

Data and preliminary code

Antoine Soetewey

28 October, 2018

- 1 Data from `relsurv` package
 - 1.1 `colrec`
 - 1.1.1 Description
 - 1.1.2 Format
 - 1.1.3 Descriptive statistics
 - 1.1.4 Kaplan-Meier
 - 1.1.5 Cox proportional hazards model
 - 1.1.6 Comparison between KM and Cox models
 - 1.2 `rdata`
 - 1.2.1 Description
 - 1.2.2 Format
 - 1.2.3 Descriptive statistics
 - 1.2.4 Kaplan-Meier
 - 1.2.5 Cox proportional hazards model
 - 1.2.6 Random forests model
 - 1.2.7 Comparison between KM, Cox and Random Forests models
 - 1.3 `slopop`
 - 1.3.1 Description
 - 1.3.2 Format
 - 1.4 Belgian population tables
- 2 Data from the Federal Planning Bureau
 - 2.1 Description
 - 2.2 Format
 - 2.3 Descriptive statistics
 - 2.3.1 Proportion of men in the population
 - 2.3.2 Mortality rates
 - 2.3.3 Life expectancy
- 3 Methods
 - 3.1 Relative survival
 - 3.1.1 Individual relative survival
 - 3.2 Esteve et al. (1990) model
 - 3.3 Poisson (Denuit and Legrand 2018)
 - 3.4 Boussari et al. (2018)
 - 3.5 Loss of expectancy
- 4 Selection of cancers
- References

1 Data from `relsurv` package

1.1 `colrec`

1.1.1 Description

Survival of patients with colon and rectal cancer diagnosed in 1994-2000.

1.1.2 Format

A data frame with 5971 observations on the following 7 variables:

- **sex:** sex (1=male, 2=female).
- **age:** age (in days).
- **diag:** date of diagnosis (in date format).
- **time:** survival time (in days).
- **stat:** censoring indicator (0=censoring, 1=death).
- stage: cancer stage. Values 1-3, code 99 stands for unknown.
- site: cancer site.

(Common variables between colrec and rdata in **bold**)

1.1.3 Descriptive statistics

```
dat <- colrec

str(dat)
```

```
## 'data.frame':    5971 obs. of  7 variables:
## $ sex   : int   1 2 1 2 2 1 2 1 2 2 ...
## $ age   : num  23004 12082 24277 29256 30416 ...
## $ diag  : 'date' num  12656 13388 12711 13971 12997 ...
## $ time  : num   16 504 22 3998 9 ...
## $ stat  : int   0 0 0 0 0 0 0 0 0 0 ...
## $ stage: Factor w/ 4 levels "1","2","3","99": 1 3 3 1 4 2 1 1 1 1 ...
## $ site  : Factor w/ 2 levels "colon","rectum": 2 2 1 1 1 1 2 2 2 2 ...
```

```
summary(dat)
```

```
##           sex           age           diag           time
## Min.      :1.000   Min.      : 4559   First :1Jan94   Min.      :    1.0
## 1st Qu.:1.000   1st Qu.:21973   Last  :30Dec2000 1st Qu.: 193.5
## Median :1.000   Median :24864                      Median : 872.0
## Mean     :1.449   Mean     :24565                      Mean     :2181.7
## 3rd Qu.:2.000   3rd Qu.:27493                      3rd Qu.:4047.5
## Max.     :2.000   Max.     :35325                      Max.     :8148.0
##           stat           stage           site
## Min.      :0.0000   1 : 889   colon :3537
## 1st Qu.:1.0000   2 :3328   rectum:2434
## Median :1.0000   3 :1361
## Mean     :0.8339   99: 393
## 3rd Qu.:1.0000
## Max.     :1.0000
```

```
datatable(dat, options = list(pageLength = 5))
```

Show

5

 entries

Search:

	sex	age	diag	time	stat	stage	site
1	1	23004	12656	16	0	1	rectum

	sex	age	diag	time	stat	stage	site
2	2	12082	13388	504	0	3	rectum
3	1	24277	12711	22	0	3	colon
4	2	29256	13971	3998	0	1	colon
6	2	30416	12997	9	0	99	colon

Showing 1 to 5 of 5,971 entries

Previous

1

2

3

4

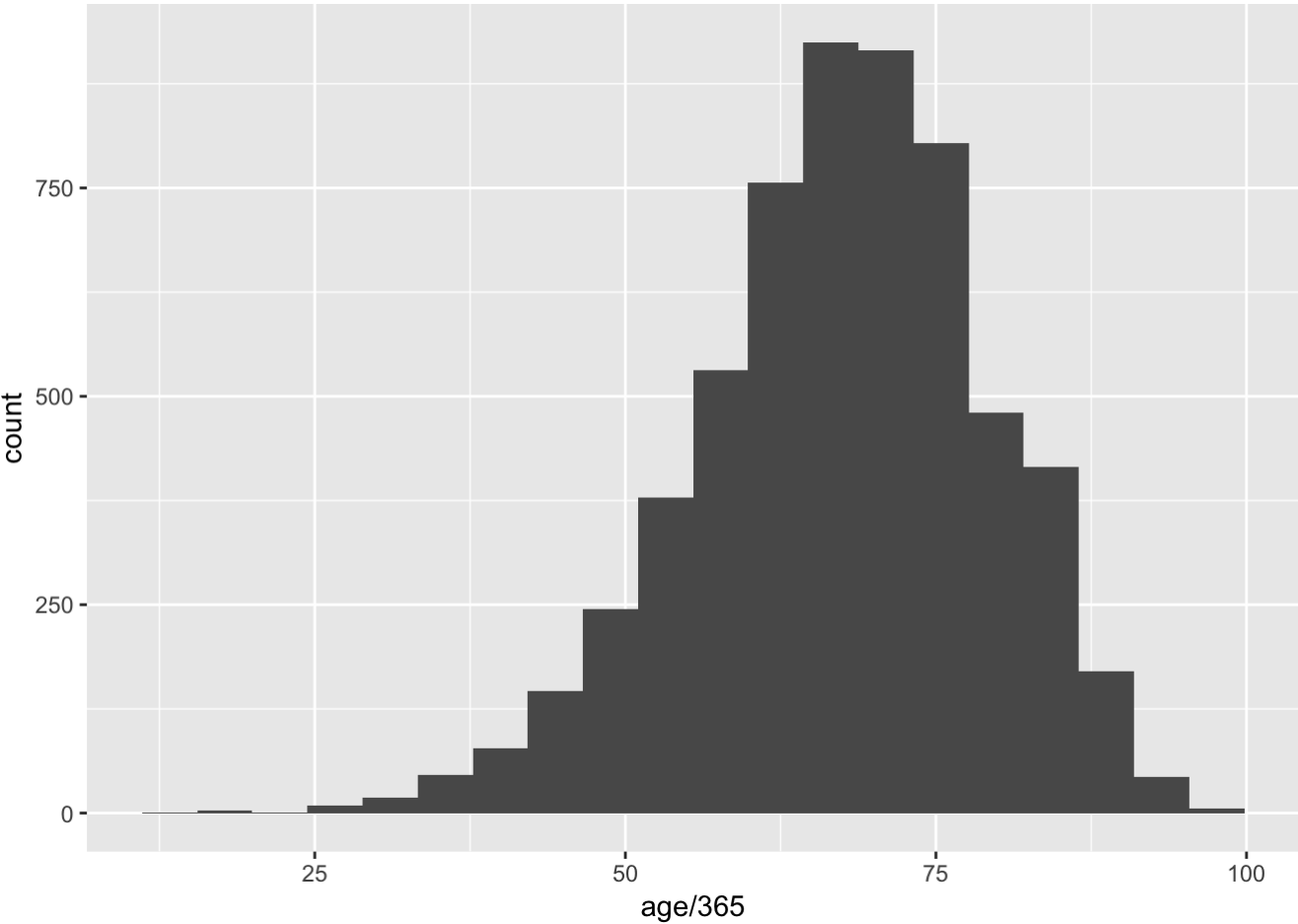
5

...

1195

Next

```
ggplot(dat, aes(age/365)) +
  geom_histogram(bins=20)
```



```
surv_object <- Surv(time = dat$time, event = dat$stat)
summary(surv_object)
```

##	time	status
##	Min. : 1.0	Min. :0.0000
##	1st Qu.: 193.5	1st Qu.:1.0000
##	Median : 872.0	Median :1.0000
##	Mean :2181.7	Mean :0.8339
##	3rd Qu.:4047.5	3rd Qu.:1.0000
##	Max. :8148.0	Max. :1.0000

1.1.4 Kaplan-Meier

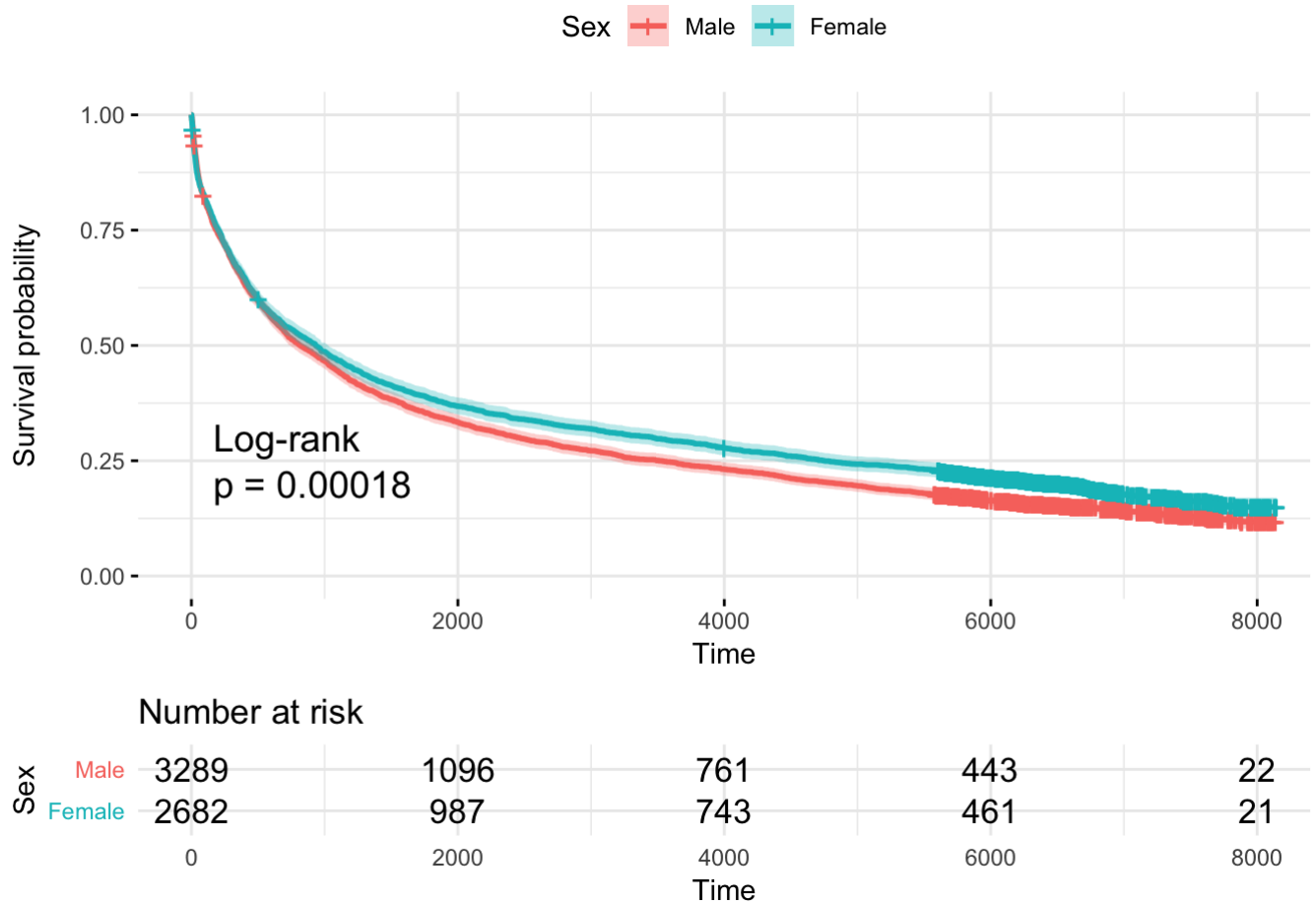
```
fit_km <- survfit(Surv(time, stat)~sex, data=dat)
fit_km
```

```
## Call: survfit(formula = Surv(time, stat) ~ sex, data = dat)
##
##           n events median 0.95LCL 0.95UCL
## sex=1 3289   2804    818    740    915
## sex=2 2682   2175    934    833   1026
```

```
summary(fit_km, times=seq(0, max(dat$time), max(dat$time)/10))
```

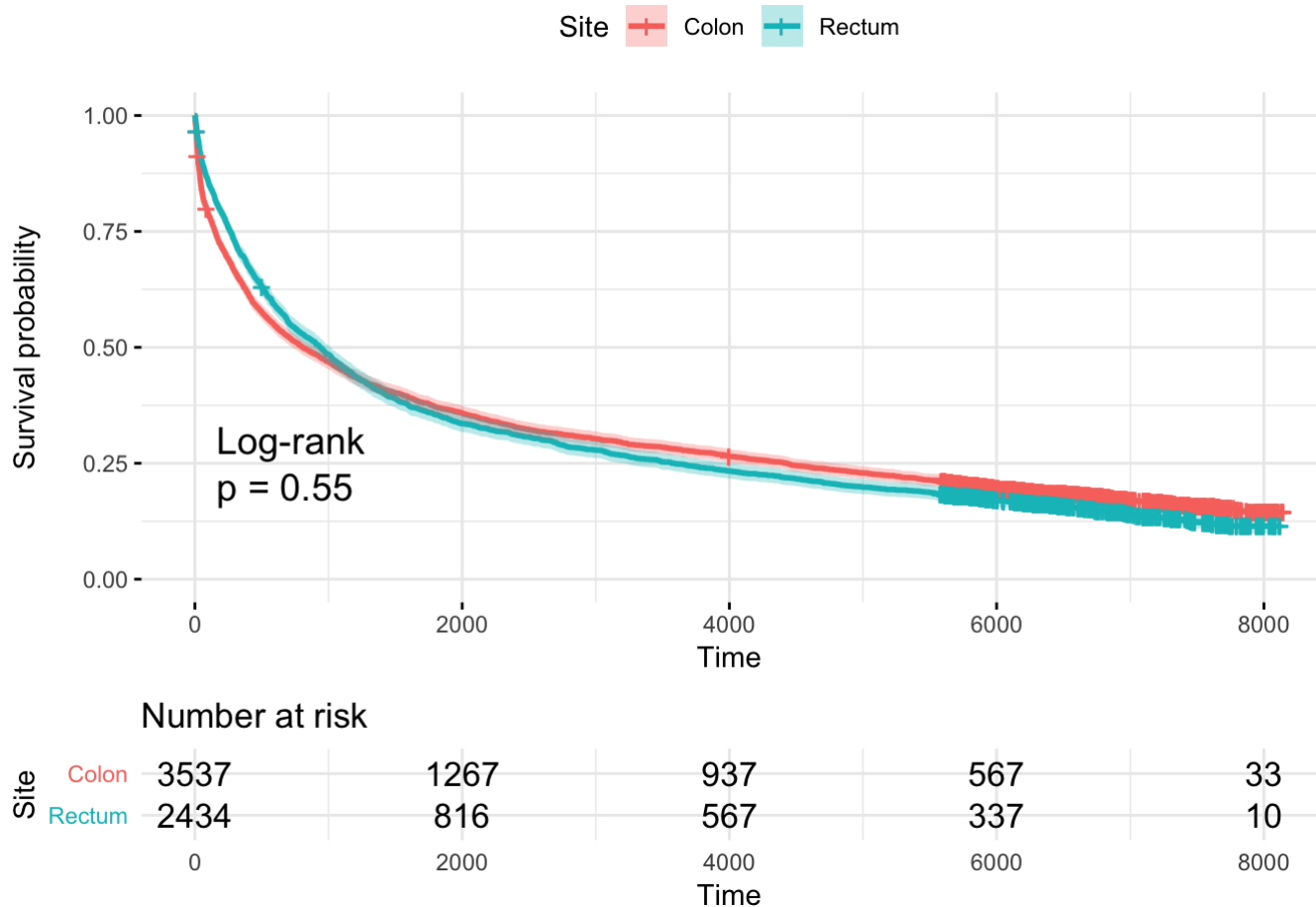
```
## Call: survfit(formula = Surv(time, stat) ~ sex, data = dat)
##
##           sex=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    0   3289     0    1.000 0.00000    1.000    1.000
##   815   1644   1642    0.500 0.00872    0.484    0.518
##  1630   1205    439    0.367 0.00841    0.351    0.384
##  2444    987    218    0.300 0.00800    0.285    0.316
##  3259    845    142    0.257 0.00762    0.243    0.273
##  4074    754     91    0.229 0.00734    0.216    0.244
##  4889    654    100    0.199 0.00697    0.186    0.213
##  5704    533     84    0.173 0.00661    0.161    0.187
##  6518    294     54    0.153 0.00641    0.140    0.166
##  7333    115     27    0.131 0.00677    0.119    0.145
##
##           sex=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    0   2682     0    1.000 0.00000    1.000    1.000
##   815   1397   1283    0.521 0.00965    0.503    0.541
##  1630   1071    326    0.400 0.00946    0.382    0.419
##  2444    915    156    0.342 0.00916    0.324    0.360
##  3259    822     93    0.307 0.00891    0.290    0.325
##  4074    734     87    0.274 0.00862    0.258    0.292
##  4889    655     79    0.245 0.00831    0.229    0.262
##  5704    561     57    0.223 0.00805    0.208    0.240
##  6518    309     50    0.199 0.00789    0.184    0.215
##  7333    116     35    0.169 0.00830    0.153    0.186
##  8148     1      9    0.150 0.00963    0.132    0.170
```

```
ggsurvplot(fit_km,
            conf.int=TRUE,
            pval=TRUE,
            pval.method=TRUE,
            risk.table=TRUE,
            legend.labs=c("Male", "Female"),
            legend.title="Sex",
            #palette=c("dodgerblue2", "orchid2"),
            main="Kaplan-Meier Curve for colrec data",
            #risk.table.height=.26,
            ggtheme = theme_minimal())
```



```
fit_km2 <- survfit(Surv(time, stat)~site, data=dat)

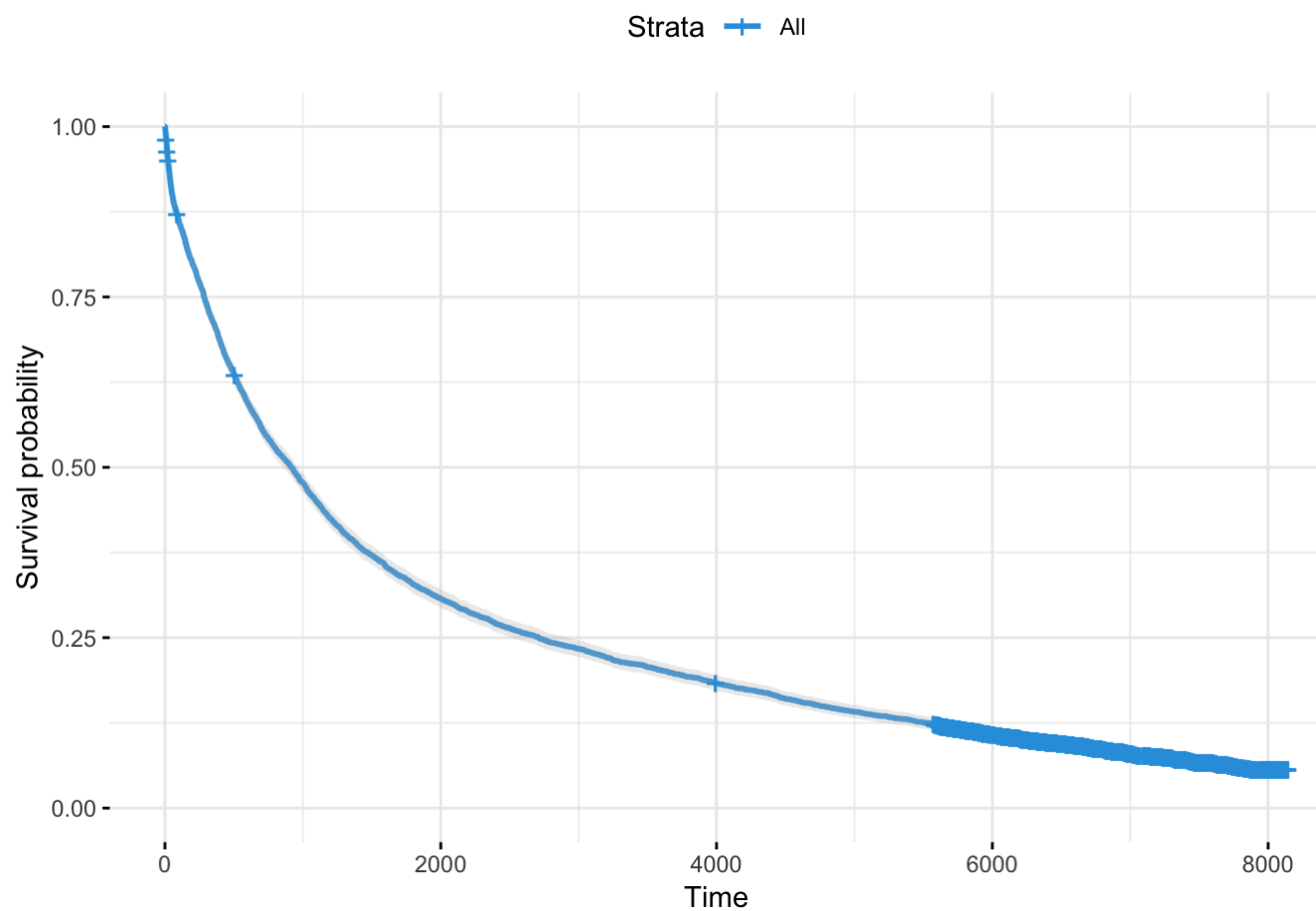
ggsurvplot(fit_km2,
  conf.int=TRUE,
  pval=TRUE,
  pval.method=TRUE,
  risk.table=TRUE,
  legend.labs=c("Colon", "Rectum"),
  legend.title="Site",
  #palette=c("dodgerblue2", "orchid2"),
  main="Kaplan-Meier Curve for colrec data",
  #risk.table.height=.26,
  ggtheme = theme_minimal())
```



Based on the p -values of the two Log-rank tests, we can say that there is a difference between survival curves for males and females, whereas there is no difference between survival curves for colon and rectum cancers.

