model_cox.R

stephane

Tue May 29 12:01:24 2018

```
#####################################################################
## load packages                                                   ##
#####################################################################
# Clear workspace variables
rm(list = ls())
cat("\014")
```

```
# options(stringsAsFactors=F)
# install.packages("survival")
# install.packages("Formula")
# install.packages("ggplot2")
# install.packages("Hmisc")
# install.packages("lattice")
# install.packages("SparseM")
# install.packages("tinytex")
# tinytex::install_tinytex(force = TRUE)
library(survival)
library(Formula)
library(ggplot2)
library(Hmisc)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(SparseM)
```

```
##
## Attaching package: 'SparseM'
```

```
## The following object is masked from 'package:base':
##
##     backsolve
```

```
library(lattice)

# install.packages("sandwich")
library(sandwich)
# install.packages("rms")
```

```r
library(rms)

#install.packages("devtools")
#library(devtools)
#install.version("rmarkdown",version=1.8)


#####################################################################
## 1. Set working directory and get data                           ##
#####################################################################
setwd("~/Dropbox/Documents/Projects/DataScience/SurvivalELSA")
#elsa_cf <- read.csv("ELSA_CF_TRUE.csv", sep=",")
#save(elsa_cf,file='elsa_cf.rdata')

load("elsa.RData")
attach(elsa_cf)

#####################################################################
## 2. Cox regression models                                        ##
#####################################################################

#####################################################################
## 2.1. Simple models                                              ##
#####################################################################

# Sex only
cox1 <- coxph(Surv(time, death)~ factor(sex), data=elsa_cf)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ factor(sex), data = elsa_cf)
##
##   n= 9282, number of events= 1802
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## factor(sex)1 -0.001538  0.998463  0.047146 -0.033    0.974
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(sex)1    0.9985      1.002    0.9103     1.095
##
## Concordance= 0.501  (se = 0.006 )
## Rsquare= 0   (max possible= 0.97 )
## Likelihood ratio test= 0  on 1 df,   p=1
## Wald test            = 0  on 1 df,   p=1
## Score (logrank) test = 0  on 1 df,   p=1
```

```r
# Sex + age
cox2 <- coxph(Surv(time, death)~ age1 + factor(sex), data=elsa_cf)
summary(cox2)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age1 + factor(sex), data = elsa_cf)
##
```

```
##   n= 9282, number of events= 1802
##
##                   coef exp(coef) se(coef)      z Pr(>|z|)
## age1          0.09048   1.09470  0.00271 33.381  < 2e-16 ***
## factor(sex)1 -0.40120   0.66952  0.04922 -8.152 3.59e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age1            1.0947     0.9135    1.0889    1.1005
## factor(sex)1    0.6695     1.4936    0.6079    0.7373
##
## Concordance= 0.714  (se = 0.007 )
## Rsquare= 0.117   (max possible= 0.97 )
## Likelihood ratio test= 1151  on 2 df,   p=<2e-16
## Wald test            = 1116  on 2 df,   p=<2e-16
## Score (logrank) test = 1248  on 2 df,   p=<2e-16
```

```
# LR test: simply use anova of the two model fits
# Example to test the effect of adding CHD into the model
cox3 <- coxph(Surv(time, death)~ age1 + factor(sex) + factor(chd1), data=elsa_cf)
summary(cox3)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age1 + factor(sex) + factor(chd1),
##     data = elsa_cf)
##
##   n= 9282, number of events= 1802
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## age1           0.087625  1.091579  0.002748 31.881  < 2e-16 ***
## factor(sex)1  -0.360043  0.697647  0.049519 -7.271 3.57e-13 ***
## factor(chd1)1  0.445509  1.561284  0.057317  7.773 7.69e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## age1             1.0916     0.9161    1.0857    1.0975
## factor(sex)1     0.6976     1.4334    0.6331    0.7688
## factor(chd1)1    1.5613     0.6405    1.3954    1.7469
##
## Concordance= 0.721  (se = 0.007 )
## Rsquare= 0.122   (max possible= 0.97 )
## Likelihood ratio test= 1207  on 3 df,   p=<2e-16
## Wald test            = 1182  on 3 df,   p=<2e-16
## Score (logrank) test = 1323  on 3 df,   p=<2e-16
```

```
# likelihood ratio test:
anova(cox2,cox3)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, death)
##  Model 1: ~ age1 + factor(sex)
```

```
##  Model 2: ~ age1 + factor(sex) + factor(chd1)
##   loglik  Chisq Df P(>|Chi|)
## 1 -15700
## 2 -15672 55.767  1 8.159e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### This returns a very small p-value, meaning that adding CHD to the model improves the likelihood.
## We can also simply see this in the p-value of the effect estimate of CHD in model 3


#####################################################################
## 2.2. COVARIATES : EXPLORATION AND SELECTION              ##
#####################################################################


# Model 4 : all potential covariates included  #
cigst1_=factor(cigst1)
educ1_=factor(educ1)
cox4 <- coxph(Surv(time, death) ~ age1 + sex + chd1 + cancer1+ educ1_ + cigst1_  +  physinact1
              + alcohol1 ,   data=elsa_cf, method="breslow")
summary(cox4)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age1 + sex + chd1 + cancer1 +
##     educ1_ + cigst1_ + physinact1 + alcohol1, data = elsa_cf,
##     method = "breslow")
##
##   n= 9282, number of events= 1802
##
##                 coef exp(coef)  se(coef)      z Pr(>|z|)
## age1        0.082631  1.086141  0.002843 29.070  < 2e-16 ***
## sex        -0.422418  0.655460  0.051610 -8.185 2.73e-16 ***
## chd1        0.325514  1.384743  0.057888  5.623 1.87e-08 ***
## cancer1     0.672062  1.958271  0.073436  9.152  < 2e-16 ***
## educ1_1     0.208939  1.232370  0.092687  2.254 0.024180 *
## educ1_2     0.323894  1.382501  0.087840  3.687 0.000227 ***
## cigst1_1    0.249609  1.283523  0.057761  4.321 1.55e-05 ***
## cigst1_2    0.656862  1.928731  0.070076  9.374  < 2e-16 ***
## physinact1  0.660724  1.936194  0.049777 13.274  < 2e-16 ***
## alcohol1    0.039824  1.040628  0.054498  0.731 0.464931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## age1           1.0861     0.9207    1.0801    1.0922
## sex            0.6555     1.5256    0.5924    0.7252
## chd1           1.3847     0.7222    1.2362    1.5511
## cancer1        1.9583     0.5107    1.6958    2.2614
## educ1_1        1.2324     0.8114    1.0277    1.4779
## educ1_2        1.3825     0.7233    1.1638    1.6422
## cigst1_1       1.2835     0.7791    1.1461    1.4374
## cigst1_2       1.9287     0.5185    1.6812    2.2127
## physinact1     1.9362     0.5165    1.7562    2.1346
```

```
