Practicals - Cox regression

Epidemiological methods in medical research 2023

2 March 2023

Exercise 1: The Bissau study

In rural Guinea-Bissau, 5274 children under 7 months of age were visited two times at home, with an interval of approximately 6 months. Information about vaccination (BCG, DTP, measles vaccine) was collected at each visit and at second visit, death during follow-up was registered. Other children move away during follow-up or survive until the second visit ('censored'). The dataset `bissau.txt` contain the available information:

```
bissau <- read.table("https://bozenne.github.io/doc/Teaching/bissau.txt",
   header=TRUE)
str(bissau)
```

```
'data.frame':        5274 obs. of  8 variables:
 $ id       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ fuptime  : int  65 161 166 166 161 161 166 166 166 166 ...
 $ fupstatus: chr  "dead" "censored" "censored" "censored" ...
 $ bcg      : chr  "yes" "yes" "no" "yes" ...
 $ dtp      : int  1 2 0 0 0 0 2 1 2 2 ...
 $ age      : int  182 125 69 96 131 26 129 90 119 146 ...
 $ agem     : int  5 4 2 3 4 0 4 2 3 4 ...
 $ dtpany   : logi  TRUE TRUE FALSE FALSE FALSE FALSE ...
```

The relevant varia bles for this practical are:

- `id` : child id.
- `fuptime` : follow-up time in days.
- `fupstatus`: survival indicator at end of follow-up
- `bcg` : whether the child received a BCG vaccine at baseline
- `agem` : age at first visit in (whole) months.
- `dtpany` whether the child received at least 1 dose of DTP vaccine at baseline

0. *Recap'*

We already analyzed this dataset in Practical 1, where we looked at the 6 months risk of death. This was performed aggregating the data into tables:

```
table(bcg = bissau$bcg, stats = bissau$fupstatus)
```

```
     stats
bcg   censored dead
  no      1876   97
  yes     3176  125
```

What was the main issue with this approch?
Would a logistic model handle this issue better?
What aspect of the treatment effect, disregarded so far, will we be able to visualize/investigate?
Hint: you can have a look to the survival curves on the next page.

In this practical we will use time to event models (e.g. Kaplan-Meier, Cox) to assess the vaccine effect.

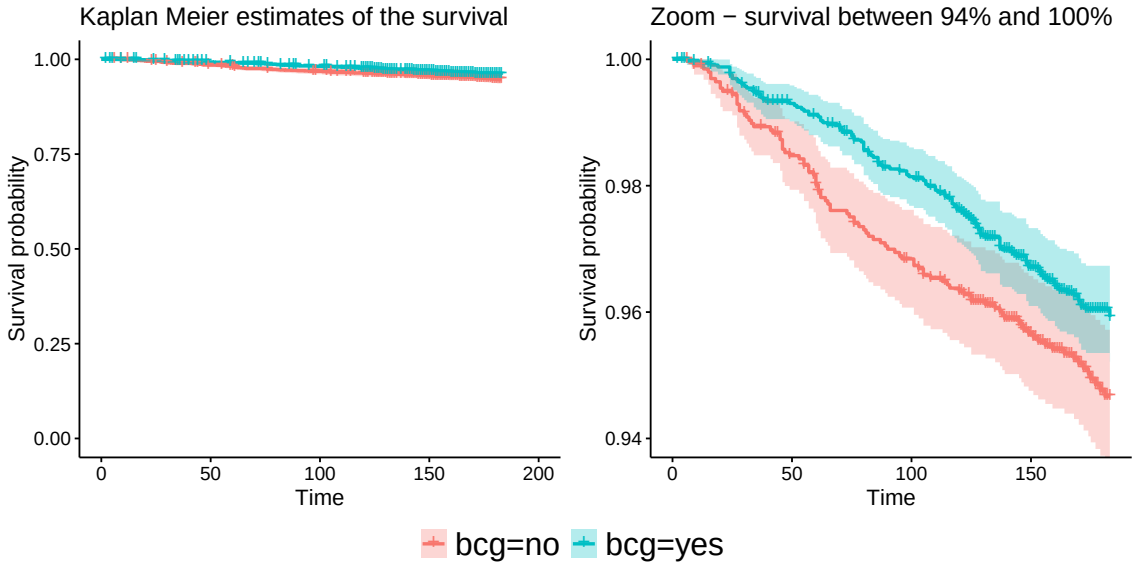
1. *Kaplan-Meier*

Denote t_i the (ordered) times where one or more child died,
d_i the number of deaths that occured at t_i,
$Y(t_i)$ the number of children at risk at t_i,
then the Kaplan Meier estimator is defined by:

$$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{Y(t_i)}\right) \tag{1}$$

We can use this estimator to visualize the survival in each BCG vaccine group:

```
library(survival)
e.KM <- survfit(Surv(fuptime, fupstatus=="dead") ~ bcg, data = bissau)
# default plot method
plot(e.KM, conf.int=TRUE, ylim=c(0.9,1.0))
lines(e.KM,lwd=3)
# alternative plot method
library(survminer)
gg <- ggsurvplot(e.KM, conf.int = TRUE)$plot
gg + coord_cartesian(ylim = c(0.94,1))
```



a) Will you be able to conclude about the (causal) effect of the vaccine by comparing the survival between groups? If not, what would be the use of such a graphical representation.

