Applied Survival Analysis

**Lab 11: Analysis of competing risks**

We analyse data involving treatment interruptions (TI) among HIV-infected patients in the presence of the competing risk of death or loss from follow-up. The data are included in file TI-data\_final.dta.

The data are as follows:

fail3 | Freq. Percent Cum.

---------------------+-----------------------------------

Alive | 10,738 75.82 75.82

TI or new regimen | 1,376 9.72 85.54

Death or lost to fup | 2,048 14.46 100.00

---------------------+-----------------------------------

Total | 14,162 100.00

In other words, there are 1,376 individuals with the event of interest and 2,048 with the competing event. The question is: “What is the cumulative incidence of treatment interruption over time?”

We will perform an analysis of these data in a number of ways. First, we will perform the naïve analysis by considering, as the estimate of the cumulative incidence function, one minus the Kaplan Meier estimate of survival at some time point , i.e.,

We will also assess the effect of perfect ARV adherence on the hazard of TI. This is done as follows:

. stset time2int , fail(fail3==1)

failure event: fail3 == 1

obs. time interval: (0, time2int]

exit on or before: failure

------------------------------------------------------------------------------

14162 total obs.

0 exclusions

------------------------------------------------------------------------------

14162 obs. remaining, representing

1376 failures in single record/single failure data

3548482 total analysis time at risk, at risk from t = 0

earliest observed entry t = 0

last observed exit t = 667

. sts graph, failure by(arvperfectad) ylab(0 0.1 0.2 0.3 0.4)

failure \_d: fail3 == 1

analysis time \_t: time2int

The result is as follows:



Now, we are told that this estimate probably overestimates the true cumulative incidence function. So let’s see what we can do in order to perform a better estimation.

First of all, we will try performing the following steps:

1. Analyze the data considering any failure other than TI as censored data. This will produce the cause-specific hazard
2. Analyze the data considering any failure other than death or loss to follow-up as censored data. This will produce the cause-specific hazard
3. Estimate the discrete survival function
4. Estimate the cumulative incidence function as

Let’s follow this approach.

STEP 1. Analyze the data considering any failure other than TI as censored data. This will produce the cause-specific hazard

. quietly stcox arvperfectad

Predict the baseline hazard of TI (this is the same as the hazard for TI among those with imperfect ARV adherence:

. predict h\_TI\_0, basehc

(12790 missing values generated)

Now we sort the data in ascending order by time and descending order by event. This will ensure that the censored times tied with event times will be placed after the event times:

. gsort \_t - \_d

. by \_t: replace h\_TI\_0 = . if \_n > 1

(994 real changes made, 994 to missing)

Note that, because we had 994 ties in the data with death or loss to follow-up as the event, we needed to remove any duplicate contributions of the hazards during all but the first of these tied event times. Now predict the hazard for TI among those with perfect adherence:

. gen h\_TI\_1 = h\_TI\_0\*exp(\_b[arvperfectad ])

(13784 missing values generated)

STEP 2. Analyze the data considering any failure other than death or loss to follow-up as censored data. This will produce the cause-specific hazard

. stset time2int , fail(fail3==2)

failure event: fail3 == 2

obs. time interval: (0, time2int]

exit on or before: failure

------------------------------------------------------------------------------

14162 total obs.

0 exclusions

------------------------------------------------------------------------------

14162 obs. remaining, representing

2048 failures in single record/single failure data

3548482 total analysis time at risk, at risk from t = 0

earliest observed entry t = 0

last observed exit t = 667

. quietly stcox arvperfectad

Predict the baseline hazard of death or loss to follow-up (DL) (this is the same as the hazard for TI among those with imperfect ARV adherence):

. predict h\_DL\_0, basehc

(12266 missing values generated)

Now we sort the data in ascending order by time and descending order by event. This will ensure that the censored times tied with event times will be placed after the event times:

. gsort \_t - \_d

. by \_t: replace h\_DL\_0 = . if \_n > 1

(1537 real changes made, 1537 to missing)

Note that, because we had 1,537 ties in the data with death or loss to follow-up as the event, we needed to remove any duplicate contributions of the hazards during all but the first of these tied event times. Now predict the hazard for death or loss to follow-up among those with perfect adherence:

. gen h\_DL\_1 = h\_DL\_0\*exp(\_b[arvperfectad ])

(13803 missing values generated)

Before going further, note that there are 513 observations with missing adherence measures:

. tab arvperfectad fail3, missing

arv perfect | fail3

adherence | Alive TI or new Death or | Total

--------------------+---------------------------------+----------

Imperfect adherence | 1,563 577 330 | 2,470

Perfect adherence | 8,818 795 1,566 | 11,179

. | 357 4 152 | 513

--------------------+---------------------------------+----------

Total | 10,738 1,376 2,048 | 14,162

STEP 2a. Generate an indicator to exclude observation without failure from any kind and replace missing hazards with zeros in cases where one or the other failure was observed.