## alcohol1      1.0406     0.9610    0.9352    1.1579
##
## Concordance= 0.756  (se = 0.007 )
## Rsquare= 0.158   (max possible= 0.97 )
## Likelihood ratio test= 1595  on 10 df,   p=<2e-16
## Wald test            = 1572  on 10 df,   p=<2e-16
## Score (logrank) test = 1759  on 10 df,   p=<2e-16
```

```
### ALL COVARIATES ARE ASSOCIATED EXCEPT ALCOHOL ###

# --> Now that we have selected our set of covariates to include in the model, we want to test the
## crude and adjusted effect of cognitive function on mortality risk

#### MODEL 5 : COGNITIVE FUNCTION SCORE (CONTINUOUS)

cox5a <- coxph(Surv(time, death) ~ cf1 + age1 + sex ,   data=elsa_cf, method="breslow") #only adjusted
cox5b <- coxph(Surv(time, death) ~ cf1 + age1 + sex + chd1 + cancer1+ educ1_ + cigst1_  +  physinact1
                ,   data=elsa_cf, method="breslow") # adjusted for covariates
summary(cox5a)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ cf1 + age1 + sex, data = elsa_cf,
##     method = "breslow")
##
##   n= 9282, number of events= 1802
##
##           coef exp(coef)  se(coef)       z Pr(>|z|)
## cf1  -0.040372  0.960433  0.002225 -18.141  < 2e-16 ***
## age1  0.070877  1.073449  0.002888  24.541  < 2e-16 ***
## sex  -0.365865  0.693597  0.049446  -7.399 1.37e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## cf1     0.9604     1.0412    0.9563    0.9646
## age1    1.0734     0.9316    1.0674    1.0795
## sex     0.6936     1.4418    0.6295    0.7642
##
## Concordance= 0.745  (se = 0.007 )
## Rsquare= 0.147   (max possible= 0.97 )
## Likelihood ratio test= 1478  on 3 df,   p=<2e-16
## Wald test            = 1450  on 3 df,   p=<2e-16
## Score (logrank) test = 1580  on 3 df,   p=<2e-16
```

```
summary(cox5b)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ cf1 + age1 + sex + chd1 +
##     cancer1 + educ1_ + cigst1_ + physinact1, data = elsa_cf,
##     method = "breslow")
##
##   n= 9282, number of events= 1802
##
```

```
##                 coef exp(coef)  se(coef)       z Pr(>|z|)
## cf1        -0.033219  0.967326  0.002301 -14.435  < 2e-16 ***
## age1        0.068654  1.071066  0.002970  23.114  < 2e-16 ***
## sex        -0.390652  0.676615  0.051587  -7.573 3.66e-14 ***
## chd1        0.281896  1.325641  0.057918   4.867 1.13e-06 ***
## cancer1     0.680085  1.974045  0.073530   9.249  < 2e-16 ***
## educ1_1     0.126269  1.134587  0.092713   1.362    0.173
## educ1_2     0.125449  1.133657  0.087978   1.426    0.154
## cigst1_1    0.256305  1.292146  0.057536   4.455 8.40e-06 ***
## cigst1_2    0.622097  1.862829  0.070074   8.878  < 2e-16 ***
## physinact1  0.543259  1.721608  0.050373  10.785  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## cf1           0.9673     1.0338    0.9630    0.9717
## age1          1.0711     0.9336    1.0648    1.0773
## sex           0.6766     1.4779    0.6115    0.7486
## chd1          1.3256     0.7544    1.1834    1.4850
## cancer1       1.9740     0.5066    1.7091    2.2801
## educ1_1       1.1346     0.8814    0.9461    1.3607
## educ1_2       1.1337     0.8821    0.9541    1.3470
## cigst1_1      1.2921     0.7739    1.1543    1.4464
## cigst1_2      1.8628     0.5368    1.6238    2.1371
## physinact1    1.7216     0.5809    1.5598    1.9003
##
## Concordance= 0.771  (se = 0.007 )
## Rsquare= 0.176   (max possible= 0.97 )
## Likelihood ratio test= 1802  on 10 df,   p=<2e-16
## Wald test            = 1780  on 10 df,   p=<2e-16
## Score (logrank) test = 1976  on 10 df,   p=<2e-16
```

```
## --> In the multivariable-adjusted model (5b), the increase in 1 point of cognitive function score
## was associated with a 3% decrease in mortality risk
## Explanation: a hazard ratio of 0.97 means a 0.97-1=-0.03 *100=-3%

## Note: all covariates remain significantly associated except for education

## We create a variable cf1/10 to interpret the HR as the reduction in risk for an increase
### of 10 points of score (instead of 1)
elsa_cf$cf1_10<-elsa_cf$cf1/10
summary(elsa_cf$cf1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.00   40.00   47.00   47.19   55.00   94.00
```

```
summary(elsa_cf$cf1_10)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.400   4.000   4.700   4.719   5.500   9.400
```

```r
cox5c <- coxph(Surv(time, death) ~ cf1_10 + age1 + sex + chd1 + cancer1+ educ1_ + cigst1_  +  physinact
               ,   data=elsa_cf, method="breslow") # adjusted for covariates
summary(cox5c)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ cf1_10 + age1 + sex + chd1 +
##     cancer1 + educ1_ + cigst1_ + physinact1, data = elsa_cf,
##     method = "breslow")
##
##   n= 9282, number of events= 1802
##
##                coef exp(coef) se(coef)       z Pr(>|z|)
## cf1_10     -0.33219   0.71735  0.02301 -14.435  < 2e-16 ***
## age1        0.06865   1.07107  0.00297  23.114  < 2e-16 ***
## sex        -0.39065   0.67661  0.05159  -7.573 3.66e-14 ***
## chd1        0.28190   1.32564  0.05792   4.867 1.13e-06 ***
## cancer1     0.68009   1.97405  0.07353   9.249  < 2e-16 ***
## educ1_1     0.12627   1.13459  0.09271   1.362    0.173
## educ1_2     0.12545   1.13366  0.08798   1.426    0.154
## cigst1_1    0.25631   1.29215  0.05754   4.455 8.40e-06 ***
## cigst1_2    0.62210   1.86283  0.07007   8.878  < 2e-16 ***
## physinact1  0.54326   1.72161  0.05037  10.785  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##            exp(coef) exp(-coef) lower .95 upper .95
## cf1_10        0.7173     1.3940    0.6857    0.7504
## age1          1.0711     0.9336    1.0648    1.0773
## sex           0.6766     1.4779    0.6115    0.7486
## chd1          1.3256     0.7544    1.1834    1.4850
## cancer1       1.9740     0.5066    1.7091    2.2801
## educ1_1       1.1346     0.8814    0.9461    1.3607
## educ1_2       1.1337     0.8821    0.9541    1.3470
## cigst1_1      1.2921     0.7739    1.1543    1.4464
## cigst1_2      1.8628     0.5368    1.6238    2.1371
## physinact1    1.7216     0.5809    1.5598    1.9003
##
## Concordance= 0.771  (se = 0.007 )
## Rsquare= 0.176   (max possible= 0.97 )
## Likelihood ratio test= 1802  on 10 df,   p=<2e-16
## Wald test            = 1780  on 10 df,   p=<2e-16
## Score (logrank) test = 1976  on 10 df,   p=<2e-16
```

```
## --> In the multivariable-adjusted model (5b), the increase in 10 points of cognitive function score
## was associated with a 28 % decrease in mortality risk

# Finally, we also look at it across quintiles
# Divide cognitive function score (quintiles)
quantile(elsa_cf$cf1, prob=c(0.20, 0.40, 0.60, 0.80))
```

```
## 20% 40% 60% 80%
##  38  45  50  57
```

```r
elsa_cf$q_cf1 <-cut(elsa_cf$cf1, breaks=c(0, 38, 45, 50, 57, 194))
cox5c <- coxph(Surv(time, death) ~ factor(q_cf1) + age1 + sex + chd1 + cancer1+ educ1_ + cigst1_  +  phy
               ,   data=elsa_cf, method="breslow") # adjusted for covariates
summary(cox5c)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ factor(q_cf1) + age1 + sex +
