Practical on Competing Risks in Survival Analysis
Revision: 1.1

Mark Lunt

September 2, 2016

Contents

List of Figures

List of Listings

1 Introduction

This practical aims to illustrate some of the problems caused by competing risks in Survival Analysis, and present some of the solutions available in Stata. It is based on [1], and we will duplicate their results and figures in the course of this practical. The data we are about to analyse concerns 329 homosexual men from the Amsterdam Cohort Studies on HIV infection and AIDS. There are two outcomes of interest, development of AIDS and development of syncytium inducing HIV phenotype (SI from now on). The data is in `aidssi.dta` and consists of the following variables:

`patnr` Subject's ID number

`time` Time at which this subject had an event of interest, or was lost to followup (in years)

`status` Subject's status at `time`: 0 = event-free, 1 = AIDS, 2 = SI

`ccrf` CCR5 genotype: 0 (labelled "WW") = wild type, 1 (labelled "WM") = mutant (a specific deletion exists on at least one chromosome: in practice, no-one in the dataset had the deletion on both chromosomes).

If you wish to work through the commands in this practical, there are a number of add-on packages to Stata that you will need to install. The commands to include these packages are given in Listing 1.

Listing 1 Installing required packages

```
ssc install stcompet.pkg
ssc install moremata.pkg
ssc install stcompadj.pkg
ssc install stpm2.pkg
```

---

2 Non-parametric Survival and cumulative incidence (CI) Curves

2.1 Overall Cumulative Incidence Curves

The first thing we would like to be able to do is to produce survival and cumulative incidence curves. The naive Kaplan-Meier estimates can be calculted with the commands in Listing 2. These commands generate the Kaplan-Meier survival functions for both AIDS and SI, and then calculates the cumulative incidence of SI as 1 - Survival(SI). The plot appears as in Figure 1, and duplicates Figure 2 in Putter *et al.*[1].

Listing 2 Naive Kaplan-Meier

```
clear
use http://personalpages.manchester.ac.uk/staff/mark.lunt/data/aidssi

stset time, f(status == 1)
sts generate aids_s = s
label variable aids_s "Aids-free survival"

stset time, f(status == 2)
sts generate si_s = s
gen si_ci = 1 - si_s
label variable si_ci "Cumulative Incidence of SI"

graph tw line si_ci aids_s time if time <= 13
graph export nkm.pdf, replace
```
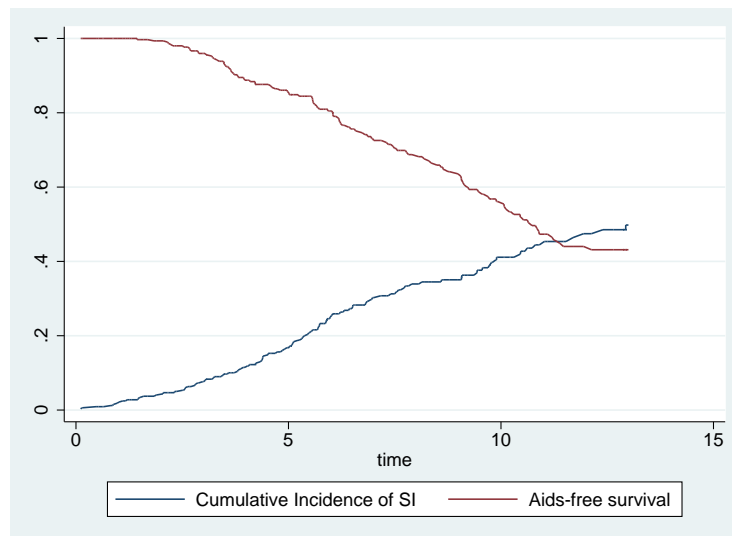

Figure 1: Survival and Cumulative Incidence: Naive Kaplan Meier

Figure 1 clearly shows the bias in the naive Kaplan-Meier estimation. Any given subject is in one of three states: either the have developed AIDS, or they have developed SI, or they have not developed either. In Figure 1, at any given time, the distance from the top of the graph to the AIDS survival curve is the proportion of subjects who developed AIDS, and the distance from the bottom of the graph to the SI cumulative incidence curve is the proportion of subjects who have developed SI. The distance between the lines is the proportion of subjects who have developed neither. Where the lines meet, everybody has developed either AIDS or SI, but the proportion of people with SI continues to increase, as does the proportion of people with AIDS. This is clearly impossible.