b) To illustrate what the Kaplan Meier estimator does, we will apply formula (1) for estimating the survival of the non-vaccinated individuals.
We first re-order the dataset by increasing follow-up time.
We then compute the number at risk (i.e. not dead and still in the study) by counting, for each follow-up time `t`, the number of individuals whose follow-up time is equal or greater than `t`:

```
## select vaccinated individuals (and only relevant columns)
bissau.no <- bissau[bissau$bcg == "no",c("id","fuptime","fupstatus")]
## reorder
bissau.no <- bissau.no[order(bissau.no$fuptime),]
## define number at risk
bissau.no$atRisk <- sapply(bissau.no$fuptime,
                           FUN = function(t){ sum(bissau.no$fuptime>=t) })
## display
head(bissau.no)
```

```
       id fuptime fupstatus atRisk
2645 2645       6  censored   1973
1415 1415       8      dead   1972
1739 1739       9  censored   1971
3364 3364       9      dead   1971
3817 3817      12  censored   1969
266   266      13      dead   1968
```

Try to compute ("by hand") the survival at time 0,6,8,9,12 days and compare it to the Kaplan-Meier estimate.

```
print(summary(e.KM, times = c(0,6,8,9,12)), digit = 7)
```

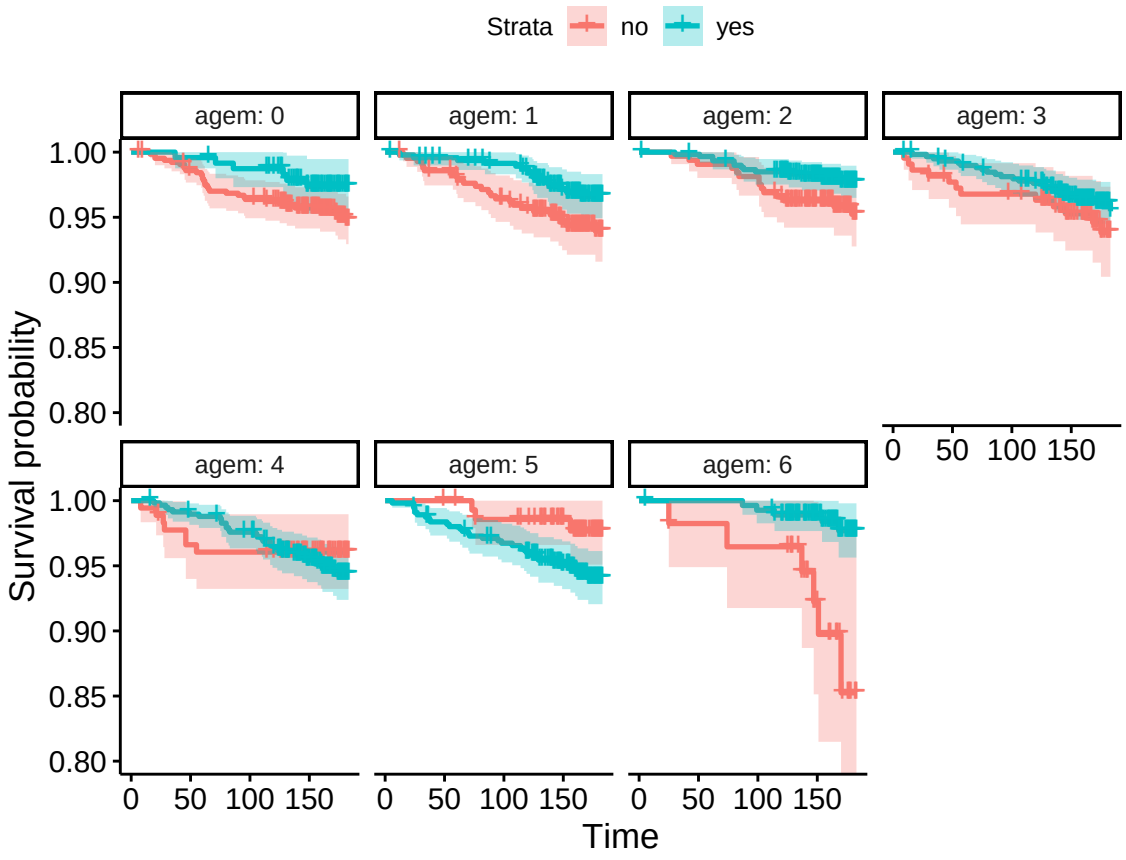
```
[...]
                bcg=no
 time n.risk n.event  survival      std.err lower 95% CI upper 95% CI
    0   1973       0 1.0000000 0.0000000000    1.0000000            1
    6   1973       0 1.0000000 0.0000000000    1.0000000            1
    8   1972       1 0.9994929 0.0005069708    0.9984997            1
    9   1971       1 0.9989858 0.0007167831    0.9975819            1
   12   1969       0 0.9989858 0.0007167831    0.9975819            1

[...]
```

We now would like to estimate the vaccine effect under the assumption that age (in months) is the only confounder.