##     chd1 + cancer1 + educ1_ + cigst1_ + physinact1, data = elsa_cf,
##     method = "breslow")
##
##   n= 9282, number of events= 1802
##
##                            coef exp(coef)  se(coef)      z Pr(>|z|)
## factor(q_cf1)(38,45]  -0.501086  0.605872  0.062803 -7.979 1.48e-15 ***
## factor(q_cf1)(45,50]  -0.582811  0.558327  0.074667 -7.805 5.93e-15 ***
## factor(q_cf1)(50,57]  -0.707011  0.493116  0.079303 -8.915  < 2e-16 ***
## factor(q_cf1)(57,194] -0.964249  0.381269  0.102172 -9.437  < 2e-16 ***
## age1                   0.071182  1.073777  0.002971 23.956  < 2e-16 ***
## sex                   -0.386460  0.679458  0.051581 -7.492 6.77e-14 ***
## chd1                   0.283569  1.327861  0.057900  4.898 9.70e-07 ***
## cancer1                0.674264  1.962587  0.073496  9.174  < 2e-16 ***
## educ1_1                0.146432  1.157696  0.092824  1.578   0.1147
## educ1_2                0.163159  1.177224  0.088153  1.851   0.0642 .
## cigst1_1               0.260279  1.297292  0.057576  4.521 6.16e-06 ***
## cigst1_2               0.635277  1.887544  0.070076  9.066  < 2e-16 ***
## physinact1             0.558993  1.748911  0.050271 11.120  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                       exp(coef) exp(-coef) lower .95 upper .95
## factor(q_cf1)(38,45]     0.6059     1.6505    0.5357    0.6852
## factor(q_cf1)(45,50]     0.5583     1.7911    0.4823    0.6463
## factor(q_cf1)(50,57]     0.4931     2.0279    0.4221    0.5760
## factor(q_cf1)(57,194]    0.3813     2.6228    0.3121    0.4658
## age1                     1.0738     0.9313    1.0675    1.0800
## sex                      0.6795     1.4718    0.6141    0.7517
## chd1                     1.3279     0.7531    1.1854    1.4874
## cancer1                  1.9626     0.5095    1.6993    2.2667
## educ1_1                  1.1577     0.8638    0.9651    1.3887
## educ1_2                  1.1772     0.8495    0.9904    1.3992
## cigst1_1                 1.2973     0.7708    1.1589    1.4523
## cigst1_2                 1.8875     0.5298    1.6453    2.1654
## physinact1               1.7489     0.5718    1.5848    1.9300
##
## Concordance= 0.769  (se = 0.007 )
## Rsquare= 0.172   (max possible= 0.97 )
## Likelihood ratio test= 1757  on 13 df,   p=<2e-16
## Wald test            = 1728  on 13 df,   p=<2e-16
## Score (logrank) test = 2004  on 13 df,   p=<2e-16
```

```r
#####################################################################
## 2.3. TESTING FOR INTERACTIONS                                   ##
#####################################################################
```

```r
# LR test: simply use anova of the two model fits
# Example to test the effect of an interaction between age and cognitive function
## !!! THE MODELS HAVE TO BE NESTED
cox6a <- coxph(Surv(time, death)~ age1 + factor(sex) + cf1, data=elsa_cf)
cox6b <- coxph(Surv(time, death)~ age1 + factor(sex) + cf1 + age1*cf1, data=elsa_cf)
summary(cox6a)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age1 + factor(sex) + cf1,
##     data = elsa_cf)
##
##   n= 9282, number of events= 1802
##
##                    coef exp(coef)  se(coef)       z Pr(>|z|)
## age1          0.070934  1.073511  0.002888  24.559  < 2e-16 ***
## factor(sex)1 -0.366040  0.693475  0.049446  -7.403 1.33e-13 ***
## cf1          -0.040408  0.960397  0.002226 -18.155  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age1            1.0735     0.9315    1.0675    1.0796
## factor(sex)1    0.6935     1.4420    0.6294    0.7640
## cf1             0.9604     1.0412    0.9562    0.9646
##
## Concordance= 0.745  (se = 0.007 )
## Rsquare= 0.147   (max possible= 0.97 )
## Likelihood ratio test= 1479  on 3 df,   p=<2e-16
## Wald test            = 1451  on 3 df,   p=<2e-16
## Score (logrank) test = 1582  on 3 df,   p=<2e-16
```

```r
summary(cox6b)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age1 + factor(sex) + cf1 +
##     age1 * cf1, data = elsa_cf)
##
##   n= 9282, number of events= 1802
##
##                     coef  exp(coef)   se(coef)      z Pr(>|z|)
## age1          0.0268762  1.0272406  0.0096647  2.781  0.00542 **
## factor(sex)1 -0.3476461  0.7063488  0.0493710 -7.042 1.90e-12 ***
## cf1          -0.1169013  0.8896730  0.0162517 -7.193 6.33e-13 ***
## age1:cf1      0.0010884  1.0010890  0.0002293  4.747 2.06e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age1            1.0272     0.9735    1.0080    1.0469
## factor(sex)1    0.7063     1.4157    0.6412    0.7781
## cf1             0.8897     1.1240    0.8618    0.9185
```

```
## age1:cf1        1.0011     0.9989    1.0006    1.0015
##
## Concordance= 0.747  (se = 0.007 )
## Rsquare= 0.149   (max possible= 0.97 )
## Likelihood ratio test= 1503  on 4 df,   p=<2e-16
## Wald test            = 1314  on 4 df,   p=<2e-16
## Score (logrank) test = 1765  on 4 df,   p=<2e-16
```

```
# likelihood ratio test:
anova(cox6a,cox6b)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, death)