1.1.5 Cox proportional hazards model

```
# Plot the baseline survival function
ggsurvplot(survfit(coxph(Surv(time, stat)~.,
                        data=dat)),
            conf.int=TRUE,
            risk.table=FALSE,
            palette = "#2E9FDF",
            ggtheme = theme_minimal(),
            data = dat)
```



```
fit_cox <- coxph(Surv(time, stat)~., data=dat)
summary(fit_cox)
```

```
## Call:
## coxph(formula = Surv(time, stat) ~ ., data = dat)
##
##      n= 5971, number of events= 4979
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## sex          -1.608e-01  8.515e-01  2.891e-02 -5.562 2.67e-08 ***
## age           1.153e-04  1.000e+00  3.902e-06 29.541 < 2e-16 ***
## diag         -1.135e-04  9.999e-01  2.012e-05 -5.640 1.70e-08 ***
## stage2        4.148e-01  1.514e+00  4.498e-02  9.223 < 2e-16 ***
## stage3        1.969e+00  7.167e+00  5.202e-02 37.858 < 2e-16 ***
## stage99       1.153e+00  3.169e+00  6.655e-02 17.329 < 2e-16 ***
## siterectum    1.003e-01  1.105e+00  2.923e-02  3.431 0.000602 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## sex              0.8515      1.1745    0.8045    0.9011
## age              1.0001      0.9999    1.0001    1.0001
## diag             0.9999      1.0001    0.9998    0.9999
## stage2           1.5141      0.6604    1.3864    1.6537
## stage3           7.1670      0.1395    6.4722    7.9363
## stage99          3.1686      0.3156    2.7811    3.6101
## siterectum       1.1055      0.9046    1.0439    1.1707
##
## Concordance= 0.718 (se = 0.004 )
## Rsquare= 0.372 (max possible= 1 )
## Likelihood ratio test= 2777 on 7 df,  p=<2e-16
## Wald test              = 2934 on 7 df,  p=<2e-16
## Score (logrank) test = 3266 on 7 df,  p=<2e-16
```

From the output above, we can conclude that all variables have highly statistically significant coefficients.

Reminder: a hazard ratio above 1 indicates a covariate that is positively associated with the event probability, and thus negatively associated with the length of survival.

In summary,

- HR = 1: No effect
- HR < 1: Reduction in the hazard
- HR > 1: Increase in the hazard

Note that in cancer studies:

- A covariate with hazard ratio > 1 is called bad prognostic factor
- A covariate with hazard ratio < 1 is called good prognostic factor.

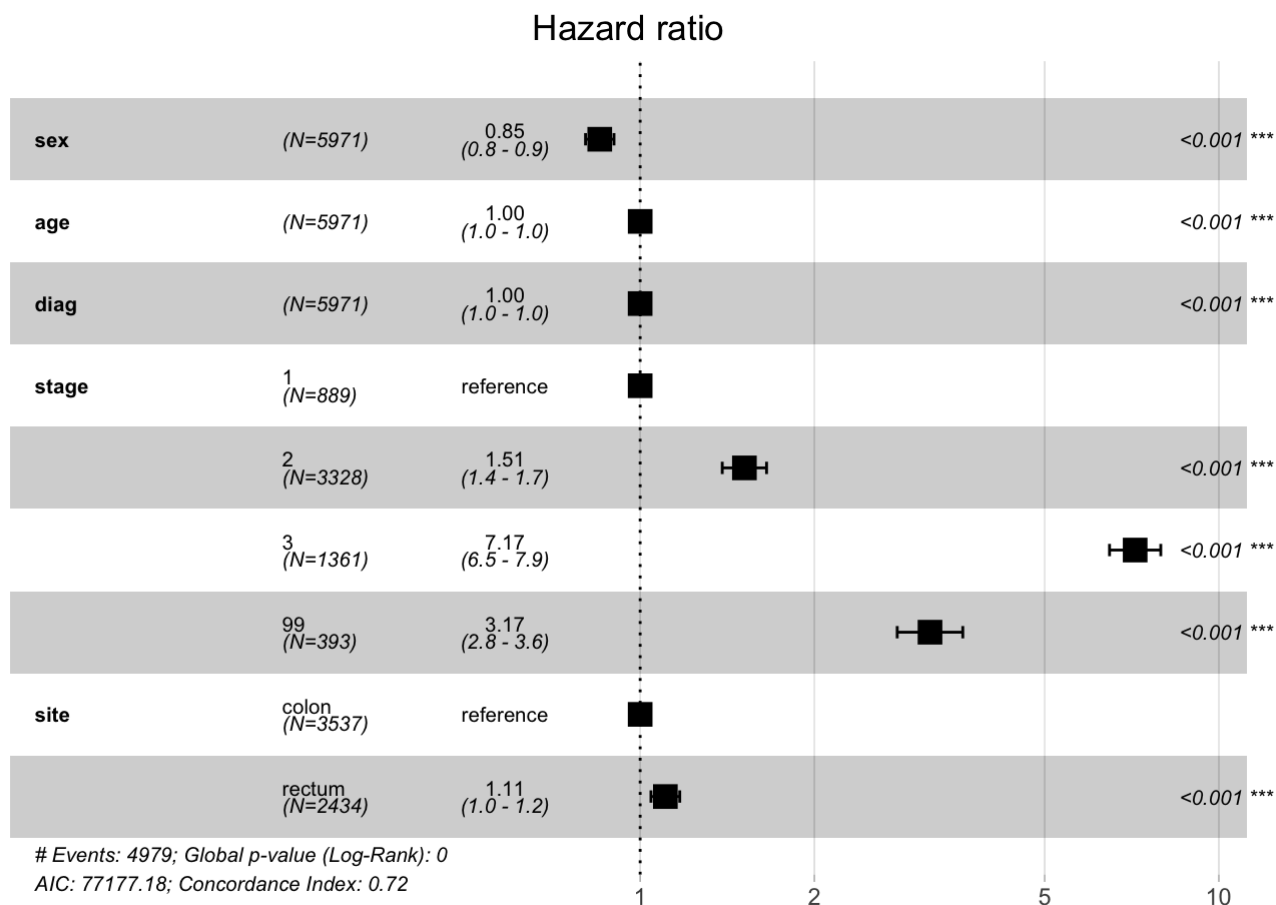
Therefore, from the output above, we can conclude that:

- being older, being diagnosed with a higher stage cancer and being diagnosed with a rectum cancer instead of a colon cancer are bad prognostic factors
- being a woman and being diagnosed later (i.e., more recently) are good prognostic factors.

For example, holding the other covariates constant, being a woman (sex=2) reduces the hazard by a factor of 0.8515, or 14.85%.

A more systematic way to look at the different covariates is to draw a forest plot.


```
# Forest plot
ggforest(fit_cox, data = dat)
```

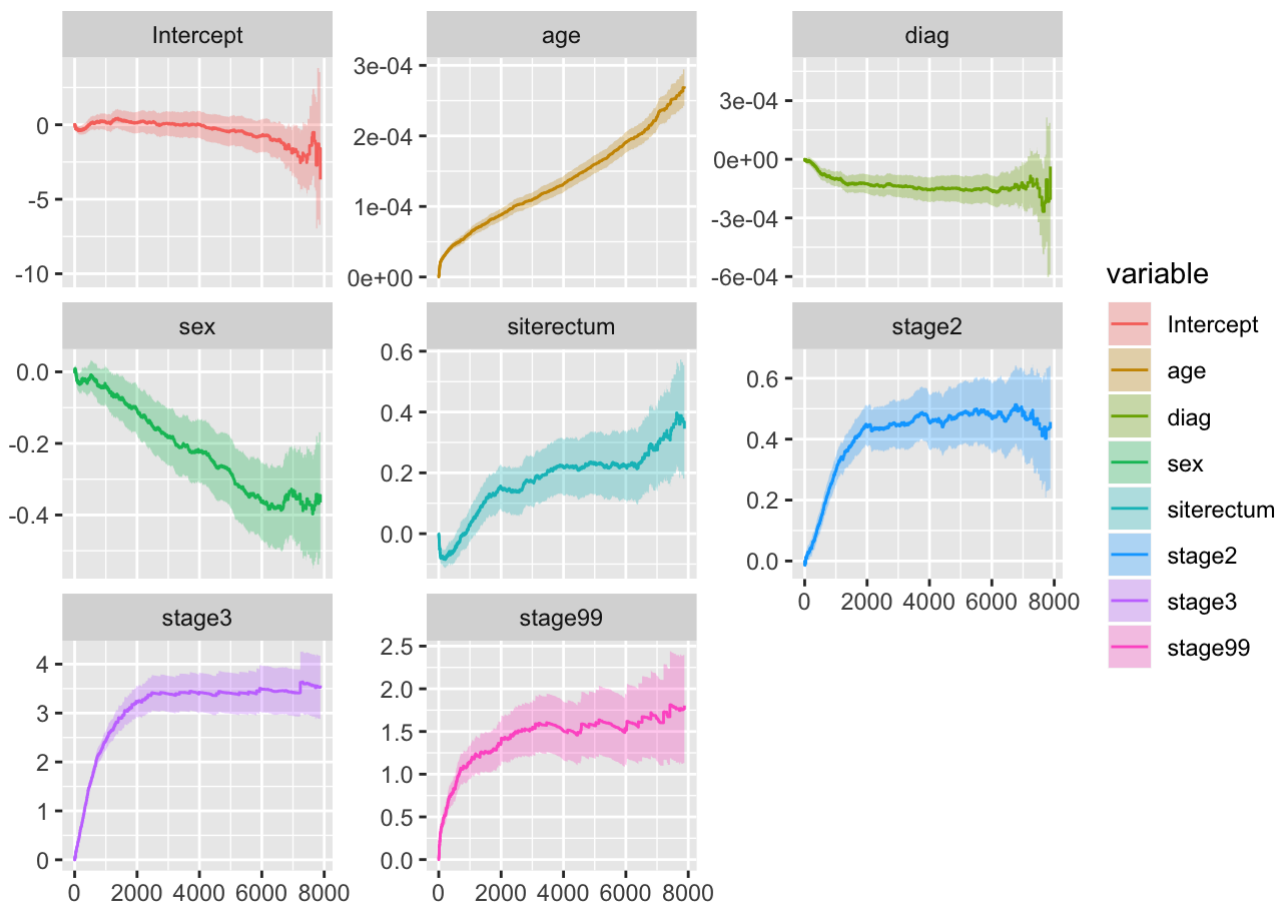


The output above gives the same results, that is, all covariates are significant and the corresponding hazard ratios.

We now fit Aalen's additive regression model for censored data to the data. The documentation states: "The Aalen model assumes that the cumulative hazard $H(t)$ for a subject can be expressed as $a(t) + XB(t)$, where $a(t)$ is a time-dependent intercept term, X is the vector of covariates for the subject (possibly time-dependent), and $B(t)$ is a time-dependent matrix of coefficients."

```
aa_fit <- aareg(Surv(time, stat) ~ .,
               data = dat)

autoplot(aa_fit)
```



The plots above show how the effects of the covariates change over time.

1.1.6 Comparison between KM and Cox models

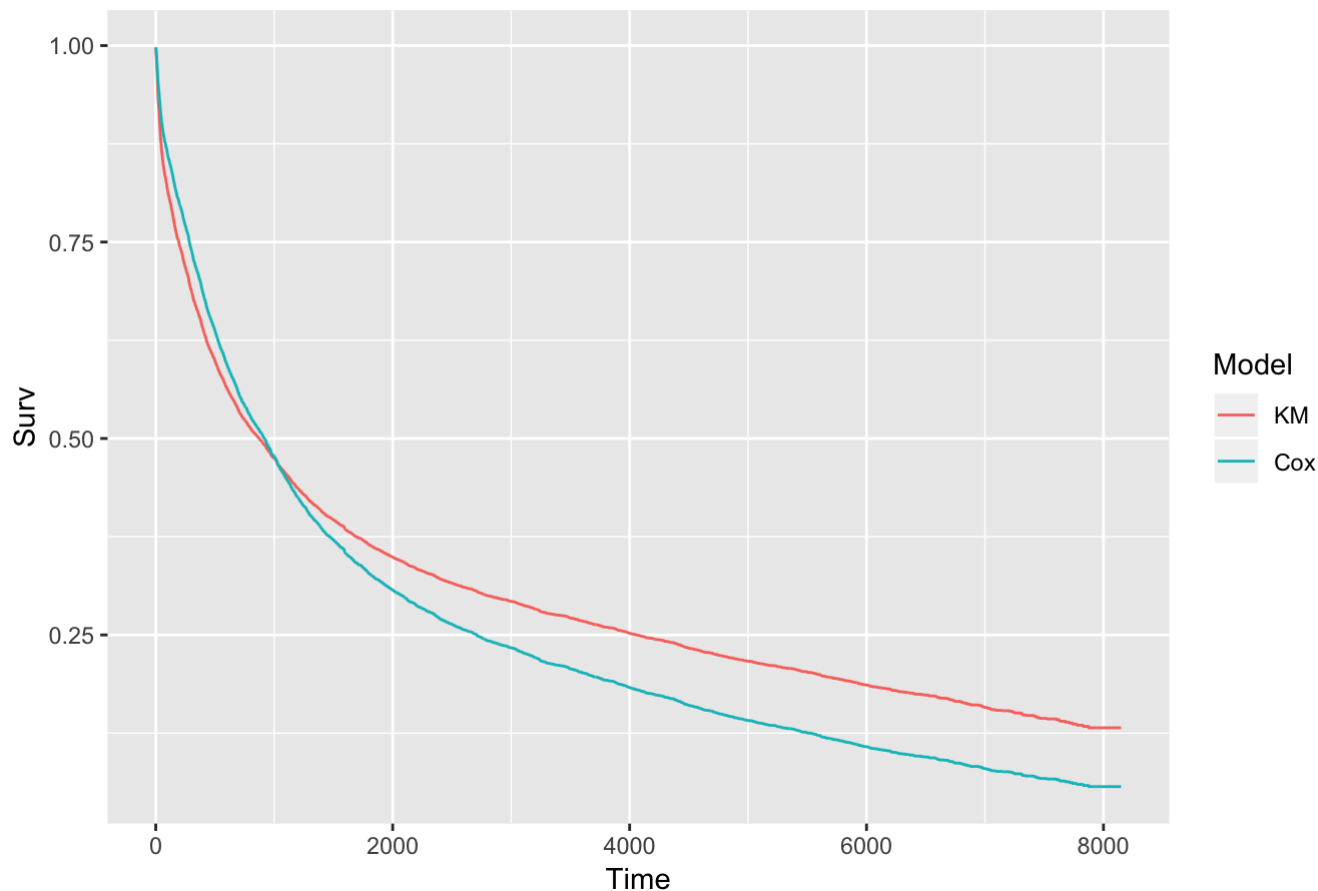
```
# Set up for ggplot
fit_km <- survfit(Surv(time, stat) ~ 1, data=dat)
km <- rep("KM", length(fit_km$time))
km_df <- data.frame(fit_km$time, fit_km$surv, km)
names(km_df) <- c("Time", "Surv", "Model")

fit_cox_comparison <- survfit(fit_cox)
cox <- rep("Cox", length(fit_cox_comparison$time))
cox_df <- data.frame(fit_cox_comparison$time, fit_cox_comparison$surv, cox)
names(cox_df) <- c("Time", "Surv", "Model")

plot_df <- rbind(km_df, cox_df)

ggplot(plot_df, aes(x = Time, y = Surv, color = Model)) +
  geom_line() +
  ggtitle("Comparison of Survival Curves")
```

Comparison of Survival Curves



1.2 rdata

1.2.1 Description

Survival data from 19Jan82 until 28Dec86.

1.2.2 Format

A data frame with 1040 observations on the following 6 variables:

- **time**: survival time (in days).
- **cens**: censoring indicator (0=censoring, 1=death).
- **age**: age (in years).
- **sex**: sex (1=male, 2=female).
- **year**: date of diagnosis (in date format).
- agegr: age group.

