```
                                                                      ___  ____  ____  ____  ____(R)
                                                                     /__    /   ____/   /   ____/
                                                                    ___/   /   /___/   /   /___/
                                                                      Statistics/Data analysis

 1 .
 2 .
 3 . use "E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8E_ADPRSPROTDEM\D

 4 .
 5 . capture mi set, clear

 6 .
 7 . save, replace
   file E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8E_ADPRSPROTDEM\DAT

 8 .
 9 . ************************************KAPLAN-MEIER SURVIVAL CURVES*****************************************
10 .
11 . capture drop LE8_TOTALSCOREtertinv

12 . gen LE8_TOTALSCOREtertinv=.
   (40,139 missing values generated)

13 . replace LE8_TOTALSCOREtertinv=1 if LE8_TOTALSCOREtert==3
   (13,311 real changes made)

14 . replace LE8_TOTALSCOREtertinv=2 if LE8_TOTALSCOREtert==2
   (12,856 real changes made)

15 . replace LE8_TOTALSCOREtertinv=3 if LE8_TOTALSCOREtert==1
   (13,971 real changes made)

16 .
17 . save, replace
   file E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8E_ADPRSPROTDEM\DAT

18 .
19 .
20 .
21 . **Total sample**
22 .
23 .
24 . set scheme sj

25 .
26 . stset Age_dementia, failure(dem_diag==1) enter(AGE) 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 AGE
        Exit on or before: failure
```

```
------------------------------------------------------------------------------
     40,139  total observations
          0  exclusions
------------------------------------------------------------------------------
     40,139  observations remaining, representing
     40,139  subjects
      1,167  failures in single-failure-per-subject data
 486,433.58  total analysis time at risk and under observation
                                                At risk from t =         0
                                     Earliest observed entry t =  50.00137
                                          Last observed exit t =  84.54757
```

```
27 .
28 .
29 . sts graph if sample_final==1 & _t<84, gwood legend(on) xlabel(65(5)90) ylabel(0.50(.10)1) xtitle("Age at diagnc
   > a,total sample") by (AD_PGStert) tmin(65) tmax(90)

          Failure _d: dem_diag==1
    Analysis time _t: Age_dementia
  Enter on or after: time AGE
        ID variable: n_eid

30 . graph save "FIGURE1A.gph", replace
   file FIGURE1A.gph saved

31 . sts test AD_PGStert if sample_final==1 & _t<84

          Failure _d: dem_diag==1
    Analysis time _t: Age_dementia
  Enter on or after: time AGE
        ID variable: n_eid

   Equality of survivor functions
   Log-rank test

              |  Observed       Expected
   AD_PGStert |    events         events
   -----------+-------------------------
            1 |       182         389.38
            2 |       280         391.86
            3 |       704         384.76
   -----------+-------------------------
        Total |      1166        1166.00

                      chi2(2) = 407.38
                      Pr>chi2 = 0.0000

32 . stcox i.AD_PGStert if sample_final==1 & _t<84

          Failure _d: dem_diag==1
    Analysis time _t: Age_dementia
  Enter on or after: time AGE
        ID variable: n_eid

   Iteration 0:  Log likelihood = -11064.899
   Iteration 1:  Log likelihood = -10875.221
   Iteration 2:  Log likelihood = -10872.056
   Iteration 3:  Log likelihood = -10872.048
   Refining estimates:
   Iteration 0:  Log likelihood = -10872.048

   Cox regression with Breslow method for ties
```

No. of subjects = **40,107** Number of obs = **40,107**
No. of failures = **1,166**
Time at risk = **485,978.412**
LR chi2(**2**) = **385.70**
Log likelihood = **-10872.048** Prob > chi2 = **0.0000**