##  Model 1: ~ age1 + factor(sex) + cf1
##  Model 2: ~ age1 + factor(sex) + cf1 + age1 * cf1
##    loglik  Chisq Df P(>|Chi|)
## 1 -15536
## 2 -15524 23.199  1 1.461e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
###### Very small p-value: evidence of a potential interaction between cognitive function and age, i.e.
## the effect of cognitive function on mortality differs according to age


# Create agegr (age in categories) and stratify analysis
elsa_cf$agegr[elsa_cf$age1 < 60] <- 1
elsa_cf$agegr[elsa_cf$age1>=60 & elsa_cf$age1 < 70] <- 2
elsa_cf$agegr[elsa_cf$age1 >= 70] <- 3

#Create subsets#
elsaage1 <-subset(elsa_cf, elsa_cf$agegr==1)
elsaage2 <-subset(elsa_cf, elsa_cf$agegr==2)
elsaage3 <-subset(elsa_cf, elsa_cf$agegr==3)

#Run models in each strata#
coxf.age1 <- coxph(Surv(time, death) ~ cf1_10 + age1  + sex + chd1 + cancer1+ factor(educ1) + factor(ci
                   data=elsaage1, method="breslow")
coxf.age2 <- coxph(Surv(time, death) ~ cf1_10 + age1  + sex + chd1 + cancer1+ factor(educ1) + factor(ci
                   data=elsaage2, method="breslow")
coxf.age3 <- coxph(Surv(time, death) ~ cf1_10 + age1  + sex + chd1 + cancer1+ factor(educ1) + factor(ci
                   data=elsaage3, method="breslow")
summary(coxf.age1)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ cf1_10 + age1 + sex + chd1 +
##     cancer1 + factor(educ1) + factor(cigst1) + physinact1, data = elsaage1,
##     method = "breslow")
##
##   n= 3481, number of events= 298
##
##                      coef exp(coef)  se(coef)      z Pr(>|z|)
## cf1_10          -0.469458  0.625341  0.051903 -9.045  < 2e-16 ***
```

```
## age1            -0.007814  0.992217  0.025216 -0.310 0.756669
## sex             -0.229694  0.794777  0.127200 -1.806 0.070953 .
## chd1             0.635006  1.887033  0.150090  4.231 2.33e-05 ***
## cancer1          1.308207  3.699536  0.178544  7.327 2.35e-13 ***
## factor(educ1)1   0.614994  1.849646  0.238515  2.578 0.009925 **
## factor(educ1)2   0.620125  1.859161  0.230376  2.692 0.007107 **
## factor(cigst1)1  0.373330  1.452564  0.152088  2.455 0.014100 *
## factor(cigst1)2  0.603552  1.828603  0.163898  3.682 0.000231 ***
## physinact1       0.718401  2.051150  0.123450  5.819 5.91e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## cf1_10             0.6253     1.5991    0.5649    0.6923
## age1               0.9922     1.0078    0.9444    1.0425
## sex                0.7948     1.2582    0.6194    1.0198
## chd1               1.8870     0.5299    1.4061    2.5324
## cancer1            3.6995     0.2703    2.6072    5.2496
## factor(educ1)1     1.8496     0.5406    1.1589    2.9520
## factor(educ1)2     1.8592     0.5379    1.1836    2.9202
## factor(cigst1)1    1.4526     0.6884    1.0781    1.9570
## factor(cigst1)2    1.8286     0.5469    1.3262    2.5213
## physinact1         2.0512     0.4875    1.6103    2.6126
##
## Concordance= 0.749  (se = 0.017 )
## Rsquare= 0.077   (max possible= 0.751 )
## Likelihood ratio test= 277.9  on 10 df,   p=<2e-16
## Wald test            = 308.6  on 10 df,   p=<2e-16
## Score (logrank) test = 338  on 10 df,   p=<2e-16
```

```
summary(coxf.age2)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ cf1_10 + age1 + sex + chd1 +
##     cancer1 + factor(educ1) + factor(cigst1) + physinact1, data = elsaage2,
##     method = "breslow")
##
##   n= 3121, number of events= 446
##
##                     coef exp(coef) se(coef)      z Pr(>|z|)
## cf1_10          -0.31501   0.72979  0.04528 -6.956 3.50e-12 ***
## age1             0.08483   1.08853  0.01735  4.888 1.02e-06 ***
## sex             -0.40844   0.66469  0.09968 -4.098 4.17e-05 ***
## chd1             0.46562   1.59300  0.11473  4.058 4.94e-05 ***
## cancer1          0.96378   2.62159  0.13692  7.039 1.93e-12 ***
## factor(educ1)1  -0.03338   0.96717  0.17337 -0.193    0.847
## factor(educ1)2  -0.07842   0.92458  0.16081 -0.488    0.626
## factor(cigst1)1  0.50469   1.65647  0.12615  4.001 6.32e-05 ***
## factor(cigst1)2  0.85938   2.36169  0.13955  6.158 7.36e-10 ***
## physinact1       0.71370   2.04153  0.09947  7.175 7.24e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
```

```
## cf1_10             0.7298     1.3703    0.6678    0.7975
## age1               1.0885     0.9187    1.0521    1.1262
## sex                0.6647     1.5045    0.5467    0.8081
## chd1               1.5930     0.6277    1.2722    1.9947
## cancer1            2.6216     0.3814    2.0046    3.4285
## factor(educ1)1     0.9672     1.0339    0.6885    1.3586
## factor(educ1)2     0.9246     1.0816    0.6746    1.2671
## factor(cigst1)1    1.6565     0.6037    1.2936    2.1211
## factor(cigst1)2    2.3617     0.4234    1.7965    3.1046
## physinact1         2.0415     0.4898    1.6799    2.4810
##
## Concordance= 0.72  (se = 0.014 )
## Rsquare= 0.094   (max possible= 0.898 )
## Likelihood ratio test= 309.3  on 10 df,   p=<2e-16
## Wald test            = 317.9  on 10 df,   p=<2e-16
## Score (logrank) test = 340.1  on 10 df,   p=<2e-16
```

```
summary(coxf.age3)
```