The reason for this bias is that the naive Kaplan-Meier approach assumes that censored subjects are still at risk, but subjects who have had a competing event are no longer at risk. In mathematical terms, the incidence of event type k at time t is being calculated as the probability of event type k at time t multiplied by the probability of not having had event type k at time t, $S_k(t)$. It should be multiplied by the probability of not having had any event at time t, $S(t)$. This can be done using the command $\texttt{stcompet}$, as shown in Listing 3

Listing 3 Generating Corrected Cumulative Incidence and Survival Curves

```
stset time, f(status == 1)
stcompet ci = ci, compet1(2)
gen si_ci2 = ci if status == 2
gen aids_ci2 = ci if status == 1
gen aids_s2 = 1 - aids_ci2
label variable aids_s2 "Aids-free survival (corrected)"
label variable aids_s "Aids-free survival (naive)"
label variable si_ci2 "Cum. Inc. of SI (corrected)"
label variable si_ci "Cum. Inc. of SI (naive)"
graph tw line aids_s2 si_ci2 aids_s si_ci time if time <= 13, ///
        lcolor(navy maroon ltblue erose)
```

Listing 3 produces the graph in Figure 2, which is the same a Figure 3 in Putter *et al.*[1]. If you prefer the stacked cumulative incidence curves they show in Figure 4, you can use the commands in Listing 4 to produce Figure 3. Note that the area between the plotted line and the x-axis is shaded, so the order of the variables is essential to avoid the cumulative incidence of AIDS and SI from hiding the cumulative incidence of AIDS.

2.2 Comparing Groups

It is also possible to produce survival curves for two different groups, and obtain a visual comparison of the groups. However, to test whether the differences between the groups are statistically significant will require a more sophisticated approach.

There is a suggestion that there is a specific deletion in the C-C chemokine receptor 5 gene (CCR5) that reduces susceptibility to AIDS. SI viruses can use a
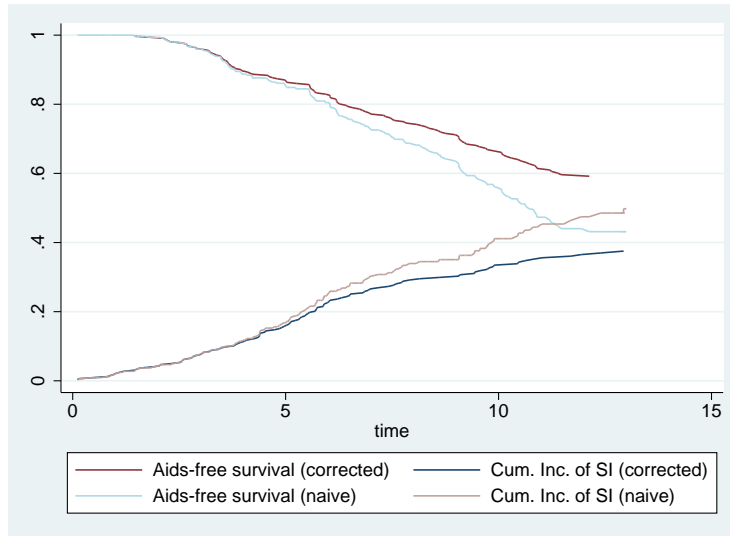
Figure 2: Survival and Cumulative Incidence: Naive and Corrected Kaplan Meier Estimates

---

Listing 4 Stacked Cumulative Incidence

```
sort time
replace aids_ci2 = 0 if _n == 1 & aids_ci2 == .
replace si_ci2 = 0 if _n == 1 & si_ci2 == .
replace aids_ci2 = aids_ci2[_n-1] if aids_ci2 == .
replace si_ci2 = si_ci2[_n-1] if si_ci2 == .
gen asi_ci2 = aids_ci2 + si_ci2
label variable aids_ci2 "Cum. Inc. Aids"
label variable asi_ci2 "Cum. Inc. SI"
graph tw area asi_ci2 aids_ci2 time if time <= 13, ///
            ylab(0 (0.2) 1) yline(1) xlab(0 (3) 9 13)
graph export stack.pdf, replace
```
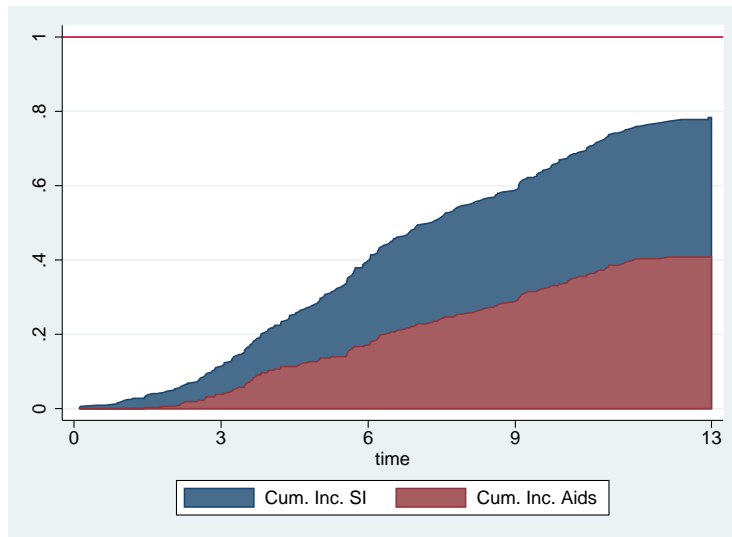
Figure 3: Stacked Cumulative Incidence Curves using Corrected Kaplan Meier Estimates

different pathway, and are not expected to be affected by this mutation. We can compare the genotypes visually using the code in Listing 5. This will reproduce the non-parametric cumulative incidence functions given in Figure 9 in Putter *et al.*[1], and reproduced here as Figure 4.

