



```
1 .
2 . //STEP 15: DETERMINE FINAL ANALYTIC SAMPLE SIZE/////
3 .
4 . use UKB_PAPER4B_LE8INFECTMRI_NODDI, clear

5 .
6 .
7 .
8 . capture drop FA_mean

9 . egen FA_mean=rmean(FA_*)
   (461,872 missing values generated)

10 .
11 .
12 .
13 . capture drop MD_mean

14 . egen MD_mean=rmean(MD_*)
   (461,872 missing values generated)

15 .
16 . capture drop ISOVF_mean

17 . egen ISOVF_mean=rmean(ISOVF_*)
   (461,874 missing values generated)

18 .
19 .
20 . capture drop ICVF_mean

21 . egen ICVF_mean=rmean(ICVF_*)
   (461,874 missing values generated)

22 .
23 .
24 . capture drop OD_mean

25 . egen OD_mean=rmean(OD_*)
   (461,874 missing values generated)

26 .
27 .
28 . save, replace
   file UKB_PAPER4B_LE8INFECTMRI_NODDI.dta saved

29 .
30 .
31 .
32 . **Sample 1**
```

```

33 .
34 . capture drop sample1

35 . gen sample1=.
    (502,389 missing values generated)

36 . replace sample1=1 if baselineage~=.
    (502,386 real changes made)

37 . replace sample1=0 if sample1~=1
    (3 real changes made)

38 .
39 . tab sample1

```

sample1	Freq.	Percent	Cum.
0	3	0.00	0.00
1	502,386	100.00	100.00
Total	502,389	100.00	

```

40 .
41 . **Sample 2: Complete data on brain sMRI and dMRI**
42 .
43 . capture drop sample2

```

```

44 . gen sample2=.
    (502,389 missing values generated)

45 . replace sample2=1 if ICV~=. & TOTALBRAIN~=. & WMH~=. & Hippocampus_Left~=. & FA_mean~=. & MD_mean~=. & ISOVF_me
    (39,989 real changes made)

46 . replace sample2=0 if sample2~=1
    (462,400 real changes made)

```

```

47 .
48 . tab sample2

```

sample2	Freq.	Percent	Cum.
0	462,400	92.04	92.04
1	39,989	7.96	100.00
Total	502,389	100.00	

```

49 .
50 . **Sample 3A: Complete data on infectionburden**
51 .
52 . capture drop sample3A

```

```

53 . gen sample3A=.
    (502,389 missing values generated)

54 . replace sample3A=1 if infectionburden~=.
    (502,389 real changes made)

55 . replace sample3A=0 if sample3A~=1
    (0 real changes made)

```

```

56 .
57 . tab sample3A

```

sample3A	Freq.	Percent	Cum.
1	502,389	100.00	100.00
Total	502,389	100.00	

```

58 .
59 . **Sample 3B: Complete data on AD_PGS, LE8 TOTAL AND SUB-SCORES
60 . capture drop sample3B

61 . gen sample3B=.
    (502,389 missing values generated)

62 . replace sample3B=1 if AD_PGS~=. & LE8_TOTALSCORE~=. & LE8_BIOLOGICAL~=. & LE8_LIFESTYLE~=.
    (485,710 real changes made)

63 . replace sample3B=0 if sample3B~=1
    (16,679 real changes made)

```

```

64 .
65 . tab sample3B

```

sample3B	Freq.	Percent	Cum.
0	16,679	3.32	3.32
1	485,710	96.68	100.00
Total	502,389	100.00	

```

66 .
67 . **Sample 3: Complete data on infectionburden and AD PGS, LE8 TOTAL AND SUB-SCORES and Brain MRI**
68 . capture drop sample3

69 . gen sample3=.
    (502,389 missing values generated)

70 . replace sample3=1 if sample2==1 & sample3A==1 & sample3B==1
    (38,806 real changes made)

```

```
71 . replace sample3=0 if sample3~=1
    (463,583 real changes made)
```

```
72 .
```

```
73 . tab sample3
```

sample3	Freq.	Percent	Cum.
0	463,583	92.28	92.28
1	38,806	7.72	100.00
Total	502,389	100.00	

```
74 .
```

```
75 . **Sample 4: final sample: exclude prevalent dementia cases and incident dementia cases within a year**
```

```
76 . stset Age_dementia, failure(dem_diag==1) enter(baselineage) id(n_eid) scale(1)
```

Survival-time data settings

```

      ID variable: n_eid
      Failure event: dem_diag==1
Observed time interval: (Age_dementia[_n-1], Age_dementia]
      Enter on or after: time baselineage
      Exit on or before: failure
```

502,389	total observations	
3	event time missing (Age_dementia>=.)	PROBABLE ERROR
228	observations end on or before enter()	

502,158	observations remaining, representing	
502,158	subjects	
7,668	failures in single-failure-per-subject data	
6,208,727	total analysis time at risk and under observation	
	At risk from t =	0
	Earliest observed entry t =	37.41821
	Last observed exit t =	87.63313

```
77 .
```

```
78 . capture drop sample4
```

```
79 . gen sample4=.
    (502,389 missing values generated)
```

```
80 . replace sample4=1 if (sample3==1 & priordem==0 & _t~=. ) | (sample3==1 & time01_1Ayears>1 & time01_1Ayears~=. &
    (38,803 real changes made)
```

```
81 . replace sample4=0 if sample4~=1
    (463,586 real changes made)
```

```
82 .
```


[illegible]

100 .