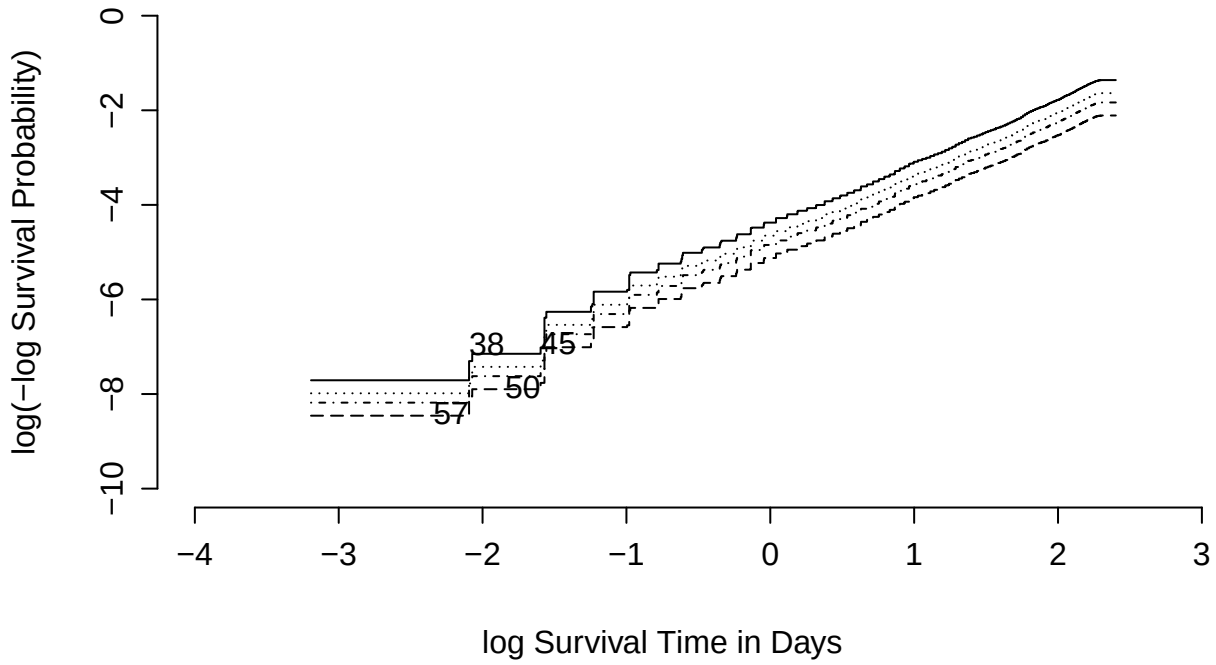
```
## Call:
## coxph(formula = Surv(time, death) ~ cf1_10 + age1 + sex + chd1 +
##     cancer1 + factor(educ1) + factor(cigst1) + physinact1, data = elsaage3,
##     method = "breslow")
##
##   n= 2680, number of events= 1058
##
##                      coef exp(coef)  se(coef)      z Pr(>|z|)
## cf1_10          -0.275977  0.758830  0.030958 -8.915  < 2e-16 ***
## age1             0.070384  1.072920  0.007206  9.767  < 2e-16 ***
## sex             -0.426267  0.652942  0.071158 -5.990 2.09e-09 ***
## chd1             0.123884  1.131884  0.074506  1.663   0.0964 .
## cancer1          0.431101  1.538950  0.100310  4.298 1.73e-05 ***
## factor(educ1)1   0.003740  1.003747  0.124312  0.030   0.9760
## factor(educ1)2   0.012809  1.012892  0.118391  0.108   0.9138
## factor(cigst1)1  0.129655  1.138436  0.072112  1.798   0.0722 .
## factor(cigst1)2  0.543218  1.721538  0.097486  5.572 2.51e-08 ***
## physinact1       0.421812  1.524721  0.065467  6.443 1.17e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## cf1_10             0.7588     1.3178    0.7142    0.8063