2. *More Kaplan-Meier*
   A first approach would to use the previous approach within each age group. Is it easy to conclude about the vaccine effect? Discuss the strength/limitations of this approach.
   (keep in mind that the confidence intervals, displayed below as a shaded area, are not adjusted for multiple comparisons across time nor across age groups).

```
e.KMagem <- survfit(Surv(fuptime, fupstatus=="dead") ~ bcg + agem,
                    data = bissau)
ggsurvplot_facet(e.KMagem, facet.by = "agem", conf.int = TRUE,
                 data = bissau, nrow = 2)
```




```
## number of individuals per age and vaccine group
table(bcg = bissau$bcg, agem = bissau$agem)
```

```
     agem
bcg     0   1   2   3   4   5   6
  no  637 421 321 218 178 141  57
  yes 237 468 598 589 581 554 274
```

3. *Cox model (single exposure)*
   Another approach is to fit a Cox regression model using age and vaccine as covariates. The Cox model decomposes the instantaneous risk of death, called hazard and denoted $\lambda(t) = \lambda_0(t)\exp(X\boldsymbol{\beta})$, into two terms:

   - the baseline hazard $\lambda_0(t)$ which represent the influence of time.
   - the linear predictor $X\boldsymbol{\beta}$ which represent the influence of the covariates X (β represents the effect of those covariates, on the log hazard scale).

   The survival can then be computed as $S(t) = \exp(-\Lambda_0(t)\exp(X\boldsymbol{\beta}))$ where $\Lambda_0(t) = \int_{s=0}^{t} \lambda_0(s)ds$, the baseline cumulative hazard, can be understood as sum of the hazard over time in the reference group (no covariate effect).

```
e.coxTime <- coxph(
    Surv( fuptime, fupstatus == "dead") ~ factor(agem) + bcg,
    data = bissau, x = TRUE )
summary(e.coxTime)
```

```
Call:
 coxph(formula = Surv(fuptime, fupstatus == "dead") ~ factor(agem) +
     bcg, data = bissau, x = TRUE)

   n= 5274, number of events= 222

                   coef exp(coef) se(coef)      z Pr(>|z|)
 factor(agem)1  0.11500   1.12187  0.23205  0.496   0.6202
 factor(agem)2 -0.25687   0.77347  0.25861 -0.993   0.3206
 factor(agem)3  0.19894   1.22011  0.24325  0.818   0.4135
 factor(agem)4  0.33252   1.39447  0.24183  1.375   0.1691
 factor(agem)5  0.33066   1.39189  0.24957  1.325   0.1852
 factor(agem)6 -0.01052   0.98953  0.35340 -0.030   0.9762
 bcgyes        -0.34720   0.70667  0.14605 -2.377   0.0174 *
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

               exp(coef) exp(-coef) lower .95 upper .95
 factor(agem)1    1.1219     0.8914    0.7119    1.7679
 factor(agem)2    0.7735     1.2929    0.4659    1.2840
 factor(agem)3    1.2201     0.8196    0.7574    1.9654
 factor(agem)4    1.3945     0.7171    0.8681    2.2401
 factor(agem)5    1.3919     0.7184    0.8534    2.2701
 factor(agem)6    0.9895     1.0106    0.4950    1.9781
 bcgyes           0.7067     1.4151    0.5308    0.9409
```

a) Under this Cox model, what can you say about the vaccine effect?

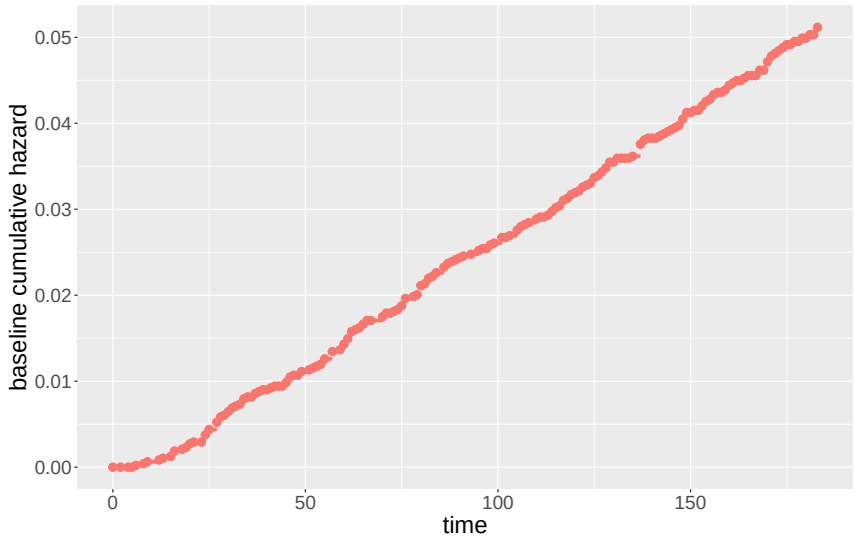
b) What assumptions are we making compared to the Kaplan Meier approach? If you are in doubt you can have a look to the modeled survival curves (see appendix A) and compare them to the Kaplan Meier curves.

