Chapter 12

Competing risks

Consider the study of the time from the start of therapy with antiretroviral medications in HIV infection until the change or treatment interruption (TI) (the event of interest).

What happens when a patient dies before he or she experiences a TI?

Competing risks are encountered in studies where the subjects under study are at risk for more than one mutually exclusive event or failure cause. They are events whose occurrence either precludes the occurrence of another event or fundamentally alters the probability of the occurrence of this other event. Note that this is different from censoring where the event happens later but we just don't know when.

12.1 The inadequacy of the naïve analysis

In the treatment interruption example, we would usually ask the question:

> *What is the probability of no TI before six months?*

In this case, death or loss from follow-up before observing a TI is a competing event[1] For this reason, it is helpful, in the competing-risk setting, to reverse the question and instead ask:

> *What is the probability that nothing happens before six months but when an event occurs it is a TI?*

Thus, in the competing-risk setting, instead of working with the survival function we instead work with the cumulative incidence function (CIF), that is, the probability that the event of interest occurs before a given time.

[1] This is not the same as censoring where the event happens at a later time but we just don't see exactly when!

12.2 Basic theory for competing risks analysis

One approach in the analysis of competing risk data assumes the existence of k failure times, one for each possible type of failure. In studies were competing risks are present we only observe the minimum of the latent failures times (T) and the corresponding cause of failure (C).

The problem with this approach is that neither the joint distribution of the failure times nor the corresponding marginal distributions are identifiable from the observed data without additional assumptions (such as independence of the different latent failure times).

The second approach considers the joint distribution of failure time T and cause of failure C, two observable random variables

Taking this approach we define the j-th cause-specific hazard at time t by

$$\lambda_j(t) = \lim_{h \to 0} \frac{Pr(t \leq T < t + h, C = j | T \geq t)}{h}$$

This is the instantaneous failure rate from cause j at time t in the presence of all other possible causes of failure.

The overall hazard $\lambda(t)$ from any cause, is equal to the sum of the cause-specific hazards over each failure type. If d_{ic} is the indicator that subject i has failed from cause c and t_i the minimum of event or censoring time of the i-th subject, the likelihood function for hazard parameters β under independent right censoring and without further assumptions is

$$\begin{aligned} L(\beta) &= \prod_{c=1}^{k} \prod_{i=1}^{n} \lambda_c(t_i)^{d_{ic}} S_c(t_i) \\ &= \prod_{c=1}^{k} \prod_{i=1}^{n} \lambda_c(t_i)^{d_{ic}} \exp\left[- \int_0^{t_i} \lambda_c(u) du \right] \end{aligned}$$

$S_c(t_i) = \exp\left[- \int_0^{t_i} \lambda_c(u) du \right]$ is a quantity which is estimable from the data, since it is a function of the identifiable cause-specific hazard $\lambda_c(u)$.

12.3 The cumulative incidence function

The CIF, or probability of failure from the event of interest j before time t in the presence of all other possible causes, is a function of the cause-specific hazards for all causes of failure:

$$\begin{aligned} F_j(t) &= Pr(T \leq t, C = j) \\ &= \int_0^t \lambda_j(u) \exp\left[- \int_0^u \sum_{c=1}^{k} \lambda_c(w) dw \right] du \end{aligned}$$

It should be noted that if, in the presence of competing risks we insist to estimate the cumulative incidence by the naïve estimator (one minus the Kaplan-Meier estimator after treating observations with failure from other causes as censored), the following inequality holds:

$$1 - \hat{S}_j(t) \geq \hat{F}_j(t)$$

Thus $1 - \hat{S}(t)$ overestimates in general the cumulative incidence of cause j.

In fact, it is possible for the sum of the naïve estimates of the cumulative incidence at time t over all possible causes to be greater than 1, which should be impossible as this sum is the cumulative probability for failure from any cause.

