Myeloid data

Terry Therneau

19 Sept 2023

1 Myeloid data

The `myeloid` data set contains data from a clinical trial in subjects with acute myeloid leukemia. To protect patient confidentiality the data set in the survival package has been slightly perturbed, but results are essentially unchanged. In this comparison of two conditioning regimens, the canonical path for a subject is initial therapy → complete response (CR) → hematologic stem cell transplant (SCT) → sustained remission, followed by relapse or death. Not everyone follows this ideal path, of course.

```
  id trt sex     flt3 futime death txtime crtime rltime
1  1   B   f ITD >=.7    235     1     NA     44    113
2  2   A   m  ITD <.7    286     1    200     NA     NA
3  3   A   f      TKD   1983     0     NA     38     NA
4  4   B   f      TKD   2137     0    245     25     NA
5  5   B   f ITD >=.7    326     1    112     56    200
```

The first few rows of data are shown above. The data set contains the follow-up time and status at last follow-up for each subject, along with the time to transplant (txtime), complete response (crtime) or relapse after CR (rltime). Subject 1 did not receive a transplant, as shown by the NA value, and subject 2 did not achieve CR.

Overall survival curves for the data are shown in figure 1. The difference between the treatment arms A and B is substantial. A goal of this analysis is to better understand this difference. The presence of mutations in the FLT3 gene was an eligibility criteria for the study; subtypes of the mutation were a stratification factor and are also significant predictors of the outcome. This is reflected in the simple Cox model below.

```
Call:
coxph(formula = Surv(futime, death) ~ trt + sex + flt3, data = myeloid)

              coef exp(coef) se(coef)    z     p
trtB         -0.37      0.69     0.11 -3.3 1e-03
sexm          0.16      1.18     0.11  1.4 0.152
flt3ITD <.7   0.42      1.52     0.16  2.7 0.008
flt3ITD >=.7  0.81      2.24     0.17  4.9 1e-06
```

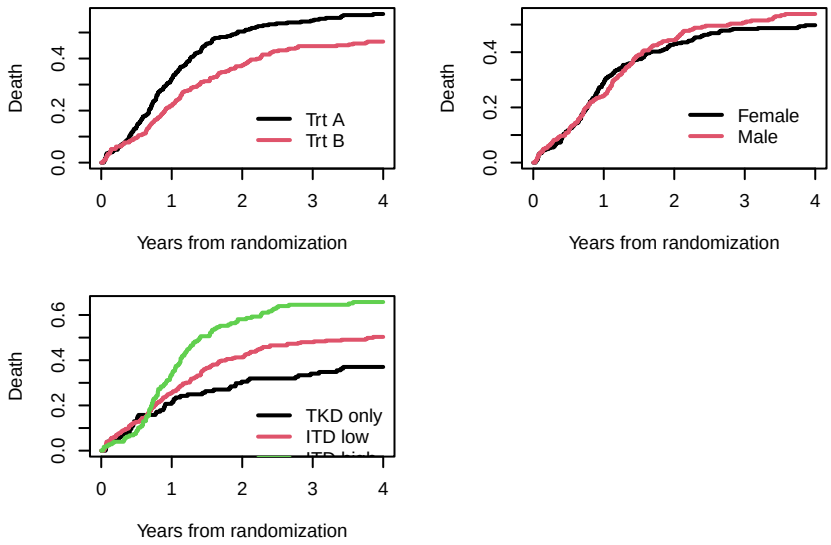

Figure 1: Overall survival curves for the myeloid study, by treatment arm, sex, and FLT3 mutation group.

```
Likelihood ratio test=36  on 4 df, p=2.8e-07
n= 646, number of events= 320
```

2 Multi-state curves

The full multi-state data set can be created with the `tmerge` routine.

```
> mdata <- tmerge(myeloid[,1:4], myeloid,  id=id,  death= event(futime, death),
                  sct = event(txtime), cr = event(crtime),
                  relapse = event(rltime))
> temp <- with(mdata, cr + 2*sct  + 4*relapse + 8*death)
> table(temp)
temp
  0   1   2   3   4   8
325 453 363   1 226 320
```