c) The cumulative baseline hazard can also be estimated from the Cox model (see figure below). What does it mean that the cumulative hazard is approximately linear?

```
library(riskRegression)
e.cumhaz <- predictCox(e.coxTime)
df.cumhaz <- data.frame(time = e.cumhaz$time,
                        cumhazard = e.cumhaz$cumhazard)
df.cumhaz[c(1:5,48,81,82),]
```

```
   time    cumhazard
1     2 0.0000000000
2     4 0.0000000000
3     5 0.0000000000
4     6 0.0002081311
5     8 0.0004163482
48   60 0.0142977094
81   99 0.0260653240
82  101 0.0267120847
```

```
library(ggplot2)
autoplot(e.cumhaz, type = "cumhaz")
```



d) Compare the results of the Cox model with the ones of the Poisson regression. Is it surprising?

```
e.glmBCG <- glm(fupstatus=="dead" ~ bcg + factor(agem),
                offset = log(fuptime), family = poisson, data = bissau)
Epi::ci.exp(e.glmBCG)
```

```
                exp(Est.)         2.5%        97.5%
(Intercept)   0.000277131 0.0001996611 0.0003846598
bcgyes        0.708006220 0.5317786614 0.9426343033
factor(agem)1 1.122316129 0.7121835942 1.7686359301
factor(agem)2 0.774247396 0.4664012673 1.2852860232
factor(agem)3 1.219971430 0.7573584948 1.9651595631
factor(agem)4 1.393708253 0.8676134569 2.2388111647
factor(agem)5 1.390276273 0.8524477945 2.2674328298
factor(agem)6 0.984781408 0.4927412072 1.9681617996
```

e) [**advanced**] Based one the cumulative baseline hazard and the estimate coefficients, we you "manually" compute the survival that is displayed in appendix A. What would be survival probability at 60 days of a 2-month old child that was not vaccinated. At 100 days? Repeated the calculations for a vaccinated child.

```
predictCox(e.coxTime, time = c(60,100),
           newdata = data.frame(bcg = c("no","yes"), agem = "2"),
           keep.newdata = TRUE)
```

```
   observation agem bcg times cumhazard survival
1:           1    2  no    60   0.01106    0.989
2:           2    2 yes    60   0.00781    0.992
3:           1    2  no   100   0.02016    0.980
4:           2    2 yes   100   0.01425    0.986
```

4. *Cox models (two exposures)* We would like now to study the effect of `dtpany` and `bcg`.

   a) We first consider a Cox model with only additive effects. What is the interpretation of the `bcg` and `dtpany` coefficients?

```
e.coxTime2 <- coxph(
    Surv( fuptime, fupstatus == "dead") ~ factor(agem) + bcg + dtpany,
    data = bissau )
summary(e.coxTime2)
```

```
 [...]
                  coef exp(coef) se(coef)      z Pr(>|z|)
 bcgyes        -0.5525    0.5755   0.1944 -2.842  0.00448 **
 dtpanyTRUE     0.3689    1.4461   0.2165  1.704  0.08838 .

               exp(coef) exp(-coef) lower .95 upper .95
 bcgyes           0.5755     1.7376    0.3932    0.8424
 dtpanyTRUE       1.4461     0.6915    0.9461    2.2103

 [...]
```

b) We now consider a Cox model with an interaction between the two vaccines. What is the interpretation of the `bcg`, `dtpany`, and `bcg:dtpany` coefficients? The tables in appendix B may be helpful to understand the parametrisation of the model.

```
e.coxTime3 <- coxph(
    Surv( fuptime, fupstatus == "dead") ~ factor(agem) + bcg * dtpany,
    data = bissau )
summary(e.coxTime3)
```

```
 [...]
                      coef exp(coef) se(coef)      z Pr(>|z|)
 bcgyes            -0.56991   0.56557  0.20445 -2.788  0.00531 **
 dtpanyTRUE         0.17430   1.19041  0.72170  0.242  0.80916
 bcgyes:dtpanyTRUE  0.21183   1.23594  0.74364  0.285  0.77576

                   exp(coef) exp(-coef) lower .95 upper .95
 bcgyes               0.5656     1.7681    0.3788    0.8443
 dtpanyTRUE           1.1904     0.8400    0.2893    4.8979
 bcgyes:dtpanyTRUE    1.2359     0.8091    0.2877    5.3086
 [...]
```

