

# BIOST537\_Project

Dante Ramirez, Nada Ali, Jiayu Sun, Machi Kaneko

2023-03-13

```
# Creating Survival Objects
s_bmt <- with(bmt_df, Surv(tdfs, deltadfs))
sfit_bmt <- survfit(s_bmt ~ 1, data = bmt_df, conf.type = "log-log")

# mean follow-up time
mean(bmt_df$tdfs)
```

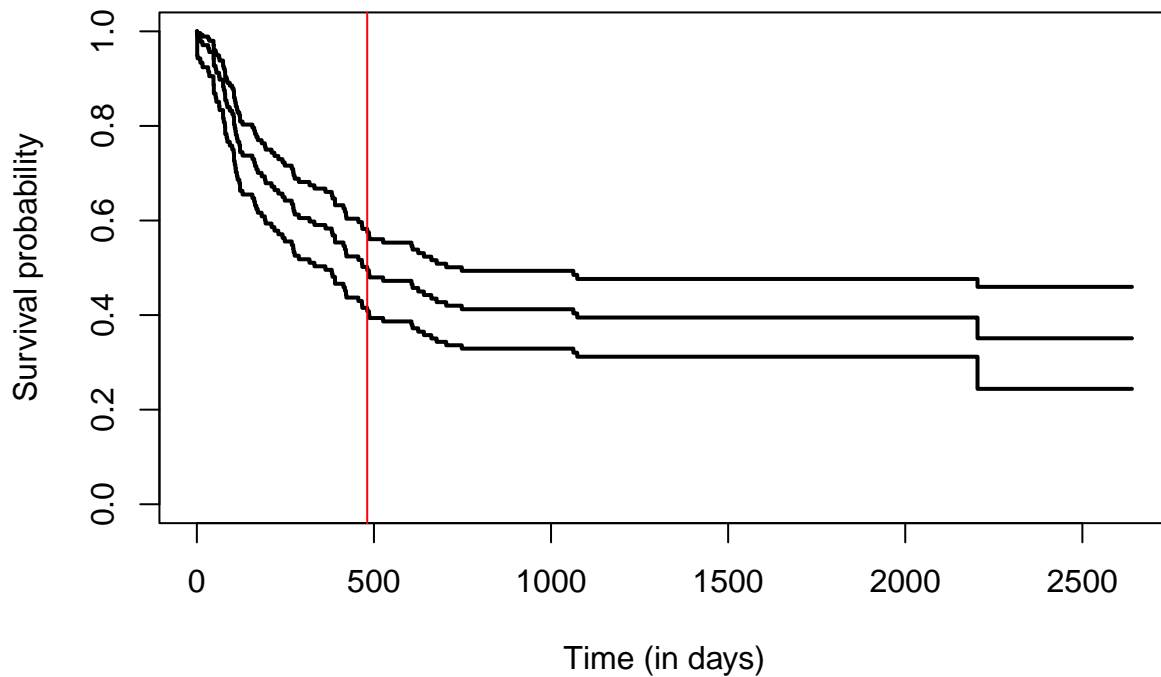
```
## [1] 782.0292
```

```
# proportion of censored of observations
1 - mean(bmt_df$deltadfs)
```

```
## [1] 0.3941606
```

```
# Kaplan Meier Plot
plot(sfit_bmt, conf.int = T,
      main="Kaplan-Meier survival estimate w/ Median Survival Time",
      xlab="Time (in days)", ylab="Survival probability", col="black",
      lty="solid", lwd=2)
abline(v = 481, col = "red")
```

## Kaplan–Meier survival estimate w/ Median Survival Time



```
# Median Survival time, C.I. for median survival time, and other summary stats
summary(sfit_bmt)$table
```

```
##   records      n.max   n.start   events    rmean se(rmean)   median  0.95LCL
## 137.0000 137.0000 137.0000   83.0000 1186.1053 100.5981  481.0000 363.0000
##   0.95UCL
## 748.0000
```

```
# Survival Object based on disease subgrouping
sfit_bmt_byDisgroup <- survfit(s_bmt ~ disgroup,
                              data = bmt_df,
                              conf.type = "log-log")
```

```
plot(sfit_bmt_byDisgroup, main="Kaplan-Meier survival estimate",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2, color = c("red", "blue", "orange"))
```

```
## Warning in plot.window(...): "color" is not a graphical parameter
```