Our check shows that there is one subject who had CR and stem cell transplant on the same day (temp=3). In the multi-state data set the software does not allow the subject to have two transitions at the same exact time, e.g., entry:cr and an immediate cr:transplant transition. To avoid length 0 intervals, we break the tie so that complete response (CR) happens first. The amount of time shift added below is completely arbitrary, you could for instance use only .1 day;

the choice will have an impact on summary statistics such as mean time in state of course, but the effect is small. (Students may be surprised to see anomalies like this, since they never appear in textbook data sets. In real data such issues always appear.)

```
> tdata <- myeloid  # temporary working copy
> tied <- with(tdata, (!is.na(crtime) & !is.na(txtime) & crtime==txtime))
> tdata$crtime[tied] <- tdata$crtime[tied] -1
> mdata <- tmerge(tdata[,1:4], tdata,  id=id,  death= event(futime, death),
                  sct = event(txtime), cr = event(crtime),
                  relapse = event(rltime),
                  priorcr = tdc(crtime), priortx = tdc(txtime),
                  priorrel = tdc(rltime))
> temp <- with(mdata, cr + 2*sct  + 4*relapse + 8*death)
> table(temp)
temp
  0   1   2   4   8
325 454 364 226 320
> mdata$event <- factor(temp, c(0,1,2,4,8),
                        c("none", "CR", "SCT", "relapse", "death"))
>
> mdata[1:7, c("id", "trt", "tstart", "tstop", "event", "priorcr", "priortx")]
  id trt tstart tstop   event priorcr priortx
1  1   B      0    44      CR       0       0
2  1   B     44   113 relapse       1       0
3  1   B    113   235   death       1       0
4  2   A      0   200     SCT       0       0
5  2   A    200   286   death       0       1
6  3   A      0    38      CR       0       0
7  3   A     38  1983    none       1       0
```

Subject 1 has a CR on day 44, relapse on day 113, death on day 235 and did not receive a stem cell transplant. The data for the first three subjects looks good. Check it out a little more thoroughly using survcheck.

```
> survcheck(Surv(tstart, tstop, event) ~1, mdata, id=id)
Call:
survcheck(formula = Surv(tstart, tstop, event) ~ 1, data = mdata,
    id = id)

Unique identifiers       Observations        Transitions
               646               1689               1364

Transitions table:
         to
from       CR SCT relapse death (censored)
  (s0)    443 106      13    55         29
```

```
  CR        0 159     168    17        110
  SCT      11   0      45   149        158
  relapse   0  99       0    99         28
  death     0   0       0     0          0

Number of subjects with 0, 1, ... transitions to each state:
         count
state       0   1   2   3  4
  CR      192 454   0   0  0
  SCT     282 364   0   0  0
  relapse 420 226   0   0  0
  death   326 320   0   0  0
  (any)    29 201 174 153 89
```

The second table shows that no subject had more than one CR, SCT, relapse, or death; the intention of the study was to count only the first of each of these, so this is as expected. Several subjects visited all four intermediate states. The transitions table shows 11 subjects who achieved CR *after* stem cell transplant and another 106 who received a transplant before achieving CR, both of which are deviations from the "ideal" pathway. No subjects went from death to another state or had further follow-up time after death (which is good).

For investigating the data we would like to add a set of alternate endpoints.

1. The competing risk of CR and death, ignoring other states. This is used to estimate the fraction who ever achieved a complete response.
2. The competing risk of SCT and death, ignoring other states.
3. An endpoint that distinguishes death after SCT from death before SCT.

Each of these can be accomplished by adding further outcome variables to the data set, we do not need to change the time intervals.