c) In the model with interactions, how would you test whether there is evidence for a combined effect of the two vaccines greater (or worse) than sum of the effects of each vaccine when used alone?
How would you test whether there is an effect of any vaccine?
How would you test whether there is an effect of bcg?

```
library(multcomp)
e.glht <- glht(e.coxTime3,
               linfct = c("bcgyes=0",
                          "dtpanyTRUE=0",
                          "bcgyes+dtpanyTRUE+bcgyes:dtpanyTRUE=0"))
summary(e.glht)
```

```
Linear Hypotheses:
                                                Estimate Std. Error z value Pr(>|z|)
 bcgyes == 0                                     -0.5699     0.2044  -2.788   0.0157 *
 dtpanyTRUE == 0                                  0.1743     0.7217   0.242   0.9926
 bcgyes + dtpanyTRUE + bcgyes:dtpanyTRUE == 0  -0.1838     0.1747  -1.052   0.6362
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
 (Adjusted p values reported -- single-step method)
```

```
e.glht2 <- glht(e.coxTime3,
                linfct = c("bcgyes=0",
                           "bcgyes+bcgyes:dtpanyTRUE=0"))
summary(e.glht2)
```

```
Linear Hypotheses:
                                   Estimate Std. Error z value Pr(>|z|)
 bcgyes == 0                        -0.5699     0.2044  -2.788   0.0106 *
 bcgyes + bcgyes:dtpanyTRUE == 0  -0.3581     0.7150  -0.501   0.8529
 ---
 Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
 (Adjusted p values reported -- single-step method)
```

5. *PH assumption*

a) We can test the proportional hazard (PH) assumption using scaled Schoenfeld residuals[1] The null hypothesis is that the PH assumption holds for each covariate (or globally). How would you interpret a rejection of the PH assumption for the vaccine effect? Would that "complicates" how the vaccine effect should be reported?

```
cox.zph(e.coxTime)
```

```
             chisq df     p
factor(agem)   7.3  6 0.294
bcg            4.9  1 0.027
GLOBAL        11.2  7 0.130
```

b) [**advanced**] A non-parametric estimator of the vaccine effect over time (see appendix C) suggests a change in the vaccination effect after 75 days. What can you conclude about the vaccination effect?

```
bissau$bcg.num <- bissau$bcg=="yes"
e.coxTime.bis <- coxph(
    Surv( fuptime, fupstatus == "dead") ~ factor(agem) + bcg.num + tt(bcg.
    num),
    tt = function(x, t,...) x*(t > 75),
    data = bissau)
summary(e.coxTime.bis)
```

```
 [...]
                    coef exp(coef)  se(coef)      z Pr(>|z|)
 bcg.numTRUE   -0.790783  0.453490  0.220106 -3.593 0.000327 ***
 tt(bcg.num)    0.752570  2.122447  0.279306  2.694 0.007051 **
 ---
               exp(coef) exp(-coef) lower .95 upper .95
 bcg.numTRUE      0.4535     2.2051    0.2946    0.6981
 tt(bcg.num)      2.1224     0.4712    1.2277    3.6693

 [...]
```

[1] what are those residuals and why they can be used to test the PH is out of the scope of this exercise.

6. *Choice of the time scale*
   Instead of using the follow-up time, we could use age to define the time scale when fitting the Cox model. Compare the estimated effects with the ones of question 4 (`e.coxTime2`).

```
bissau$outage <- bissau$age + bissau$fuptime
e.coxAge <- coxph(
    Surv(age, outage, fupstatus == "dead") ~ bcg * dtpany,
    data = bissau)
summary(e.coxAge)
```

```
[...]
                     coef exp(coef) se(coef)      z Pr(>|z|)
bcgyes            -0.5763    0.5620   0.2023 -2.848   0.0044 **
dtpanyTRUE         0.1273    1.1357   0.7178  0.177   0.8592
bcgyes:dtpanyTRUE  0.2200    1.2461   0.7429  0.296   0.7671

                  exp(coef) exp(-coef) lower .95 upper .95
bcgyes                0.562     1.7795    0.3780    0.8355
dtpanyTRUE            1.136     0.8805    0.2782    4.6370
bcgyes:dtpanyTRUE     1.246     0.8025    0.2905    5.3447
[...]
```

Exercise 2: IHD data from Clayton & Hills

The study is described by Clayton & Hills, Ch. 13. The dataset diet.txt contains one record for each of the 337 subjects in the data set (and variable names in the first record). Note that energy intake is given as a quantitative variable. The data set has the following variables:

`id`	Person id
`doe`	Date of entry (format: MM/DD/YYYY)
`dox`	Date of exit (format: MM/DD/YYYY)
`chd`	Coronary Heart Disease status at exit: 0-no, 1-yes
`dob`	Date of birth (format: MM/DD/YYYY)
`job`	Not used
`month`	Not used
`energy`	Daily energy intake (Mcal)
`height`	Height (cm)
`weight`	Weight (kg)
`fat`	Daily fat intake (g)
`fibre`	Daily fibre intake (g)

1. Read the individual diet data records from the file (SAS users may use the program `ihdindiv.sas`) and create variables for *age at entry* by subtracting date of birth from date of entry and for the *person-years* by subtracting date of entry from date of exit. Also create a variable with the log(person-years).

2. Use `chd` as outcome variable in a Poisson regression model with the log(person-years) as offset, using `energy` as a linear explanatory variable and adjusting for age at entry as a linear variable. Is there an effect of `energy` on mortality?