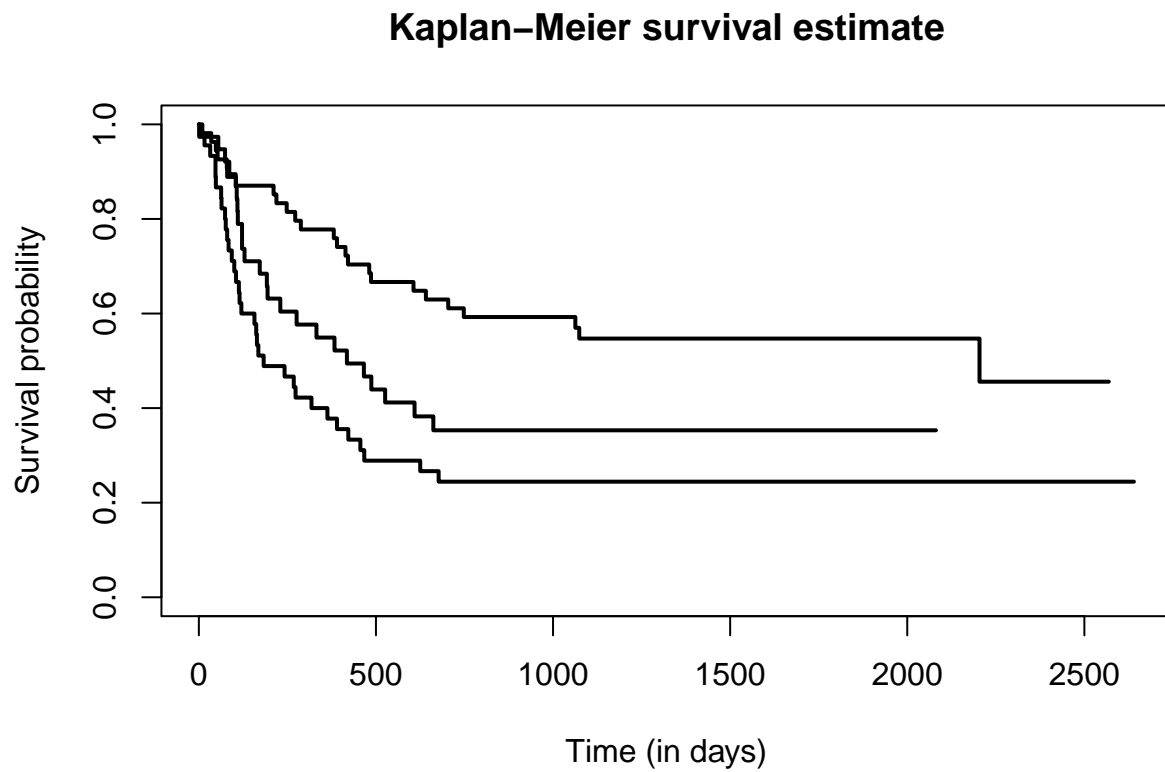
```
## Warning in plot.xy(xy, type, ...): "color" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "color" is not a
## graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "color" is not a
## graphical parameter
```

```
## Warning in box(...): "color" is not a graphical parameter
```

```
## Warning in title(...): "color" is not a graphical parameter
```



```
pander(survdiff(s_bmt ~ disgroup, data = bmt_df))
```

Table 1: Call: `s_bmt ~ disgroup` Chisq = 13.803722 on 2 degrees of freedom, p = 0.001006

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
<b>disgroup=1</b>	38	24	21.85	0.2112	0.2893
<b>disgroup=2</b>	54	25	39.97	5.604	11.01
<b>disgroup=3</b>	45	34	21.18	7.756	10.53

```
comp(ten(sfit_bmt_byDisgroup))
```

```
##           chiSq df pChisq
## 1          13.8037 2      5
## n          16.2407 2      1
## sqrtN       15.6529 2      4
## S1          15.7260 2      3
## S2          15.7781 2      2
```

```
## FH_p=1_q=1  9.9331  2      6
## $tft
##           Q          Var          Z pNorm
## 1          -10.6695    42.7801 -1.63127    5
## n         -1294.0000  439987.8847 -1.95081    1
## sqrtN       -118.1769   4202.2583 -1.82302    4
## S1          -9.2667    23.2023 -1.92379    3
## S2          -9.1996    22.7588 -1.92839    2
## FH_p=1_q=1   -1.0948     1.4957 -0.89516    6
##
## $scores
## [1] 1 2 3
```