Listing 5 Producing Cumulative Incidence Curves For Each Genotype

```
stcompet ci_ccr = ci, compet1(2) by(ccr5)
gen aids_ci_ccr0 = ci_ccr if status == 1 & ccr5 == 0
gen aids_ci_ccr1 = ci_ccr if status == 1 & ccr5 == 1
gen si_ci_ccr0 = ci_ccr if status == 2 & ccr5 == 0
gen si_ci_ccr1 = ci_ccr if status == 2 & ccr5 == 1

label variable aids_ci_ccr0 "WW"
label variable aids_ci_ccr1 "WM"
label variable si_ci_ccr0 "WW"
label variable si_ci_ccr1 "WM"

graph tw line aids_ci_ccr0 aids_ci_ccr1 time if time <= 12,  ///
      lcolor(maroon navy) title("AIDS") ytitle("Probability")   ///
      xtitle("Years from HIV infection") xlab(0(2)12)        ///
      connect(stairstep stairstep) name("npaids")

graph tw line si_ci_ccr0 si_ci_ccr1 time if time <= 12,  ///
      lcolor(maroon navy) title("SI") ytitle("Probability")   ///
      xtitle("Years from HIV infection") xlab(0(2)12)        ///
      connect(stairstep stairstep) name("npsi")

graph combine npaids npsi
graph export "npci.pdf", replace
```


Figure 4: Cumulative incidence by genotype

3   Modelling Hazards

3.1   Fitting Cox Regression Models

The above commands can produce descriptive statistics (and graphs) for survival and cumulative incidence in the presence of competing risks, but cannot be used for modelling survival. Fortunately, the standard models for survival analysis give unbiased estimates of the hazard in the presence of competing risks.

By default, subjects with a competing risk are treated as censored, which for computing hazards is appropriate (the hazard is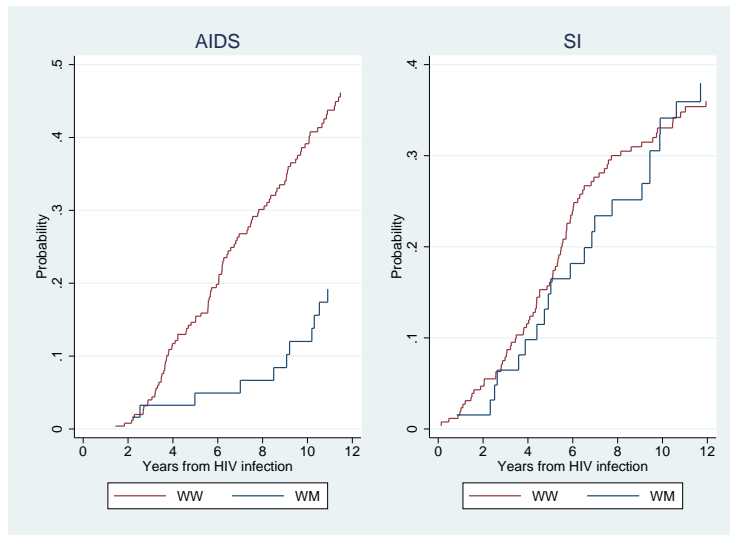 the risk of having the event among those still at risk, and those who have had a competing risk are no longer at risk). So to get the cause specific hazard for AIDS, we merely need to call `stcox`, since the data is already `stset`. The results are shown in Listing 6: as expected, this risk of AIDS is lower in the subjects with the mutant genotype (HR = 0.29).