12.4 Modeling of the CIF by Cox regression

Modeling the CIF with Cox regression in the TI example above would require the following steps:

1. Perform a Cox regression of time until the occurrence of a TI considering deaths and losses from follow-up occurring before a TI as censored observations
2. Produce the estimated hazard from all observation from the regression in Step 1
3. Fit a Cox regression with death or loss from follow-up as the event of interest considering the occurrence of a TI before death or loss from follow-up as a censoring event
4. Predict the hazard from the model in Step 3
5. Calculate the estimated CIF manually as

$$\hat{F}_{\text{TI}}(t) = \sum_{i:t_i \leq t} \hat{h}_{\text{TI}}(t_i)\hat{S}(t_{j-1})$$

   where $\hat{S}(t) = \prod_{i:t_i \leq t} \left\{1 - \hat{h}_{\text{TI}}(t_i) - \hat{h}_{\text{death}}(t_i)\right\}$ and $\hat{h}_j(t)$ is the cause-specific hazard defined earlier.
6. Plot the resulting CIF function

12.5 The Fine & Gray model of the cumulative incidence function

Fine & Gray (1999) proposed a framework for regression modeling with cumulative incidence functions.

Their method makes use of the so-called *subdistribution hazard*, which is a function of the cumulative incidence for the corresponding cause of failure and is defined as

$$\lambda_j^{\text{sub}}(t; \mathbf{Z}) = \lim_{h \to 0} \frac{1}{h} Pr\left[t \leq T < t + h, C = j | T \geq t \cup (T \leq t, C \neq j), \mathbf{Z}\right]$$

The idea behind the Fine & Gray model is that the risk set at time t includes not only subjects who have not yet experienced the event of interest (e.g., TI in our example), but also subjects who have failed from other causes before t (e.g., they have died or were lost to follow-up), and are not physically at risk for the event of interest at t!

In other words, patients who have failed from cause other than j remain in the risk set for cause j.

Fine and Gray proposed a semiparametric proportional hazards model for the subdistribution hazard of the event of interest:

$$\lambda_j^{\text{sub}}(t) = \lambda_{0,j}^{\text{sub}}(t) \exp(\mathbf{x}\beta)$$

where $\lambda_{0,j}^{\text{sub}}(t)$ is the baseline subhazard.

The subhazard $\lambda_j^{\text{sub}}(t)$, does not properly addresss the probability of failure from any cause. However, once the subhazard has been estimated, the CIF for the event of interest is readily available as

$$F_j(t) = 1 - \exp\left[\Lambda_j^{\text{sub}}(t)\right]$$

where $\Lambda_j^{\text{sub}}(t) = \int_0^t \lambda_j^{\text{sub}}(t)dt$ is the cumulative subhazard.

Parameter estimation for this model depends on the right-censoring mechanism.

When censoring is administrative (i.e., caused by the end of the study or the freezing of the database), the censoring time is known even for subjects who fail for other causes before the administrative censoring time.

Unfortunately, for general censoring at random, the time for which a patient who has failed from a competing event remains "at risk" for cause j is not known.

12.5.1 Implementation of the Fine& Gray model by STATA

Stata uses a weighted procedure, which maximizes a weighted version of the Breslow log likelihood of the Cox model, i.e.,

$$\log L(\beta) = \sum_{i=1}^{n} \delta_i \left[\mathbf{Z}_i\beta - \log \left\{ \sum_{j \in R_j} w_{ji} \exp(\mathbf{Z}_j\beta) \right\} \right]$$

where,

$$w_{ji} = \begin{cases} \frac{\hat{S}_c(t_i)}{\hat{S}_c\{\min(t_j,t_i)\}} & \text{if subject } j \text{ has a competing event at time } t_j \\ 1 & \text{otherwise} \end{cases}$$

and $\hat{S}_c(t)$ is the Kaplan-Meier estimate of the *censoring* distribution (i.e., the distribution where censoring is the event of interest).

12.6 Example: Antiretroviral treatment interruption in HIV

Consider the situation where a number of patients infected with the human immunodeficiency virus commence antiretroviral therapy (ART). Due to a number of reasons like toxicity, developing of viral resistance to their drug regimen or personal reasons, these patients may change or interrupt their treatment (we will consider this as a combined endpoint, which we call "treatment interruption" or TI).

In addition, these patients may also die from complications arising from their HIV infection or stop attending the clinic and be lost to follow-up (LTFU; we will consider death or

lost-to-follow-up as a combined competing event).