3. Does this change if the effect of age at entry is modeled using a linear spline?

4. Is there any evidence of a non-linear effect of energy, when using linear splines with knots at say 2, 2.5 and 3? (these numbers are approximately the quartiles in the energy-distribution).

Extra questions to be used if time permits:

5. The Poisson models you just fitted implicitly assume that the rates of CHD are constant over time. Try to relax this assumption by fitting the Cox model corresponding to question 2, with time since study entry (i.e., person-years) as time variable. Do the regresssion parameters change much?

6. Try to use current age as underlying time variable in the Cox model instead. Does the exposure effect change much?
   Hint: You must compute age at exit and use this plus age at entry when fitting the Cox model.

Appendix A: Display of the survival curves in a Cox model

We first load a few packages

```
library(riskRegression) ## ease extraction of the survival values
library(colorspace) ## work with colors
```

We then create a dataset containing all possible combinations of age and vaccine status:

```
df.grid <- unique(bissau[,c("bcg","agem")])
df.grid <- df.grid[order(df.grid$bcg,df.grid$agem),] ## re-order lines
df.grid
```

```
     bcg agem          bcg agem          bcg agem
78    no    0     20   no    5      4   yes    3
42    no    1     25   no    6      2   yes    4
3     no    2     6   yes    0      1   yes    5
17    no    3     29  yes    1     22  yes    6
100   no    4     8   yes    2
```

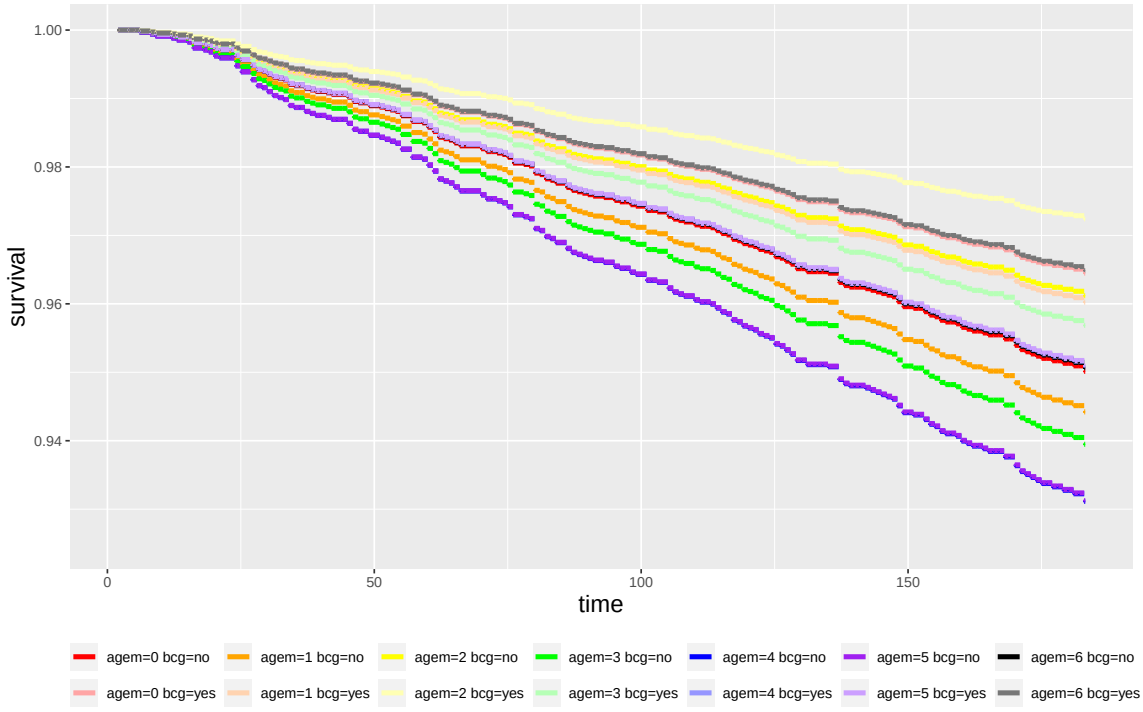
We call the `predictCox` function to extract the survival values at each time point corresponding to each combination of age and vaccine status:

```
pred.cox <- predictCox(e.coxTime,
                       newdata = df.grid,
                       times = sort(unique(bissau$fuptime)),
                       keep.newdata = TRUE)
pred.cox
```

```
      observation agem bcg times cumhazard survival
   1:           1    0  no     2    0.0000    1.000
   2:           2    1  no     2    0.0000    1.000
   3:           3    2  no     2    0.0000    1.000
   4:           4    3  no     2    0.0000    1.000
   5:           5    4  no     2    0.0000    1.000
  ---
2250:          10    2 yes   183    0.0280    0.972
2251:          11    3 yes   183    0.0441    0.957
2252:          12    4 yes   183    0.0504    0.951
2253:          13    5 yes   183    0.0503    0.951
2254:          14    6 yes   183    0.0358    0.965
```