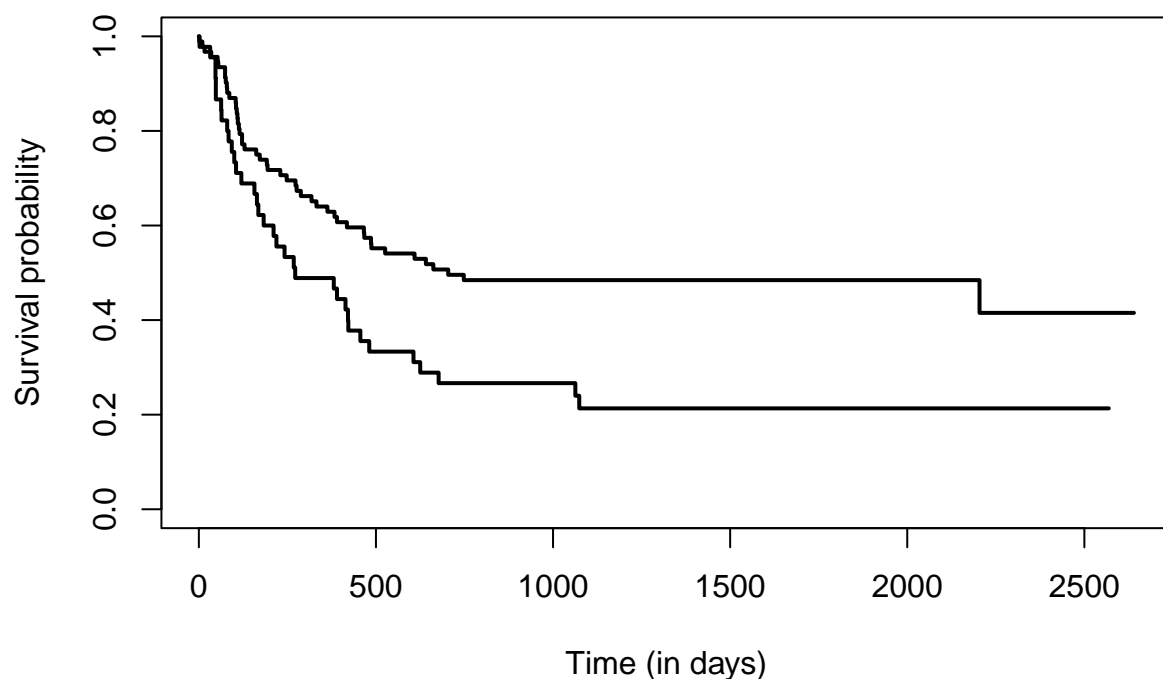
```
pchisq(q = (-1.95081)^2, df=1, lower.tail=FALSE)
```

```
## [1] 0.05107965
```

```
# Survival Object based on "When we was FAB" classification
sfit_bmt_byFAB <- survfit(s_bmt ~ fab,
                        data = bmt_df,
                        conf.type = "log-log")

plot(sfit_bmt_byFAB, conf.int = F, main="Kaplan-Meier survival estimate",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

### Kaplan-Meier survival estimate



```
survdifff(s_bmt ~ fab, data = bmt_df)
```

```
## Call:
## survdifff(formula = s_bmt ~ fab, data = bmt_df)
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## fab=0 92      48      59.8      2.34      8.44
## fab=1 45      35      23.2      6.03      8.44
##
## Chisq= 8.4  on 1 degrees of freedom, p= 0.004
```