Listing 6 Hazard Ratio For AIDS From Cox Model

```
. stcox ccr5

        failure _d:  status == 1
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           324                  Number of obs   =        324
No. of failures =           113
Time at risk    =  2261.959996
                                                 LR chi2(1)      =      21.98
Log likelihood  =   -555.37301                   Prob > chi2     =     0.0000

------------------------------------------------------------------------------
         _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
       ccr5 |   .2906087   .0892503    -4.02   0.000     .1591812     .530549
------------------------------------------------------------------------------
```

If we want to calculate the cause-specific hazard of developing SI, we need to re-`stset` the data to make `status = 2` the failure outcome, and rerun the same `stcox` command. In this case there is no effect of the CCR5 mutation, as shown in Listing 7: the hazard ratio is below 1, but the effect is not statistically significant. These `stcox` models are exactly the same as the models fitted by Putter *et al.*[1]

What about if we want to compare the hazard ratios for the two different outcomes ? We can do this, but we need to use a little trick that depends on the fact that the cause-specific hazards for different causes are independent. We can expand the dataset to contain two records for each subject, one for their AIDS outcome and one for their SI outcome. We will create a new variable `event_type`, which contains 1 for the AIDS outcome and 2 for the SI outcome,

9

Listing 7 Hazard Ratio For SI From Cox Model

```
. stset time, fail(status = 2)

-------------------------------------------------------------------------------
      329  total obs.
        0  exclusions
-------------------------------------------------------------------------------
      329  obs. remaining, representing
      108  failures in single record/single failure data
 2274.551  total analysis time at risk, at risk from t =         0
                                earliest observed entry t =         0
                                   last observed exit t =    13.936

. stcox ccr5

        failure _d:  status == 2
  analysis time _t:  time

Cox regression -- no ties

No. of subjects =          324                  Number of obs   =        324
No. of failures =          107
Time at risk    =  2261.959996
                                                LR chi2(1)      =       1.19
Log likelihood  =   -549.73443                  Prob > chi2     =     0.2748

-------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+-------------------------------------------------------------------
      ccr5 |   .7755334    .1846031    -1.07   0.286     .4863914    1.23656
-------------------------------------------------------------------------------
```

and a new variable `fail` which contains 0 if that subjects did not have that
outcome (either they were censored or they had the other outcome) and 1 if
they had that outcome. The values of the relevant variables before and after
expansion is given in Table 1: note that a given subject can have at most one
"1" in the fail column. The code to perform the transformation is in Listing 8.
The command to generate the fail variable is a bit tricky: will be 1 whenever
`status == event_type` and 0 otherwise, which is exactly what we want.

Listing 8 Expanding Dataset

```
use aidssi
expand 2
sort patnr
by patnr: gen event_type = _n
gen fail = status == event_type
```

Patno	CCR5	Time	Before status	After event_type	fail
1	WW	9.106	AIDS	1	1
				2	0
2	WM	11.039	Event Free	1	0
				2	0
3	WW	2.234	AIDS	1	1
				2	0
4	WM	9.878	SI	1	0
				2	1
5	WW	3.819	AIDS	1	1
				2	0

Table 1: Effect of expanding data

Now we can fit the same Cox regression models as we did with the one
observation per subject data. We can fit them separately with the first two
commands in Listing 9, or fit them as a single model, which is what the third
command does. The third command uses the factor notation, which you may
want to type `help fvvarlist` in Stata to remind yourself about these.

Listing 9 Fitting Cox Model to long data

```
stcox ccr5 if event_type == 1
stcox ccr5 if event_type == 2
stcox 1.ccr5#1.event_type 1.ccr5#2.event_type, strata(event_type)
```

I have not shown the output from these commands, but it it is functionally
identical to the output we got from the same models above. These models are

equivalent to those fitted by Putter *et al.* at the bottom of page 2404. To test whether the effect of the CCR5 mutation differs between the output types, we need to fit an interaction between `ccr5` and `event_type`, as shown in Listing 10.

Listing 10 Comparing Hazard Ratios For Different Outcomes

```
stcox i.ccr5##i.event_type, strata(event_type)
```

3.2 Producing Cumulative Incidence Curves from a Cox Regression Model

Producing cumulative incidence curves from a Cox regression model requires the `stcompadj` command[1]. The syntax is somewhat complicated, since it allows for different variables to predict the different competing outcomes. I have used the `showmod` option to show that it is based on exactly the same Cox regression model as we have seen previously. The cumulative incidence curves are shown in Figure 5, which is equivalent to Figure 5 in Putter *et al.*[1], and the commands to produce it are in Listing 11

Listing 11 Cumulative Incidence Curves from a Cox Regression Model

---

```
clear
use aidssi
stset time, fail(status=1)
stcompadj ccr=0 , compet(2) maineffect(ccr) competeffect(ccr) gen(Main0 Compet0) showmod
stcompadj ccr=1 , compet(2) maineffect(ccr) competeffect(ccr) gen(Main1  Compet1) showmod

graph tw line Main* _t if _t < 13, connect(J J)  yscale(range(0 0.5))  ///
            sort title(AIDS) ytitle(Cumulative Incidence)              ///
            xtitle(analysis time) name(cox_ci1, replace)               ///
            legend(label (1 "WW") label(2 "WM"))

graph tw line Compet** _t if _t < 13, connect(J J)  yscale(range(0 0.5)) ///
            sort title(SI) ytitle(Cumulative Incidence)                ///
            xtitle(analysis time) name(cox_ci2, replace)               ///
            legend(label (1 "WW") label(2 "WM"))

graph combine cox_ci1 cox_ci2
graph export cox-ci.pdf, replace
```