(Common variables between `colrec` and `rdata` in **bold**)

1.2.3 Descriptive statistics

```
dat <- rdata  
  
str(dat)
```

```
## 'data.frame': 1040 obs. of 6 variables:
## $ time : int 2657 1097 3764 3724 5076 139 4940 5078 596 4635 ...
## $ cens : int 1 1 1 1 0 1 1 0 1 1 ...
## $ age : int 68 63 60 66 57 57 67 59 43 50 ...
## $ sex : num 2 2 1 2 2 2 1 1 1 2 ...
## $ year : 'date' num 8210 8278 8254 8054 8224 ...
## $ agegr: Factor w/ 4 levels "<54","54-61",...: 3 3 2 3 2 2 3 2 1 1 ...
```

```
summary(dat)
```

```
##      time      cens      age      sex
## Min.   : 9   Min.   :0.000   Min.   :24.00   Min.   :1.000
## 1st Qu.:1206 1st Qu.:0.000   1st Qu.:53.00   1st Qu.:1.000
## Median :2872 Median :1.000   Median :61.00   Median :1.000
## Mean   :2681 Mean   :0.526   Mean   :61.67   Mean   :1.278
## 3rd Qu.:4126 3rd Qu.:1.000   3rd Qu.:70.00   3rd Qu.:2.000
## Max.   :5345 Max.   :1.000   Max.   :95.00   Max.   :2.000
##      year      agegr
## First :19Jan82  <54 :270
## Last  :28Dec86  54-61:256
##                62-70:255
##                71-95:259
##
##
```

```
datatable(dat, options = list(pageLength = 5))
```

Show

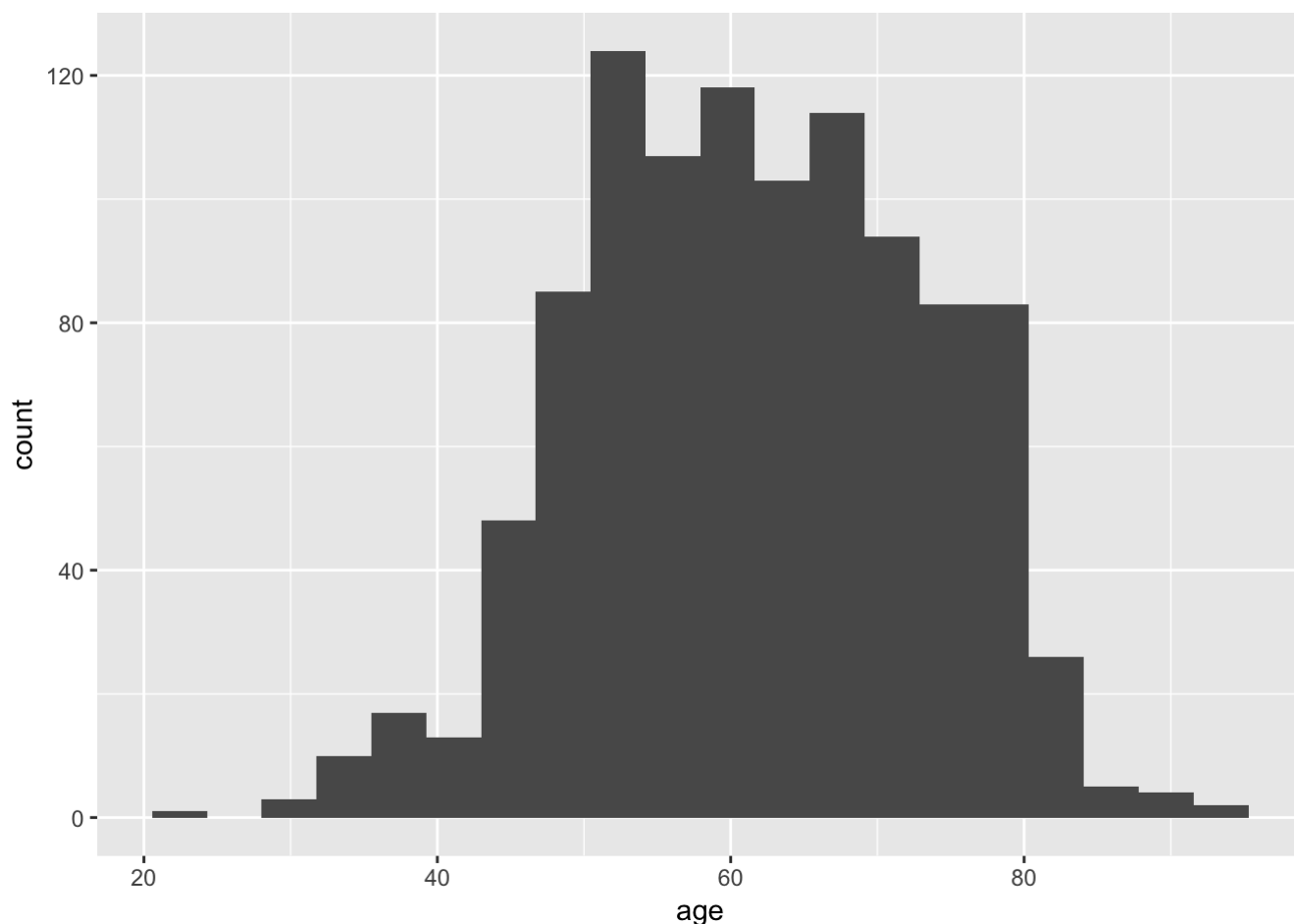
5

 entries

Search:

	time	cens	age	sex	year	agegr
1	2657	1	68	2	8210	62-70
2	1097	1	63	2	8278	62-70
3	3764	1	60	1	8254	54-61
4	3724	1	66	2	8054	62-70
5	5076	0	57	2	8224	54-61

```
ggplot(dat, aes(age)) +
  geom_histogram(bins=20)
```



```
surv_object <- Surv(time = dat$time, event = dat$cens)
summary(surv_object)
```

```
##      time      status
##  Min.   : 9      Min.   :0.000
## 1st Qu.:1206    1st Qu.:0.000
## Median :2872    Median :1.000
## Mean   :2681    Mean   :0.526
## 3rd Qu.:4126    3rd Qu.:1.000
## Max.   :5345    Max.   :1.000
```

1.2.4 Kaplan-Meier

```
fit_km <- survfit(Surv(time, cens)~sex, data=dat)
fit_km
```

```
## Call: survfit(formula = Surv(time, cens) ~ sex, data = dat)
##
##           n events median 0.95LCL 0.95UCL
## sex=1 751   360   4204   3810   4596
## sex=2 289   187   2203   1720   2823
```

```
summary(fit_km, times=seq(0, max(dat$time), max(dat$time)/10))
```

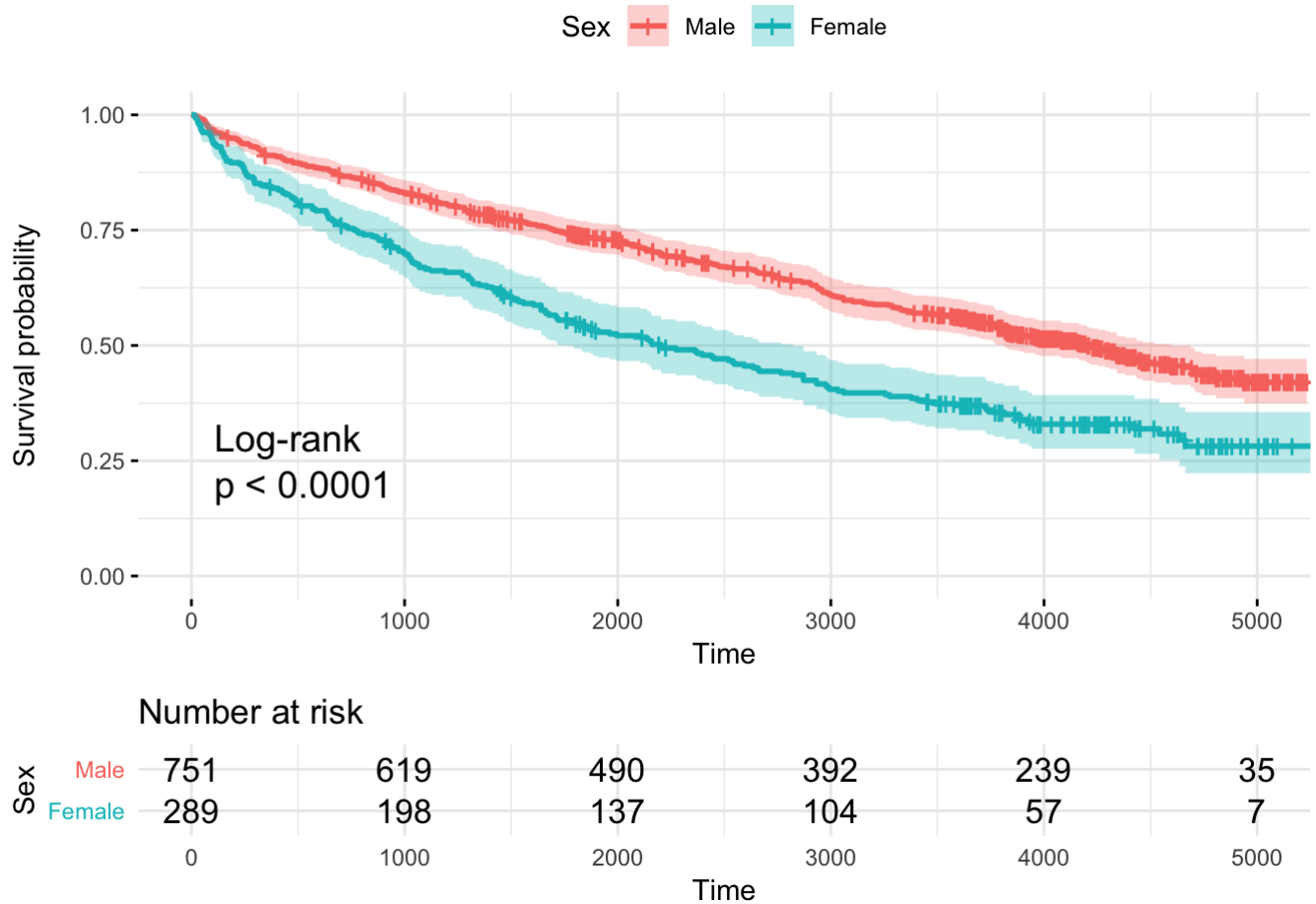
```
## Call: survfit(formula = Surv(time, cens) ~ sex, data = dat)
##
##
```

	sex=1							
##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
##	0	751	0	1.000	0.0000	1.000	1.000	
##	534	668	81	0.892	0.0113	0.870	0.915	
##	1069	612	50	0.825	0.0139	0.798	0.853	
##	1604	543	46	0.762	0.0156	0.732	0.793	
##	2138	471	38	0.707	0.0169	0.675	0.741	
##	2672	427	33	0.657	0.0178	0.623	0.693	
##	3207	379	44	0.589	0.0187	0.553	0.626	
##	3742	307	25	0.548	0.0191	0.512	0.587	
##	4276	165	26	0.493	0.0201	0.455	0.534	
##	4810	73	16	0.429	0.0234	0.385	0.477	

```
##
##
```

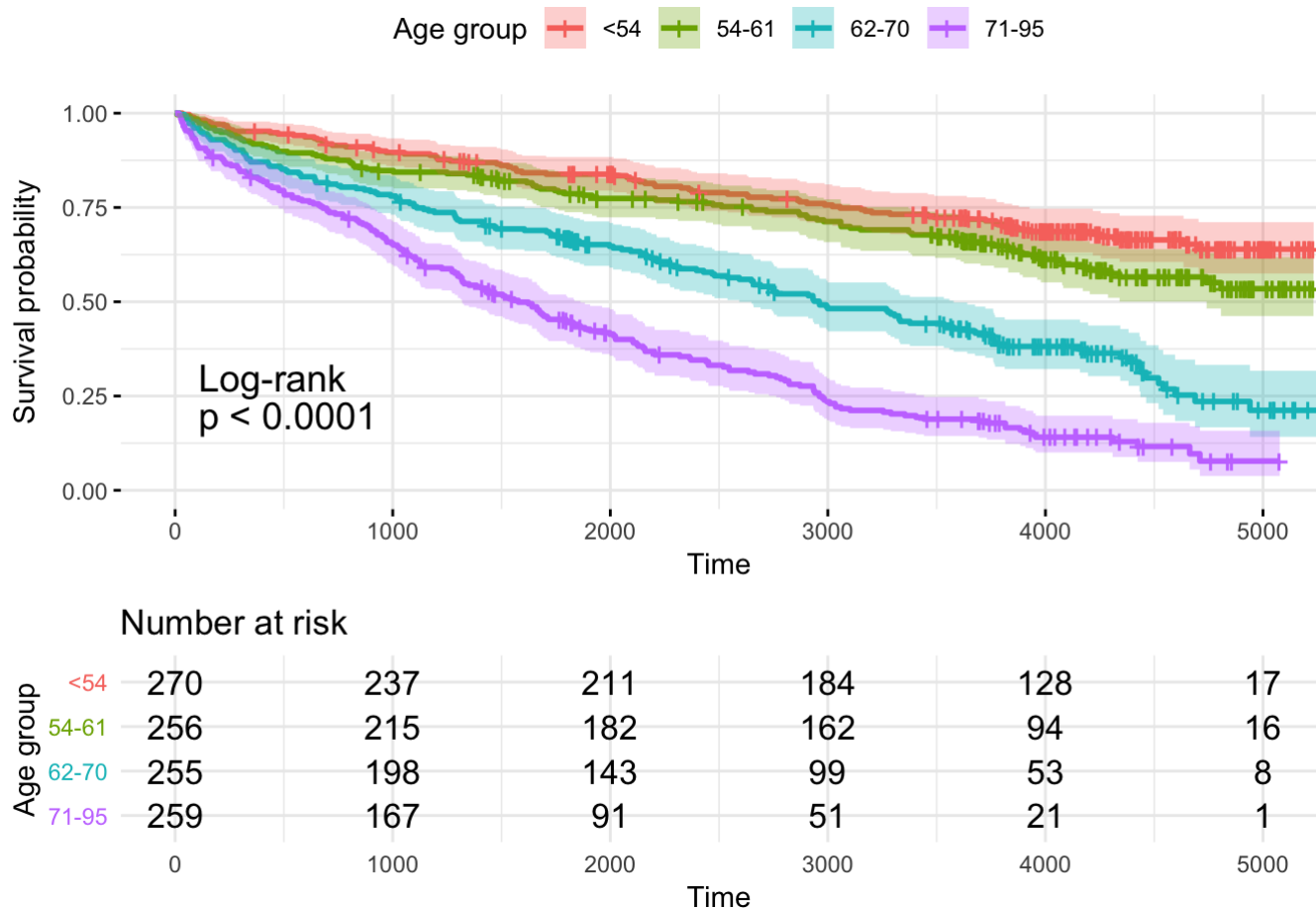
	sex=2							
##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
##	0	289	0	1.000	0.0000	1.000	1.000	
##	534	230	57	0.803	0.0234	0.758	0.850	
##	1069	190	38	0.669	0.0278	0.617	0.726	
##	1604	164	22	0.591	0.0291	0.536	0.651	
##	2138	134	21	0.514	0.0298	0.458	0.575	
##	2672	114	18	0.444	0.0300	0.389	0.507	
##	3207	102	12	0.397	0.0297	0.343	0.460	
##	3742	75	9	0.360	0.0294	0.307	0.422	
##	4276	42	6	0.329	0.0295	0.276	0.392	
##	4810	15	4	0.282	0.0336	0.223	0.356	
##	5345	1	0	0.282	0.0336	0.223	0.356	

```
ggsurvplot(fit_km,
            conf.int=TRUE,
            pval=TRUE,
            pval.method=TRUE,
            risk.table=TRUE,
            legend.labs=c("Male", "Female"),
            legend.title="Sex",
            #palette=c("dodgerblue2", "orchid2"),
            main="Kaplan-Meier Curve for rdata",
            #risk.table.height=.26,
            ggtheme = theme_minimal())
```



```
fit_km2 <- survfit(Surv(time, cens)~agegr, data=dat)

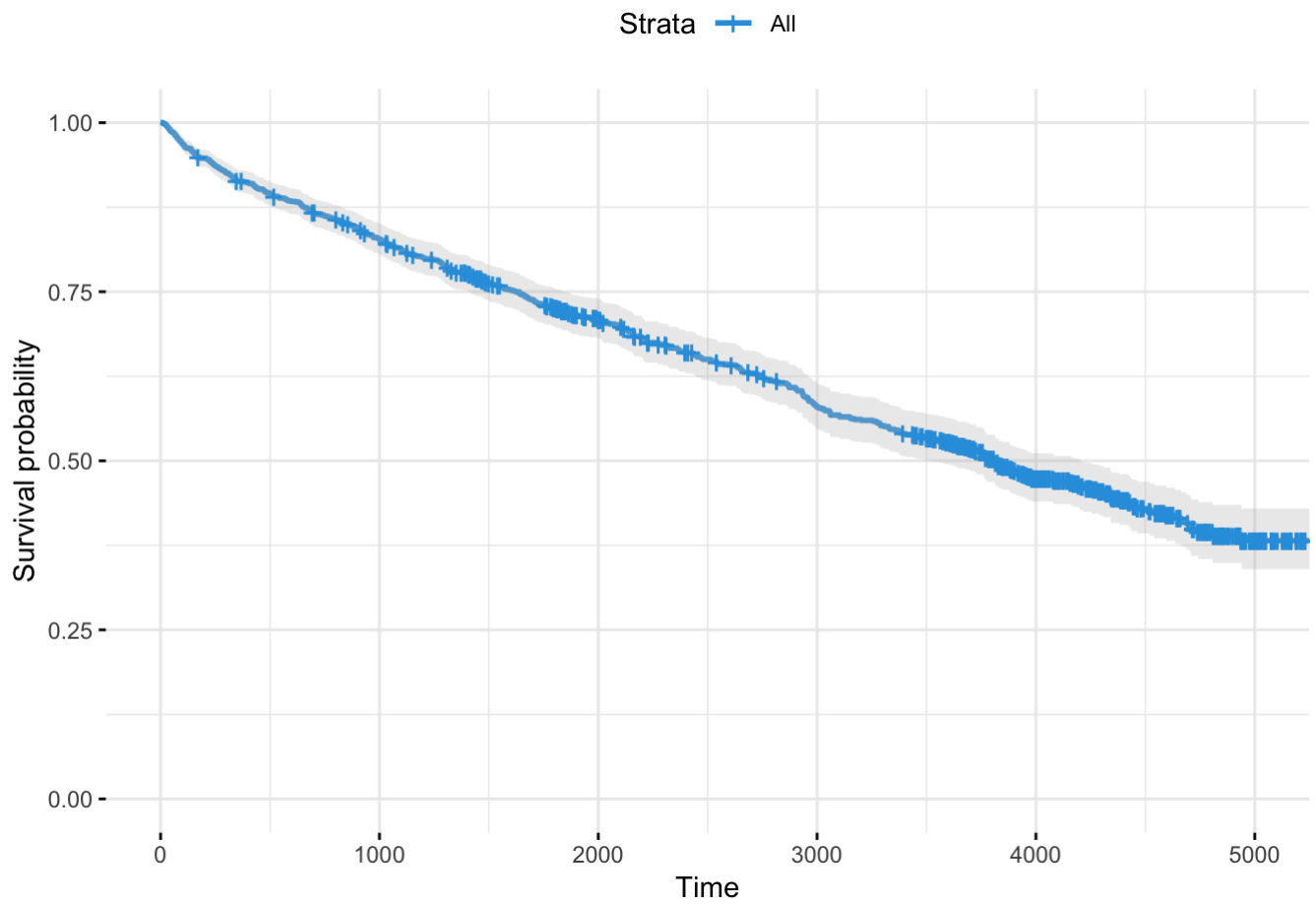
ggsurvplot(fit_km2,
  conf.int=TRUE,
  pval=TRUE,
  pval.method=TRUE,
  risk.table=TRUE,
  legend.labs=c("<54", "54-61", "62-70", "71-95"),
  legend.title="Age group",
  #palette=c("dodgerblue2", "orchid2"),
  main="Kaplan-Meier Curve for rdata",
  risk.table.height=.35,
  ggtheme = theme_minimal())
```



Based on the p -values of the two Log-rank tests, we can say that there is a difference between survival curves for males and females, and there is a difference between survival curves for each age group.

1.2.5 Cox proportional hazards model

```
# Plot the baseline survival function
ggsurvplot(survfit(coxph(Surv(time, cens)~.,
                        data=dat)),
            conf.int=TRUE,
            risk.table=FALSE,
            palette = "#2E9FDF",
            ggtheme = theme_minimal(),
            data = dat)
```

```
fit_cox <- coxph(Surv(time, cens)~., data=dat)
summary(fit_cox)
```

```
## Call:
## coxph(formula = Surv(time, cens) ~ ., data = dat)
##
##      n= 1040, number of events= 547
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## age          5.525e-02  1.057e+00  1.201e-02  4.600 4.22e-06 ***
## sex          8.265e-02  1.086e+00  9.680e-02  0.854 0.39318
## year        -2.764e-04  9.997e-01  9.316e-05 -2.966 0.00301 **
## agegr54-61 -2.258e-01  7.979e-01  1.834e-01 -1.232 0.21808
## agegr62-70 -2.668e-02  9.737e-01  2.572e-01 -0.104 0.91737
## agegr71-95  3.156e-02  1.032e+00  3.663e-01  0.086 0.93132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age          1.0568      0.9463      1.0322      1.0820
## sex          1.0862      0.9207      0.8985      1.3131
## year          0.9997      1.0003      0.9995      0.9999
## agegr54-61    0.7979      1.2534      0.5570      1.1429
## agegr62-70    0.9737      1.0270      0.5882      1.6118
## agegr71-95    1.0321      0.9689      0.5034      2.1158
##
## Concordance= 0.684 (se = 0.013 )
## Rsquare= 0.21 (max possible= 0.999 )
## Likelihood ratio test= 244.7 on 6 df, p=<2e-16
## Wald test = 241.4 on 6 df, p=<2e-16
## Score (logrank) test = 266.1 on 6 df, p=<2e-16
```

From the output above, we can conclude that only the variables `age` and `year` have statistically significant coefficients.