The event time, **time2int**, is the time from ART start to either TI or death and the event is **fail3** (0=Alive; 1=TI or regimen change; 2=Death or LTFU).

12.6.1 The TI data set

The data set, and the relevant events are as follows: In other words, we have 14,162 individ-

Table 12.1: Frequency table of relevant entpoints

Endpoint	Frequency	Percent
Alive	10,738	75.82
TI or new regimen	1,376	9.72
Death or lost to fup	2,048	14.46
Total	14,162	100.00

uals who started ART, of whom, 10,738 are alive without either event, while 1,376 (9.7%) experienced a treatment interruption *before* being lost to follow-up or dying, while 2,048 (14.5%) were confirmed deceased or were lost to follow-up.

12.6.2 Risk factors for TI

We will consider the following risk factors as possibly being related with the risk of TI or death/loss-to-follow-up are as follows:

- **traveltime**: This is the time it takes to reach the clinic (1=<30 minutes; 2=30-60 minutes; 3=1-2 hours; 4=> 2 hours)
- **male**: Male gender (1=Male, 0=Female)
- **Arvperfectad**: Self-reported perfect adherence (1=Perfect adherence, 0=Imperfect adherence)
- **cd4_cat**: CD4 category (0=<50 cells/μl; 1=50-100; 2=100-200; 3=> 200)
- **whoatarvstart**: World Health Organization (WHO) stage at the start of ART (1-4: higher means more serious disease)
- **urban**: Urban/referral hospital (=1) versus non-urban (rural) hospital (=0)

12.6.3 Analysis of cause-specific hazards: TI

First we carry out an analysis of the cause-specific hazard of TI in the presence of death or LTFU:

```
stset time2int, f(fail3==1)
```

Output omitted

```
xi: stcox i.traveltime i.male i.arvperfectad i.cd4_cat i.whoatarvstart i.urban
```

Output omitted

```
Cox regression -- Breslow method for ties

No. of subjects =        13201                     Number of obs   =     13201
No. of failures =         1270
Time at risk    =      3262128
                                                   LR chi2(3)      =     21.75
Log likelihood  =   -11362.848                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  traveltime |
           2 |   1.160588   .1173503     1.47   0.141      .951943    1.414965
           3 |   1.367276   .1396009     3.06   0.002     1.119302    1.670187
           4 |   1.290106   .1473656     2.23   0.026     1.031323    1.613823
      1.male |   .9347786   .0691897    -0.91   0.362     .8085469    1.080718
1.arvperfe~d |   .3430612   .0243896   -15.05   0.000     .2984394    .3943547
     cd4_cat |
           1 |    .777618   .0758362    -2.58   0.010     .6423236    .9414097
           2 |   .7285993   .0646583    -3.57   0.000     .6122803    .8670162
           3 |   .6327438   .0797538    -3.63   0.000     .4942409    .8100598
whoatarvst~t |
           2 |   1.068188   .1254769     0.56   0.574     .8485149    1.344732
           3 |   1.144179   .1180587     1.31   0.192     .9346845    1.400627
           4 |   1.345382   .1947441     2.05   0.040     1.013058    1.786722
       urban |   .6842188    .050356    -5.16   0.000     .5923107     .790388
------------------------------------------------------------------------------
```

Comments

- Travel time, especially above one hour, is associated with a significant increase in the risk of TI
- There is no significant difference in TI hazard between males and females
- Increasing CD4 count is associated with progressively lower hazard of TI
- More serious illnesses (particularly WHO-stage-4 diseases) are associated with higher risk of TI
- Attending an urban/referral hospital is associated with significant decreases in TI hazard

12.6.4 Analysis of cause-specific hazards: Death

Then we carry out an analysis of the cause-specific hazard of death or LTFU in the presence of TI:

```
stset time2int, f(fail3==2)
```

Output omitted

```
xi: stcox i.traveltime i.male i.arvperfectad i.cd4_cat i.whoatarvstart i.urban
```