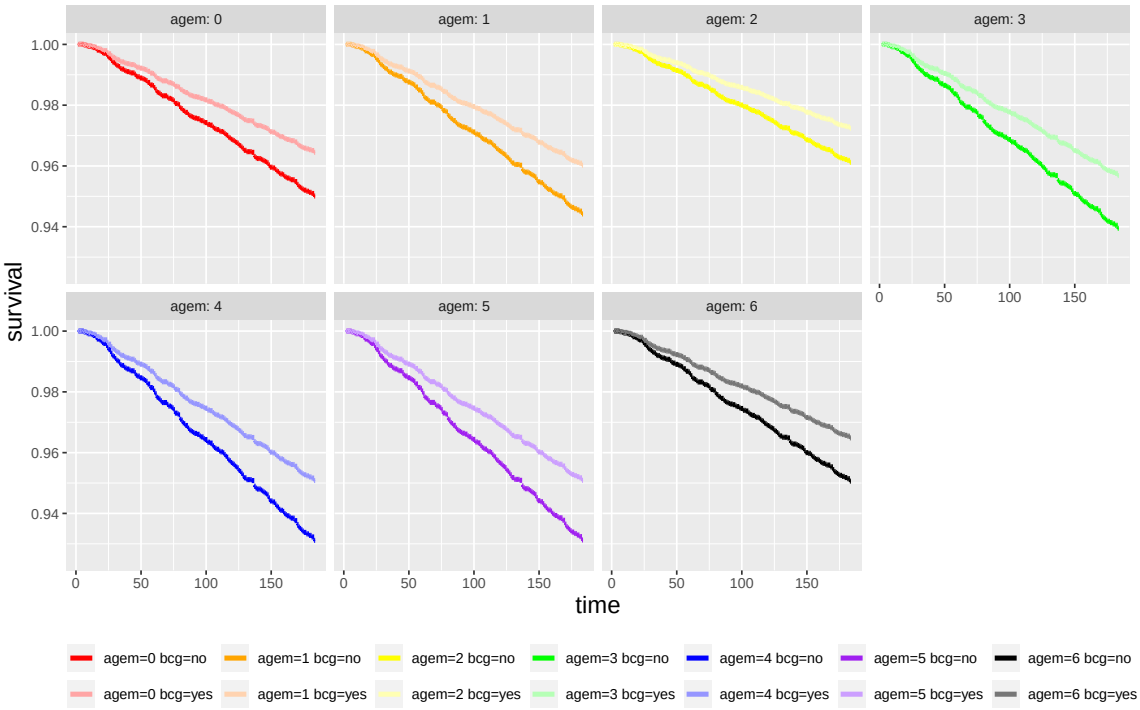
We then display the values either on a single graph:

```
gg <- autoplot(pred.cox, group.by = "covariate", type = "survival",
               size.point = 0, plot = FALSE)$plot
## fix y scale
gg <- gg + coord_cartesian(ylim = c(0.925,1))
## put colors specific to each age group (dark v)
col <- c("red","orange","yellow","green","blue","purple","black")
gg <- gg + scale_colour_manual(values = c(col, lighten(col,0.5)))
## make the caption and axis labels more readable
gg <- gg + guides(color = guide_legend(ncol = 7, byrow = TRUE))
gg <- gg + labs(color = "")
gg <- gg + theme(legend.position = "bottom",
                 text = element_text(size=11),
                 axis.title = element_text(size=15))
gg
```



or using a graph per age group:

```
gg + facet_wrap(~agem, labeller = label_both, nrow = 2)
```



Appendix B: Parametrisation of the hazard in the various Cox models

We give here the expression of the instantaneous hazard $\lambda(t) = \lambda_0(t)\exp(X\boldsymbol{\beta})$ for various models.

- `e.coxTime`: Cox model with an additive age $(\alpha_1, \ldots, \alpha_6)$ and bcg (β) effect

	no dtp		dtp	
agem	no bcg	bcg	no bcg	bcg
0	$\lambda_0(t)$	$\lambda_0(t)\exp(\beta)$	$\lambda_0(t)$	$\lambda_0(t)\exp(\beta)$
1	$\lambda_0(t)\exp(\alpha_1)$	$\lambda_0(t)\exp(\alpha_1+\beta)$	$\lambda_0(t)\exp(\alpha_1)$	$\lambda_0(t)\exp(\alpha_1+\beta)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	$\lambda_0(t)\exp(\alpha_6)$	$\lambda_0(t)\exp(\alpha_6+\beta)$	$\lambda_0(t)\exp(\alpha_6)$	$\lambda_0(t)\exp(\alpha_6+\beta)$