. gen exclude=(missing(h\_TI\_0) & missing(h\_DL\_0))

Note that the following observations will be excluded:

. tab exclude fail3

| fail3

exclude | Alive TI or new Death or | Total

-----------+---------------------------------+----------

0 | 0 378 359 | 737

1 | 10,738 998 1,689 | 13,425

-----------+---------------------------------+----------

Total | 10,738 1,376 2,048 | 14,162

In addition to the 10,738 observations which are censored (since they have not experienced either type of failure) there are 994 ties and 4 observations with missing adherence measures among subjects with a TI and 1537 ties and 152 observations with missing adherence measures among patients who died or were lost to follow-up. These will also be excluded.

Now replace the remaining observations with missing only one of the two types of failure with zero hazard.

. replace h\_TI\_0 =0 if h\_TI\_0 ==. & h\_DL\_0 ~=.

(359 real changes made)

. replace h\_TI\_1 =0 if h\_TI\_1 ==. & h\_DL\_1 ~=.

(359 real changes made)

. replace h\_DL\_0 =0 if h\_DL\_0 ==. & h\_TI\_0 ~=.

(378 real changes made)

. replace h\_DL\_1 =0 if h\_DL\_1 ==. & h\_TI\_1 ~=.

(378 real changes made)

Note that there were 359 changes made among those experiencing a TI and 378 changes among those experiencing death or loss to follow-up.

STEP 3. Estimate the discrete survival function

. gsort -exclude \_t

Note that we sorted the excluded variables first and the non-excluded last, so we can produce a running sum over them[[1]](#footnote-3):

. gen S\_0 = exp(sum(log(1- h\_DL\_0 - h\_TI\_0))) if exclude==0

(13425 missing values generated)

. gen S\_1 = exp(sum(log(1- h\_DL\_1 - h\_TI\_1))) if exclude ==0

(13425 missing values generated)

Note the trick where, instead of performing the product we sum the log survival and exponentiate, i.e., we use the property that,

Finally, we calculate the estimated CIFs and graph:

. gen cif\_TI\_0 = sum(S\_0[\_n-1]\*h\_TI\_0) if exclude==0

(13425 missing values generated)

. label var cif\_TI\_0 "CIF Imperfect adherence"

. gen cif\_TI\_1 = sum(S\_1[\_n-1]\*h\_TI\_1) if exclude==0

(13425 missing values generated)

. label var cif\_TI\_1 "CIF perfect adherence"

. twoway line cif\_TI\* \_t , connect(J J) sort ylab(0 0.1 0.2 0.3 0.4) title(TI)

> ytitle(Cumulative Incidence) xtitle(analysis time)

This results in the following graph:



Now we will generate the Aalen-Johansen estimator of the cumulative incidence function. First we will perform the analysis of the effect of ARV adherence on TI in the presence of death or loss to follow-up.

. stset time2int , fail(fail3==1)

failure event: fail3 == 1

obs. time interval: (0, time2int]

exit on or before: failure

------------------------------------------------------------------------------

14162 total obs.

0 exclusions

------------------------------------------------------------------------------

14162 obs. remaining, representing

1376 failures in single record/single failure data

3548482 total analysis time at risk, at risk from t = 0

earliest observed entry t = 0

last observed exit t = 667

. stcrreg arvperfectad , compete(fail3 == 2)

failure \_d: fail3 == 1

analysis time \_t: time2int

Iteration 0: log pseudolikelihood = -12326.422

Iteration 1: log pseudolikelihood = -12326.263

Iteration 2: log pseudolikelihood = -12326.263

Competing-risks regression No. of obs = 13649

No. of subjects = 13649

Failure event : fail3 == 1 No. failed = 1372

Competing event: fail3 == 2 No. competing = 1896

No. censored = 10381

Wald chi2(1) = 418.43

Log pseudolikelihood = -12326.263 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

| Robust

\_t | SHR Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

arvperfectad | .3258656 .0178622 -20.46 0.000 .2926713 .3628248

------------------------------------------------------------------------------

We note that perfect adherence is associated with an almost 2/3 reduction of the sub-hazard for TI adjusted for the effect of death and loss to follow-up. The CIF estimate based on the Aalen-Johansen estimator is produced as follows:

. sts gen St\_KM=s, by(arvperfectad )

gen CIF\_KM\_0=1-St\_KM if arvperfectad ==0

(11692 missing values generated)

. gen CIF\_KM\_1=1-St\_KM if arvperfectad ==1

(2983 missing values generated)

. label var CIF\_KM\_1 "1-St\_KM (perfect adherence)"

. label var CIF\_KM\_0 "1-St\_KM (imperfect adherence)"

. stcurve, cif at1(arvperfectad =0) at2(arvperfectad =1) title(TI) ylab(0 0.1 0.2 0.3 > 0.4) addplot( line CIF\_KM\_0 \_t , connect(J)||line CIF\_KM\_1 \_t, connect(J))



We see how the Kaplan Meier estimate of the CIF for TI overestimates the true CIF (adjusted for death or loss to follow-up).

On the other hand, plotting the Aalen-Johansen estimator and the one manually produced by the Cox model produces the following graph:

. stcurve, cif at1(arvperfectad =0) at2(arvperfectad =1) title(TI) ylab(0 0.1 0.2 0.3 > 0.4) addplot( line cif\_TI\_0 \_t , connect(J)||line cif\_TI\_1 \_t, connect(J))



The agreement is remarkable!

1. If we did not do this, only values with successively available observations would be summed resulting in a virtually zero result! [↑](#footnote-ref-3)