Output omitted

```
Cox regression -- Breslow method for ties

No. of subjects =         7757                     Number of obs   =      7757
No. of failures =         1093
Time at risk    =      1989374
                                                   LR chi2(12)     =    206.19
Log likelihood  =   -9244.5968                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  traveltime |
           2 |   .9901785   .0804957    -0.12   0.903     .8443368    1.161211
           3 |   .8948088   .0782826    -1.27   0.204     .7538114    1.062179
           4 |   1.096808   .1023531     0.99   0.322      .913476    1.316934
        male |   1.213482    .074637     3.15   0.002     1.075669     1.36895
  arvperfe~d |    1.20441     .09472     2.36   0.018     1.032363     1.40513
     cd4_cat |
           1 |   .6688478   .0546056    -4.93   0.000     .5699466     .784911
           2 |   .5416667    .041827    -7.94   0.000     .4655895     .630175
           3 |    .580659   .0608404    -5.19   0.000     .4728613    .7130312
whoatarvst~t |
           2 |   .8467284   .0960886    -1.47   0.143     .6778723    1.057646
           3 |   1.355472   .1265788     3.26   0.001     1.128762    1.627717
           4 |   2.016113   .2363358     5.98   0.000     1.602264    2.536855
       urban |   .8649403   .0530121    -2.37   0.018     .7670366    .9753403
------------------------------------------------------------------------------
```

Comments

- Travel time, is not significantly associated with the hazard of mortality of loss from follow-up
- Men have significant higher hazard of death or loss to follow-up[2]
- Increasing CD4 count is associated with progressively lower hazard of TI (by up to five-fold)
- More serious illnesses (particularly WHO-stage-4 diseases) are associated with higher risk of death or loss from follow-up
- Attending an urban/referral hospital is associated with a slight decrease in hazard of mortality or loss from follow-up

[2]This is widely published. However, in a recent analysis, we published that, once the vital status of LTFU patients is ascertained, this difference in the hazards between men and women disappears!

12.6.5 Analysis of cumulative incidences: TI

First we carry out an analysis of the cumulative incidence of TI in the presence of death or LTFU as a competing risk:

```
stset time2int, f(fail3==1)
```

Output omitted

```
xi: stcrreg i.traveltime i.male i.arvperfectad i.cd4_cat i.whoatarvstart i.urban,
> compete(fail3==2)
```

Output omitted

```
Competing-risks regression                          No. of obs       =      7757
                                                    No. of subjects  =      7757
Failure event  : fail3 == 1                         No. failed       =       810
Competing event: fail3 == 2                         No. competing    =      1093
                                                    No. censored     =      5854
                                                    Wald chi2(12)    =    321.93
Log pseudolikelihood = -6780.3656                   Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
          _t |        SHR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  traveltime |
           2 |   1.160351   .1182214     1.46   0.144     .9503105    1.416816
           3 |   1.383711   .1414842     3.18   0.001     1.132427    1.690754
           4 |   1.288156   .1477912     2.21   0.027     1.028749    1.612975
             |
      1.male |   .9173906   .0674747    -1.17   0.241     .7942328    1.059646
1.arvperfe~d |   .3278192   .0233185   -15.68   0.000     .2851587    .3768619
             |
     cd4_cat |
           1 |   .8157998   .0796878    -2.08   0.037     .6736554    .9879374
           2 |   .7876888    .069449    -2.71   0.007     .6626831     .936275
           3 |   .6925262   .0876629    -2.90   0.004     .5403654    .8875338
             |
whoatarvst~t |
           2 |   1.081854   .1264687     0.67   0.501     .8603259    1.360423
           3 |   1.112823   .1135711     1.05   0.295      .911077    1.359242
           4 |   1.240271   .1753043     1.52   0.128     .9401668     1.63617
             |
     1.urban |   .6918406   .0507005    -5.03   0.000     .5992762    .7987026
------------------------------------------------------------------------------
```

Comments

In the presence of death or LTFU, the analysis of TI cumulative incidence results in similar results with some important exceptions:

- More serious illnesses are no longer associated with higher risk of death or loss from follow-up. This is likely the result of the fact that WHO stage is much more strongly related with the risk of death or LTFU and/or the risk for TI according to WHO stage is adequately summarized by the remaining risk factors (particularly CD4 category).

