

BMT-cure-short.R

hputter

2024-11-22

This file contains the code supporting Section 2 of the paper “Cure models in survival analysis: Issues with identifiability and competing risks” by Hein Putter and Per Kragh Andersen.

It uses the BMT data as used in the book by Andersen & Ravn (2023), available at “<https://multi-state-book.github.io/companion/data/bmt.csv>”.

```
# Packages used in this analysis
```

```
library(survival)
```

```
library(prodlim)
```

```
library(smcurve)
```

```
library(intsurv)
```

```
# Read in the data
```

```
bmt <- read.csv("https://multi-state-book.github.io/companion/data/bmt.csv")
```

```
# Have a look at the events of interest, here relapse and death
```

```
head(bmt)
```

```
##      id team timedead death timerel rel timegvhd gvhd timeanc500 anc500 sex  age all bmonly
## 1 1 224    13.75    1    6.94  1      NA    0      0.66      1  1  9.51  0      1
## 2 2 248   103.32    0      NA    0      NA    0      0.53      1  0 25.86  0      1
## 3 3 218    44.70    0      NA    0      NA    0      0.60      1  1 49.18  0      1
## 4 4 5      3.65    1      NA    0    0.43    1      0.36      1  0 53.62  0      0
## 5 5 86     3.52    1      NA    0      NA    0      0.70      1  0 45.02  0      1
## 6 6 36    75.86    0      NA    0      NA    0      0.40      1  1 32.76  0      0
```

```
table(bmt$death)
```

```
##
```

```
##      0      1
```

```
## 1272  737
```

```
table(bmt$rel)
```

```
##
```

```
##      0      1
```

```
## 1750  259
```

```
table(bmt$rel, bmt$death)
```

```
##
##      0      1
##  0 1245  505
##  1   27  232
```

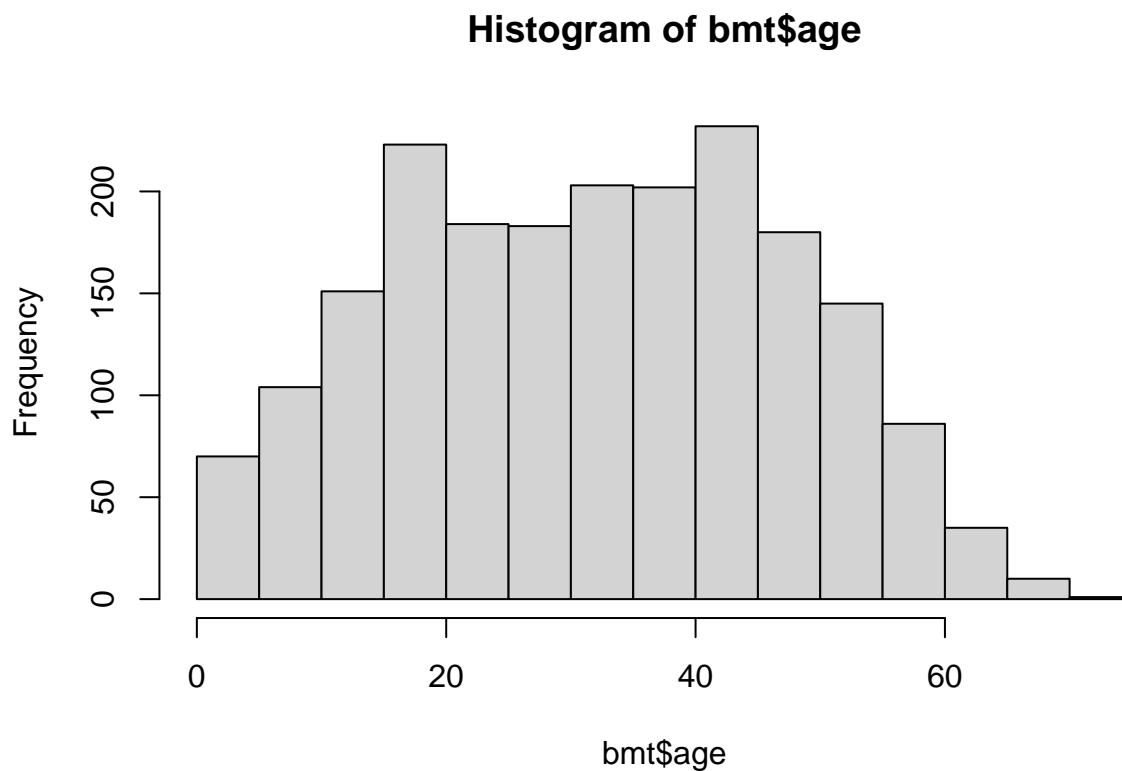
```
# The only covariates that we will use are all (selecting AML, all=0) and age
table(bmt$all)
```

```
##
##      0      1
## 1406   603
```

```
summary(bmt$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.00  18.99   32.35   31.91  44.06   70.60
```

```
hist(bmt$age)
```



A cure model would focus on relapse and fit a cure model where subjects that died are treated as censored observations. We favour a competing risks model where death without prior relapse is considered as a competing risk.