```
comp(ten(sfit_bmt_byFAB))
```

```
##           Q      Var      Z pNorm
## 1      1.1825e+01 1.6590e+01 2.9033      1
## n      1.0830e+03 1.6628e+05 2.6559      6
## sqrtN      1.1217e+02 1.6047e+03 2.8001      2
## S1      7.9035e+00 8.7832e+00 2.6668      4
## S2      7.8227e+00 8.6118e+00 2.6657      5
## FH_p=1_q=1 2.1652e+00 6.0024e-01 2.7948      3
##           maxAbsZ      Var      Q pSupBr
## 1      1.2047e+01 1.6590e+01 2.9578      1
## n      1.0850e+03 1.6628e+05 2.6608      6
## sqrtN      1.1283e+02 1.6047e+03 2.8167      3
## S1      7.9834e+00 8.7832e+00 2.6938      4
## S2      7.8946e+00 8.6118e+00 2.6902      5
## FH_p=1_q=1 2.2184e+00 6.0024e-01 2.8633      2
```

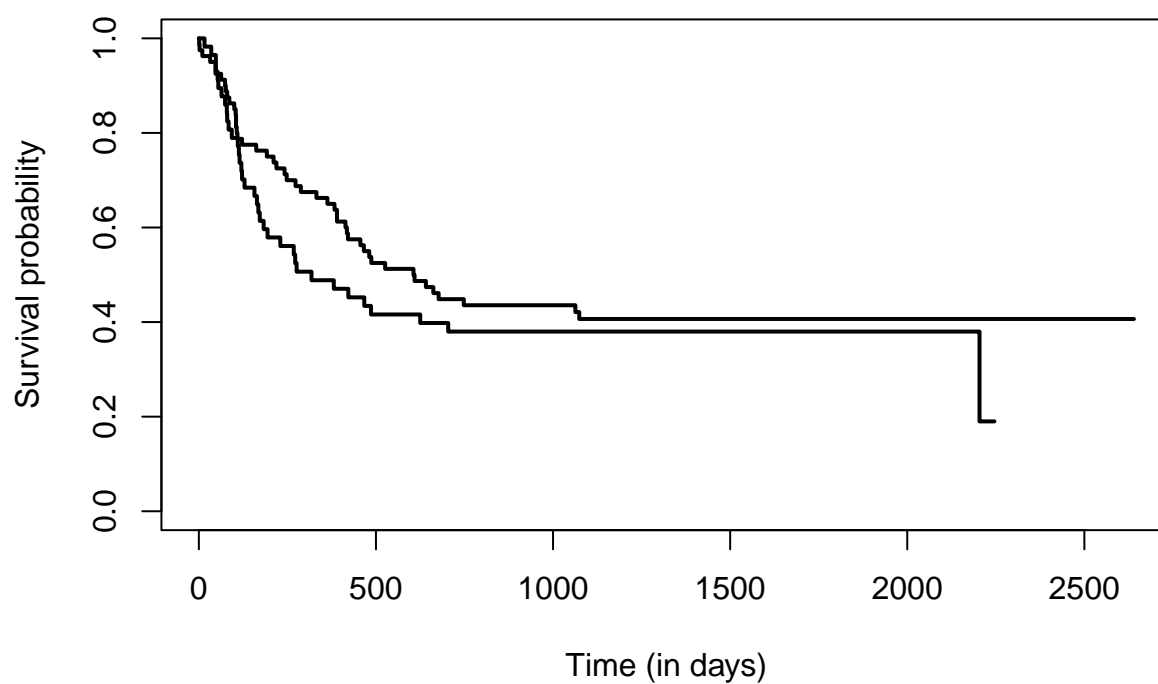
```
pchisq(q = (2.6559)^2, df=1, lower.tail=FALSE)
```

```
## [1] 0.007909707
```

```
# Survival Object based on sex subgrouping
sfit_bmt_byMale <- survfit(s_bmt ~ male,
                           data = bmt_df,
                           conf.type = "log-log")

plot(sfit_bmt_byMale, conf.int = F, main="Kaplan-Meier survival estimate, by Sex",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