12.6.6 Analysis of cumulative incidences: Death or LTFU

Then carry out an analysis of the failure of death or LTFU in the presence of the competing risk of TI:

```
stset time2int, f(fail3==2)
```

Output omitted

```
xi: stcrreg i.traveltime i.male i.arvperfectad i.cd4_cat i.whoatarvstart i.urban,
> compete(fail3==1)
```

Output omitted

```
Competing-risks regression                          No. of obs       =      7757
                                                    No. of subjects  =      7757
Failure event  : fail3 == 2                         No. failed       =      1093
Competing event: fail3 == 1                         No. competing    =       810
                                                    No. censored     =      5854

                                                    Wald chi2(12)    =    198.39
Log pseudolikelihood = -9315.0866                   Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
          _t |        SHR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  traveltime |
           2 |   .9872015   .0801714    -0.16   0.874     .8419359    1.157531
           3 |   .8767503    .076832    -1.50   0.133      .738385    1.041044
           4 |   1.083248   .1010022     0.86   0.391     .9023224    1.300451
             |
      1.male |   1.219167   .0752425     3.21   0.001     1.080265     1.37593
1.arvperfe~d |   1.310567   .1028708     3.45   0.001     1.123687    1.528526
             |
     cd4_cat |
           1 |   .6860827   .0561352    -4.60   0.000     .5844283    .8054186
           2 |   .5578133   .0430267    -7.57   0.000     .4795476    .6488525
           3 |   .6002537   .0631109    -4.85   0.000     .4884714    .7376162
             |
whoatarvst~t |
           2 |   .8370448   .0950569    -1.57   0.117     .6700143    1.045715
           3 |   1.337329   .1248547     3.11   0.002     1.113702    1.605859
           4 |   1.969626   .2323303     5.75   0.000     1.563072    2.481925
             |
     1.urban |   .8874493   .0547164    -1.94   0.053     .7864336     1.00144
------------------------------------------------------------------------------
```

Comments

In the presence of TI, the hazard of death or LTFU is associated with the same risk factors; with some important exceptions:

- The statistical significance of perfect adherence, as a predictor of death or LTFU has increased.

- The estimate of the hazard ratio of death or LTFU between subjects treated at an urban/referral hospital versus a non-urban/rural clinic, is no longer statistically significant at the 5% alpha level. This has the implication that the main difference between the two types of clinics is in the way they handle treatment interruptions (perhaps urban clinics are better equiped to recognize drug failure and/or can implement regimen changes more easily than less well-resourced health facilities).

Of interest is that, in the presence of TI, perfect adherence is associated with a higher (!) and not a lower hazard of death or LTFU.

Separate analyses (not shown here) showed that, while self-reported perfect adherence is not associated with the hazard of death, it is significantly associated, in the presence of TI, with the hazard of being lost from follow-up.

Why that may be is not clear, but it may be associated with an over-compensation by the model of the huge reduction in TI hazard associated with perfect adherence as observed in the TI-specific CIF estimation described previously.