Also from the output above, we can conclude that:

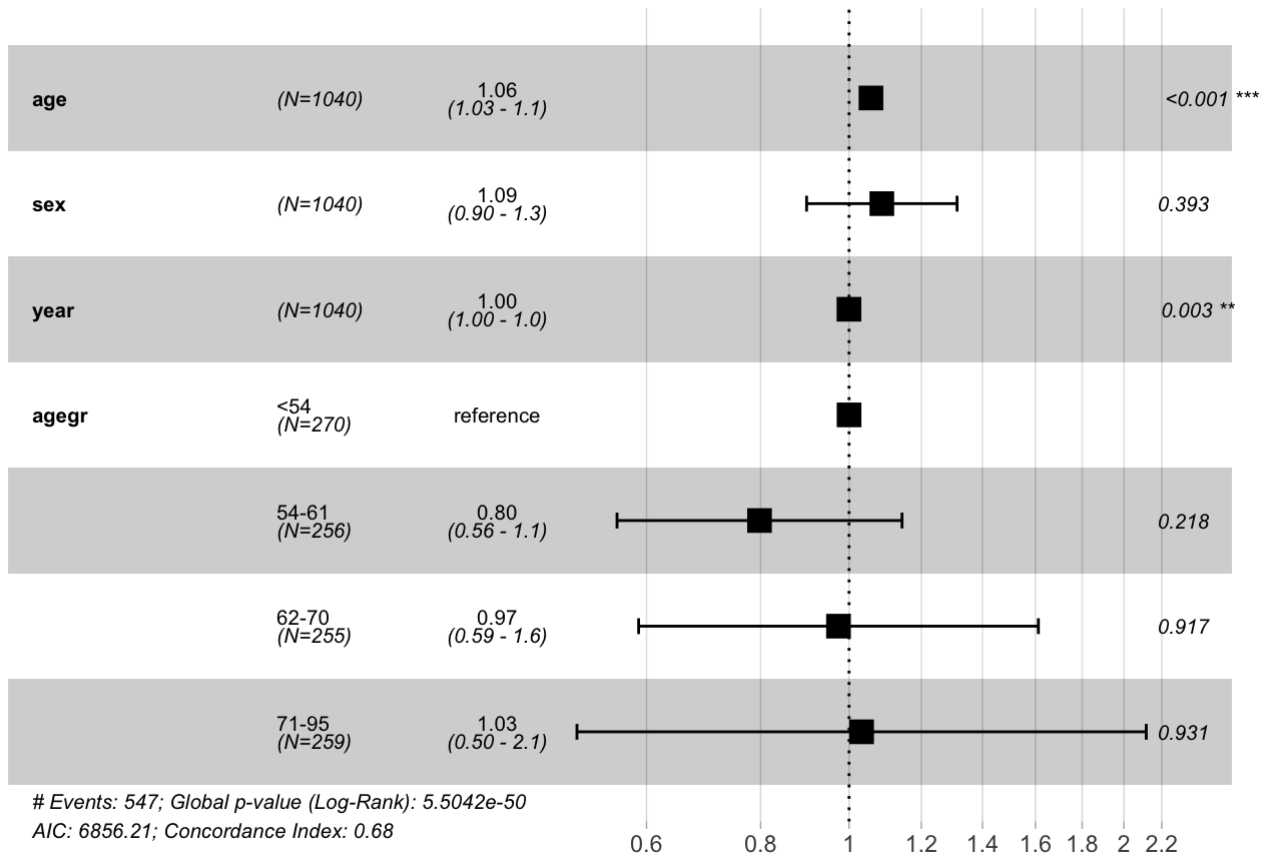
- being older is a bad prognostic factor
- being diagnosed later (i.e., more recently) is a good prognostic factor.

For example, holding the other covariates constant, being 1 year older at time of diagnosis increases the hazard by a factor of 1.0568, or 5.68%.

Forest plot:

```
# Forest plot
ggforest(fit_cox, data = dat)
```

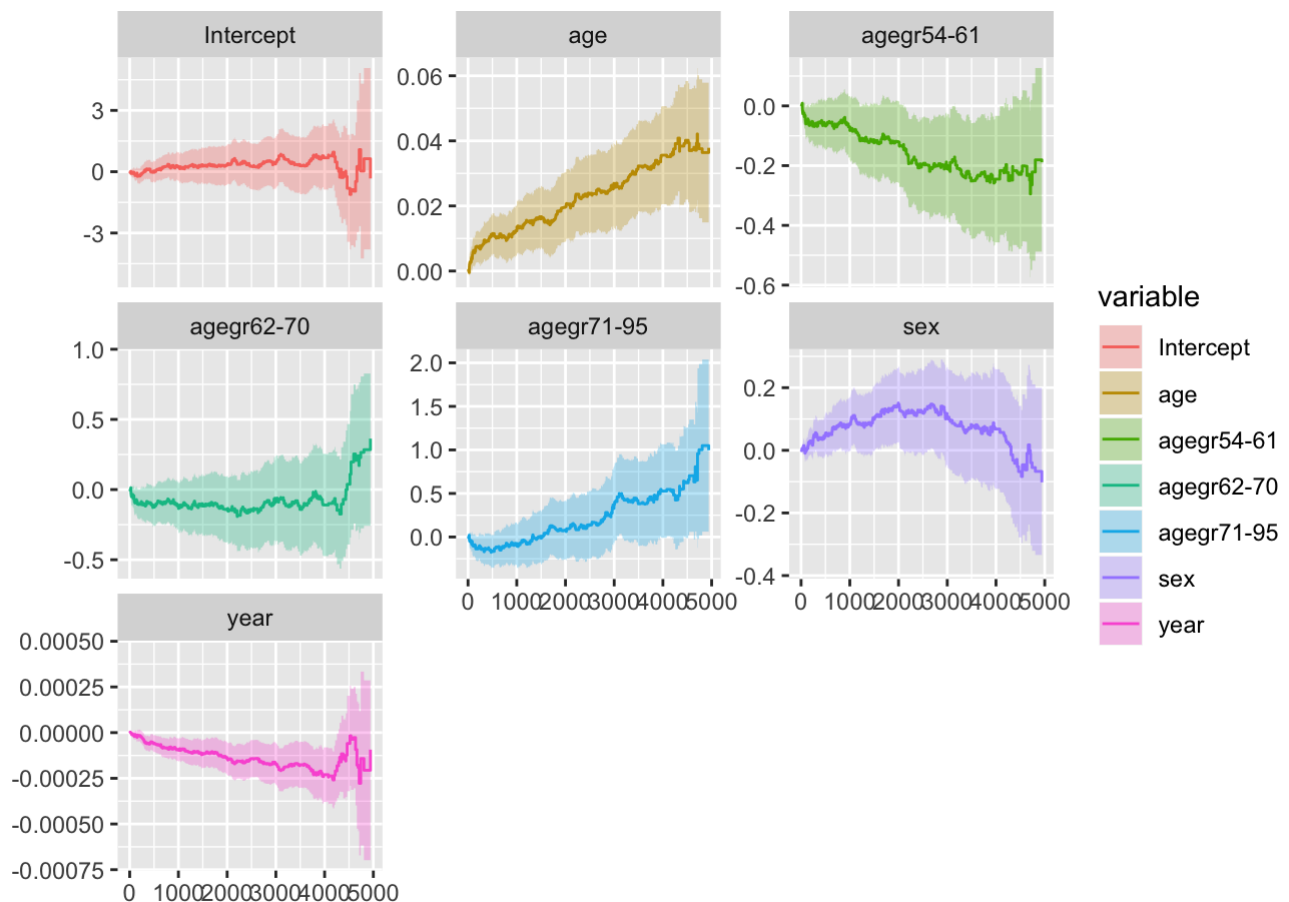
Hazard ratio



We now fit Aalen's additive regression model for censored data to the data. The documentation states: "The Aalen model assumes that the cumulative hazard $H(t)$ for a subject can be expressed as $a(t) + XB(t)$, where $a(t)$ is a time-dependent intercept term, X is the vector of covariates for the subject (possibly time-dependent), and $B(t)$ is a time-dependent matrix of coefficients."

```
aa_fit <- aareg(Surv(time, cens) ~ .,
               data = dat)

autoplot(aa_fit)
```



The plots above show how the effects of the covariates change over time.

1.2.6 Random forests model

As a final example, we now use the `ranger()` function to fit a Random Forests Ensemble model to the data. `ranger()` builds a model for each observation in the data set. The next block of code builds the model using the same variables used in the Cox model above, and plots 50 random curves, along with a curve that represents the global average for all of the patients.

```
# ranger model
r_fit <- ranger(Surv(time, cens) ~ .,
               data = dat,
               # mtry = 4,
               importance = "permutation",
               splitrule = "extratrees",
               verbose = TRUE)

r_fit
```

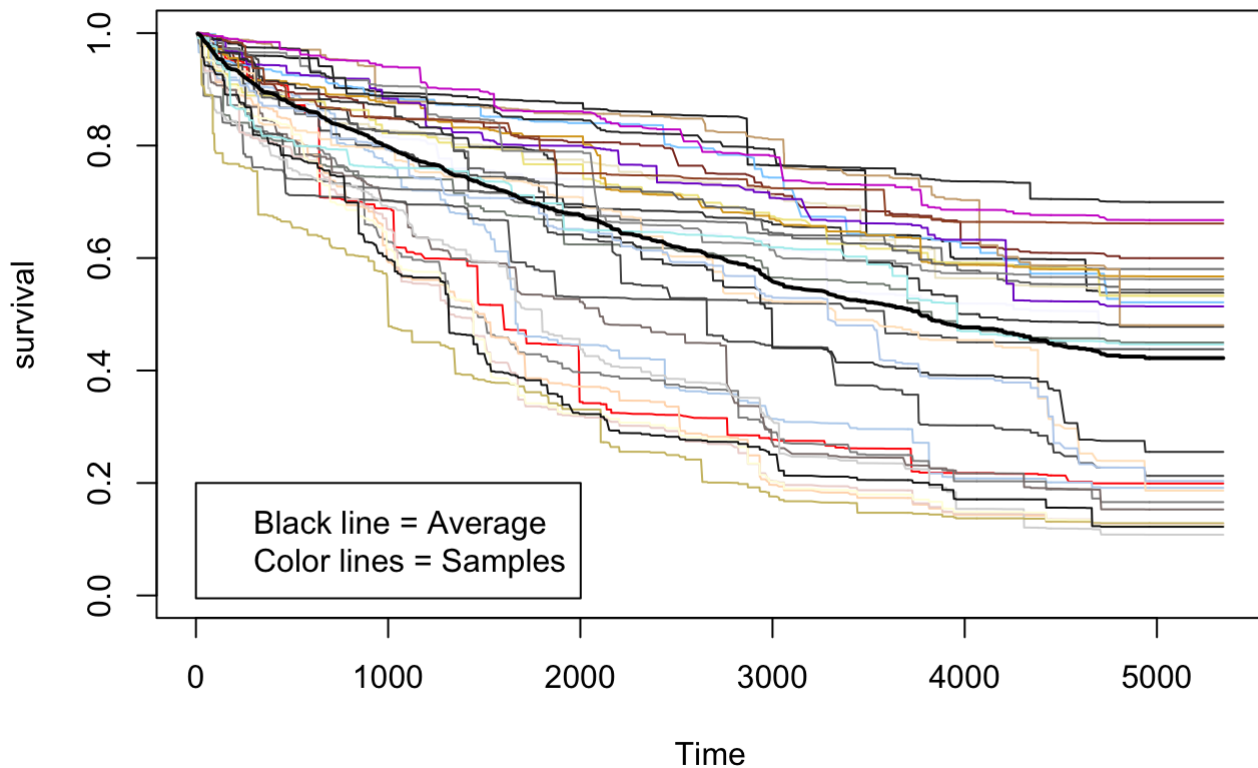
```
## Ranger result
##
## Call:
## ranger(Surv(time, cens) ~ ., data = dat, importance = "permutation",      splitru
le = "extratrees", verbose = TRUE)
##
## Type:                      Survival
## Number of trees:           500
## Sample size:               1040
## Number of independent variables: 4
## Mtry:                      2
## Target node size:          3
## Variable importance mode:   permutation
## Splitrule:                 extratrees
## Number of unique death times: 945
## OOB prediction error (1-C): 0.3285898
```

```
# Average the survival models
death_times <- r_fit$unique.death.times
surv_prob <- data.frame(r_fit$survival)
avg_prob <- sapply(surv_prob, mean)

# Plot the survival models for each patient
plot(r_fit$unique.death.times, r_fit$survival[1,],
     type = "l",
     ylim = c(0, 1),
     col = "red",
     xlab = "Time",
     ylab = "survival",
     main = "Patient Survival Curves")

#
cols <- colors()
for (n in sample(c(2:dim(dat)[1]), n_colors)){
  lines(r_fit$unique.death.times, r_fit$survival[n,], type = "l", col = cols[n])
}
lines(death_times, avg_prob, lwd = 2)
legend(0, 0.2, legend = c("Black line = Average", "Color lines = Samples"))
```

Patient Survival Curves



The next block of code illustrates how `ranger()` ranks variable importance.

```
vi <- data.frame(sort(round(r_fit$variable.importance, 4), decreasing = TRUE))
names(vi) <- "importance"
head(vi)
```

```
##      importance
## age      0.0685
## agegr    0.0616
## year     0.0069
## sex      0.0008
```

Notice that `ranger()` flags `age` as the most important; the same variable with the smallest p-value in the Cox model.

1.2.7 Comparison between KM, Cox and Random Forests models

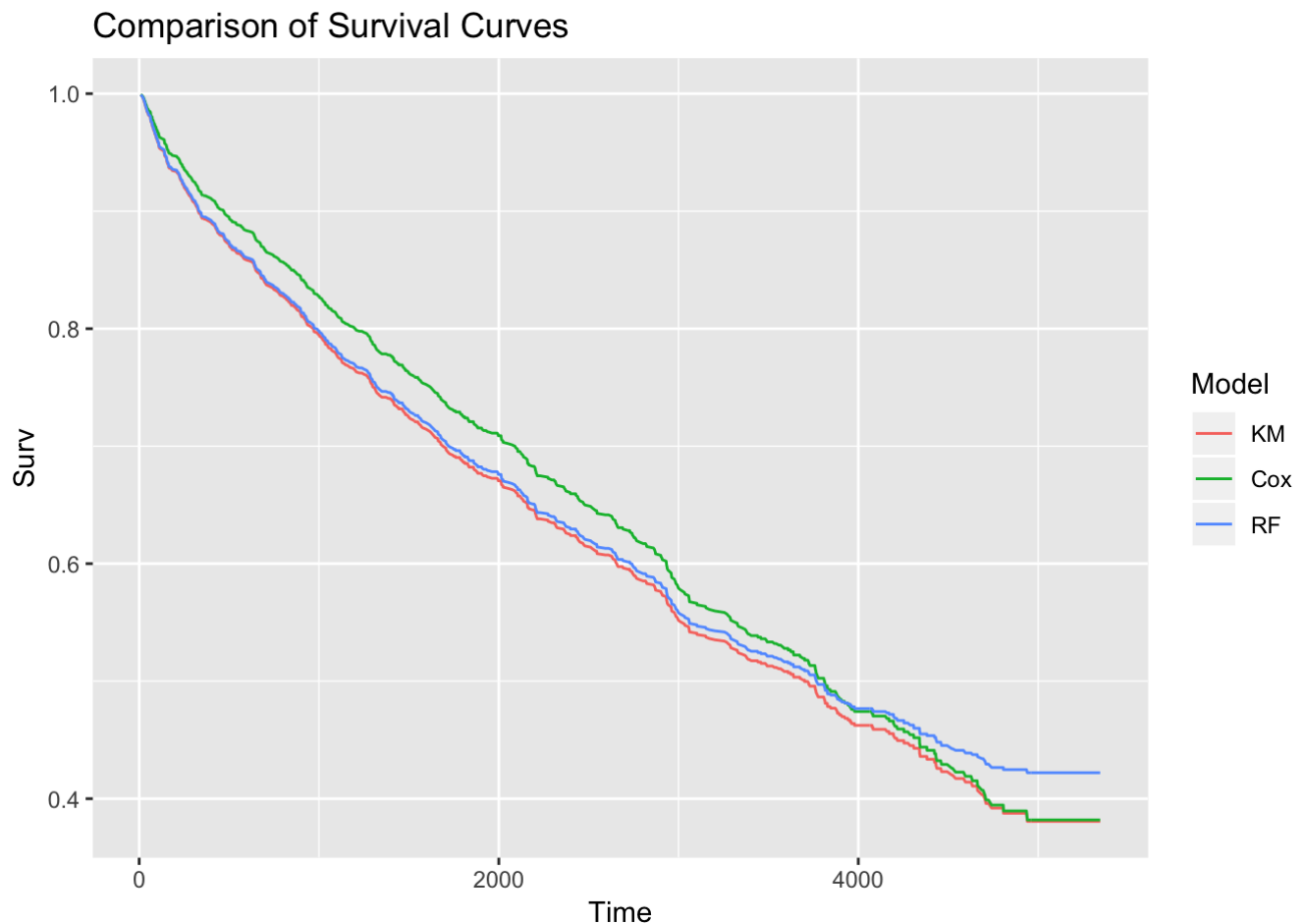
```
# Set up for ggplot
fit_km <- survfit(Surv(time, cens) ~ 1, data=dat)
km <- rep("KM", length(fit_km$time))
km_df <- data.frame(fit_km$time, fit_km$surv, km)
names(km_df) <- c("Time", "Surv", "Model")

fit_cox_comparison <- survfit(fit_cox)
cox <- rep("Cox", length(fit_cox_comparison$time))
cox_df <- data.frame(fit_cox_comparison$time, fit_cox_comparison$surv, cox)
names(cox_df) <- c("Time", "Surv", "Model")

rfi <- rep("RF", length(r_fit$unique.death.times))
rf_df <- data.frame(r_fit$unique.death.times, avg_prob, rfi)
names(rf_df) <- c("Time", "Surv", "Model")

plot_df <- rbind(km_df, cox_df, rf_df)

ggplot(plot_df, aes(x = Time, y = Surv, color = Model)) +
  geom_line() +
  ggtitle("Comparison of Survival Curves")
```



1.3 slopop

1.3.1 Description

Census data set for the Slovene population. This table corresponds to expected or population survival and is to be compared with observed survival (e.g., `colrec` and `rdata` tables).

1.3.2 Format

```
dat <- slopop  
is.ratetable(dat)
```

```
## [1] TRUE
```

```
summary(dat)
```

```
## Rate table with 3 dimensions:  
## age ranges from 0 to 37619.82; with 104 categories  
## year ranges from -10957 to 20454; with 41 categories  
## sex has levels of: male female
```

```
head(dat, n = 10)
```

```
## [1] 4.429994e-04 7.145309e-05 3.341402e-05 2.163265e-05 1.757899e-05  
## [6] 1.347629e-05 1.091866e-05 8.857789e-06 8.281024e-06 7.594557e-06
```

1.4 Belgian population tables

```
# Population tables  
belpop <- transrate.hld(c("BEL1994.txt", "BEL1995.txt", "BEL1996.txt", "BEL1997.txt", "BEL1998.txt", "BEL1999.txt", "BEL2000.txt", "BEL2001.txt", "BEL2002.txt", "BEL2003.txt", "BEL2004.txt", "BEL2005.txt", "BEL2006.txt", "BEL2007.txt", "BEL2008.txt", "BEL2009.txt", "BEL2010.txt", "BEL2011.txt", "BEL2012.txt"))  
  
is.ratetable(belpop)
```

```
## [1] TRUE
```

```
summary(belpop)
```

```
## Rate table with 3 dimensions:  
## age ranges from 0 to 38350.31; with 106 categories  
## year ranges from 12419 to 18993; with 19 categories  
## sex has levels of: male female
```

```
head(belpop, n = 10)
```

```
## [1] 2.248827e-05 1.999410e-06 1.067996e-06 6.024083e-07 6.571792e-07  
## [6] 5.750232e-07 3.833354e-07 7.119513e-07 3.559525e-07 4.654857e-07
```

```
#attributes(belpop)
```

2 Data from the Federal Planning Bureau

2.1 Description

The file can be downloaded from the following link (<https://www.plan.be/databases/data-50-en-prospective+mortality+rates+2017+2070>). The data contained in the file are the results of the projection by gender and unisex of the mortality rates (q-type rate) for Belgium made within the framework of the population projections 2017-2070 (according to the method described in WP 18-09).