First prepare the data for a competing risks analysis by defining time and status variables.

```
bmt$relcr <- bmt$rel
bmt$relcr[bmt$death == 1 & bmt$rel==0] <- 2
table(bmt$relcr)
```

```
##
##      0      1      2
## 1245  259  505
```

```
bmt$timecr <- bmt$timedeath
bmt$timecr[bmt$relcr == 1] <- bmt$timerel[bmt$relcr == 1]
```

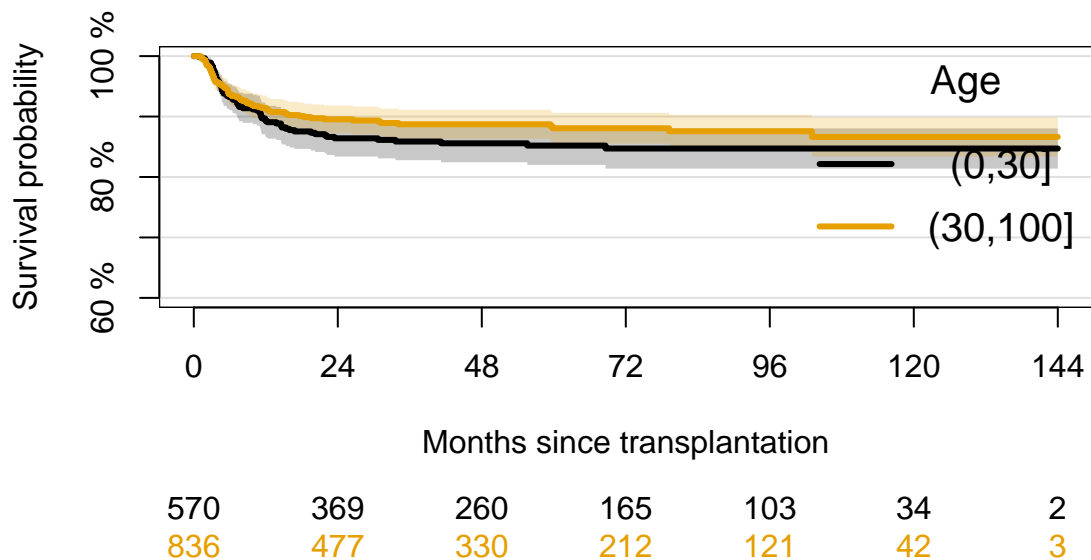
We will use age very crudely as above or below 30 years. Define this variable, select the AML patients, retain only the relevant variables and look at the first six lines.

```
bmt$Age <- cut(bmt$age, c(0, 30, 100))
aml <- subset(bmt, all==0)
table(aml$Age, aml$relcr)
```

```
##
##              0      1      2
## (0,30]      410    73    87
## (30,100]    499    81   256
```

Start with the incorrect naive Kaplan-Meiers for relapse.

```
kmfit <- prodlim(Hist(timecr, rel) ~ Age, data = aml)
plot(kmfit,
     xlim = c(0, 144), xlab = "Months since transplantation", ylim = c(0.6, 1),
     axis1.at = seq(0, 144, 24), axis2.at = seq(0.6, 1, 0.1),
     atrisk.at = seq(0, 144, 24), atrisk.title = "",
     legend.x = "topright", legend.cex = 1.25)
```



There are the plateaus of the KM's.

```
summary(kmfit, times=120)
```

```
##      Age time n.risk n.event n.lost  surv se.surv lower upper
## 1  (0,30] 120    34      0      0 0.847  0.0170 0.814 0.880
## 2 (30,100] 120    42      0      0 0.866  0.0166 0.834 0.899
```