Figure 12.1: Cause-specific cumulative incidence of TI or death/LTFU

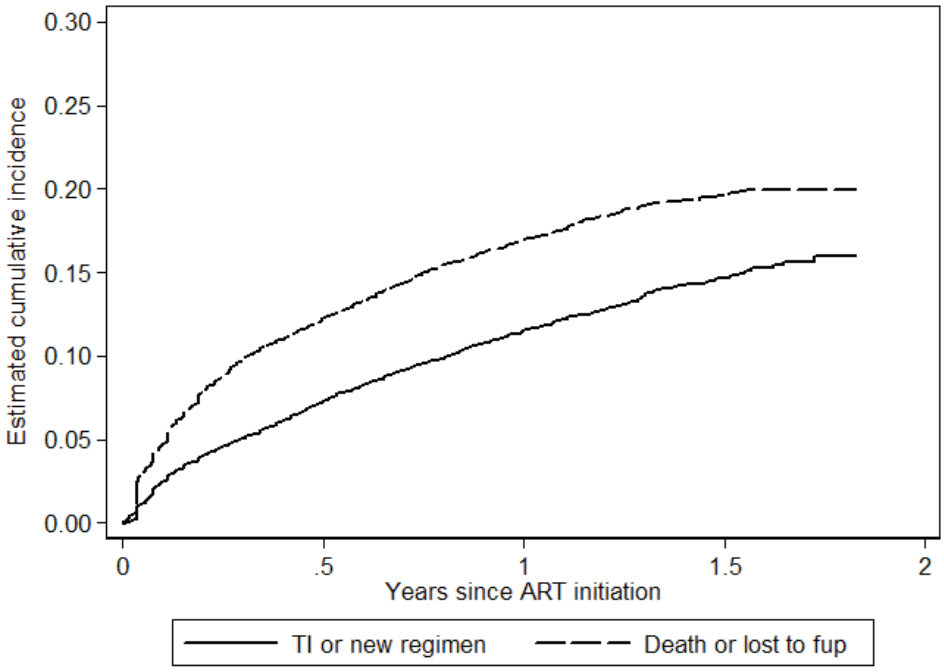
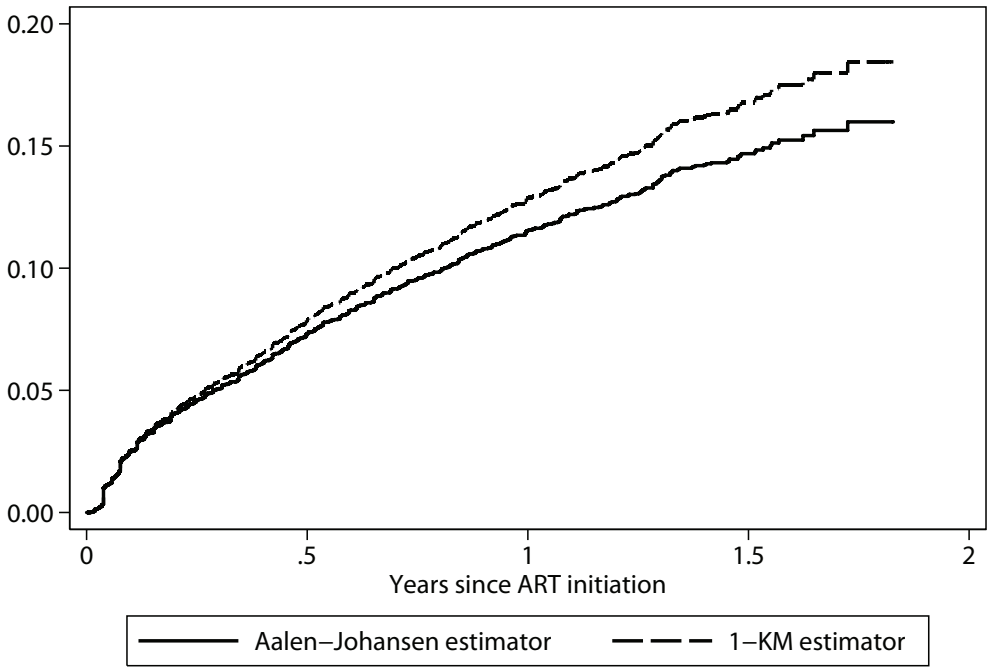


Figure 12.1 presents the cumulative probability of TI or initiating a new regimen, in the presence of (or under the possibility of) death or loss to follow up. The cumulative incidence of the competing event (death or lost to follow up) is higher than that of TI or initiating a new regimen.

Figure 12.2 depicts the naïve (1-Kaplan-Meier) and the Aalen-Johansen estimate (which accounts for the competing risks) of the cumulative incidence of TI or regimen change accounting for the competing risk of death or LTFU. It is obvious from the above figure that

Figure 12.2: Estimated cumulative incidence of TI or new regimen



the naïve estimator (after treating deaths or losses to follow-up as censored observations) overestimates, potentially significantly, the true cumulative incidence of TI or of initiating a new regimen, compared to consistent Aalen-Johansen estimator which accounts for the competing risks.

12.7 References

1. 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

2. Pintilie, M. 2006. Competing Risks: A Practical Perspective. Chichester, UK: Wiley.

3. Putter, H., M. Fiocco, and R. B. Geskus. 2007. Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine 26: 23892430.