The above cumulative incidence curves are based on the Cox regression model, in which the baseline hazard is recalculated at every event, and can take any value at these times. An alternative approach allowed by `stcompadj` is

---

[1]There is an alternative manual method, outlined in the `stcrreg` entry in the ST manual, but that his tedious. It is used in Listing 11, since it is more flexible than `stcompadj`
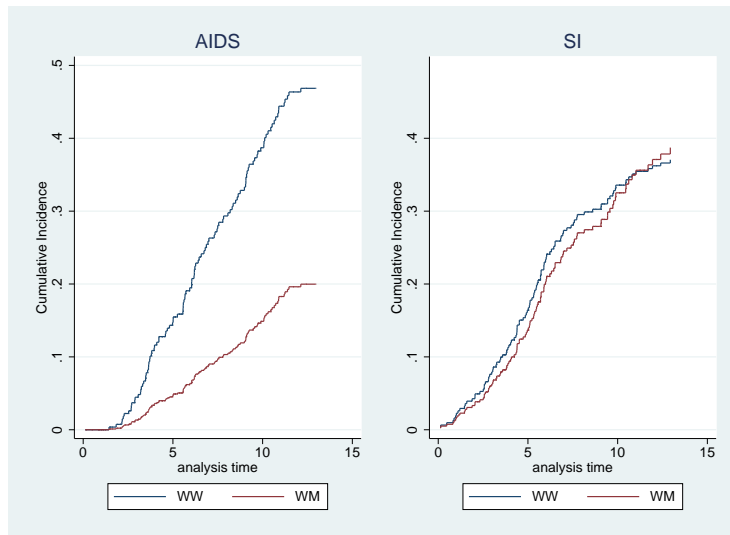
Figure 5: Cumulative Incidence Curves from Cox Model

to use the `stpm2` package to model the baseline hazard. This produces a smooth curve for the baseline hazard, rather than a stairstep function: the resulting cumulative incidence curves can be seen in Figure 6, the code to produce them is in Listing 12.

It is important to note that the effect of a given variable on the cumulative incidence of a given outcome depends not only on the hazard ratio for that outcome, but also on the hazard ratios for the competing risks, and indeed on the baseline hazards for the competing risks. The cumulative incidence will decrease if the hazard of a competing risk increases, since there will be fewer surviving subjects who could possibly have the outcome of interest.

Putter *et al.*[1] demonstrate this by showing how the cumulative incidence curves for the different genotype groups change as the hazard of AIDS changes. The 6 graphs in Figure 7 (which corresponds to Figure 7 Putter *et al.*[1]) show the cumulative incidence of SI in the two genotype groups, and all that differs between the graphs is the hazard of AIDS. The hazard observed in the dataset we are using is multiplied by a factor before calculating the overall survival (needed to calculate the cumulative incidence of SI): the factor was allowed to vary from 0 to 4.

The code used to generate Figure 7 is relatively long and complicated, and not something you would want to do in an analysis, so I have relegated it to Appendix A. However, it does illustrate a very important feature of cumulative incidence curves based on Cox regression. All that differs between the graphs in this figure is the baseline hazard of AIDS, which increases as the factor gets bigger. However, this leads to changes in the differences between genotypes for the cumulative incidence of SI. This is because as more subjects develop

13

Listing 12 Cumulative Incidence Curves from a Cox Regression Model

```
stcompadj ccr=0 , compet(2) maineffect(ccr) competeffect(ccr)  ///
               flexible gen(Mainf0 Competf0) showmod
stcompadj ccr=1 , compet(2) maineffect(ccr) competeffect(ccr)  ///
               flexible gen(Mainf1  Competf1) showmod

graph tw line Mainf* _t if _t < 13,  yscale(range(0 0.5))  ///
            sort title(AIDS) ytitle(Cumulative Incidence)           ///
            xtitle(analysis time) name(stpm_ci1, replace)            ///
            legend(label (1 "WW") label(2 "WM"))

graph tw line Competf* _t if _t < 13,  yscale(range(0 0.5)) ///
            sort title(SI) ytitle(Cumulative Incidence)             ///
            xtitle(analysis time) name(stpm_ci2, replace)            ///
            legend(label (1 "WW") label(2 "WM"))

graph combine stpm_ci1 stpm_ci2
graph export stpm-ci.pdf, replace
```
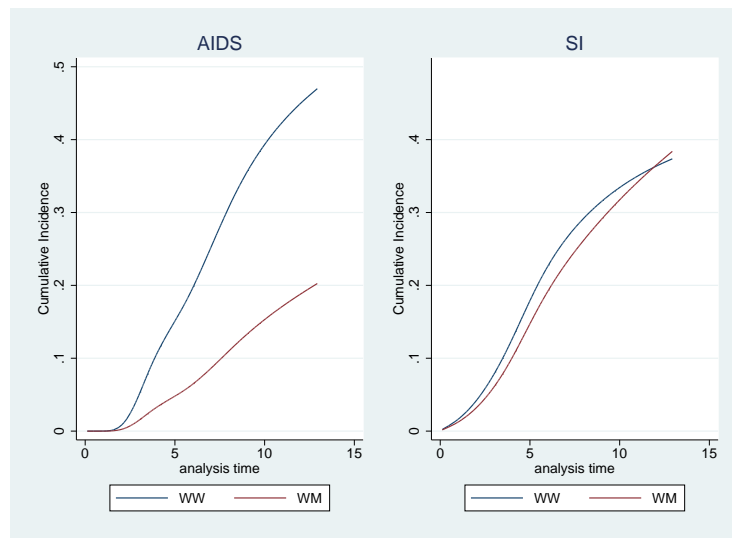