Fit a cure model with `smcure`.

```
curefit <- smcure(formula = Surv(timecr, rel) ~ Age,
                  cureform = ~ Age, data = aml,
                  model = "ph", Var = TRUE)
```

```
## Program is running..be patient... done.
## Call:
## smcure(formula = Surv(timecr, rel) ~ Age, cureform = ~Age, data = aml,
##       model = "ph", Var = TRUE)
##
## Cure probability model:
##      Estimate      Std.Error      Z value  Pr(>|Z|)
## (Intercept) 2.656607e+01 4.948331e-06 5.368692e+06 0.0000000
## Age        -4.642854e-09 3.252908e-06 -1.427293e-03 0.9988612
##
##
```

```
## Failure time distribution model:
##           Estimate Std. Error   Z value  Pr(>|Z|)
## Age(30,100] -0.1131185    0.28904 -0.3913595 0.6955315
```

I tried the same thing now using another package that can fit cure models, called `intsurv`. Adapting some code from the help from `intsurv`, I was able to obtain much more plausible estimates, at least of the cure probability model.

```
mm <- model.matrix(~ Age, data = aml)[, -1, drop=FALSE] # without intercept
set.seed(2024) # To get reproducible results for the variance (using bootstrap)
fit1 <- cox_cure.fit(mm, mm, aml$timecr, aml$rel, bootstrap = 100)
summary(fit1)
```

```
## Call:
## cox_cure.fit(surv_x = mm, cure_x = mm, time = aml$timecr, event = aml$rel,
##             bootstrap = 100)
##
## Coefficient estimates from the survival part:
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Age(30,100] 0.048222  1.049403 0.149673 0.3222  0.7473
##
## Coefficient estimates from the cure rate part:
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## (Intercept) -1.65617   0.19087 0.08506 -19.4706 <2e-16 ***
## Age(30,100] -0.24251   0.78466 0.14783 -1.6404  0.1009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of observations: 1406
## Number of events: 154
## Weighted concordance index: 50.99%
## Observed data log-likelihood: -1193.320
```

Checking whether the KM plateaus agree with the cure probabilities.

```
b <- coef(fit1)$cure
1 - plogis(b[1]) # very close to the KM plateau of AML
```

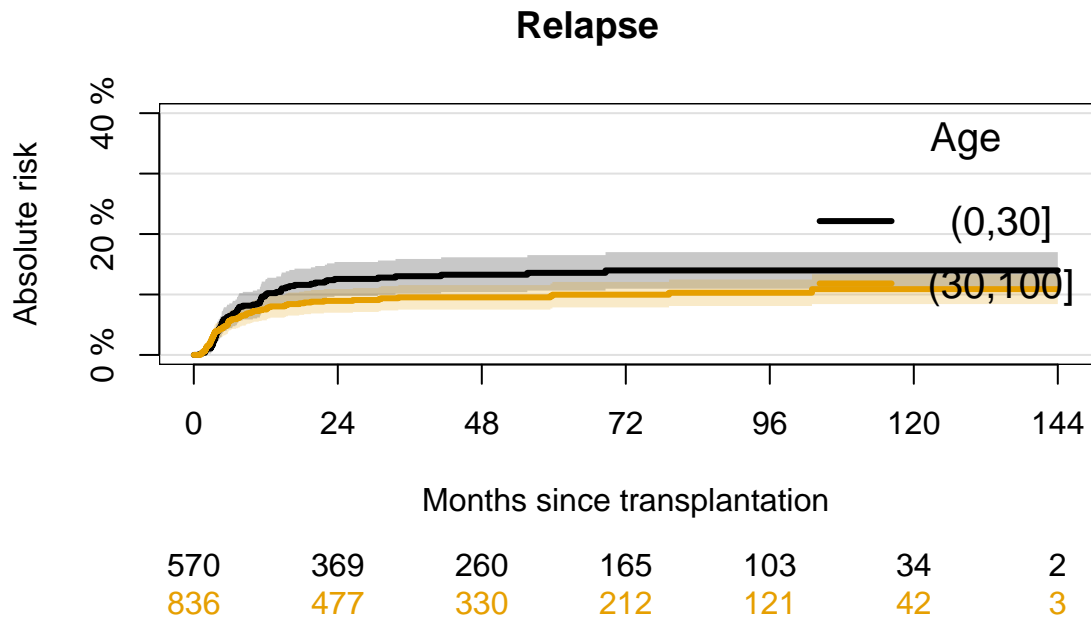
```
## (Intercept)
##      0.8397229
```

```
1 - plogis(b[1] + b[2]) # very close to the KM plateau of ALL
```