This data set is the basis for our first set of curves, which are shown in figure 2. The plot overlays three separate `survfit` calls: standard survival until death, complete response with death as a competing risk, and transplant with death as a competing risk. For each fit we have shown only one state: the fraction who have died, fraction ever in CR, and fraction ever to receive transplant, respectively. Most of the CR events happen before 2 months (the green vertical line) and most of the additional CRs conferred by treatment B appear to occur between months 2 and 8. This visual impression is however contradicted by a simple tabulation of the CR events shown below.

```
   CR month
trt (0,1] (1,2] (2,3] (3,6] (6,20]
  A    40   125    28    12      1
  B    41   152    38    12      5
```

Most transplants happen after 2 months, which is consistent with the clinical guide of transplant after CR. The survival advantage for treatment B begins between 4 and 6 months, which argues that it could be at least partially a consequence of the additional CR events.

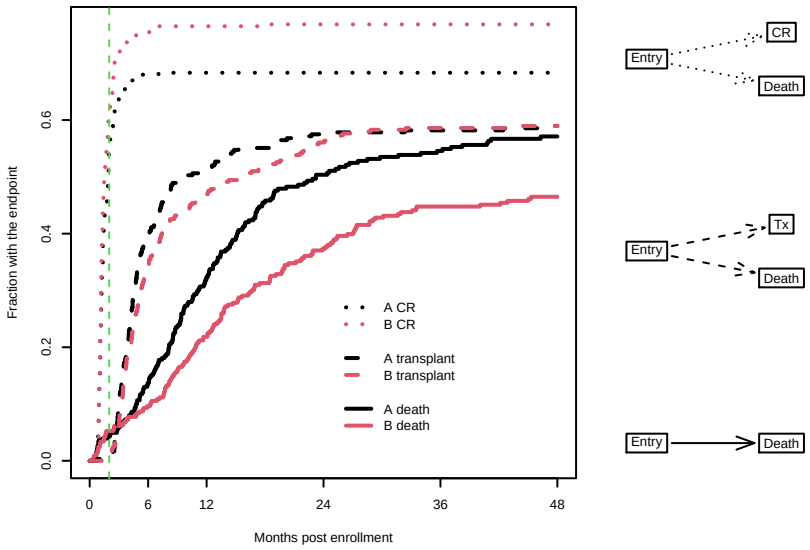

Figure 2: Overall survival curves: time to death, to transplant (Tx), and to complete response (CR). Each shows the estimated fraction of subjects who have ever reached the given state. The vertical line at 2 months is for reference. The curves were limited to the first 48 months to more clearly show early events. The right hand panel shows the state-space model for each pair of curves.

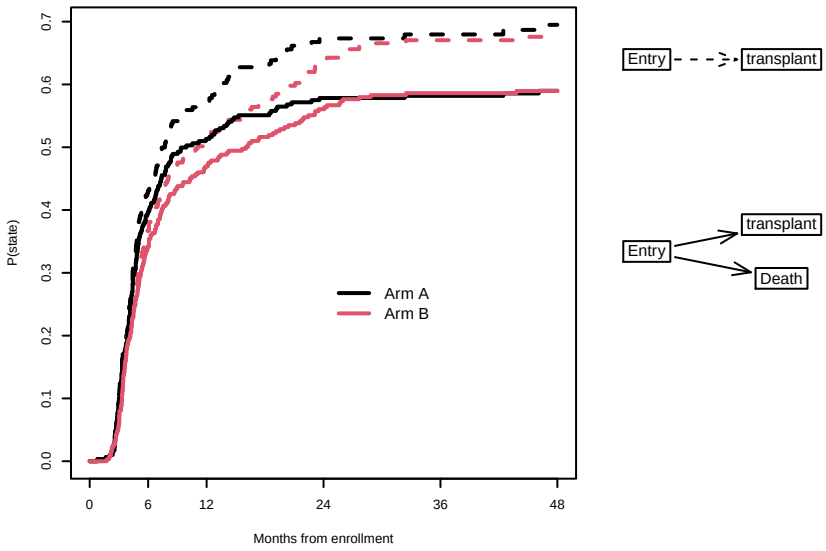

Figure 3: Correct (solid) and invalid (dashed) estimates of the number of subjects transplanted.