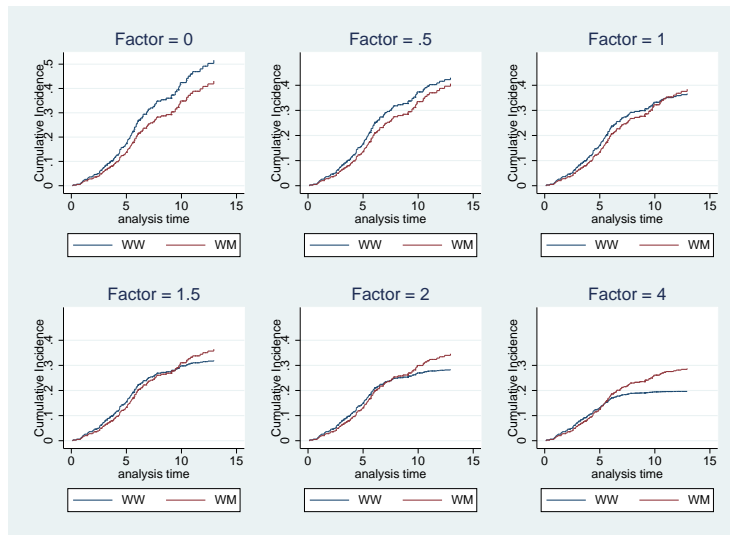Figure 6: Cumulative Incidence Curves from Cox Model

Figure 7: Effect of Changing Hazard of Aids Cumulative Incidence Curves from Cox Model

AIDS, there are fewer subjects who can develop SI, and hence the cumulative incidence must go down. Since AIDS is more common in the WW genotype, the cumulative incidence of SI reduces more in this genotype than in the WM genotype as the hazard of AIDS increases, so this genotype appears to have a strongly protective effect. This appears to contradict the fact that the hazard ratio for SI in this group is not statistically significant, but in fact the effect on the incidence is due to the increased prevalence of AIDS in this group, which reduces the number of subjects susceptible to SI.

4   Modelling Survival and Cumulative Incidence Curves

An alternative to using Cox regression and modelling the hazard ratios is to use the Fine and Gray regression model[2] for competing risks, implemented as `stcrreg` in Stata. In this model, the cumulative incidence depends only on the cause_specific hazard, so the cumulative incidence graph reflects the hazard ratio more directly. However, the hazard ratio calculated in this model has no real interpretation as a hazard ratio, since subjects who have a competing event are still considered to be at risk. This also changes the proportional hazards assumption: the hazards are still assumed to be proportional, but the definition of the hazard has changed. If the proportional hazards assumption is true for the Cox model, it need not be true for this model, and vice-versa.

The commands to fit this model in Stata and produce a cumulative incidence

graph are given in Listing 13. The option `nohr` is given to duplicate the results of Putter et al at the top of page 2410, and the graph in Figure 8 is equivalent to Figure 8 in Putter et al.