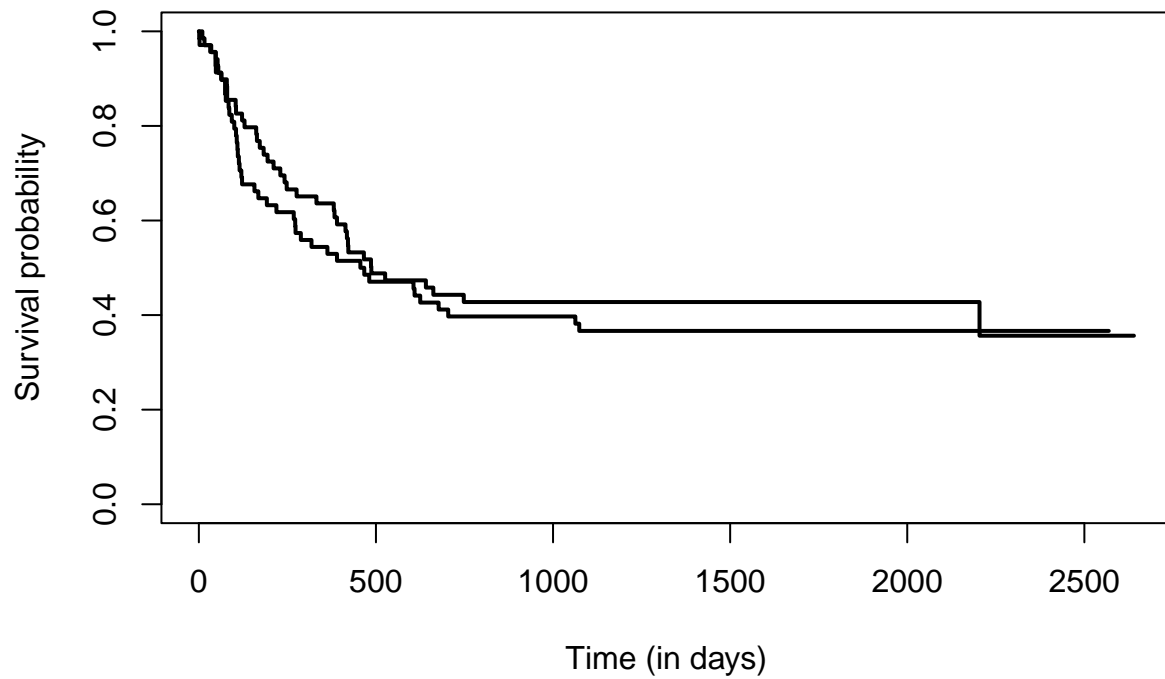
## Kaplan–Meier survival estimate, by Sex



```
# Survival Object based on CMV subgrouping
sfit_bmt_byCMV <- survfit(s_bmt ~ cmv,
                          data = bmt_df,
                          conf.type = "log-log")

plot(sfit_bmt_byCMV, conf.int = F, main="Kaplan-Meier survival estimate, by CMV",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

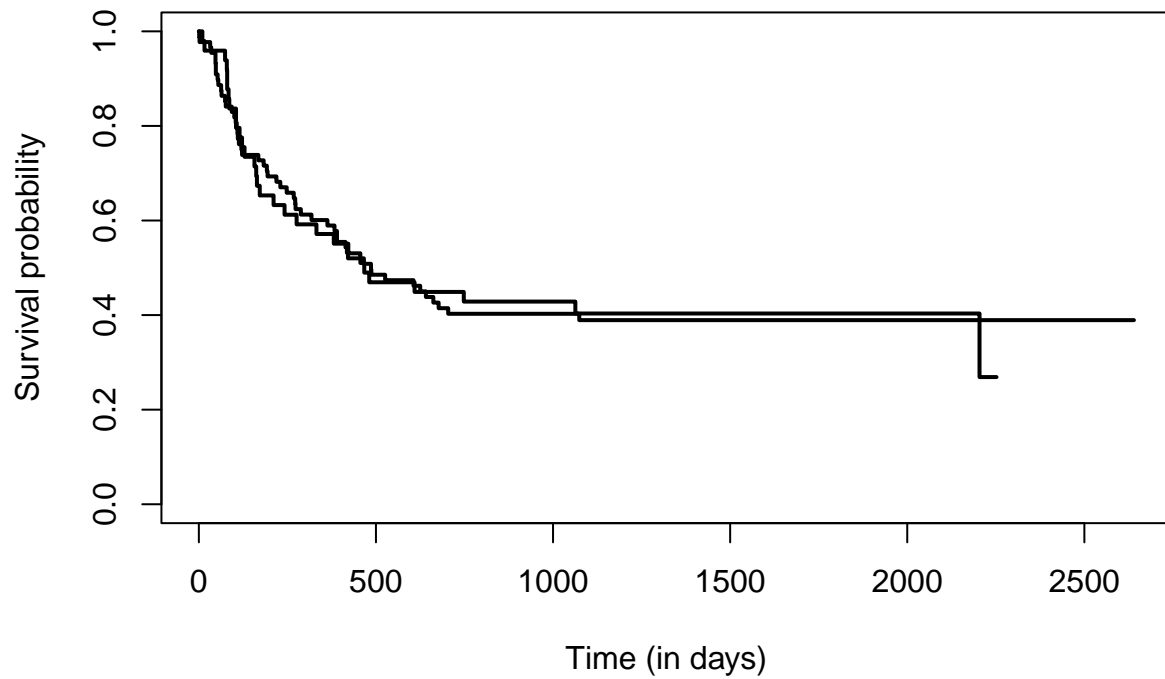
## Kaplan–Meier survival estimate, by CMV