The association between a particular curve and its corresponding state space diagram is critical. As we will see below, many different models are possible and it is easy to get confused. Attachment of a diagram directly to each curve, as was done above, will not necessarily be day-to-day practice, but the state space should always be foremost. If nothing else, draw it on a scrap of paper and tape it to the side of the terminal when creating a data set and plots.

Figure 3 shows the transplant curves overlaid with the naive KM that censors subjects at death. There is no difference in the initial portion as no deaths have yet intervened, but the final portion of the curve overstates the transplant outcome by more than 10%.

1. The key problem with the naive estimate is that subjects who die can never have a transplant. The result of censoring them is an estimate of the "fraction who would be transplanted, if death before transplant were abolished". This is not a real world quantity.

2. In order to estimate this fictional quantity one needs to assume that death is uninformative with respect to future disease progression. The early deaths in months 0–2, before transplant begins, are however a very different class of patient. Non-informative censoring is untenable.

We are left with an unreliable estimate of an uninteresting quantity. Mislabeling any true state as censoring is always a mistake, one that will not be repeated here. (There are cases where the "what if" curve would be of interest for policy or other reasons, e.g., expected survival if lung cancer were eliminated. The necessary assumptions to estimate such a curve are however very strong, and often untenable.)

Complete response is a goal of the initial therapy; figure 4 looks more closely at this. As was noted before arm B has an increased number of responses. The duration of response is also

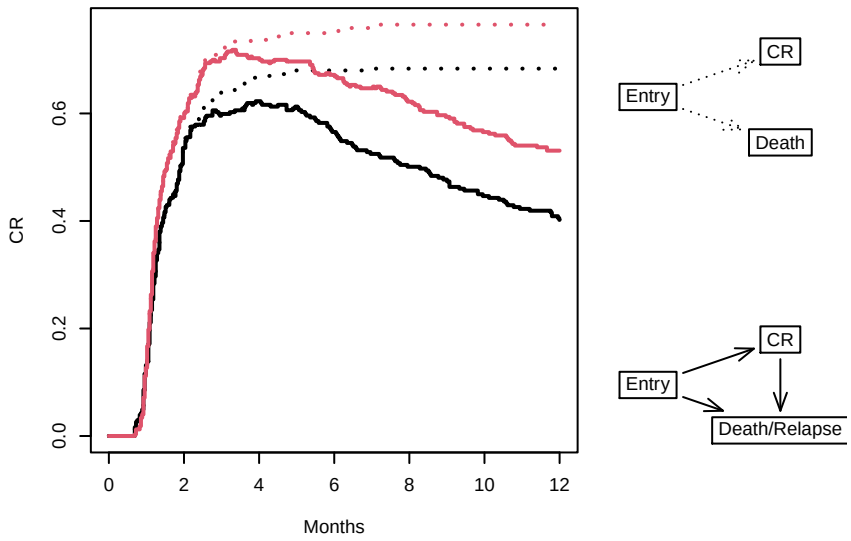

Figure 4: Models for 'ever in CR' and 'currently in CR'; the only difference is an additional transition. Both models ignore transplant.

increased: the solid curves show the number of subjects still in response, and we see that they spread apart as much as the dotted "ever in response" curves. The figure shows only the first year in order to better visualize the details, but continuing the curves out to 48 months reveals a similar pattern ($.77 - .68 = .11$ vs $.53 - .40 = .13$ at 12 months, .11 vs .09 at 48).

```
> print(crsurv, rmean=48, digits=2)
Call: survfit(formula = Surv(mstart, mstop, cr2) ~ trt, data = mdata,
    id = id, influence = TRUE)

                       n nevent rmean se(rmean)*
trt=A, (s0)          693      0   7.1       0.78
trt=B, (s0)          739      0   5.6       0.65
trt=A, CR            693    206  16.3       1.13
trt=B, CR            739    248  21.2       1.12
trt=A, death/relapse 693    194  24.6       1.13
trt=B, death/relapse 739    184  21.1       1.07
   *restricted mean time in state (max time = 48 )
```