Listing 13 Producing cumulative incidence Curves from Fine-Gray model

---

```
clear
use aidssi
stset time, fail(status=1)
stcrreg i.ccr5, compet(status=2) nohr
stcurve, cif at1(ccr5 = 0) at2(ccr5=1) range(0 13) title("AIDS") ///
        legend(label (1 "WW") label(2 "WM")) name(craids)

stset time, fail(status=2)
stcrreg i.ccr5, compet(status=1) nohr
stcurve, cif at1(ccr5 = 0) at2(ccr5=1) range(0 13) title("SI") ///
        legend(label (1 "WW") label(2 "WM")) name(crsi)

graph combine craids crsi
graph export "crci.pdf", replace
```
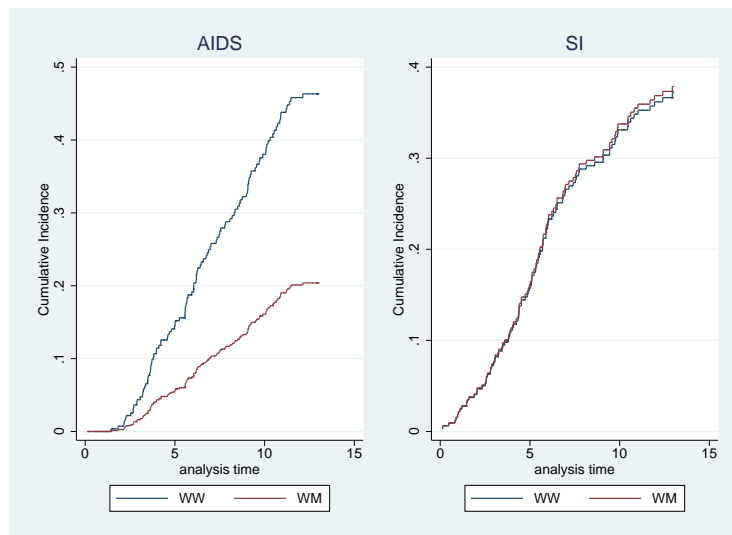

Figure 8: Cumulative incidence by genotype from Fine-Gray Model

5 Further Exercises

5.1 Data

In this section, you will perform all of the same the analyses on a different dataset. This dataset consists of 541 patients with early disease stage follicular cell lymphoma, treated with radiation alone (chemo = 0) or radiation and chemotherapy (chemo = 1). The outcome of interest is time to relapse, with death in remission as a competing risk. The outcome is coded in cause, with 1 = relapse, 2 = death, 0 = censored, and the time of the event (relapse, death or censoring) is in the variable time. Other potential predictors of the outcome are age and stage (which can take the values 1 or 2). The command to read in the data is

use http://personalpages.manchester.ac.uk/staff/mark.lunt/data/follic

5.1   How many subjects had a relapse, and how many died ?

....

5.2   What was the mean age of those who relapsed, and of those who died ?

....

5.3   How many patients were in stage 1, and how many at stage 2 ?

....

5.4   How many patients received chemotherapy ?

....

5.2 Non-parametric Curves

5.5   Use stset to prepare the data for analysis, and use sts generate to create a new variable relapse_s1 containing the the probabilites of not having relapsed at any given time, using the naive Kaplan Meier method.

5.6   Create the cumulative incidence of relapse, calculated using the naive Kaplan Meier method, as 1 - survival probability.

5.7   Use stcompet to calculate the cumulative incidence of relapse accounting for the competing risk of dying.

5.8   Plot both cumulative incidence curves on the same figure, to see the
        diffence in estimated cumulative incidence caused by the Kaplan Meier
        method not accounting for competing risks appropriately.

5.9   Use `stcompet` to graph the cumulative incidence of relapse in those at
        stage 1 and those at stage 2 separately. Which group has the higher

        incidence ?   ......

5.10   Use `stcompet` to graph the cumulative incidence of mortality in those
        at stage 1 and those at stage 2 separately (you will need to rerun `stset`

        first). Which group has the higher mortality ?   ......

5.3   Proportional Hazards Models

5.11   Run `stset` again to make relapse the outcome of interest, and fit a Cox
        regression model to assess the effects of age, stage and chemotherapy.
        Which variables are significant predictors of relapsing ?

        ....

5.12   Run `stset` again to make mortality the outcome of interest, and fit a Cox
        regression model to assess the effects of age, stage and chemotherapy.
        Which variables are significant predictors of mortality ?

        ....

5.13   Expand the dataset so that there are two records for each subject, one
        with relapse as the outcome and one with mortality as the outcome (see
        Listing 8. Fit a Cox regression model with age, stage, chemotherapy
        and their interactions with cause as predictors, and use `lincom` to verify
        that you get the same results as in the previous two questions.

5.14   Does the effect of chemotherapy on mortality differ significantly from its

        effect on relapse ?   ......

5.15   Does the effect of age on mortality differ significantly from its effect on

        relapse ?   ......

5.16   Use the `stcompadj` command to calculate the cumulative incidenec of relapse for subjects at stage 1 and stage 2 separately, controlling for age and chemotherapy. On the same graph, CI curves for stage 1 and stage 2, both with and without controlling for age and chemotherapy. What effect has controlling for age and chemotherapy had on the effect of stage ?

....

5.17   Confirm your answer to the previous question by fitting a Cox regression model with stage as the only predictor. How does the unadjusted hazard ratio differ from the adjusted hazard ratio ?

....