The resulting transversal and longitudinal life expectancies are also included in the file.¹

2.2 Format

The file is divided in several sheets:

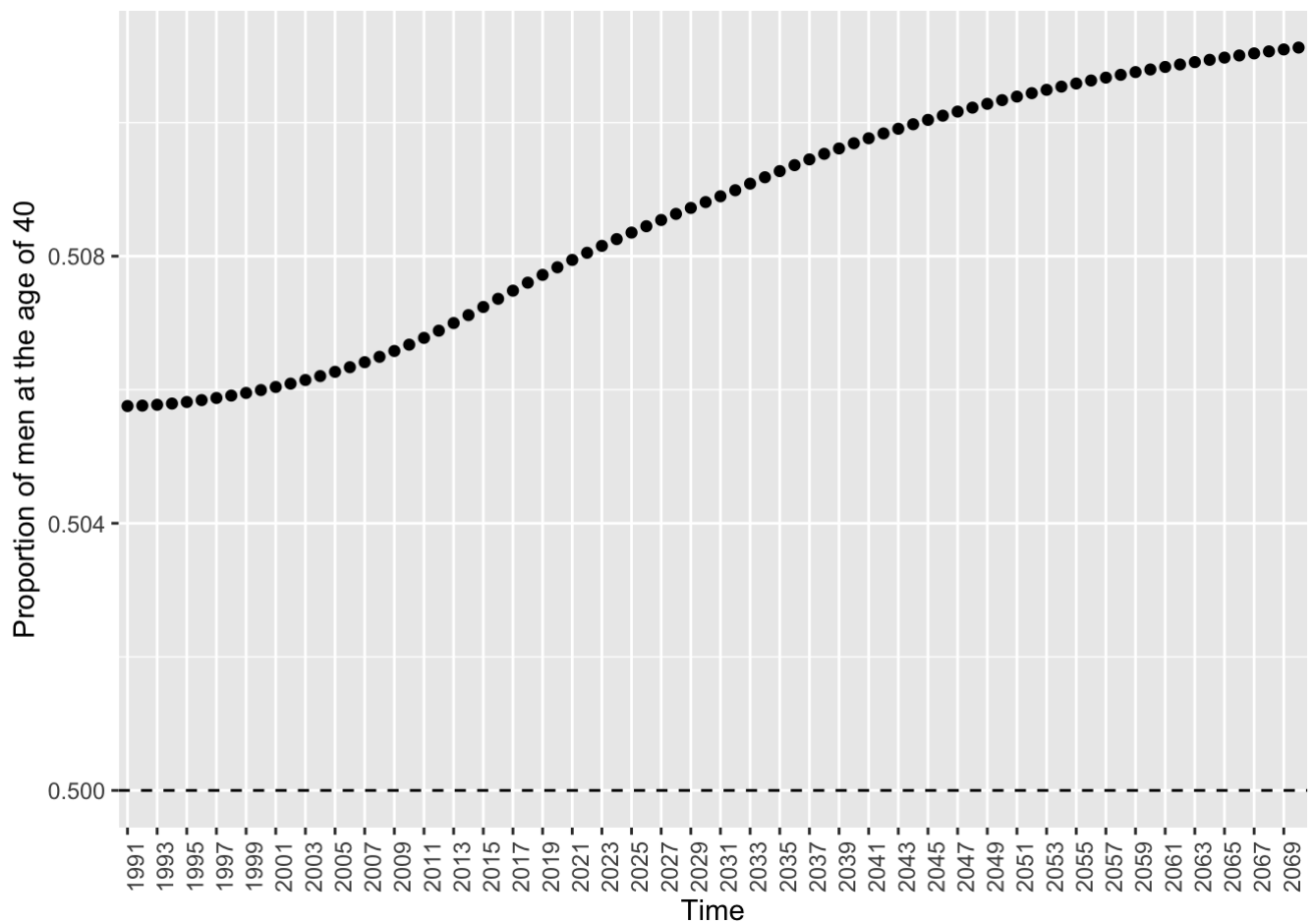
- The sheets QxObs-{F,M,U} contain observed mortality rates (q-type rate) in complete years for women (F), men (M) and unisex (U) from 1991 to 2016
- the sheets ExObs-{F,M,U} contain observed transversal life expectancy for women (F), men (M) and unisex (U) from 1991 to 2016
- the two next sheets (QxCalc-F et QxCalc-M) present the calculated α_x and β_x as well as mortality rates (q-type rate) projected by gender between 1991 and 2070
- the sheet KxCalc shows the proportion of men in the population between 1991 and 2070 that was calculated according to the methodology described in section 4.3 of WP 18-09
- the sheet QxCalc-U contains the results of the unisex projection calculated using the proportional method between 1991 and 2070
- the sheets ExCalc-{F,M,U} present the transversal life expectancy that was projected for women (F), men (M) and unisex (U) between 1991 and 2070
- the sheets EGxCalc-{F,M,U} present the longitudinal life expectancy that was projected for women (F), men (M) and unisex (U) between 1991 and 2070

Each sheet covers the same age range: birth to 120 years.

2.3 Descriptive statistics

2.3.1 Proportion of men in the population

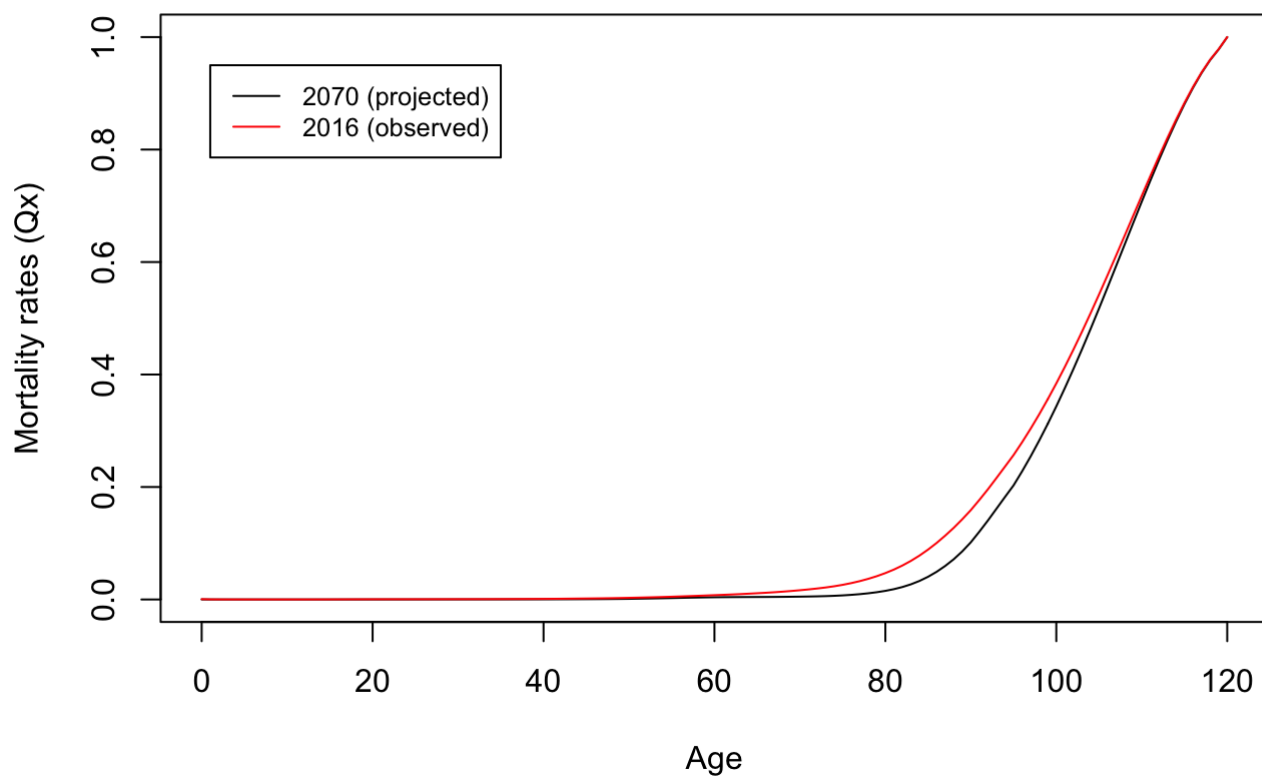
```
# Evolution of the proportion of men at the age of 40 years old
ggplot(KxCalc_40yrs, aes(y = prop_men, x = KxCalc_40yrs$years)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_x_discrete(breaks = seq(min(KxCalc_40yrs$years), max(KxCalc_40yrs$years), by
= 2)) +
  labs(y = "Proportion of men at the age of 40", x = "Time") +
  geom_hline(yintercept=0.5, linetype="dashed")
```



Note that points between 1991 and 2016 are observed proportions and points between 2017 and 2070 are estimated proportions.

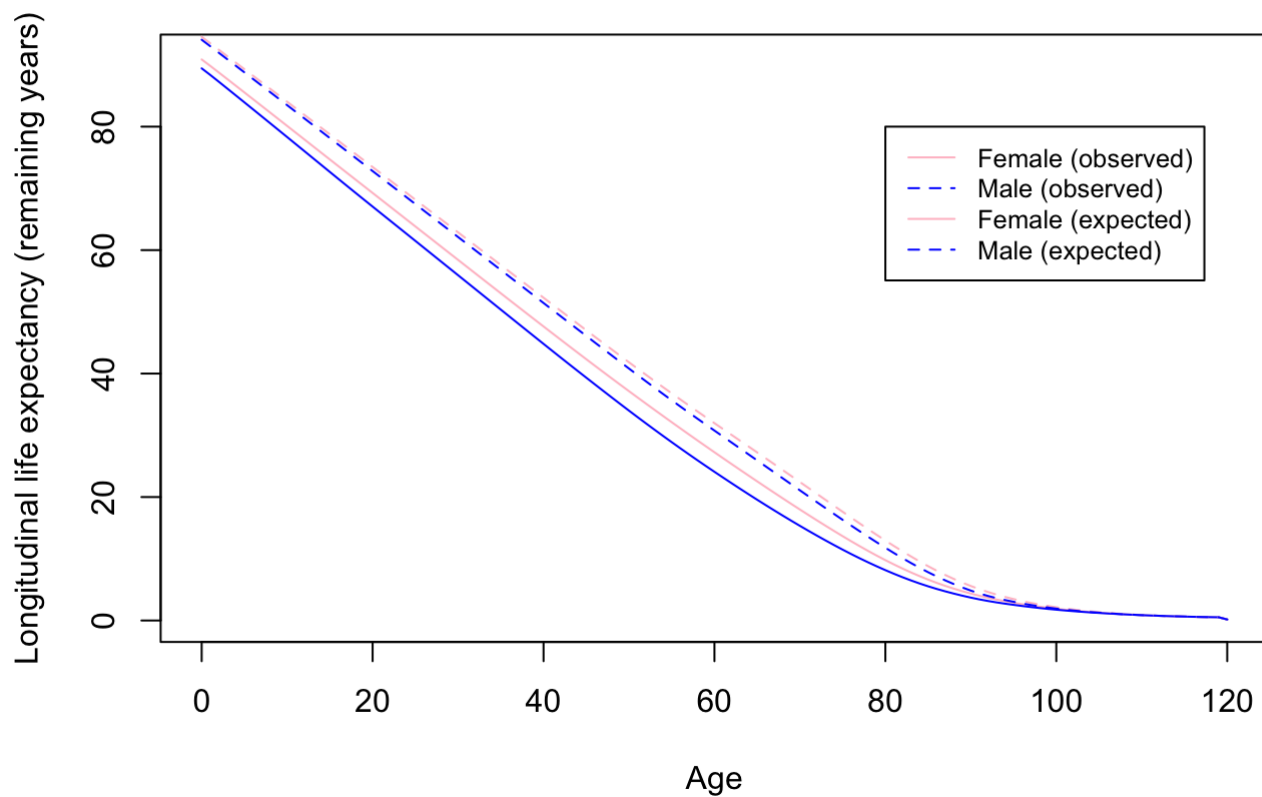
2.3.2 Mortality rates

```
plot(x = QxCalc_U$X__1, y = QxCalc_U$`2070`, type = "line", ylab = "Mortality rates
(Qx)", xlab = "Age")
lines(x = QxCalc_U$X__1, y = QxCalc_U$`2016`, col = "red")
legend(1, 0.95, legend=c("2070 (projected)", "2016 (observed)"),
      col=c("black", "red"), lty = 1, cex = 0.8)
```

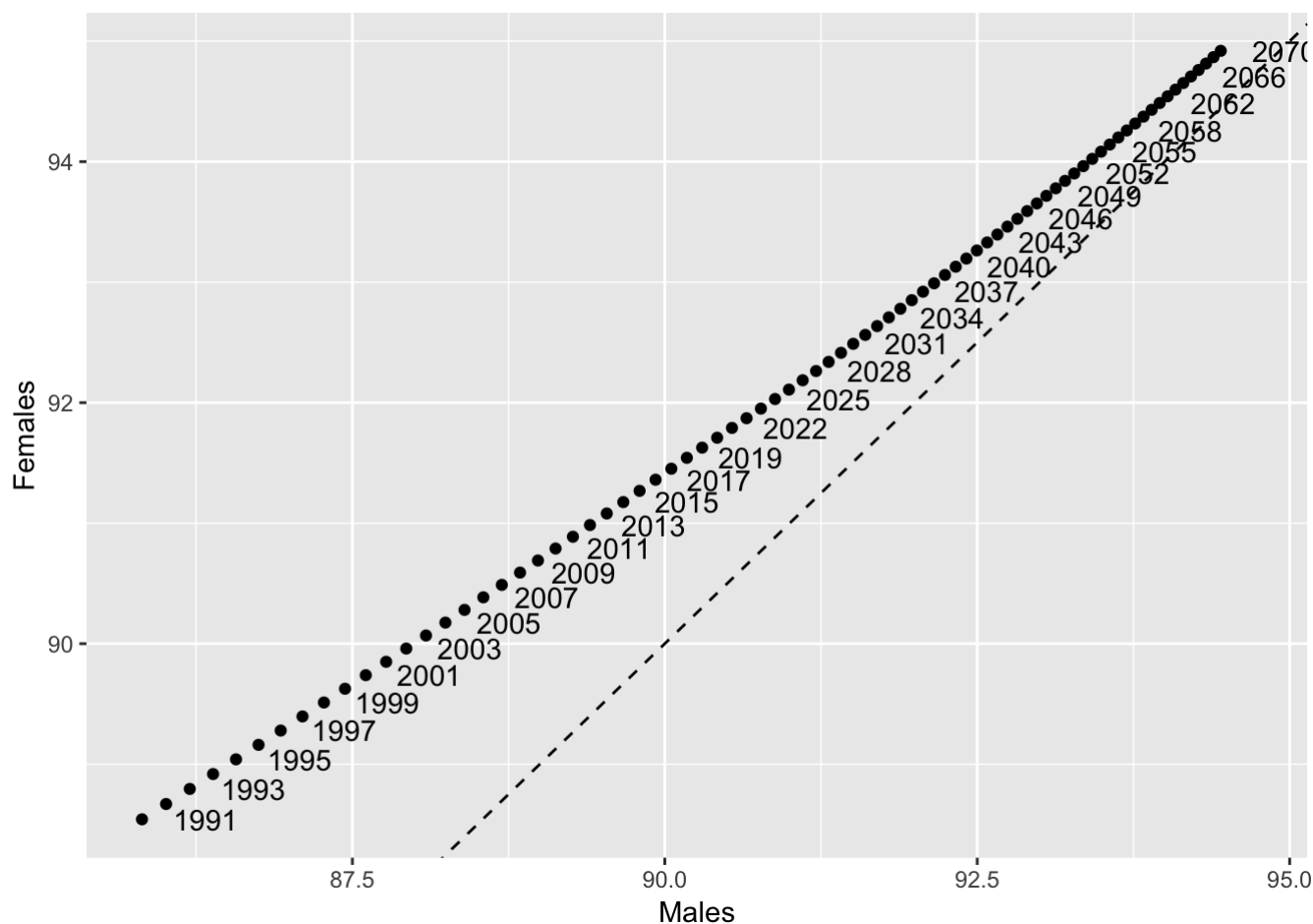


2.3.3 Life expectancy

```
plot(x = EGxCalc_F$X__1, y = EGxCalc_F$`2016`, type = "line", ylab = "Longitudinal li
fe expectancy (remaining years)", xlab = "Age", col = "pink")
lines(x = EGxCalc_M$X__1, y = EGxCalc_M$`2016`, col = "blue")
lines(x = EGxCalc_F$X__1, y = EGxCalc_F$`2070`, col = "pink", lty = 2)
lines(x = EGxCalc_M$X__1, y = EGxCalc_M$`2070`, col = "blue", lty = 2)
legend(80,80, legend=c("Female (observed)", "Male (observed)", "Female (expected)",
"Male (expected)"),
      col=c("pink", "blue"), lty = c(1,2), cex = 0.8)
```



```
# Life expectancy: differences between males and females
ggplot(life_exp_at_birth, aes(x=Males, y=Females, label = rownames(life_exp_at_birt
h))) +
  geom_point() +
  geom_abline(intercept=0, slope = 1, linetype = 2) +
  geom_text(hjust = 0, nudge_x = 0.25, check_overlap = TRUE)
```



Note again that points between 1991 and 2016 are observed life expectancies and points between 2017 and 2070 are estimated life expectancies.