## age1               1.0729     0.9320    1.0579    1.0882
## sex                0.6529     1.5315    0.5679    0.7507
## chd1               1.1319     0.8835    0.9781    1.3099
## cancer1            1.5390     0.6498    1.2643    1.8733
## factor(educ1)1     1.0037     0.9963    0.7867    1.2807
## factor(educ1)2     1.0129     0.9873    0.8031    1.2774
## factor(cigst1)1    1.1384     0.8784    0.9884    1.3113
## factor(cigst1)2    1.7215     0.5809    1.4221    2.0840
## physinact1         1.5247     0.6559    1.3411    1.7335
##
## Concordance= 0.667  (se = 0.009 )
## Rsquare= 0.133   (max possible= 0.998 )
```

```
## Likelihood ratio test= 383.2  on 10 df,   p=<2e-16
## Wald test            = 382.6  on 10 df,   p=<2e-16
## Score (logrank) test = 390.3  on 10 df,   p=<2e-16
```

```
#####################################################################
## 3. CHECKING PROPORTIONAL HAZARDS ASSUMPTION                     ##
#####################################################################

# Plot 1- log (-log survival)
coxa <- cph(Surv(time, death) ~ cf1+age1 + sex + chd1,
            data=elsa_cf, x=TRUE,y=TRUE, method="breslow")

survplot(coxa, cf1=c(38,45,50,57), age1=mean(age1), sex=0, chd1=0,
         logt=TRUE, loglog=TRUE, xlim=c(-4, 2.5), ylim=c(-10, 0))
```

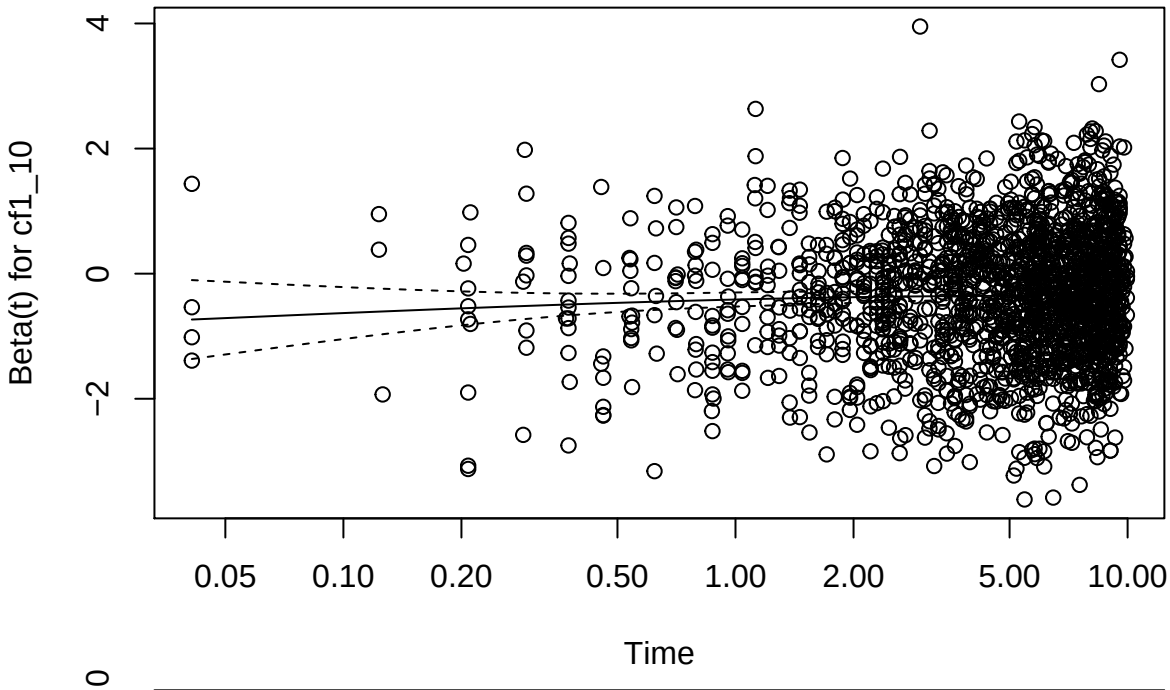


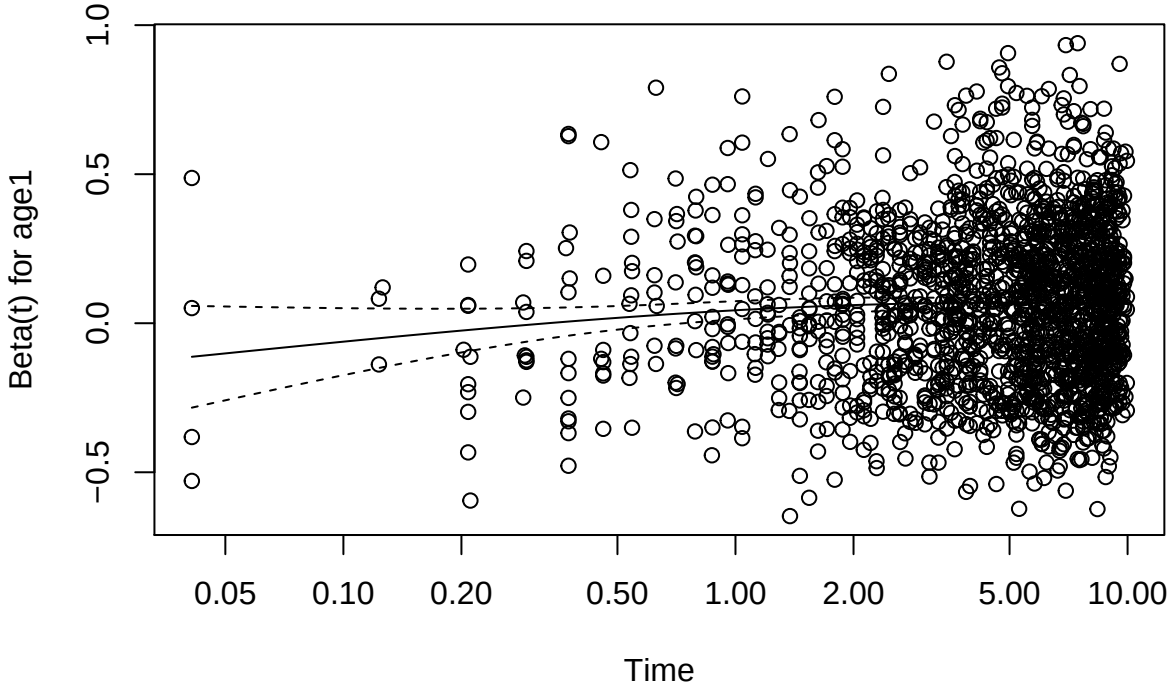

```
####
#Create a time-dependent model and test for interactions with time for all covariates
time.dep <- coxph (Surv(time, death)~ cf1_10 + age1  + sex + chd1 + cancer1 + physinact1 + educ1_ + cigs
                   strata(agegr), data=elsa_cf, method="breslow",  na.action=na.exclude)
time.dep.zph <- cox.zph(time.dep, transform = 'log')
time.dep.zph
```

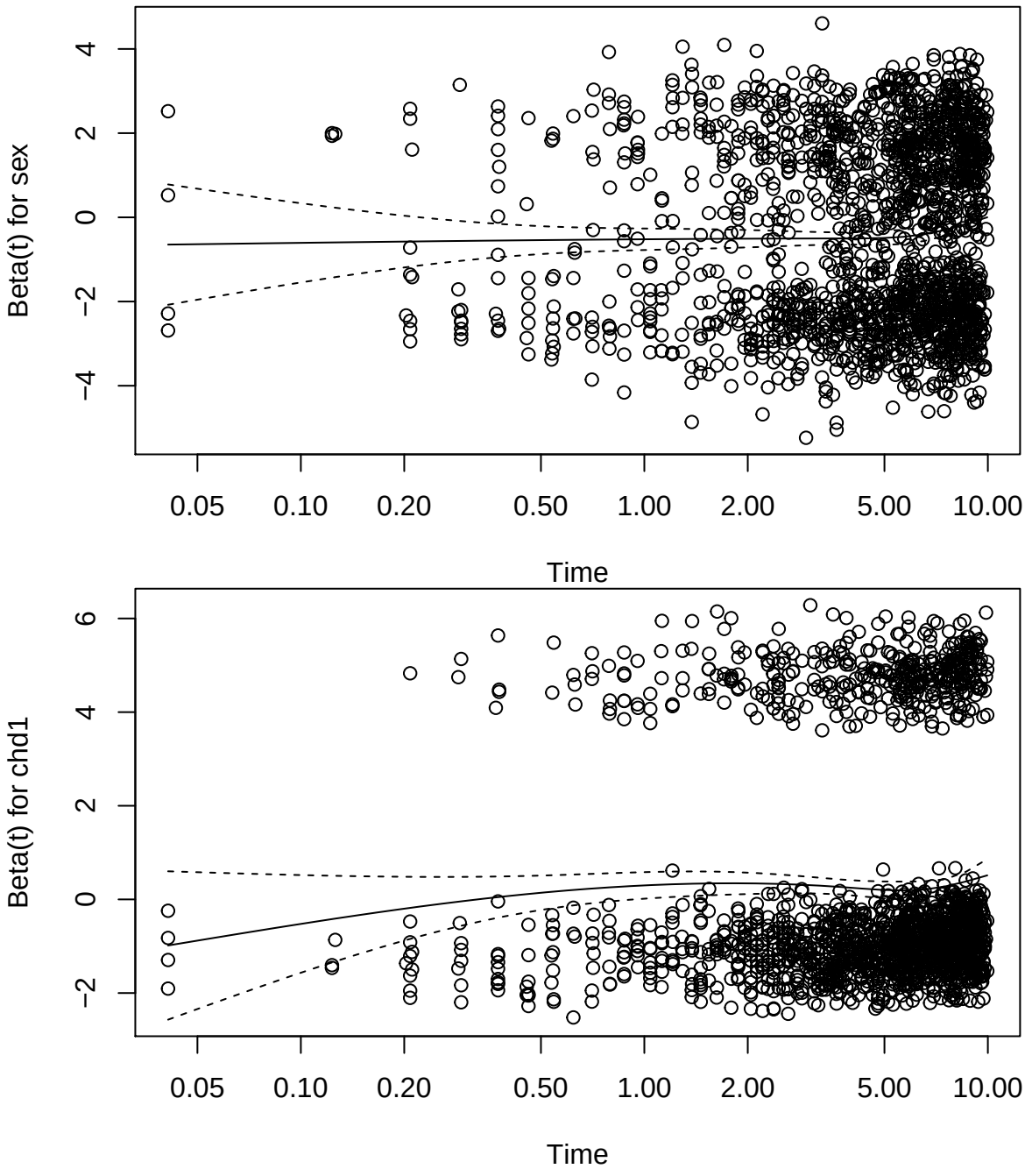
```
##                rho    chisq        p
## cf1_10      0.05421  6.09538 1.36e-02
## age1        0.04702  4.15477 4.15e-02
## sex         0.04764  4.12605 4.22e-02
## chd1        0.02271  0.95357 3.29e-01
## cancer1    -0.08866 14.43839 1.45e-04
## physinact1 -0.09891 17.91530 2.31e-05
## educ1_1     0.00666  0.08061 7.76e-01
## educ1_2     0.01800  0.58677 4.44e-01
```

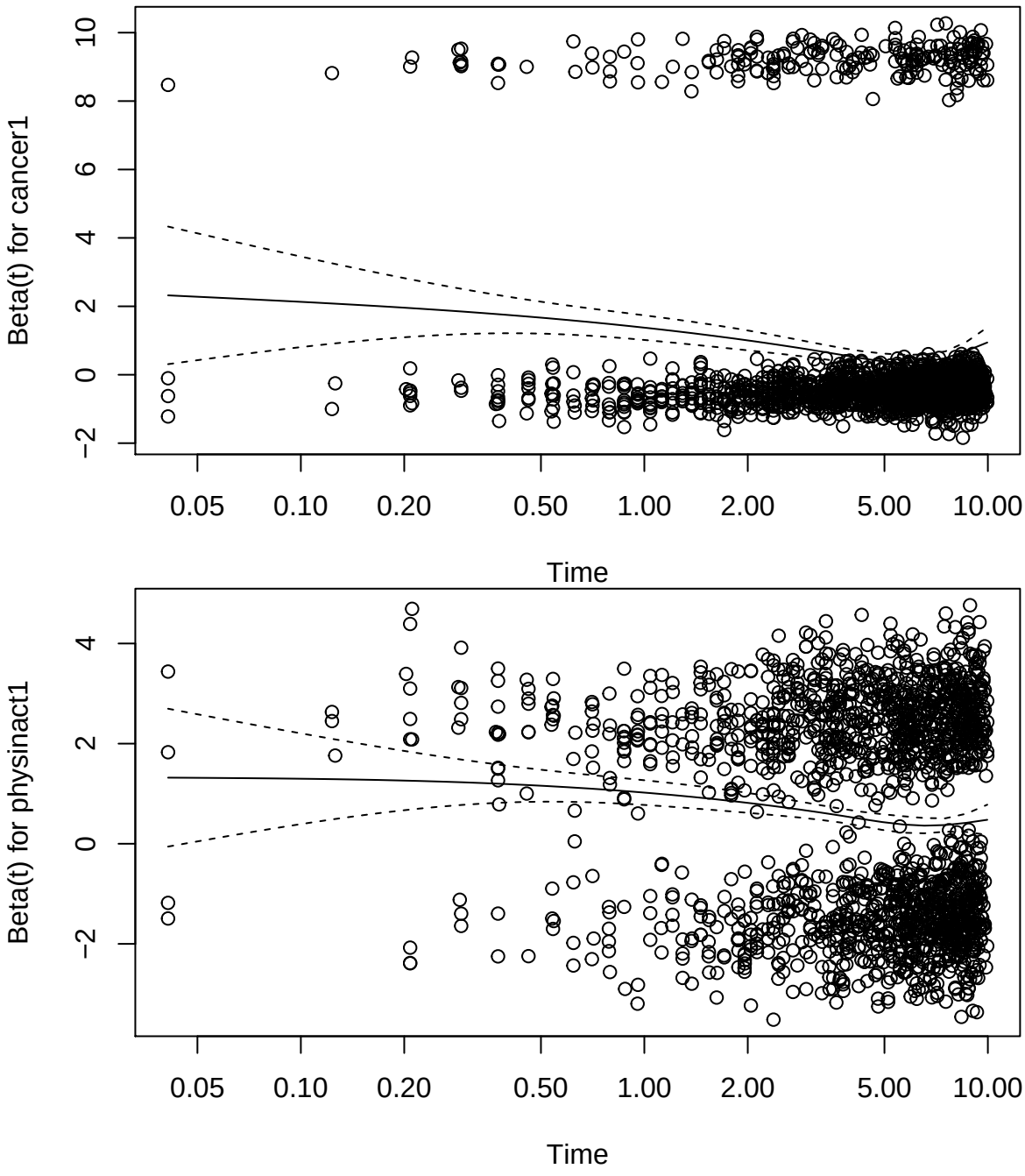
```