5.18   Use `stcompadj` to produce CI curves for stages 1 and 2 at ages 30 and 70, and plot all four curves on a single graph. Which has the greater effect on incidence: changing stage from 1 to 2, or changing age from 30 to 70 ?

....

5.19   Use `lincom` to calculate the hazard ratio for a 40-year increase in age. Is this larger or smaller than the hazard ratio for the difference between

stage 1 and stage2 ?   ......

5.20   Use `lincom` to test whether the difference between the hazard ratio for a 40-year increase in age and the hazard ratio for a change from stage 1 to stage 2 is statistically significant. What do you conclude ?

....

5.4   Cumulative Incidence Models

5.21   Use `stcrreg` to fit a Fine-Gray model with relapse as the outcome of interest and age, stage and chemotherapy as predictors. Which variables are statistically significant predictors ?

....

5.22   Use `stcurve` to produce CI curves for subjects at stages 1 and 2, and at ages 30 and 70 (i.e. replicating the figure you drew in question 5.18). Which has the large effect: a 40 year increase in age or a change from stage 1 to stage 2 ?

....

5.23   Use `lincom` to test whether the difference between the hazard ratio for a 40-year increase in age and the hazard ratio for a change from stage 1 to stage 2 is statistically significant. What do you conclude ?

....

References

[1] Putter H, Fiocco M, Geskus RB Tutorial in biostatistics: Competing risks and multi-state models Statistics in Medicine 2007;26:2389–2430.

[2] Fine JP, Gray RJ A proportional hazards model for the subdistribution of a competing risk Journal of the American Statistical Association 1999; 94:496–509.

A   Cumulative Incidence Curves after Cox Regression

Listing 14

```
use aidssi, clear
stset time, failure(status == 2)
stcox ccr5
predict h_si_0, basehc
gen h_si_1 = h_si_0*exp(_b[ccr5])

stset time, failure(status == 1)
stcox ccr5
predict h_aids_0, basehc
gsort _t -_d
by _t: replace h_aids_0 = . if _n > 1
gen h_aids_1 = h_aids_0*exp(_b[ccr5])

drop if missing(h_si_0) & missing(h_aids_0)
replace h_aids_0 = 0 if missing(h_aids_0)
replace h_aids_1 = 0 if missing(h_aids_1)
replace h_si_0 = 0 if missing(h_si_0)
replace h_si_1 = 0 if missing(h_si_1)

sort _t
gen S_0 = exp(sum(log(1- h_aids_0 - h_si_0)))
gen S_1 = exp(sum(log(1- h_aids_1 - h_si_1)))

gen cif_si_0 = sum(S_0[_n-1]*h_si_0)
label var cif_si_0 "WW"
gen cif_si_1 = sum(S_1[_n-1]*h_si_1)
label var cif_si_1 "WM"

twoway line cif_si* _t if _t<13, connect(J J) yscale(range(0 0.5))  ///
       sort title(SI) ytitle(Cumulative Incidence)           ///
       xtitle(analysis time) name(cox_ci1, replace)

gen cif_aids_0 = sum(S_0[_n-1]*h_aids_0)
label var cif_aids_0 "WW"
gen cif_aids_1 = sum(S_1[_n-1]*h_aids_1)
label var cif_aids_1 "WM"

twoway line cif_aids* _t if _t<13, connect(J J) yscale(range(0 0.5))  ///
       sort title("AIDS") ytitle(Cumulative Incidence)      ///
       xtitle(analysis time) name(cox_ci2, replace)

graph combine cox_ci1 cox_ci2
graph export cox_ci.pdf
```

Listing 15 Cumulative incidence functions for SI, as effect of CCR5 on hazard of AIDS varies

```
foreach num of numlist 0 1 2 3 4 8 {
gen h_aids_0_`num' = h_aids_0 * `num'/2
gen h_aids_1_`num' = h_aids_1 * `num'/2
gen S_0_`num' = exp(sum(log(1- h_aids_0_`num' - h_si_0)))
gen S_1_`num' = exp(sum(log(1- h_aids_1_`num' - h_si_1)))
gen cif_si_0_`num' = sum(S_0_`num'[_n-1]*h_si_0)
label var cif_si_0_`num' "WW"
gen cif_si_1_`num' = sum(S_1_`num'[_n-1]*h_si_1)
label var cif_si_1_`num' "WM"

local factor = `num' / 2
twoway line cif_si_0_`num' cif_si_1_`num' _t if _t<13,  ///
   connect(J J) yscale(range(0 0.5))  ///
           sort title("Factor = `factor'") ytitle(Cumulative Incidence)      ///
           xtitle(analysis time) name(cox_ci_`num', replace)
}

graph combine cox_ci_0 cox_ci_1 cox_ci_2 cox_ci_3 cox_ci_4 cox_ci_8
```