2
165    rename kategori category
166    drop _merge
167    sort category
168    merge m:1 category using "$apnyls\SamplingPlan\Data\samplefrac1_5.dta"
169    rename category kategori
170
171
172    /*new variables*/
173    do "$apnyls\Analysedata\Program\variable_nye.do"
174
175    /*indicator variables, used in stamdata*/
176    do "$apnyls\Analysedata\Program\indikator_stam.do"
177
178    order lbgravnr kategori kat1num particind_motor particind_cat_motor
       particind_cat ind_motor excl_bef fo_status ///
179        sampfrac sampfrac_p malder eduindex premarital marital paritet
```

```stata
           paritet_extra presmoke postsmoke postsmoke_p ///
180           n_ik_me3 bmi_before_int1 bmi_extra binge nbinge timing_int ///
181           gender alder_vtest healthindex homeindex home_dic nosports hearing
       vision vagt vagt_extra ga_dage ///
182           m_abcsco abc_pp_11 abc_ps_5 abc_cs_6 abc_tb_4 abc_ga_3 abc_1b_11
       abc_hs_4 abc_hs_25
183
184    drop _merge
185
186    save "$apnyls/Analysedata/Data/analysedata_alle", replace
187
188
189    log cl
190
191
192
193
194
195
196
197
198
199
200
```