```
# Survival Object based on sex subgrouping of donor
sfit_bmt_byDonerMale <- survfit(s_bmt ~ donormale,
                                data = bmt_df,
                                conf.type = "log-log")

plot(sfit_bmt_byDonerMale, conf.int = F, main="Kaplan-Meier survival estimate, by Donor Sex",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

### Kaplan–Meier survival estimate, by Donor Sex

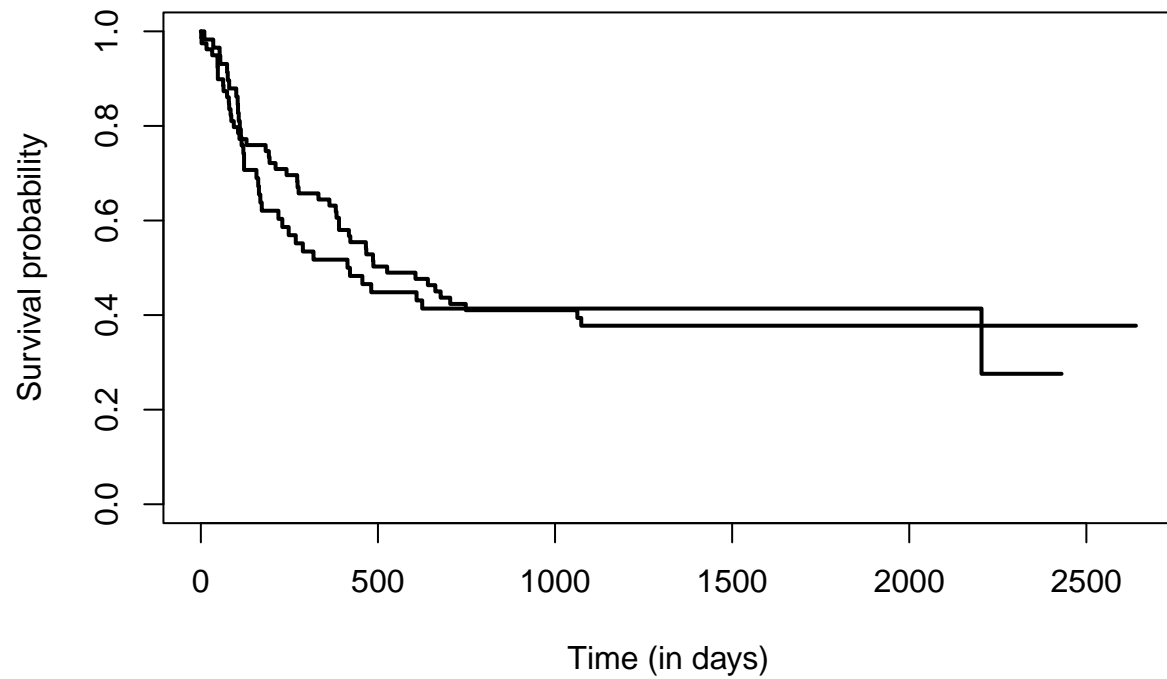


```
# Survival Object based on CMV subgrouping of donor
sfit_bmt_byDonerCMV <- survfit(s_bmt ~ donorcmv,
                               data = bmt_df,
                               conf.type = "log-log")

plot(sfit_bmt_byDonerCMV, conf.int = F, main="Kaplan-Meier survival estimate, by Donor CMV",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```



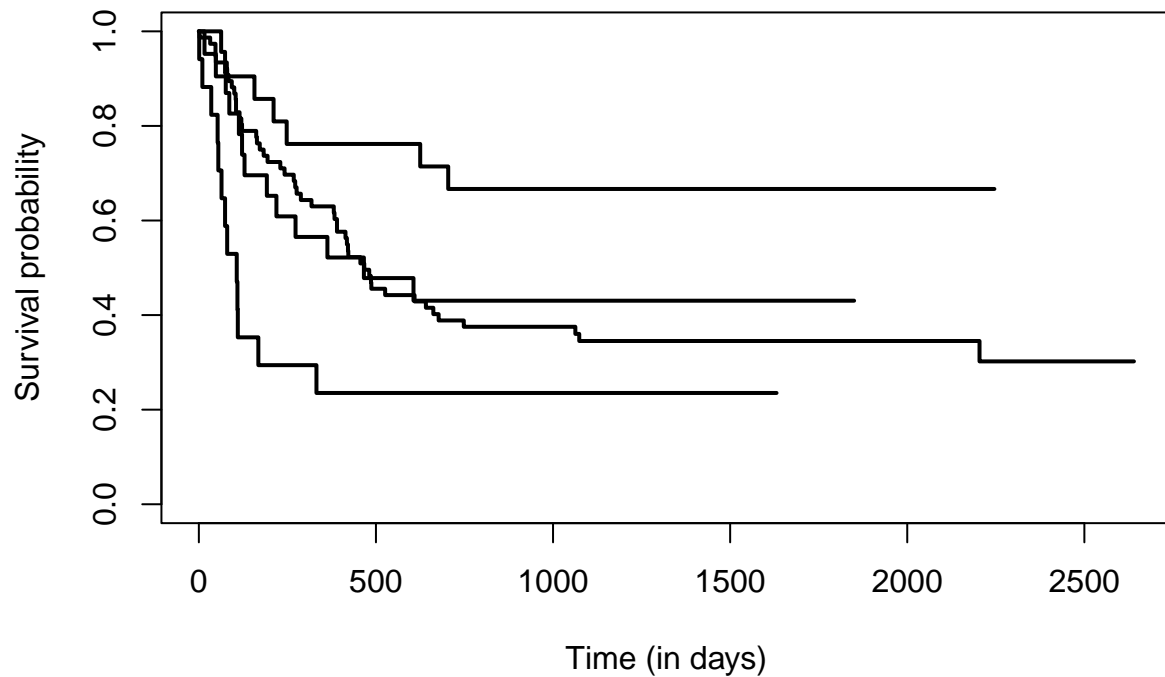
## Kaplan–Meier survival estimate, by Donor CMV