@ The restricted mean time in the CR state is extended by 21.2 - 16.3 = 4.9 months. A question which immediately gets asked is whether this difference is "significant", to which there are two answers. The first and more important is to ask whether 5 months is an important gain from either a clinical or patient perspective. The overall restricted mean survival for the study is approximately 30 of the first 48 months post entry; on this backdrop an extra 5 months in

CR might or might not be viewed as an meaningful advantage. The less important answer is to test whether the apparent gain is sufficiently rare from a mathematical point of view, i.e., "statistical" significance. The standard errors of the two values are 1.1 and 1.1, and since they are based on disjoint subjects the values are independent, leading to a standard error for the difference of $\sqrt{1.1^2 + 1.1^2} = 1.6$. The 5 month difference is more than 3 standard errors, so highly significant.

The code for the figure created yet another event variable so as to ignore transitions to the transplant state. For the new `cr2` variable 'transplant' becomes 'none', and relapse and death are combined. Transplant is essentially ignored. Below we show a small part of the data set. The `event` variable is the overall code for the subject's progression through states, while `crstat` and `txstat` deal with the competing risk for CR/death and SCT/death, respectively, which are shown in figure 2. The `crstat` variable pick up the first of CR or death, and codes all other lines as 'none'. The `cr2` event is used for the duration of CR curves in figure 4, it captures CR and death/relapse, with all other rows as 'none'.

```
   id tstart tstop   event crstat           cr2 txstat
1   1      0    44      CR     CR            CR censor
2   1     44   113 relapse   none death/relapse censor
3   1    113   235   death   none          none  death
4   2      0   200     SCT   none          none    SCT
5   2    200   286   death  death death/relapse censor
6   3      0    38      CR     CR            CR censor
7   3     38  1983    none   none          none censor
8   4      0    25      CR     CR            CR censor
9   4     25   245     SCT   none          none    SCT
10  4    245  2137    none   none          none censor
11  5      0    56      CR     CR            CR censor
12  5     56   112     SCT   none          none    SCT
13  5    112   200 relapse   none death/relapse censor
14  5    200   326   death   none          none censor
```

For the estimates themselves, do we need to remove redundant lines. That is should the CR/death curve be one of the two lines below and not the other? It turns out that the results are the same. None of the "extra" rows after CR contain a transition and hence they all get ignored, since they are not in the risk set for any possible transition.

```
> curve1 <- survfit(Surv(tstart, tstop, crstat) ~ trt, mdata, id=id)
> curve2 <- survfit(Surv(tstart, tstop, crstat) ~ trt, mdata, id=id,
                    subset = (priorcr ==0))
```

In summary

- Arm B adds further complete responses (about 10%); 206/317 = 65% achieve CR in arm A vs. 248/329 = 75% in arm B.
- The difference in 4 year survival is about 6%.
- There is approximately 2 months longer average duration of CR (of 48).

```
strata states
     2      6
[1] "(s0)"          "CR"            "SCT"           "relapse"
[5] "death"         "SCT after CR"
```