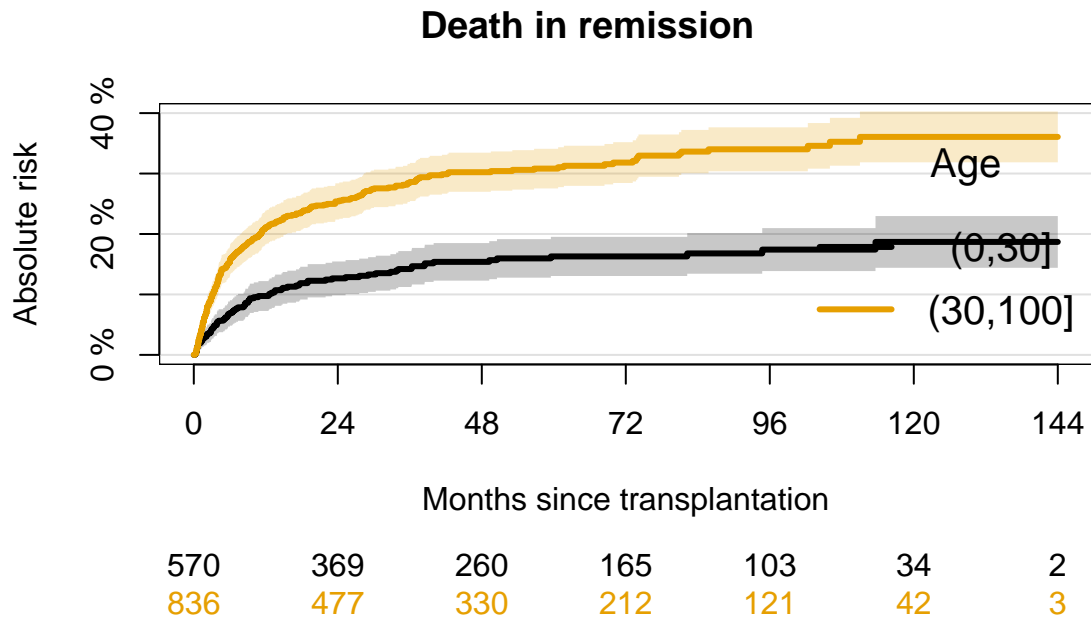
```
## (Intercept)
##      0.8697416
```

Let's stick with the competing risks analysis and plot the non-parametric estimates of the cumulative incidence curves, using `prodlim`.

```
ajfit <- prodlim(Hist(timecr, relcr) ~ Age, data = aml)
plot(ajfit, # relapse
     xlim = c(0, 144), xlab = "Months since transplantation", ylim = c(0, 0.4),
     axis1.at = seq(0, 144, 24), axis2.at = seq(0, 0.4, 0.1),
     atrisk.at = seq(0, 144, 24), atrisk.title = "",
     legend.x = "topright", legend.cex = 1.25)
title(main = "Relapse")
```



```
plot(ajfit, cause=2, # death in remission
     xlim = c(0, 144), xlab = "Months since transplantation", ylim = c(0, 0.4),
     axis1.at = seq(0, 144, 24), axis2.at = seq(0, 0.4, 0.1),
     atrisk.at = seq(0, 144, 24), atrisk.title = "",
     legend.x = "bottomright", legend.cex = 1.25)
title(main = "Death in remission")
```



The following fits proportional cause-specific hazards models in the AML data with age (categorized older or younger than 30) as sole covariate.

Cause 1, relapse

```
coxph(Surv(timecr, relcr==1) ~ Age, data = aml)
```

Call:

```
## coxph(formula = Surv(timecr, relcr == 1) ~ Age, data = aml)
```

##

```
##               coef exp(coef) se(coef)      z      p
## Age(30,100] -0.2029    0.8164   0.1614 -1.257 0.209
```

##

```
## Likelihood ratio test=1.57 on 1 df, p=0.2099
```

```
## n= 1406, number of events= 154
```

Cause 2, non-relapse mortality

```
coxph(Surv(timecr, relcr==2) ~ Age, data = aml)
```

Call:

```
## coxph(formula = Surv(timecr, relcr == 2) ~ Age, data = aml)
```

##

```
##               coef exp(coef) se(coef)      z      p
## Age(30,100] 0.7741    2.1686   0.1241  6.236 4.5e-10
```

##

```
## Likelihood ratio test=43.36 on 1 df, p=4.55e-11
```

```
## n= 1406, number of events= 343
```