## cigst1_1   -0.00198  0.00708 9.33e-01
## cigst1_2   -0.01444  0.37023 5.43e-01
## GLOBAL           NA 49.74832 2.97e-07
```

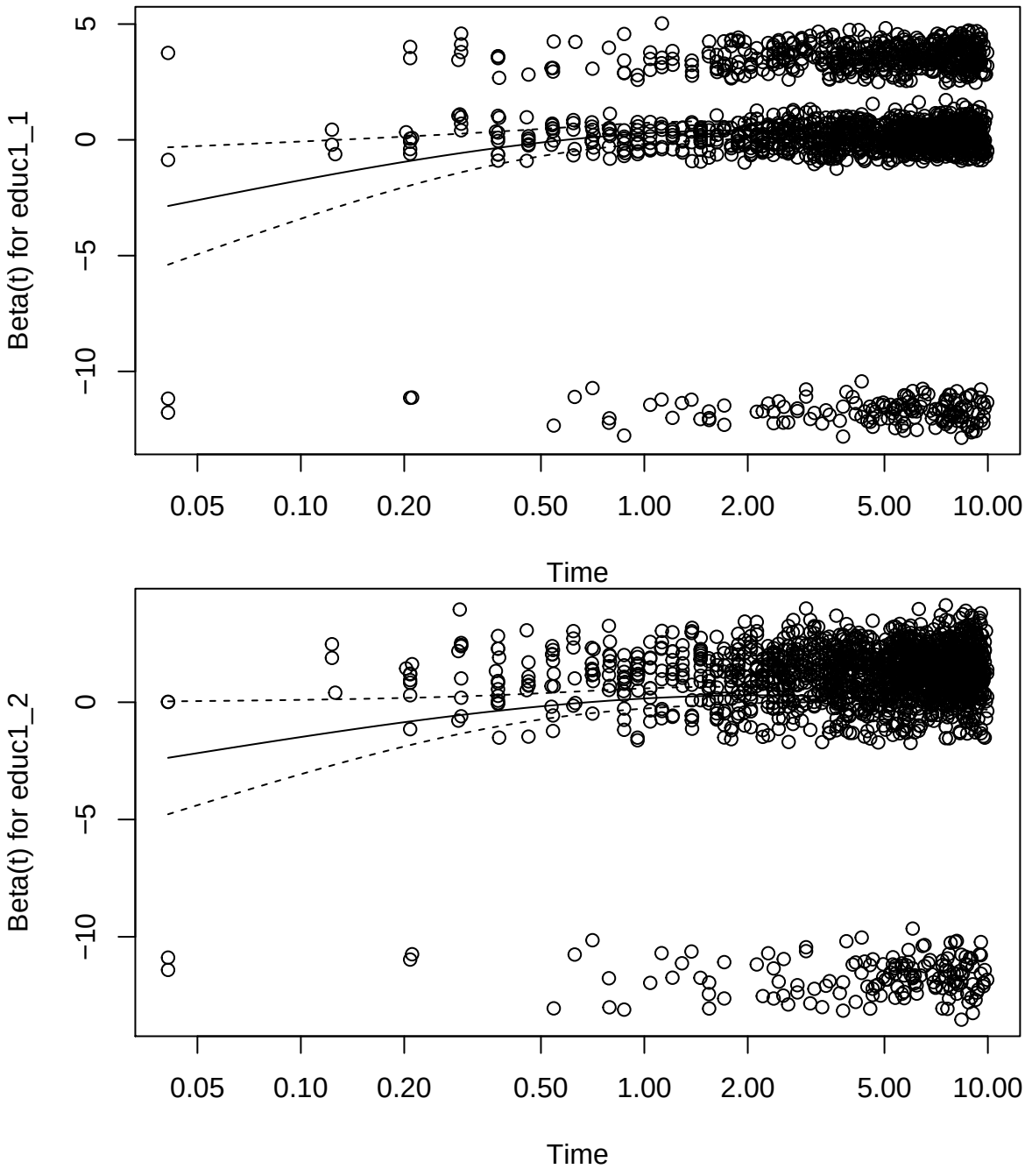
```
# plots of residuals for all predictors (use arrows to go from one plot to the other)
plot(time.dep.zph)
```

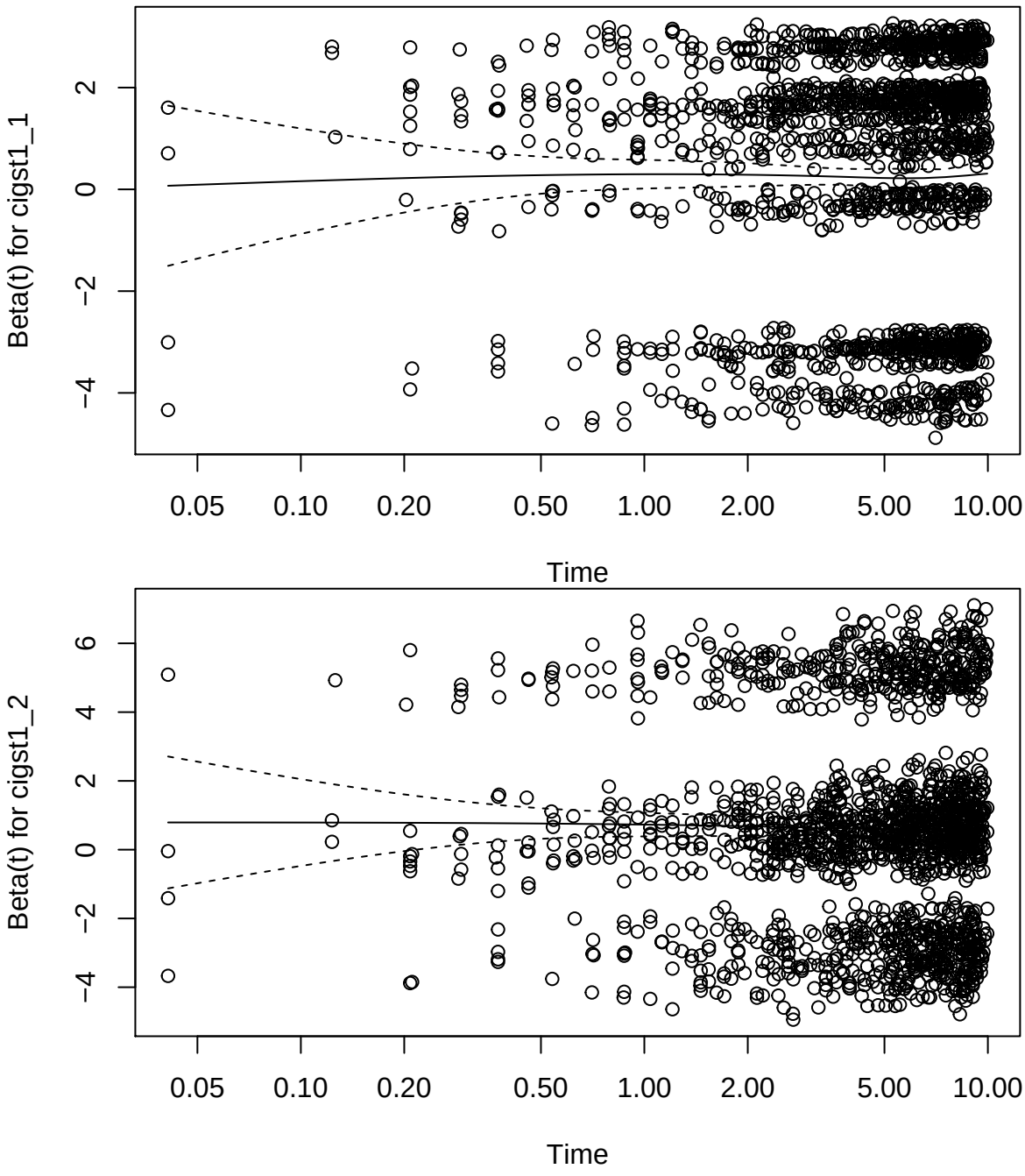




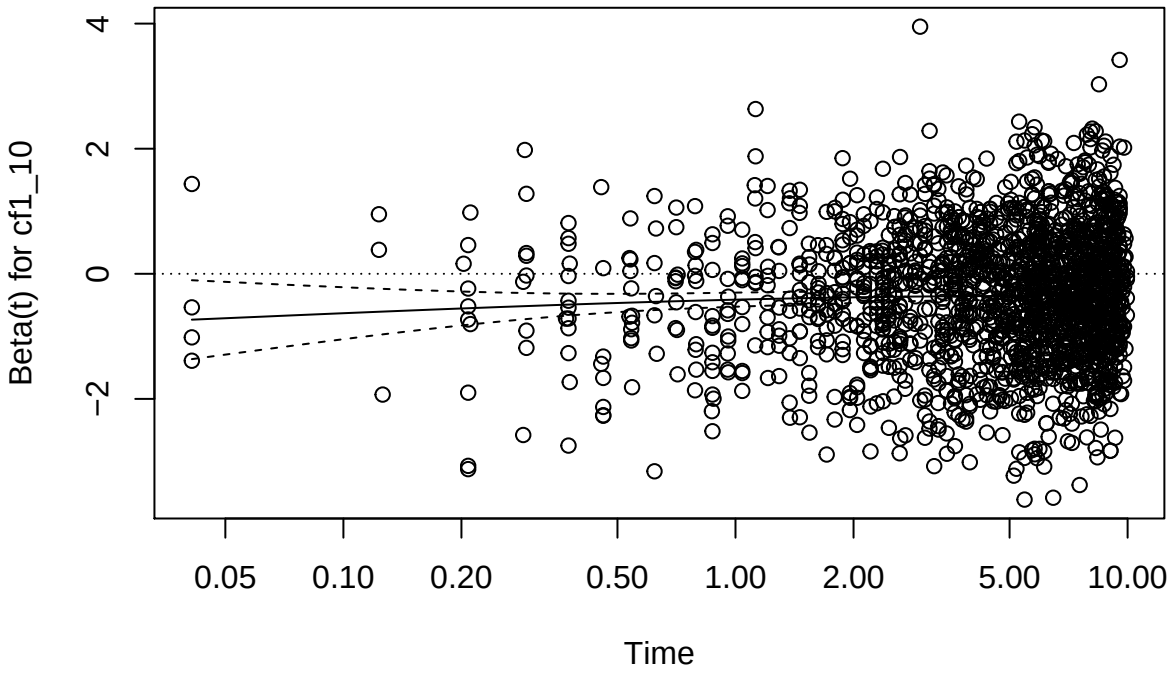

Beta(t) for sex
4
2
0
−2
−4
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time
Beta(t) for chd1
6
4
2
0
−2
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time


Beta(t) for cancer1
10
8
6
4
2
0
−2
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time
Beta(t) for physinact1
4
2
0
−2
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time


Beta(t) for educ1_1
5
0
−5
−10
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time
Beta(t) for educ1_2
0
−5
−10
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time




```
#plots for the predictor number in bracket. In the model, cf1_10 is the first predictor so we are plott
plot(time.dep.zph[1])
abline(h=0, lty=3)
```


Beta(t) for cf1_10
4
2
0
−2
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
Time