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
AD_PGStert						
2	**1.52845**	**.1455355**	**4.46**	**0.000**	**1.268241**	**1.842047**
3	**3.915689**	**.3256319**	**16.41**	**0.000**	**3.326761**	**4.608873**

```
33 .
34 . sts graph if sample_final==1 & _t<84, gwood legend(on) xlabel(65(5)90) ylabel(0.50(.10)1) xtitle("Age at diagnc
   > a,total sample") by (AD_PGStert) tmin(65) tmax(90) risktable

          Failure _d: dem_diag==1
   Analysis time _t: Age_dementia
  Enter on or after: time AGE
        ID variable: n_eid

35 .
36 .
37 . **Men**
38 .
39 . set scheme sj

40 .
41 . stset Age_dementia, failure(dem_diag==1) enter(AGE) 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 AGE
     Exit on or before: failure

------------------------------------------------------------------------------
     40,139  total observations
          0  exclusions
------------------------------------------------------------------------------
     40,139  observations remaining, representing
     40,139  subjects
      1,167  failures in single-failure-per-subject data
 486,433.58  total analysis time at risk and under observation
                                        At risk from t =         0
                             Earliest observed entry t =  50.00137
                                  Last observed exit t =  84.54757
```

```
42 .
43 .
44 . sts graph if sample_final==1 & _t<84 & SEX==1, gwood legend(on) xlabel(65(5)90) ylabel(0.50(.10)1) xtitle("Age
   > . dementia, Men") by (AD_PGStert) tmin(65) tmax(90)

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
     Enter on or after: time AGE
           ID variable: n_eid

45 . graph save "FIGURE1B.gph", replace
   file FIGURE1B.gph saved

46 . sts test AD_PGStert if sample_final==1 & _t<84 & SEX==1

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
     Enter on or after: time AGE
           ID variable: n_eid

   Equality of survivor functions
   Log-rank test

             |  Observed       Expected
   AD_PGStert|    events         events
   ----------+-------------------------
             1|       109         205.06
             2|       169         207.01
             3|       338         203.93
   ----------+-------------------------
         Total|       616         616.00

                       chi2(2) = 140.20
                       Pr>chi2 = 0.0000

47 . stcox i.AD_PGStert if sample_final==1 & _t<84 & SEX==1

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
     Enter on or after: time AGE
           ID variable: n_eid

   Iteration 0:  Log likelihood = -5381.9582
   Iteration 1:  Log likelihood = -5315.4208
   Iteration 2:  Log likelihood = -5314.3109
   Iteration 3:  Log likelihood = -5314.3096
   Refining estimates:
   Iteration 0:  Log likelihood = -5314.3096

   Cox regression with Breslow method for ties

   No. of subjects =      18,549                     Number of obs = 18,549
   No. of failures =         616
   Time at risk    = 221,184.688
                                                     LR chi2(2)    = 135.30
   Log likelihood = -5314.3096                       Prob > chi2   = 0.0000
```

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
AD_PGStert						
2	**1.535505**	**.1886432**	**3.49**	**0.000**	**1.206918**	**1.953552**
3	**3.119538**	**.3436564**	**10.33**	**0.000**	**2.513736**	**3.871336**

```
48 .
49 .
50 . **Women**
51 .
52 . set scheme sj

53 .
54 . stset Age_dementia, failure(dem_diag==1) enter(AGE) 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 AGE
        Exit on or before: failure

   ------------------------------------------------------------------------------
        40,139  total observations
             0  exclusions
   ------------------------------------------------------------------------------
        40,139  observations remaining, representing
        40,139  subjects
         1,167  failures in single-failure-per-subject data
    486,433.58  total analysis time at risk and under observation
                                                  At risk from t =         0
                                       Earliest observed entry t =  50.00137
                                            Last observed exit t =  84.54757

55 .
56 .
57 . sts graph if sample_final==1 & _t<84 & SEX==2, gwood legend(on) xlabel(65(5)90) ylabel(0.50(.10)1) xtitle("Age
   > . dementia, Women") by (AD_PGStert) tmin(65) tmax(90)

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
     Enter on or after: time AGE
           ID variable: n_eid

58 . graph save "FIGURE1C.gph", replace
   file FIGURE1C.gph saved

59 . sts test AD_PGStert if sample_final==1 & _t<84 & SEX==2

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
     Enter on or after: time AGE
           ID variable: n_eid

   Equality of survivor functions
   Log-rank test
```

AD_PGStert	Observed events	Expected events
1	**73**	**184.14**
2	**111**	**184.82**
3	**366**	**181.04**
Total	**550**	**550.00**

chi2(2) = **285.60**
Pr>chi2 = **0.0000**

```
60 . stcox i.AD_PGStert if sample_final==1 & _t<84 & SEX==2

           Failure _d: dem_diag==1
     Analysis time _t: Age_dementia
   Enter on or after: time AGE
         ID variable: n_eid

   Iteration 0:  Log likelihood = -4868.9886
   Iteration 1:  Log likelihood = -4737.2953
   Iteration 2:  Log likelihood = -4735.4674
   Iteration 3:  Log likelihood = -4735.4582
   Iteration 4:  Log likelihood = -4735.4582
   Refining estimates:
   Iteration 0:  Log likelihood = -4735.4582

   Cox regression with Breslow method for ties

   No. of subjects =      21,558                           Number of obs = 21,558
   No. of failures =         550
   Time at risk    = 264,793.724
                                                           LR chi2(2)    = 267.06
   Log likelihood = -4735.4582                             Prob > chi2   = 0.0000
```

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
AD_PGStert						
2	**1.514807**	**.2282717**	**2.76**	**0.006**	**1.127422**	**2.035298**
3	**5.101202**	**.653925**	**12.71**	**0.000**	**3.967863**	**6.558256**

```
61 .
62 . graph combine "FIGURE1A.gph" "FIGURE1B.gph" "FIGURE1C.gph"

63 . graph save "FIGURE1.gph", replace
   file FIGURE1.gph saved

64 .
65 .
66 . save, replace
   file E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8E_ADPRSPROTDEM\DAT

67 .
68 . capture log close
```