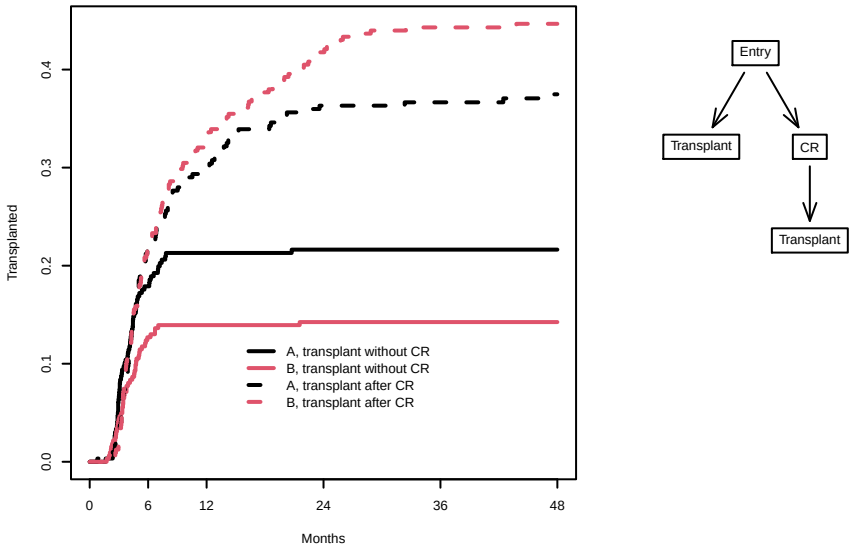

Figure 5: Transplant status of the subjects, broken down by whether it occurred before or after CR.

```
        to
from       CR SCT relapse death (censored)
  (s0)    443 106      13    55         29
  CR        0 159     168    17        110
  SCT      11   0      45   149        158
  relapse   0  99       0    99         28
  death     0   0       0     0          0
```

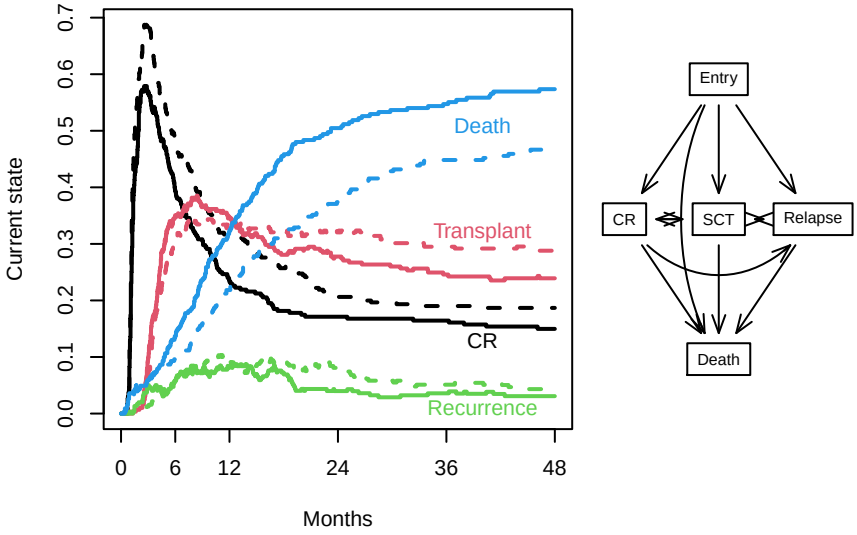

Figure 6: The full multi-state curves for the two treatment arms.

CR → transplant is the target treatment path for a patient; given the improvements listed above why does figure 2 show no change in the number transplanted? Figure 5 shows the transplants broken down by whether this happened before or after complete response. Most of the non-CR transplants happen by 10 months. One possible explanation is that once it is apparent to the patient/physician pair that CR is not going to occur, they proceed forward with other treatment options. The extra CR events on arm B lead to a consequent increase in transplant as well, but at a later time of 12–24 months: for a subject in CR we can perhaps afford to defer the transplant date.

Computation is again based on a manipulation of the event variable: in this case dividing the transplant state into two sub-states based on the presence of a prior CR. The code makes use of the time-dependent covariate `priorcr`. (Because of scheduling constraints within a hospital it is unlikely that a CR that is within a few days prior to transplant could have affected the decision to schedule a transplant, however. An alternate breakdown that might be useful would be "transplant without CR or within 7 days after CR" versus those that are more than a week later. There are many sensible questions that can be asked.)

Figure 6 shows the full set of state occupancy probabilities for the cohort over the first 4

years. At each point in time the curves estimate the fraction of subjects currently in that state. The total who are in the transplant state peaks at about 9 months and then decreases as subjects relapse or die; the curve rises whenever someone receives a transplant and goes down whenever someone leaves the state. At 36 months treatment arm B (dashed) has a lower fraction who have died, the survivors are about evenly split between those who have received a transplant and those whose last state is a complete response (only a few of the latter are post transplant). The fraction currently in relapse – a transient state – is about 5% for each arm. The figure omits the curve for "still in the entry state". The reason is that at any point in time the sum of the 5 possible states is 1 — everyone has to be somewhere. Thus one of the curves is redundant, and the fraction still in the entry state is the least interesting of them.