3 Methods

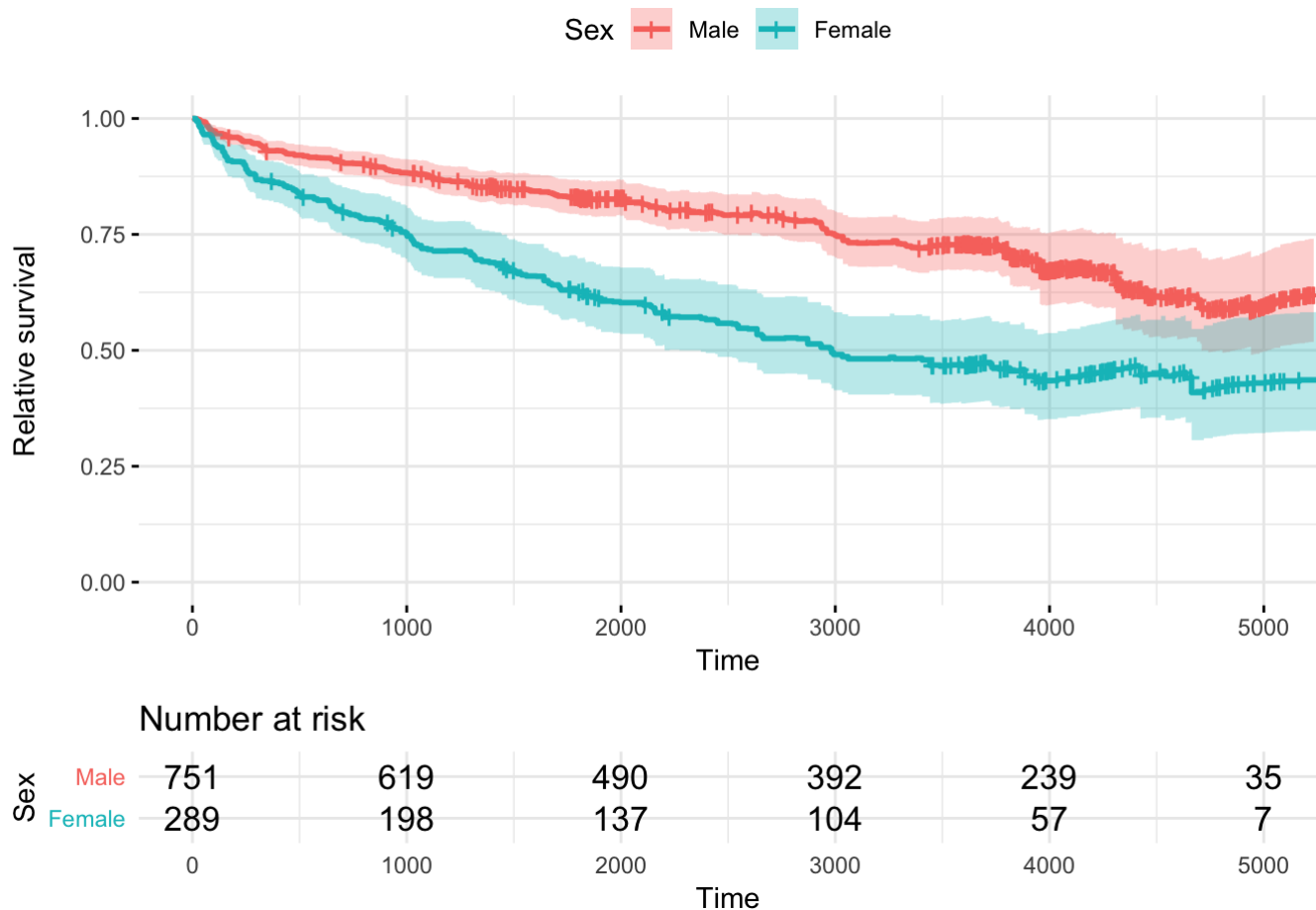
3.1 Relative survival

```
#calculate the relative survival curve
#note that the variable year is in the date format - the number of days since 01.01.1
960 and that age must be multiplied by 365.241 in order to be expressed in days.
```

```
fit <- rs.surv(Surv(time,cens)~sex+ratetable(age=age*365.241,sex=sex,year=year),ratet
able=belpop,data=rdata)
fit
```

```
## Call: rs.surv(formula = Surv(time, cens) ~ sex + ratetable(age = age *
##      365.241, sex = sex, year = year), data = rdata, ratetable = belpop)
##
##           n events median 0.95LCL 0.95UCL
## sex=1 751      360      NA      4710      NA
## sex=2 289      187    2983      2273      NA
```

```
ggsurvplot(fit,
  conf.int=TRUE,
  risk.table=TRUE,
  legend.labs=c("Male", "Female"),
  legend.title="Sex",
  #palette=c("dodgerblue2", "orchid2"),
  main="Relative survival by sex for rdata",
  ylab="Relative survival",
  ggtheme = theme_minimal())
```

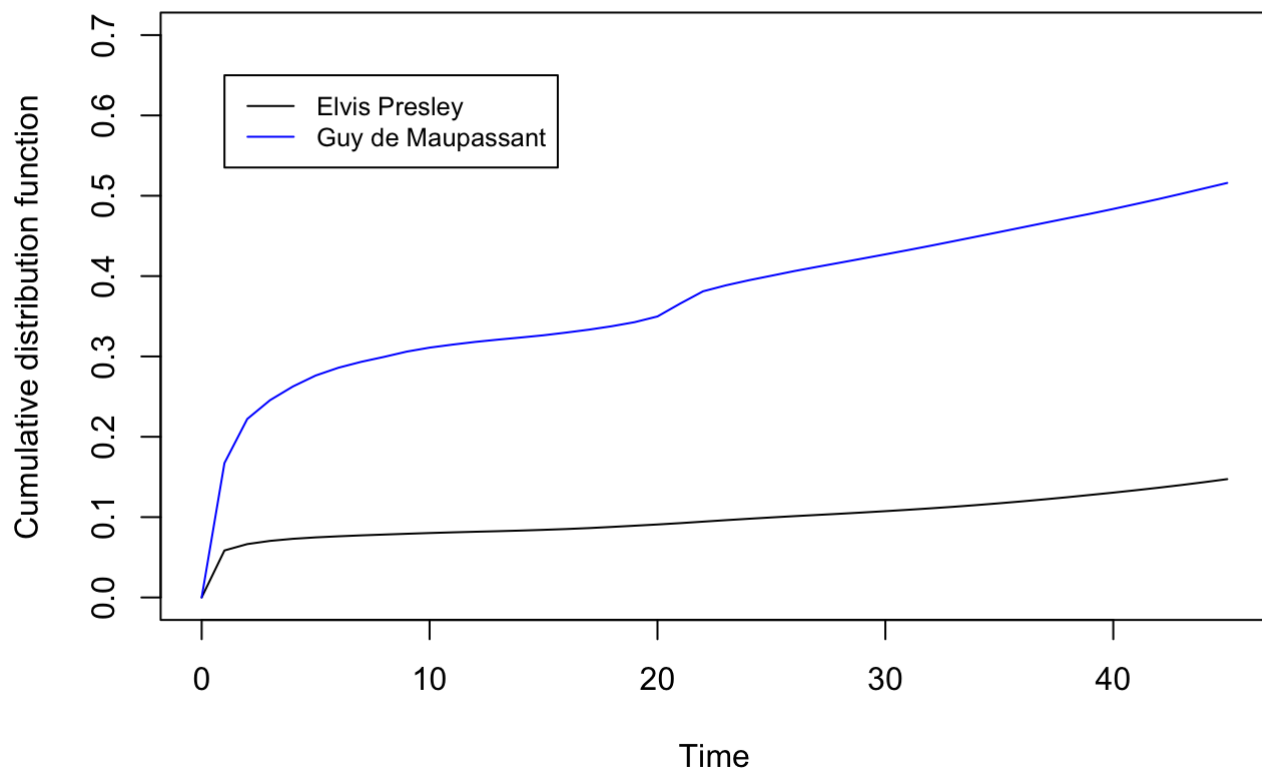


3.1.1 Individual relative survival

```
# Individual relative survival of Elvis Presley and Guy de Maupassant
# Elvis died in 1977 at the age of 42.6 (average was 69.4 at that time in the US), and
# Guy died in 1983 at the age of 42.9 (average was 42.03 at that time in France)
frpop <- transrate.hld("FRA18061997.txt")
uspop <- transrate.hld("USA19011999.txt")

elvispresley <- 1-survexp(~ratetable(age = 0,sex = "male",year = as.date("8Jan1935"
)), times = (0:45)*365.24,ratetable = uspop)$surv
guydemaupassant <- 1-survexp(~ratetable(age = 0,sex = "male",year = as.date("5Aug185
0")), times = (0:45)*365.24,ratetable = frpop)$surv

plot(x = 0:45,y = elvispresley,type = "l", ylim = c(0,0.7), ylab = "Cumulative distri
bution function", xlab = "Time")
lines(x = 0:45,y = guydemaupassant,type = "l", col="blue")
legend(1, 0.65, legend=c("Elvis Presley", "Guy de Maupassant"),
  col=c("black", "blue"), lty = 1, cex = 0.8)
```



This figure shows the cumulative distribution functions for lifetime of men born in France in 1850 (Maupassant) and in USA in 1935 (Elvis).

3.2 Esteve et al. (1990) model

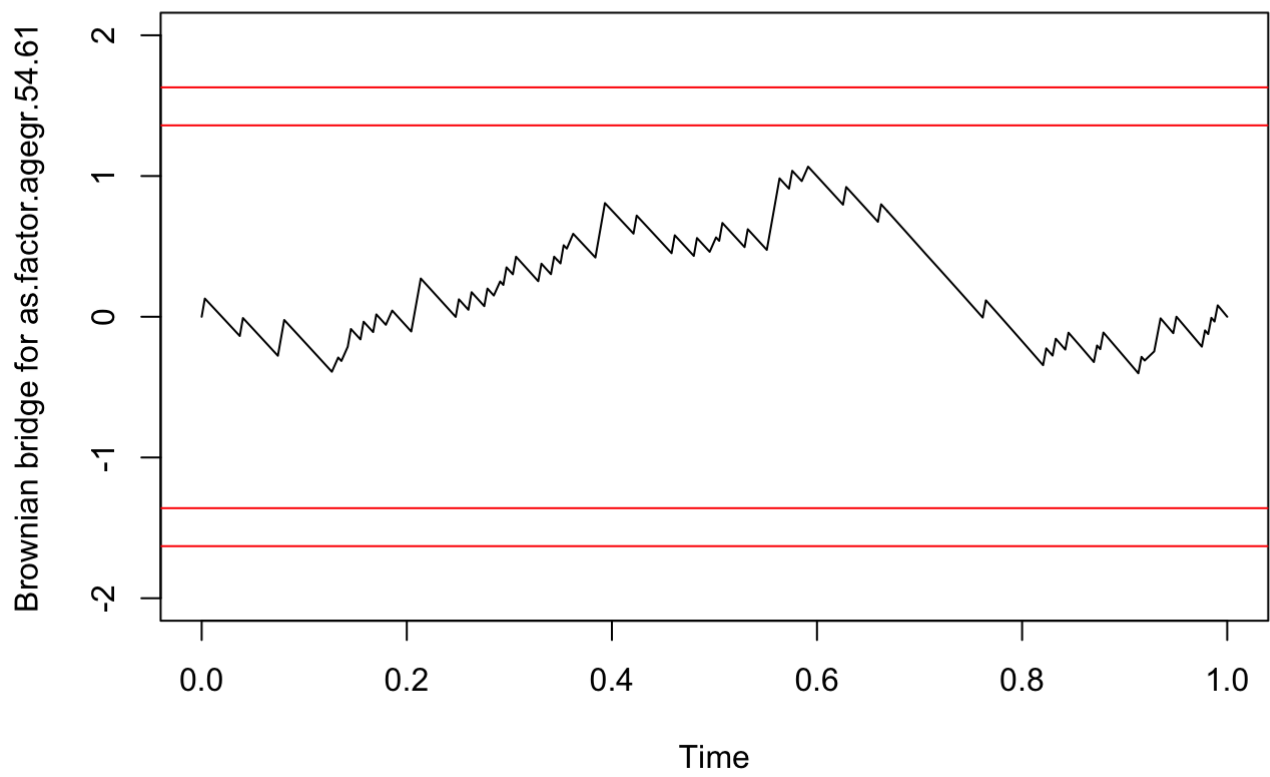
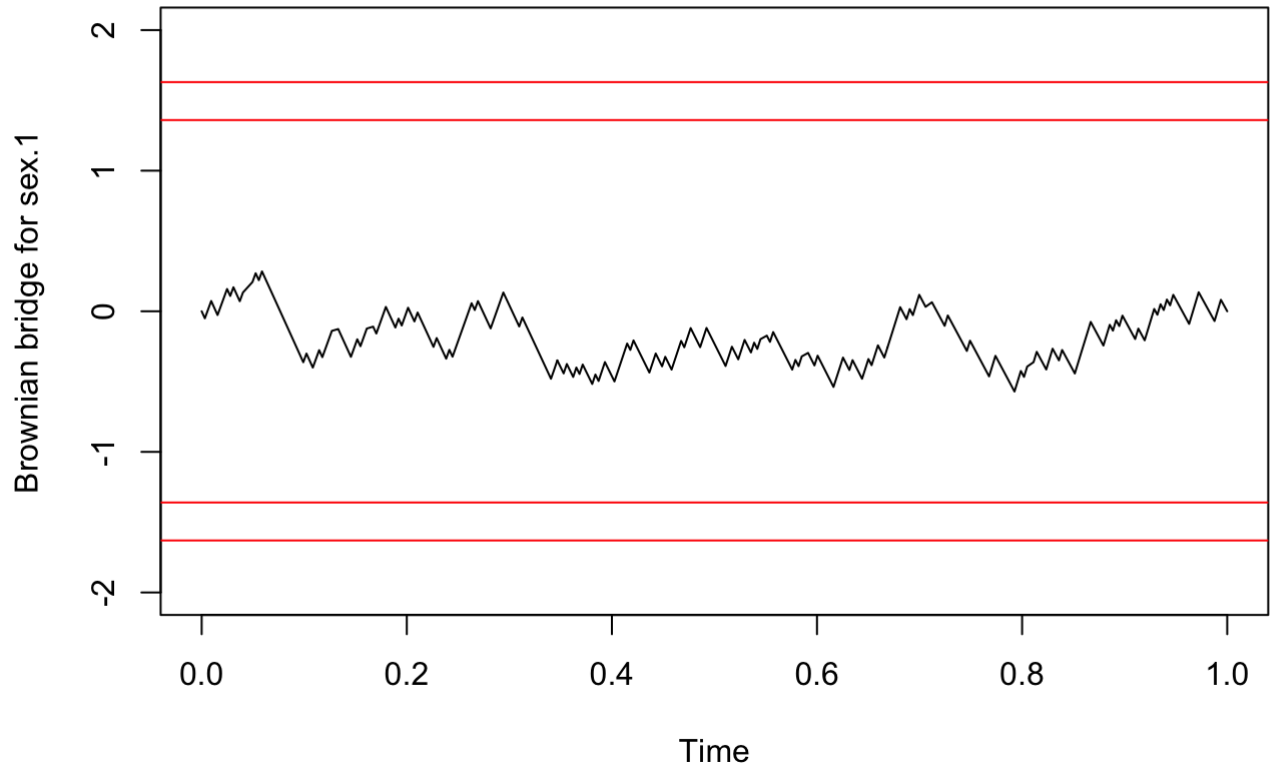
```
data(slopop)
data(rdata)

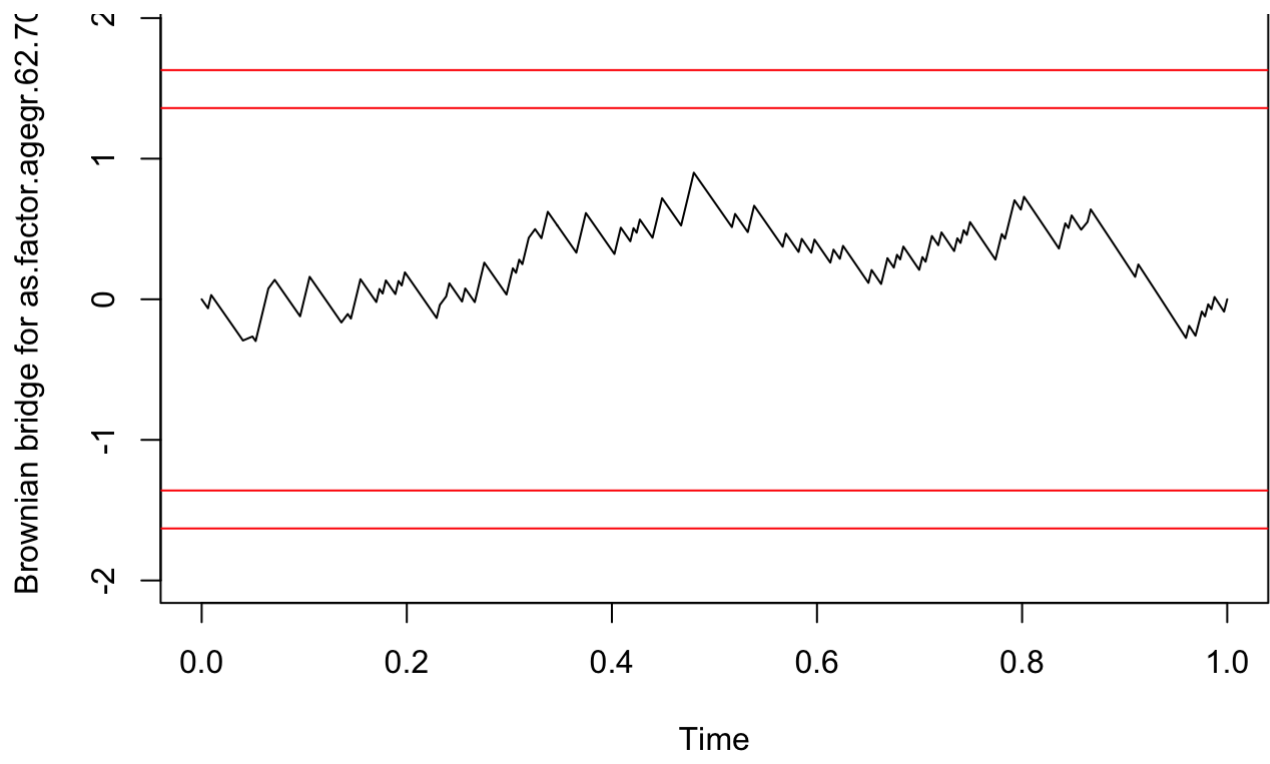
#fit an additive model
fit <- rsadd(Surv(time,cens)~sex+as.factor(agegr)+ratetable(age=age*365.241,sex=sex,year=year),ratetable=belpop,data=rdata,int=5)
summary(fit)
```