101 .

```
102 . capture drop zISOVF_*
```

```
103 . foreach x4 of varlist ISOVF_* {
104     2.     egen z`x4'=std(`x4') if sample_final==1
105     3. }
```

[illegible]

[illegible]

104 .

105 .

```
106 . capture drop zICVF_*
```

```
107 . foreach x5 of varlist ICVF_* {
108     2.     egen z`x5'=std(`x5') if sample_final==1
109     3. }
```

[illegible]


```
(463,586 missing values generated)
(463,586 missing values generated)
(463,586 missing values generated)
(463,586 missing values generated)
(463,586 missing values generated)
(463,586 missing values generated)
```

[illegible]

```

112 .
113 .
114 .
115 . save, replace
    file UKB_PAPER4B_LE8INFECTMRI_NODDI.dta saved

116 .
117 . //STEP 16: DETERMINE DIFFERENCES IN FINAL ANALYTIC SAMPLE VS. EXCLUDED SAMPLE BY BASIC SOCIO-DEMOGRAPHICS and g
118 .
119 . capture drop RACE_ETHN

120 . gen RACE_ETHN=ethnicity2

121 .
122 .
123 . logistic sample_final baselineage sex i.RACE_ETHN

```

```

Logistic regression
Log likelihood = -135349.27

Number of obs = 502,386
LR chi2(5)     = 2569.10
Prob > chi2    = 0.0000
Pseudo R2     = 0.0094

```

sample_final	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
baselineage	.9728901	.0006293	-42.49	0.000	.9716575	.9741242
sex	.9302991	.0098809	-6.80	0.000	.9111331	.9498682
RACE_ETHN						
1	.311369	.0205112	-17.71	0.000	.2736548	.3542809
2	.4252869	.022222	-16.36	0.000	.3838886	.4711497
3	.5522226	.0235199	-13.94	0.000	.507996	.6002996
_cons	.45635	.0183634	-19.50	0.000	.4217412	.493799

Note: **_cons** estimates baseline odds.

```

124 .
125 .
126 . save, replace
    file UKB_PAPER4B_LE8INFECTMRI_NODDI.dta saved

127 .
128 .
129 . xi:probit sample_final baselineage sex i.RACE_ETHN
    i.RACE_ETHN      _IRACE_ETHN_0-3      (naturally coded; _IRACE_ETHN_0 omitted)

Iteration 0:  log likelihood = -136633.82
Iteration 1:  log likelihood = -135338.87
Iteration 2:  log likelihood = -135324.55
Iteration 3:  log likelihood = -135324.51
Iteration 4:  log likelihood = -135324.51

Probit regression
Log likelihood = -135324.51

Number of obs = 502,386
LR chi2(5)     = 2618.61
Prob > chi2    = 0.0000
Pseudo R2     = 0.0096

```

sample_final	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
baselineage	-.0139835	.0003257	-42.93	0.000	-.0146219	-.0133451
sex	-.0382865	.0052504	-7.29	0.000	-.0485771	-.0279958
_IRACE_ETHN_1	-.5441425	.0282896	-19.23	0.000	-.5995891	-.4886959
_IRACE_ETHN_2	-.4052211	.023323	-17.37	0.000	-.4509333	-.3595088
_IRACE_ETHN_3	-.284311	.0196648	-14.46	0.000	-.3228532	-.2457688
_cons	-.5575898	.0204354	-27.29	0.000	-.5976424	-.5175372

```

130 .
131 . capture drop p1MRIINF

132 . predict p1MRIINF, xb
    (3 missing values generated)

133 .
134 . capture drop phiMRIINF

135 . capture drop caphiMRIINF

136 . capture drop invmillsMRIINF

137 .
138 . gen phiMRIINF=(1/sqrt(2*pi))*exp(-(p1MRIINF^2/2))
    (3 missing values generated)

139 .
140 . egen caphiMRIINF=std(p1MRIINF)
    (3 missing values generated)

141 .
142 . capture drop invmillsMRIINF

143 . gen invmillsMRIINF=phiMRIINF/caphiMRIINF
    (3 missing values generated)

144 .
145 .
146 . su invmillsMRIINF

```

Variable	Obs	Mean	Std. dev.	Min	Max
invmillsMR~F	502,386	.6947878	75.01703	-584.8448	9007.412

```

147 .
148 . save , replace
    file UKB_PAPER4B_LE8INFECTMRI_NODDI.dta saved

149 .
150 .
151 .
152 . capture log close

```