3 Multi-state models

We now repeat some of the above using MSH models. An advantage in principle is that we can jointly adjust for treament and FLT3 group, though the fact that the study is randomized and thus balanced for FLT3 suggests that any changes in results will be modest. The MLT also assumes proportional hazards, i.e., a constant coefficient or hazard ratio over time; figure 1 suggests that this may be questionable, for the first half year at least, for both treatment and FLT3. A formal test of PH is not rejected, however.

```
Call:
coxph(formula = Surv(mstart, mstop, crstat) ~ trt + flt3, data = mdata,
    id = id)


1:2              coef exp(coef) se(coef) robust se    z    p
  trtB          0.234     1.264    0.095     0.095  2.5 0.01
  flt3ITD <.7  -0.067     0.935    0.117     0.116 -0.6 0.56
  flt3ITD >=.7 -0.120     0.887    0.130     0.128 -0.9 0.35


1:3               coef exp(coef) se(coef) robust se    z   p
  trtB         -0.1005    0.9044   0.1848    0.1831 -0.5 0.6
  flt3ITD <.7  -0.0021    0.9979   0.2421    0.2596  0.0 1.0
  flt3ITD >=.7 -0.1003    0.9046   0.2660    0.2755 -0.4 0.7

 States:  1= (s0), 2= CR, 3= death

Likelihood ratio test=7.5  on 6 df, p=0.28
n= 1689, number of events= 573
Call:
coxph(formula = Surv(mstart, mstop, cr2) ~ trt + flt3, data = mdata,
    id = id)
```

```
1:2              coef exp(coef) se(coef) robust se    z    p
  trtB          0.221     1.248    0.095     0.095  2.3 0.02
  flt3ITD <.7  -0.064     0.938    0.117     0.116 -0.6 0.58
  flt3ITD >=.7 -0.125     0.883    0.130     0.128 -1.0 0.33


1:3              coef exp(coef) se(coef) robust se    z   p
  trtB         -0.140     0.869    0.183     0.182 -0.8 0.4
  flt3ITD <.7   0.026     1.027    0.237     0.250  0.1 0.9
  flt3ITD >=.7 -0.151     0.860    0.263     0.268 -0.6 0.6


2:3              coef exp(coef) se(coef) robust se    z     p
  trtB         -0.30      0.74     0.13      0.13 -2.4  0.02
  flt3ITD <.7   0.69      1.99     0.18      0.18  3.8 1e-04
  flt3ITD >=.7  1.19      3.29     0.19      0.19  6.2 5e-10

 States:  1= (s0), 2= CR, 3= death/relapse

Likelihood ratio test=55  on 9 df, p=1e-08
n= 1689, number of events= 832
```

The above are fits for the two models in the right margin of figure 4. It is interesting that FLT3 group has minimal effect on the probability of reaching CR, or on the rate of death without CR, but a major effect on the transition from CR to death. We can see this clearly in figure 7, and in the sojourn time table below. The range between the three FLT3 levels is smaller for time in the entry state (1.6 months) than for time in the CR state (12.1), while the difference between treatments A and B is fairly constant across FLT3.

```
          Entry A Entry B diff      CR A CR B diff
TKD           6.4     5.3 -1.1      22.9 27.8  4.9
ITD <.7       6.8     5.7 -1.0      16.0 21.2  5.1
ITD >=.7      8.0     6.7 -1.3      10.8 15.4  4.7
```

I was puzzled at first why the coefficients for the entry:CR transition in the first fit are different than the the second. The answer is the 13 subjects who go directly from s0 to relapse without passing throug CR: in the first fit they remain at risk for a CR event for just a bit longer. The resulting small increase in a few risk sets moves the coefficients just a smidge. (Per the study design, subjects were supposed to only be counted as a relapse if they had experienced CR, so this was not anticipated.)