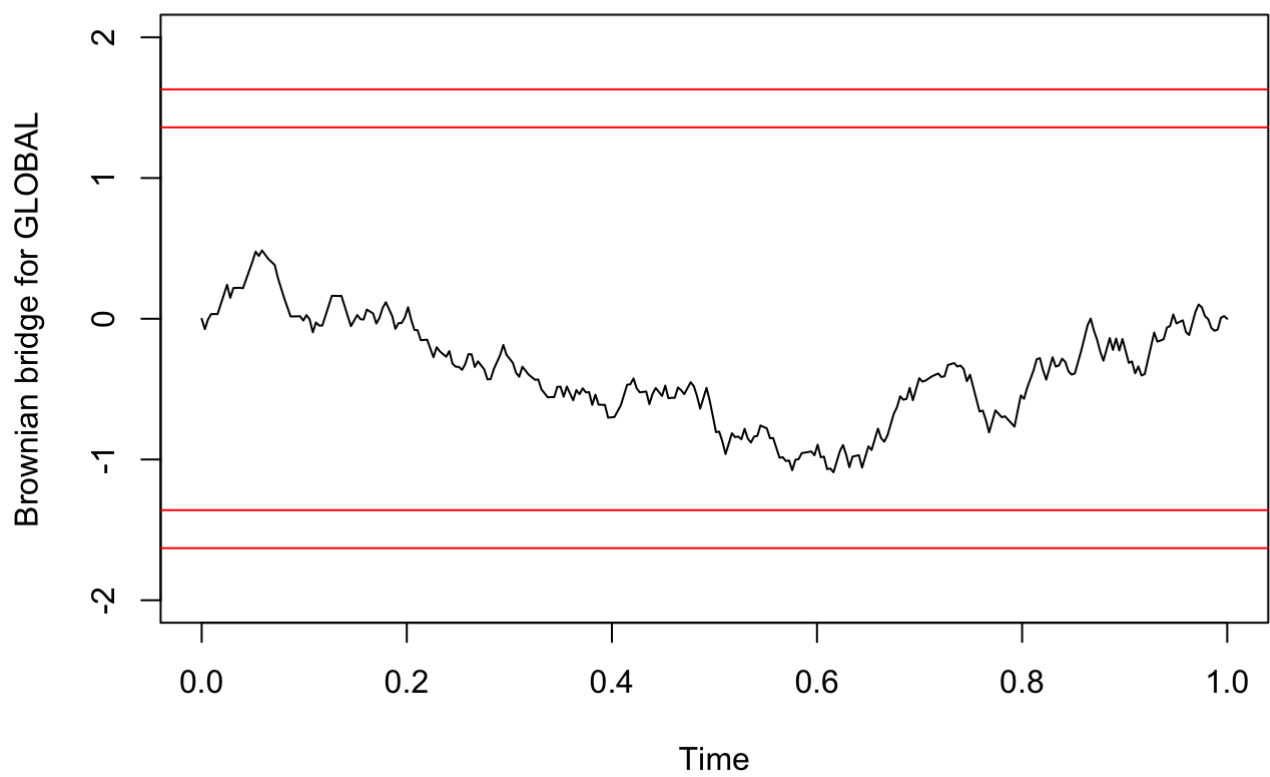
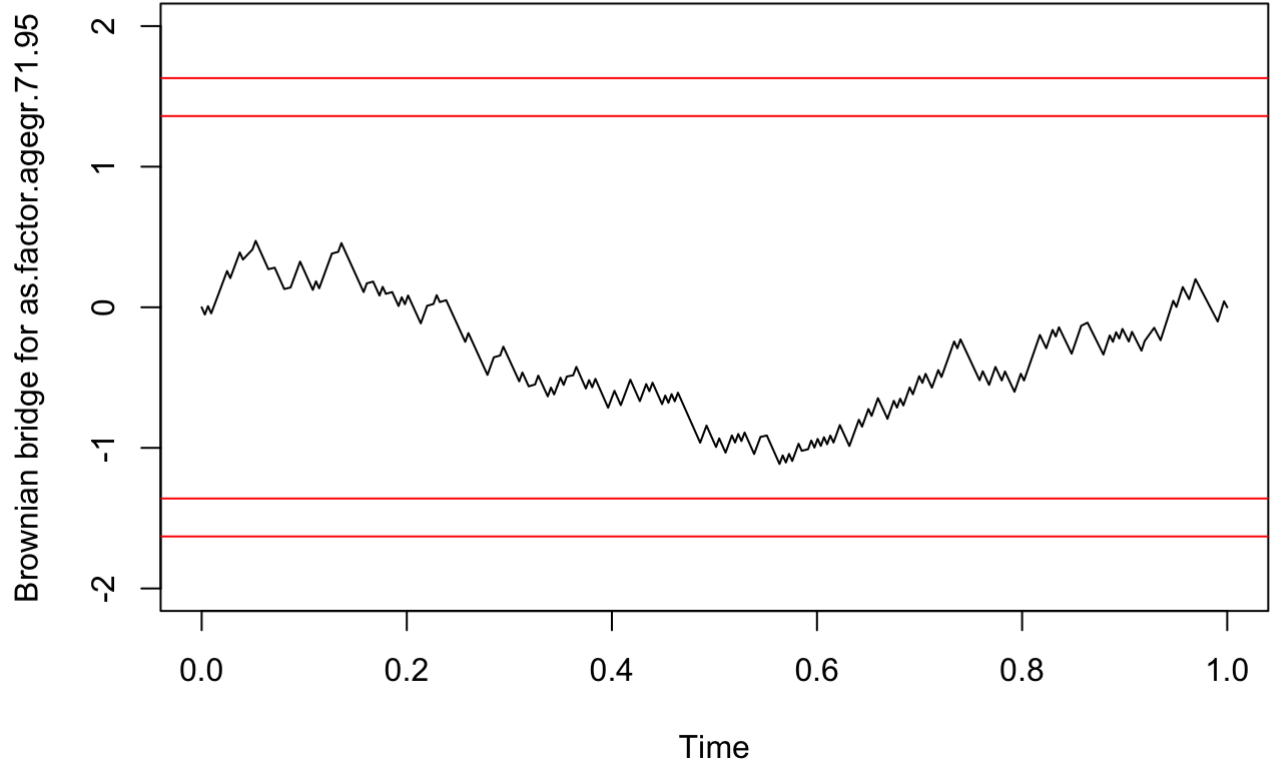
```
##
## Call:
## rsadd(formula = Surv(time, cens) ~ sex + as.factor(agegr) + ratetable(age = age *
##      365.241, sex = sex, year = year), data = rdata, ratetable = belpop,
##      int = 5)
##
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## sex.1            0.6486    0.1702   3.810 0.000139 ***
## as.factor.agegr.54.61 0.1806    0.2535   0.712 0.476343
## as.factor.agegr.62.70 0.5706    0.2389   2.388 0.016924 *
## as.factor.agegr.71.95 0.8608    0.2438   3.530 0.000415 ***
## fu [0,1)         -3.7588    0.2943 -12.772 < 2e-16 ***
## fu [1,2)         -4.4217    0.3245 -13.628 < 2e-16 ***
## fu [2,3)         -4.4074    0.3293 -13.384 < 2e-16 ***
## fu [3,4)         -4.6530    0.3740 -12.442 < 2e-16 ***
## fu [4,5)         -4.4200    0.3409 -12.966 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The positive sex coefficient (0.649) implies that the survival of men is relatively better. Moreover, only the two oldest age groups differ significantly from the youngest.

```
#check the goodness of fit
rsbr <- rs.br(fit)
plot(rsbr)
```





The processes for sex and all age groups lie within the 95 and 99 per cent critical values indicating a good fit (see Stare, Pohar, and Henderson 2005).

3.3 Poisson (Denuit and Legrand 2018)

3.4 Boussari et al. (2018)

```
data(brcancer)
```

```
summary(fit <- stpm2(Surv(rectime,censrec==1)~hormon,data=brcancer,df=3))
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = negll, start = coef, eval.only = TRUE, vecpar = TRUE,
##      gr = function (beta)
##      {
##          localargs <- args
##          localargs$init <- beta
##          localargs$return_type <- "gradient"
##          return(.Call("model_output", localargs, PACKAGE = "rstpm2"))
##      }, control = list(parscale = c(`(Intercept)` = 1, hormon = 1,
##      `nsx(log(rectime), df = 3)1` = 1, `nsx(log(rectime), df = 3)2` = 1,
##      `nsx(log(rectime), df = 3)3` = 1), maxit = 300), lower = -Inf,
##      upper = Inf)
##
## Coefficients:
##              Estimate Std. Error  z value    Pr(z)
## (Intercept)    -7.25672    0.67302 -10.7823 < 2.2e-16 ***
## hormon          -0.36140    0.12488  -2.8940  0.003804 **
## nsx(log(rectime), df = 3)1  4.76100    0.41836  11.3803 < 2.2e-16 ***
## nsx(log(rectime), df = 3)2 11.57138    1.30800   8.8466 < 2.2e-16 ***
## nsx(log(rectime), df = 3)3  4.56781    0.30890  14.7872 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 5215.915
```

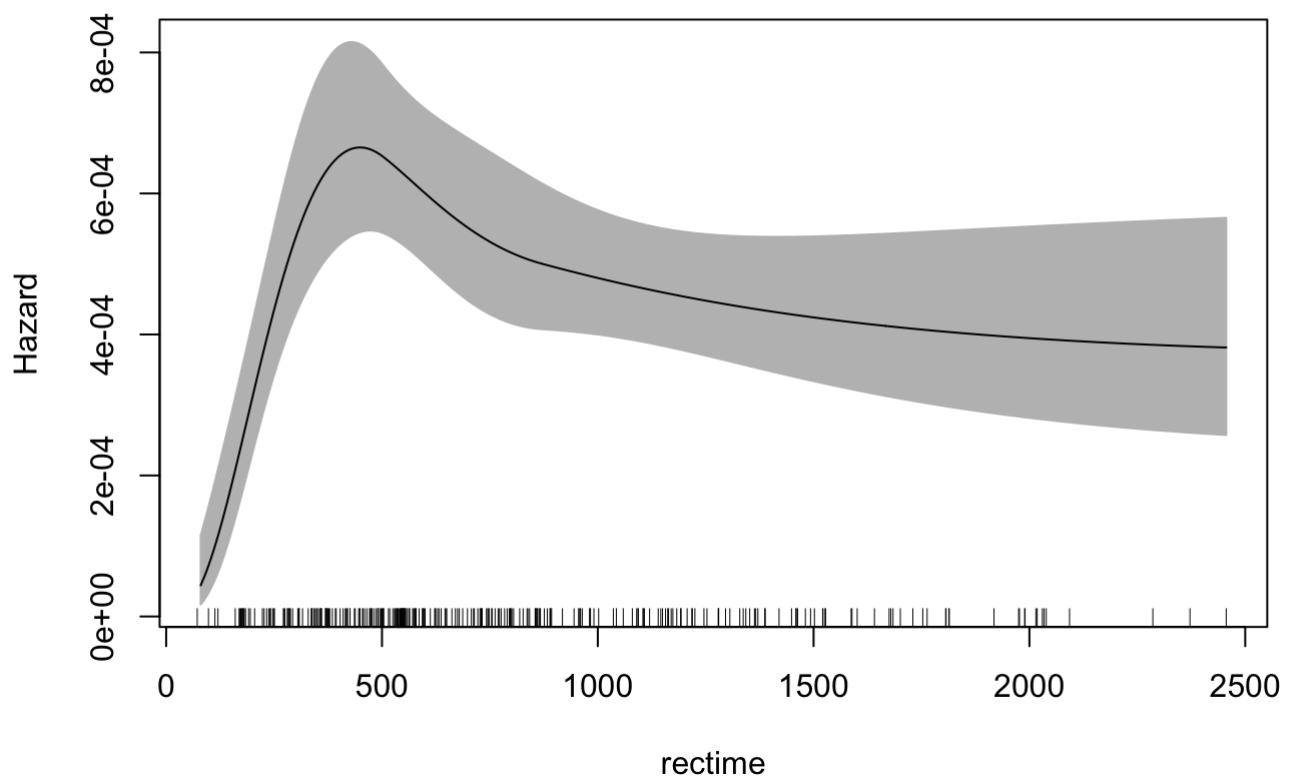
```
## some predictions
head(predict(fit,se.fit=TRUE,type="surv")) # predict survival probabilities
```

```
##      Estimate      lower      upper
## 1 0.4414689 0.3893046 0.4923020
## 2 0.5345699 0.4629419 0.6009135
## 3 0.8016169 0.7578334 0.8383377
## 4 0.5668354 0.4986300 0.6293291
## 5 0.7050672 0.6644352 0.7417639
## 6 0.8553697 0.8252622 0.8806682
```

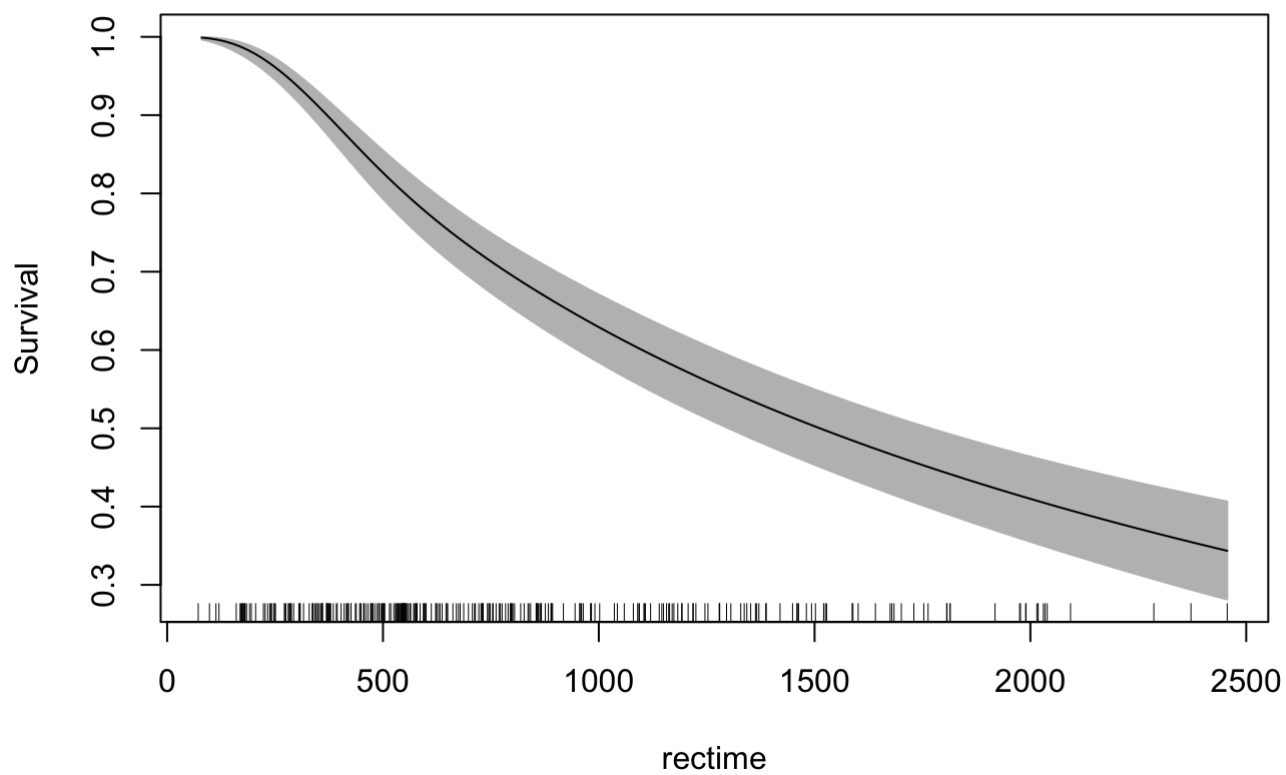
```
head(predict(fit,se.fit=TRUE,type="hazard")) # predict hazard
```

##	Estimate	lower	upper
## 1	0.0004034361	0.0002974048	0.0005472698
## 2	0.0002745182	0.0001918477	0.0003928129
## 3	0.0003803607	0.0002931354	0.0004935409
## 4	0.0002813338	0.0002031753	0.0003895587
## 5	0.0005241065	0.0004222318	0.0006505613
## 6	0.0006652879	0.0005449594	0.0008121854

```
## some plots
plot(fit,newdata=data.frame(hormon=0),type="hazard")
```



```
plot(fit,newdata=data.frame(hormon=0),type="surv")
```



```
## the same model using logH.formula  
summary(stpm2(Surv(rectime,censrec==1)~hormon,data=brcancer,logH.formula=~ns(log(rectime),df=3)))
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = negll, start = coef, eval.only = TRUE, vecpar = TRUE,
##      gr = function (beta)
##      {
##          localargs <- args
##          localargs$init <- beta
##          localargs$return_type <- "gradient"
##          return(.Call("model_output", localargs, PACKAGE = "rstpm2"))
##      }, control = list(parscale = c(`(Intercept)` = 1, hormon = 1,
##      `ns(log(rectime), df = 3)1` = 1, `ns(log(rectime), df = 3)2` = 1,
##      `ns(log(rectime), df = 3)3` = 1), maxit = 300), lower = -Inf,
##      upper = Inf)
##
## Coefficients:
##
##              Estimate Std. Error  z value    Pr(z)
## (Intercept)      -7.25672    0.67302 -10.7823 < 2.2e-16 ***
## hormon           -0.36140    0.12488  -2.8940  0.003804 **
## ns(log(rectime), df = 3)1  4.76100    0.41836  11.3803 < 2.2e-16 ***
## ns(log(rectime), df = 3)2 11.57138    1.30800   8.8466 < 2.2e-16 ***
## ns(log(rectime), df = 3)3  4.56781    0.30890  14.7872 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 5215.915
```

```
## time-varying coefficient
summary(fit.tvc <- stpm2(Surv(rectime,censrec==1)~hormon,data=brcancer,df=3,
                        tvc=list(hormon=3)))
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = negll, start = coef, eval.only = TRUE, vecpar = TRUE,
##      gr = function (beta)
##      {
##          localargs <- args
##          localargs$init <- beta
##          localargs$return_type <- "gradient"
##          return(.Call("model_output", localargs, PACKAGE = "rstpm2"))
##      }, control = list(parscale = c(`(Intercept)` = 1, hormon = 1,
##      `nsx(log(rectime), df = 3)1` = 1, `nsx(log(rectime), df = 3)2` = 1,
##      `nsx(log(rectime), df = 3)3` = 1, `hormon:nsx(log(rectime), df = 3)1` = 1,
##      `hormon:nsx(log(rectime), df = 3)2` = 1, `hormon:nsx(log(rectime), df = 3)3` =
1
##      ), maxit = 300), lower = -Inf, upper = Inf)
##
## Coefficients:
##
##                                     Estimate Std. Error z value      Pr(z)
## (Intercept)                        -7.03607    0.75318  -9.3418 < 2.2e-16
## hormon                           -1.30956    1.65722  -0.7902    0.4294
## nsx(log(rectime), df = 3)1         4.61694    0.46884   9.8477 < 2.2e-16
## nsx(log(rectime), df = 3)2        11.16932    1.46730   7.6122 2.695e-14
## nsx(log(rectime), df = 3)3         4.43796    0.34946  12.6994 < 2.2e-16
## hormon:nsx(log(rectime), df = 3)1  0.61780    1.03133   0.5990    0.5491
## hormon:nsx(log(rectime), df = 3)2  1.73437    3.22005   0.5386    0.5902
## hormon:nsx(log(rectime), df = 3)3  0.51817    0.75191   0.6891    0.4907
##
## (Intercept)                        ***
## hormon
## nsx(log(rectime), df = 3)1         ***
## nsx(log(rectime), df = 3)2         ***
## nsx(log(rectime), df = 3)3         ***
## hormon:nsx(log(rectime), df = 3)1
## hormon:nsx(log(rectime), df = 3)2
## hormon:nsx(log(rectime), df = 3)3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 5215.322
```

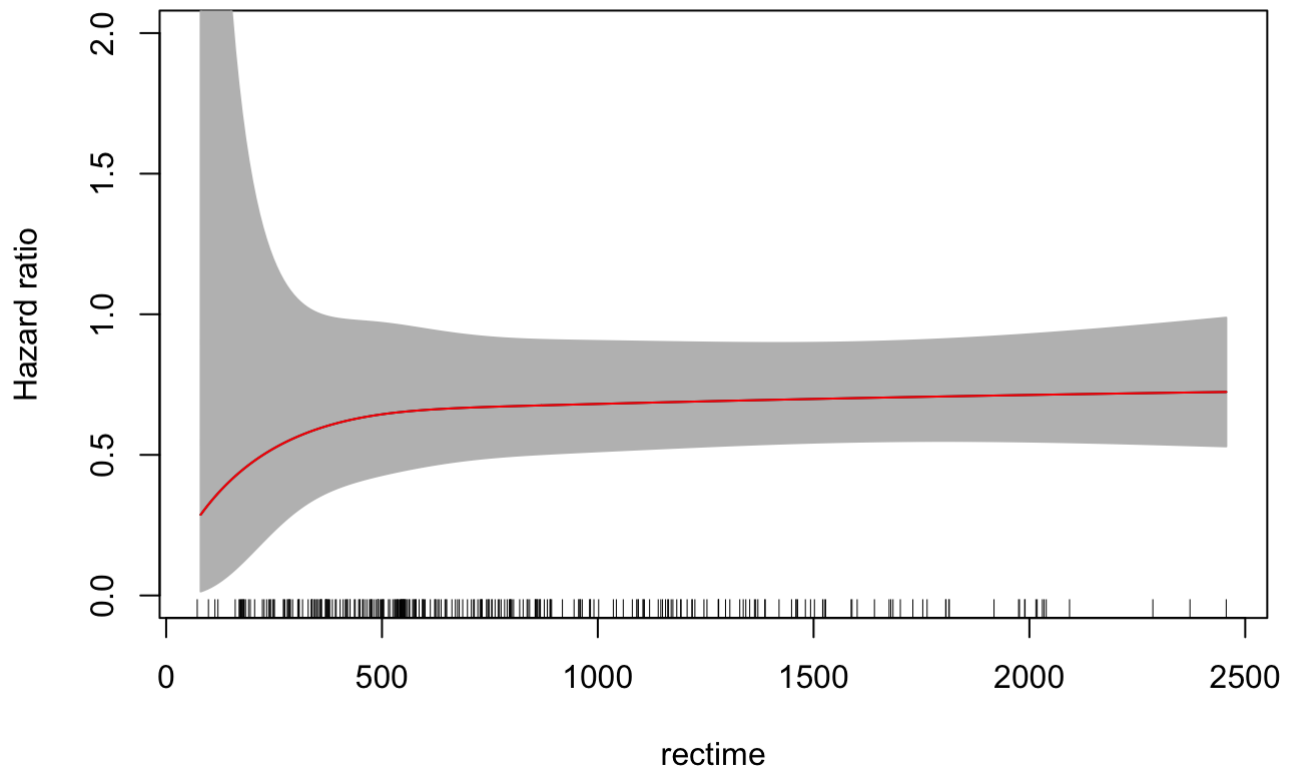
```
anova(fit,fit.tvc) # compare with and without tvc
```