```
# Survival Object based on hospital subgrouping
sfit_bmt_byHospital <- survfit(s_bmt ~ hospital,
                              data = bmt_df,
                              conf.type = "log-log")

plot(sfit_bmt_byHospital, conf.int = F, main="Kaplan-Meier survival estimate, by Hospital",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

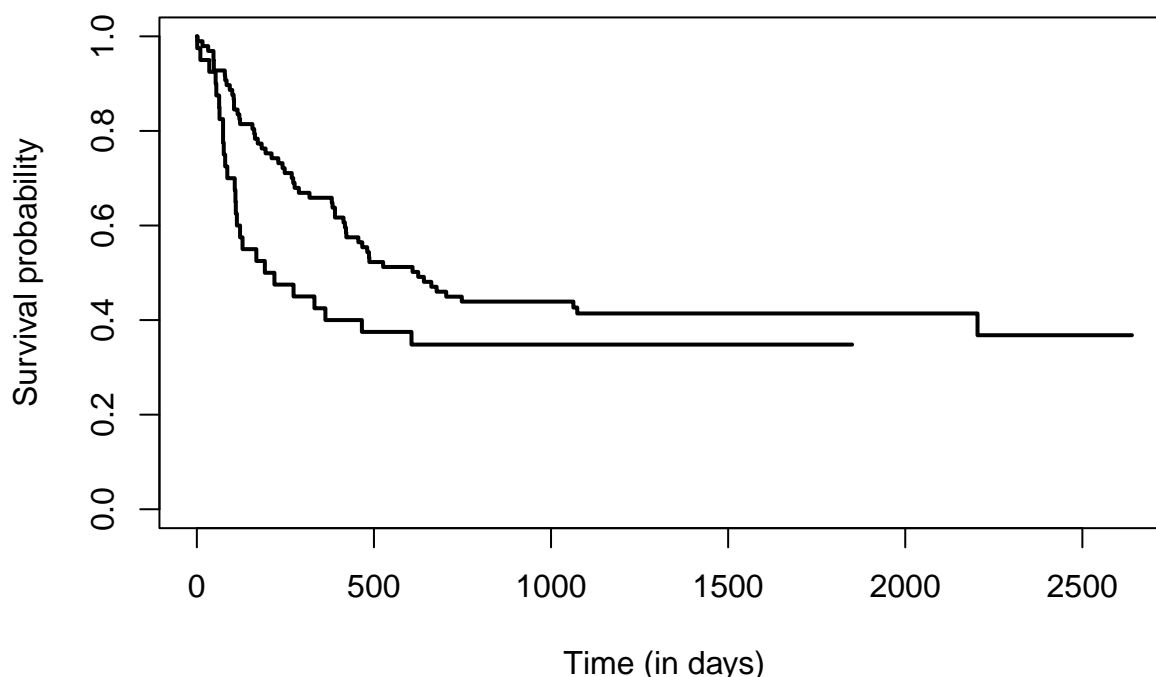
## Kaplan–Meier survival estimate, by Hospital



```
# Survival Object based on mtx subgrouping
sfit_bmt_byMTX <- survfit(s_bmt ~ mtx,
                          data = bmt_df,
                          conf.type = "log-log")

plot(sfit_bmt_byMTX, conf.int = F, main="Kaplan-Meier survival estimate, by MTX",
     xlab="Time (in days)", ylab="Survival probability", col="black",
     lty="solid", lwd=2)
```