For AJ modeling transplant we had divided transplant into two states: before and after CR. There are then 6 states, the Aalen-Johansen estimate with treatment and FLC3 as covariates creates 6 x 6= 36 curves, we ignored all but the two SCT states when plotting and printing. An MPH model is focused instead on the transitions: for this splitting the SCT state only adds complexity, replacing the CR:SCT transition with CR:(SCT after CR), a 1:1 substitution, but also splitting the 3 transitions from SCT to CR, relapse, and death into 2.

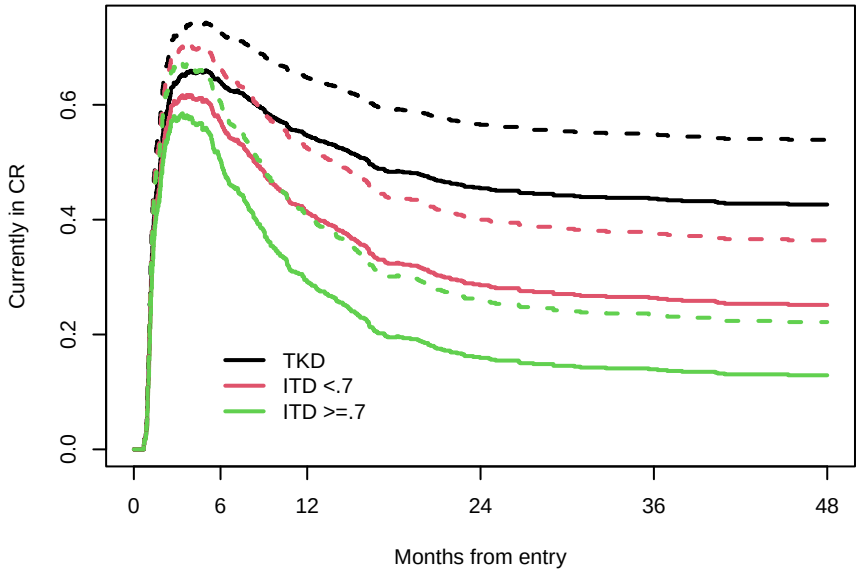

Figure 7: Estimated probability of being in the CR state. Treatment A = solid line, B = dashed.

Instead start by focusing our question. We already know that more non-CR transplants come from arm A than B, that more CR transplant come from B than A, and that FLT3 appears to have little effect on CR rates. What further effect might FLT3 have on transplant rates? One possible effect, both before and after CR, would be acceleration of transplant, for instance if the procedure is being delayed until failure seems immenent. For non CR patients the delay may be in hope of CR, for those in CR there may be a desire to let the patient enjoy their treatment holiday for a bit longer. For this question, simplify the states by ignoring relapse. The table below shows coefficients.

```
       to
from     CR SCT death (censored)
  (s0)  443 111    63         29
  CR      0 253    81        120
  SCT    11   0   176        176
  death   0   0     0          0
                     coef exp(coef)       z Pr(>|z|)
trtB_1:3         -0.32000      0.73 -1.6000   0.0990
flt3ITD <.7_1:3   0.03800      1.00  0.1400   0.8900
flt3ITD >=.7_1:3  0.12000      1.10  0.4400   0.6600

trtB_2:3         -0.00037      1.00 -0.0029   1.0000
flt3ITD <.7_2:3   0.50000      1.60  3.0000   0.0024
flt3ITD >=.7_2:3  0.58000      1.80  3.2000   0.0016
```

The FLT3 level appears to have little or no effect on the rate of entry:SCT transitions, but a strong effect on CR:SCT transitions. This may simply be an aspect of shorter duration of CR for the more severe mutations. Treatment B has a somewhat lower rate of entry:SCT transitions. We can only speculate, but perhaps this may be an anticipatory effect of the higher CR rate, i.e., subjects on B not yet declared to be CR are nevertheless less likely to be seen as a clear failure and promoted to early transplant.