```
## Likelihood Ratio Tests
## Model 1: fit, [negll]: (Intercept)+hormon+nsx(log(rectime), df = 3)1+
##      nsx(log(rectime), df = 3)2+nsx(log(rectime), df = 3)3
## Model 2: fit.tvc, [negll]: (Intercept)+hormon+nsx(log(rectime), df =
##      3)1+nsx(log(rectime), df = 3)2+nsx(log(rectime), df = 3)3+
##      hormon:nsx(log(rectime), df = 3)1+hormon:nsx(log(rectime), df =
##      3)2+hormon:nsx(log(rectime), df = 3)3
## Tot Df Deviance Chisq Df Pr(>Chisq)
## 1      5    5215.9
## 2      8    5215.3 0.593 3      0.898
```

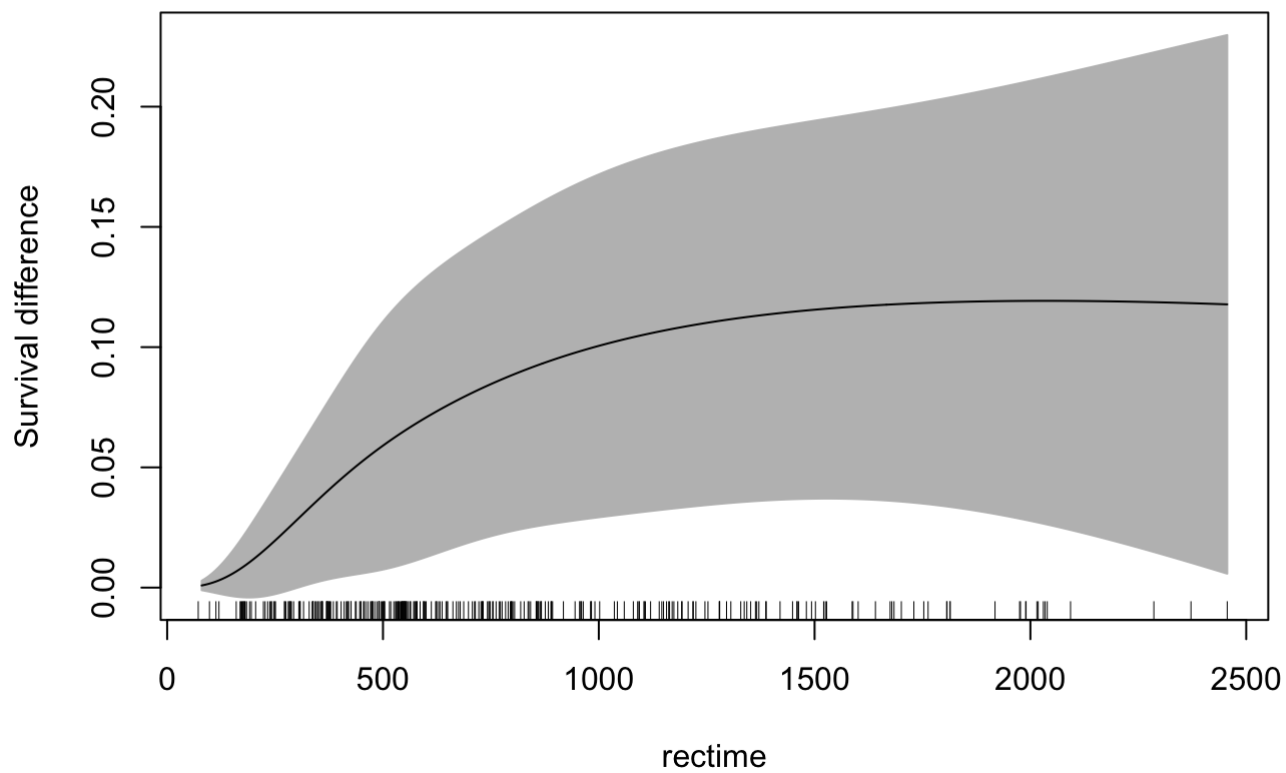


```
## some more plots
plot(fit.tvc,newdata=data.frame(hormon=0),type="hr",var="hormon", ylim=c(0,2))

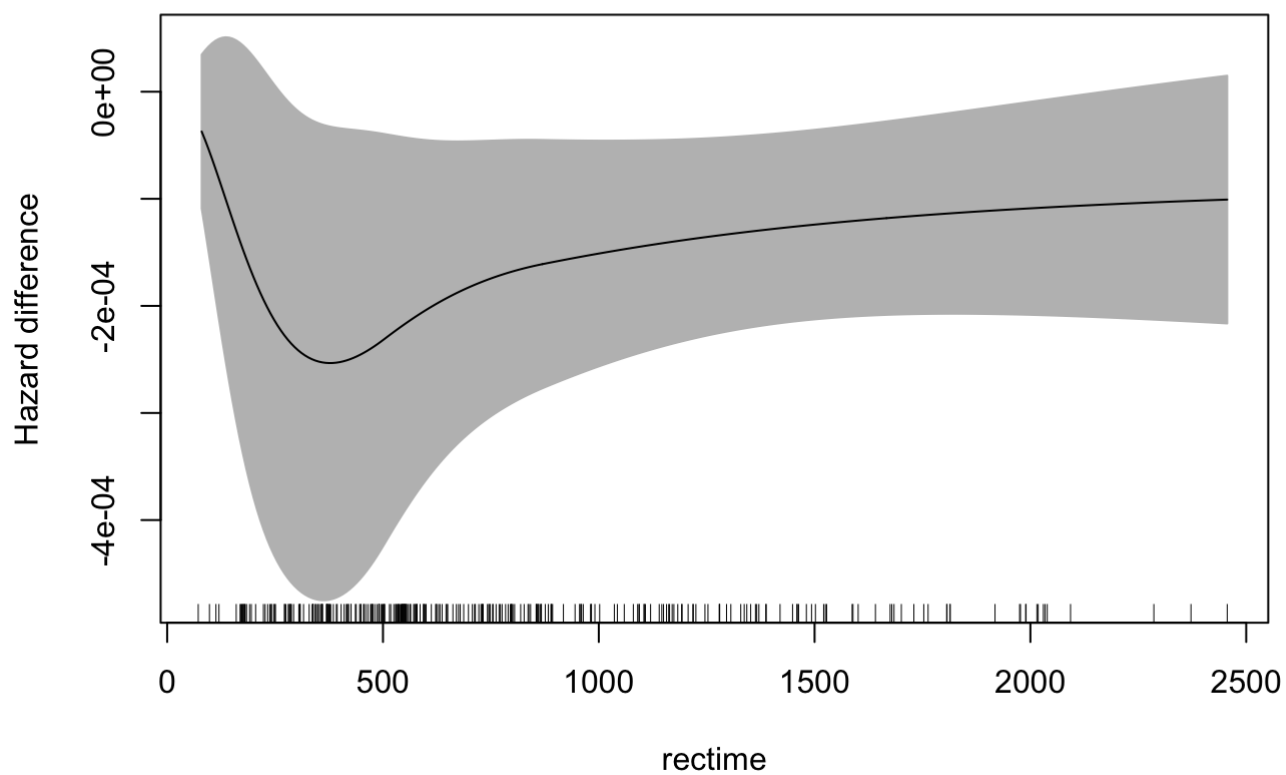
# no lines method: use add=TRUE
plot(fit.tvc,newdata=data.frame(hormon=1),type="hr",var="hormon",
      add=TRUE,ci=FALSE,line.col=2)
```



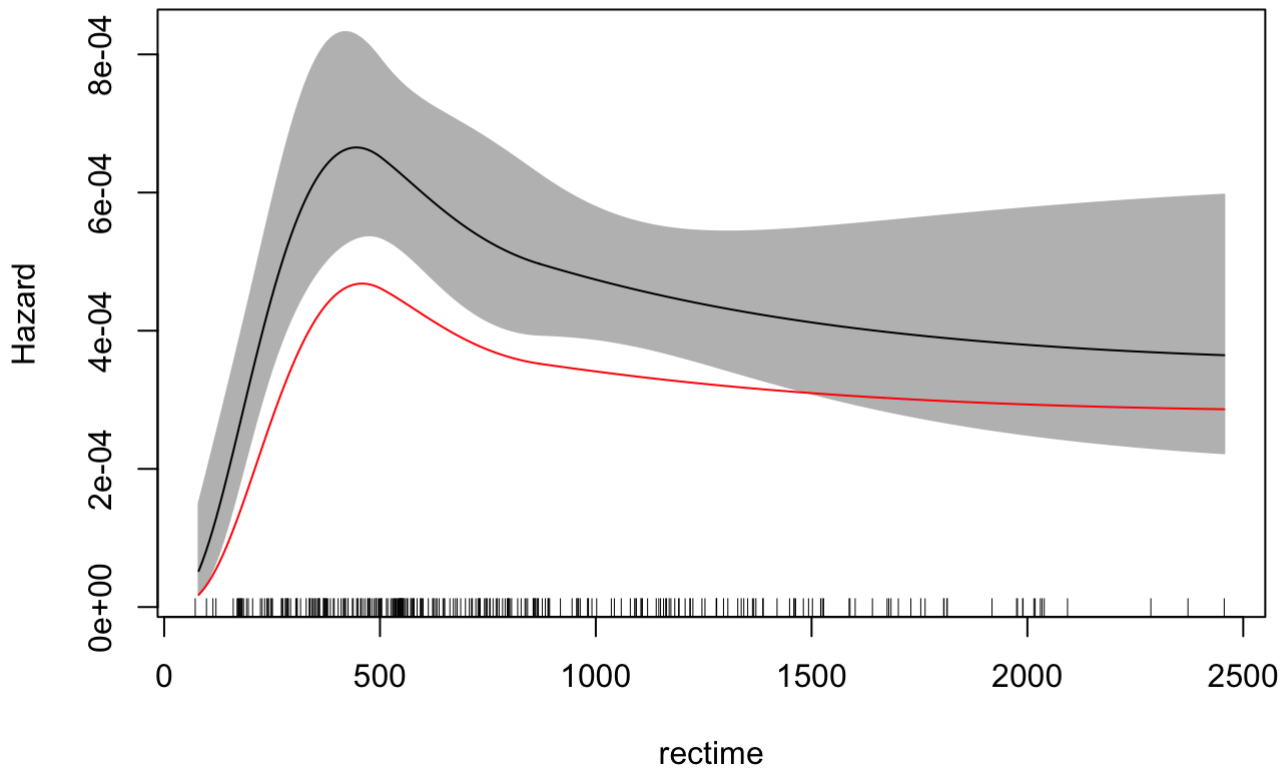
```
plot(fit.tvc,newdata=data.frame(hormon=0),type="sdiff",var="hormon")
```



```
plot(fit.tvc,newdata=data.frame(hormon=0),type="hdiff",var="hormon")
```



```
plot(fit.tvc,newdata=data.frame(hormon=0),type="hazard")
plot(fit.tvc,newdata=data.frame(hormon=1),type="hazard",line.col=2,ci=FALSE,add=TRUE)
```



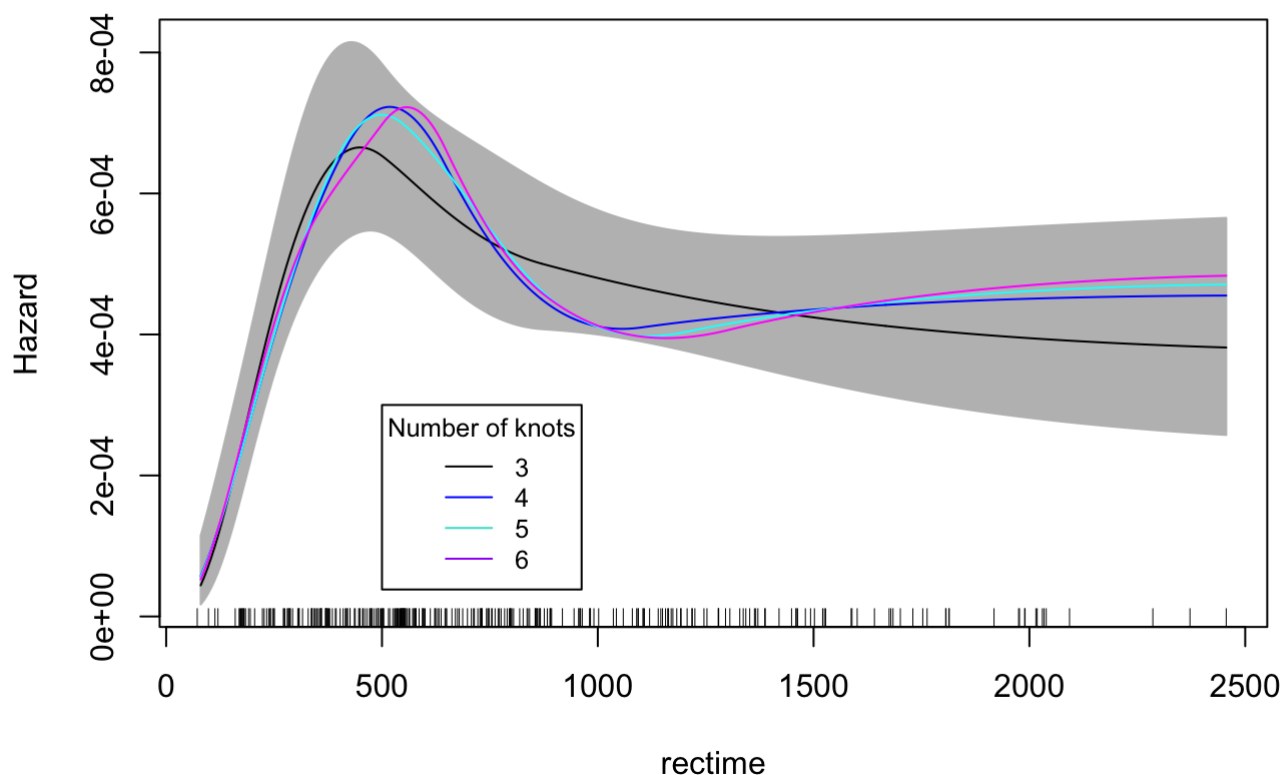
```
## compare number of knots
hormon0 <- data.frame(hormon=0)
plot(fit,type="hazard",newdata=hormon0)
AIC(fit)
```

```
## [1] 5225.915
```

```
for (df in 4:6) {
  fit.new <- stpm2(Surv(rectime,censrec==1)~hormon,data=brcancer,df=df)
  plot(fit.new,type="hazard",newdata=hormon0,add=TRUE,ci=FALSE,line.col=df)
  print(AIC(fit.new))
}
```

```
## [1] 5224.943
## [1] 5227.05
## [1] 5228.692
```

```
legend(500, 3*10^(-4), title = "Number of knots", legend=c("3", "4", "5", "6"),
      col=c("black", "blue", "turquoise", "purple"), lty = 1, cex = 0.8)
```



3.5 Loss of expectancy

The loss of expectancy is: $YLL * u$, where YLL is the expected years of life lost and u a measure of utility or value of one year for one person (e.g., annual per capita income, etc.).

YLL is based on comparing the age of death to an external standard life expectancy curve, and can incorporate time discounting and age weighting (Aragon et al. 2008).

```
# Set the function to compute YLL
yll <- function(number.deaths, average.age.death, model.life.expectancy,
                discount.rate = 0.03, beta.constant = 0.04,
                modulation.constant = 0, adjustment.constant = 0.1658){
  ##abbreviate inputs
  N <- number.deaths;          a <- average.age.death
  L <- model.life.expectancy; r <- discount.rate
  b <- beta.constant;          K <- modulation.constant
  CC <- adjustment.constant
  ##do calculations
  if(discount.rate==0){
    N*(K*CC*((exp(-b*a))/b^2)*((exp(-b*L))*
                                (-b*(L+a)-1)-(-b*a-1))+((1-K)*L))
  } else {
    N*(K*((CC*exp(r*a))/(-(r+b)^2))*((exp(-(r+b)*(L+a))*(-(r+b)*
    (L+a)-1))-(exp(-(r+b)*a)*(-(r+b)*a-1)))+((1-K)/r)*((1-exp(-r*L))))
  }
}
```

For example, for the years 2003-2004 in San Francisco, for men ages 50 to 54 years, there were 434 deaths with an average age of death of 52.5 years. The interpolated model life expectancy for that age interval was 28.6 years.

```
number.deaths <- 434
average.age.death <- 52.5
model.life.expectancy <- 28.6

yll(number.deaths, average.age.death, model.life.expectancy)
```

```
## [1] 8332.666
```

```
yll(number.deaths, average.age.death, model.life.expectancy, discount.rate=0)
```

```
## [1] 12412.4
```

```
yll(number.deaths, average.age.death, model.life.expectancy)/number.deaths
```

```
## [1] 19.19969
```

The expected years of life lost for San Francisco male deaths, ages 50 to 54 years, during years 2003-2004, was 8332.7 years with discounting and 1.24124×10^4 years without discounting. However, the YLLs are influenced by the large number of deaths among older people. The average YLL, which highlights premature causes of death and brings attention to preventable deaths that contribute most to the mortality burden, was 19.2 years.

4 Selection of cancers

The most appropriate cancer type(s) for our research are selected based on several criteria. First, the target group is people in need of long-term financial services, so we will ideally restrict ourselves to people under 40. Therefore, we are looking for cancers with a significant number of incidences occurring before the age of 40. Second, we are looking for cancer patients who have a chance of survival close to cancer-free patients. We are thus also looking for cancers with a relatively high survival rate or high cured rate. Third, the number of incidences must be statistically significant. Last but not least, our research is partly intended for banks and insurance companies so we will avoid rare cancers and prefer cancers that are largely known to the general public.

Based on our analyses we believe that the following cancers are appropriate for our research:

Cancer	Pros	Cons
Melanoma	High relative survival (M: 87%, F: 93%) ² and early incidences (start at 10-15 years old)	
Breast (women only)	High and early incidences (10,695 new diagnoses in 2013 and start at 20-25), high relative survival (F: 90%) and very common cancer among women	
Testis (men only)	Very high relative survival (M: 97%) and early incidences (start at 10 and peak at 30)	Very low incidences (353 new diagnoses in 2013)

Cancer	Pros	Cons
Thyroid	High relative survival (M: 91%, F: 96%) and early incidences (start at 5-10)	Low incidences (941 new diagnoses in 2013)
Cervix (women only)	Good relative survival (F: 69%) and early incidences (start at 15)	Low incidences (633 new diagnoses in 2013)
Leukemia	High relative survival (between 83% and 89%), very early incidences (distinctive peak around the age of 3 years) and common cancer among children	Very low incidences (123 new diagnoses in 2013)

The final selection will depend on the number of cancers that we choose to study.

References

- Aragon, Tomas J, Daphne Y Lichtensztajn, Brian S Katcher, Randy Reiter, and Mitchell H Katz. 2008. "Calculating Expected Years of Life Lost for Assessing Local Ethnic Disparities in Causes of Premature Death." *BMC Public Health* 8 (1). BioMed Central: 116.
- Boussari, Olayide, Gaelle Romain, Laurent Remontet, Nadine Bossard, Morgane Mounier, Anne-Marie Bouvier, Christine Binquet, Marc Colonna, and Valerie Jooste. 2018. "A New Approach to Estimate Time-to-Cure from Cancer Registries Data." *Cancer Epidemiology* 53. Elsevier: 72–80.
- Denuit, Michel, and Catherine Legrand. 2018. "Risk Classification in Life and Health Insurance: Extension to Continuous Covariates." *European Actuarial Journal*. Springer, 1–11.
- Esteve, J, E Benhamou, M Croasdale, and L Raymond. 1990. "Relative Survival and the Estimation of Net Survival: Elements for Further Discussion." *Statistics in Medicine* 9 (5). Wiley Online Library: 529–38.
- Stare, Janez, Maja Pohar, and Robin Henderson. 2005. "Goodness of Fit of Relative Survival Models." *Statistics in Medicine* 24 (24). Wiley Online Library: 3911–25.

1. Note that the transversal life expectancy corresponds to the average number of years that would remain to live to an individual of age x , in year t , if he was confronted for his remaining life with mortality quotients at different ages of the year. Longitudinal life expectancy takes into account the evolution of mortality quotients during the individual's future life. ↩
2. Relative survival corresponds to the 5-year relative survival proportion for the Belgian 2009-2013 cohort. F = Female. M = Male. ↩