- `e.coxTime2`: Cox model with an additive age $(\alpha_1, \ldots, \alpha_6)$, bcg (β), and dtp (γ) effect

	no dtp		dtp	
agem	no bcg	bcg	no bcg	bcg
0	$\lambda_0(t)$	$\lambda_0(t)\exp(\beta)$	$\lambda_0(t)\exp(\gamma)$	$\lambda_0(t)\exp(\beta+\gamma)$
1	$\lambda_0(t)\exp(\alpha_1)$	$\lambda_0(t)\exp(\alpha_1+\beta)$	$\lambda_0(t)\exp(\alpha_1+\gamma)$	$\lambda_0(t)\exp(\alpha_1+\beta+\gamma)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	$\lambda_0(t)\exp(\alpha_6)$	$\lambda_0(t)\exp(\alpha_6+\beta)$	$\lambda_0(t)\exp(\alpha_6+\gamma)$	$\lambda_0(t)\exp(\alpha_6+\beta+\gamma)$

- `e.coxTime3`: Cox model with an additive age $(\alpha_1, \ldots, \alpha_6)$, bcg (β), dtp (γ) effect, and an interaction (δ):

	no dtp		dtp	
agem	no bcg	bcg	no bcg	bcg
0	$\lambda_0(t)$	$\lambda_0(t)\exp(\beta)$	$\lambda_0(t)\exp(\gamma)$	$\lambda_0(t)\exp(\beta+\gamma+\delta)$
1	$\lambda_0(t)\exp(\alpha_1)$	$\lambda_0(t)\exp(\alpha_1+\beta)$	$\lambda_0(t)\exp(\alpha_1+\gamma)$	$\lambda_0(t)\exp(\alpha_1+\beta+\gamma+\delta)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	$\lambda_0(t)\exp(\alpha_6)$	$\lambda_0(t)\exp(\alpha_6+\beta)$	$\lambda_0(t)\exp(\alpha_6+\gamma)$	$\lambda_0(t)\exp(\alpha_6+\beta+\gamma+\delta)$

Note: the survival $S(t) = \exp(-\Lambda_0(t)\exp(X\boldsymbol{\beta})) = S_0(t)^{\exp(X\boldsymbol{\beta})}$ is a function of:

- $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ which is the "sum" over time of the instantaneous hazard for the reference group. It is often called cumulative baseline hazard.

- $X\boldsymbol{\beta}$ which is the sum of all covariate effects (on the log scale, e.g. $X\boldsymbol{\beta} = \alpha_1 + \beta$). It is often called linear predictor.

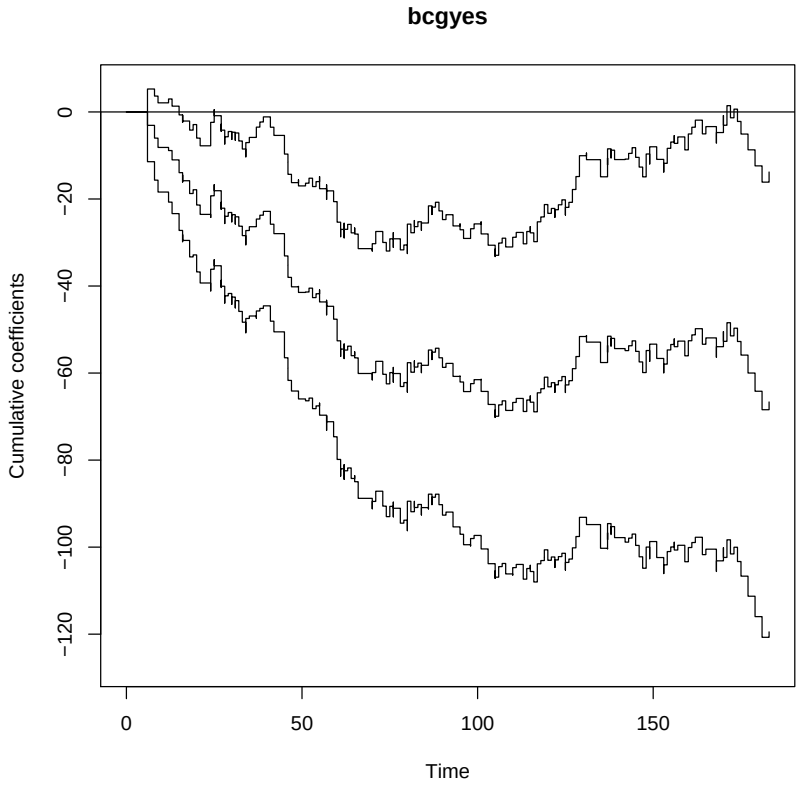
Appendix C: Cox model with time varying coefficient

We use the `timecox` function of the *timereg* package to fit a Cox model with a time varying bcg coefficient estimated non-parametrically. We force the age effect to be constant over time using `const`:

```
library(timereg)
e.tcoxTime <- timecox(
    Surv( fuptime, fupstatus == "dead") ~ const(factor(agem)) + bcg,
    data = bissau, max.time = 150)
```

We can then display the estimated bcg effect over time:

```
plot(e.tcoxTime)
```



The corresponding Cox model is:

$$\lambda(t) = \lambda_0(t) \exp\left(\text{age}\,\alpha + \text{bcg}\,\beta(t)\right)$$

and the graph displays the cumulative regression function:

$$B(t) = \int_0^t \beta(s) ds$$