## Kaplan–Meier survival estimate, by MTX



Two of the hospitals appear to be significant, but since each has little data, the power might be low, the SE might be high, and the CI may cross.

```
#nonparametric survival function survfit.bmt <- survfit(s.bmt~1, data=bmt, conf.type="log-log")#1.
estimate median disease free survival time print(survfit.bmt)#2How do patients in different disease
groups or in different FAB classifications compare to each other with respect to other available baseline
measurements? #table 1: columns = disease groups, rows= baseline characteristics #table 2: columns =
FAB classifications, rows = baseline characteristics#3 Are any of the measured baseline variables associated
with differences in disease-free survival?#4 Is occurrence of aGVHD after transplantation associated with
improved disease-free survival? summary(coxph(s.bmt ~deltaa + age + cmv + donorcmv + strata(hospital),
data=bmt))#Is it associated with a decreased risk of relapse? summary(coxph(s.relapse~deltaa + age
+ cmv + donorcmv, data=bmt))#5 Among the patients who develop aGVHD, are any of the measured
baseline factors associated with differences in disease-free survival? gvhd <- survfit(s.gvhd ~ mtx, data =
bmt, conf.type = "log-log") plot(gvhd, conf.int = F, main="Kaplan-Meier GVHD survival estimate, by
MTX", xlab="Time (in days)", ylab="Survival probability", col="black", lty="solid", lwd=2)gvhdcmv <-
survfit(s.gvhd ~ cmv, data = bmt, conf.type = "log-log") plot(gvhdcmv, conf.int = F, main="Kaplan-Meier
GVHD survival estimate, by recipient CMV status", xlab="Time (in days)", ylab="Survival probability",
col="black", lty="solid", lwd=2)gvhdhospital <- survfit(s.gvhd ~ hospital, data = bmt, conf.type =
"log-log") plot(gvhdhospital, conf.int = F, main="Kaplan-Meier GVHD survival estimate, by hospital",
xlab="Time (in days)", ylab="Survival probability", col="black", lty="solid", lwd=2)gvhddonorcmv
<- survfit(s.gvhd ~ donorcmv, data = bmt, conf.type = "log-log") plot(gvhddonorcmv, conf.int =
F, main="Kaplan-Meier GVHD survival estimate, by donor CMV status", xlab="Time (in days)",
ylab="Survival probability", col="black", lty="solid", lwd=2)#6 Is prophylactic use of methotrexate asso-
ciated with an increased or decreased risk of developing aGVHD? **incude confounders s.gvhd <- with(bmt,
Surv(ta, deltaa==1)) summary(coxph(s.gvhd~mtx + donorcmv + strata(hospital), data=bmt))#Provide
an estimate of the survival function of time from transplant until onset of aGVHD separatefor patients
either administered methotrexate or not. In doing so, consider the importance of accounting for relevant
```

```

confounding factors. s.gvhdm <- with(bmt, Surv(ta, deltaa==1)) survfit.gvhdmx <- survfit(s.gvhdm~mtx,
data=bmt, conf.type="log-log" ) plot(survfit.gvhdmx) summary(survfit.gvhdmx, times=c(7, 14, 21, 28,
35, 42, 49, 56))#7 Is recovery of normal platelet levels associated with improved disease-free survival? -
yes summary(coxph(s.bmt ~deltap + age + donorcmv + strata(hospital), data=bmt))#Is it associated
with a decreased risk of relapse? - no s.relapse <- with(bmt, Surv(agediagnosis, ageevent, deltar==1))
summary(coxph(s.relapse~deltap + age + donorcmv + strata(hospital), data=bmt)

```