Epidemiologic data analysis using R

Practicals 9

Faculty of Social Sciences, University of Tampere

–

Janne Pitkaniemi / Karri Seppä

23.3.2020

# Topics of practical 9

Learning objectives of this practical

* Compare estimates and interpretations of different survival measures (overall, cause-specific and relative survival, and 1-cumulative incidence)
* Age standardisation of relative survival
* Piecewise constant excess hazard model for relative survival

# Survival analysis: Rectal cancer patients

## Description of the data

Data frame sire includes observations of a simulated cohort of 8243 female Finnish rectal cancer patients diagnosed between 1993-2012. The dataset contains the following variables:

|  |  |
| --- | --- |
| variable | description |
| sex | 0 = Male, 1 = Female |
| bi\_date | date of birth |
| dg\_date | date of cancer diagnosis |
| ex\_date | date of exit from follow-up (death or censoring) |
| status | status at the end of follow-up (numeric): 0 = alive, 1 = death from rectal cancer, 2 = death from other causes. |
| dg\_age | age at diagnosis (in fractional years) |

## Loading the packages and the data

Load the R packages Epi and popEpi needed in this exercise.

library(Epi)  
library(popEpi)  
#mortality rates of the population of Finland (included in popEpi package)  
pm <- popmort

First, let’s estimate the different measures of survival by using survtab function in popEpi package.

## Lexis object with multi-state set-up

sire$agegr <- cut(sire$dg\_age,c(0,45,55,65,75,Inf),right=F)  
  
Lex <- Lexis(entry = list(FUT = 0, AGE = dg\_age,   
 CAL = cal.yr(dg\_date)),  
 exit = list(CAL = cal.yr(ex\_date)),  
 data = sire,  
 exit.status = factor(status, levels = 0:2,  
 labels = c("alive", "canD", "othD")),  
 merge = TRUE)

NOTE: entry.status has been set to "alive" for all.  
NOTE: Dropping 16 rows with duration of follow up < tol

## Overall survival

In overall survival, outcome event is death from any cause.

#piecewise constant hazard using monthly intervals  
BL <- list(FUT=seq(0, 15, by = 1/12))  
#overall survival  
st.obs <- survtab(  
 Surv(time=FUT, event=lex.Xst%in%c("canD", "othD")) ~ 1,   
 data = Lex,surv.type = "surv.obs",breaks = BL)  
#summary(st.obs,t=c(1,5,10,15))

## 1 - Cumulative incidence of rectal cancer death

One minus cumulative incidence = Proportion of patients who avoid death from rectal cancer in the presence of other causes of death. Rectal cancer death is avoided if a patient died from other causes than cancer.

st.cif <- survtab(Surv(time = FUT, event = lex.Xst) ~ 1,   
 data = Lex,   
 surv.type = "cif.obs",breaks = BL)  
#summary(st.cif,t=c(1,5,10,15))

## Relative survival

Relative survival gives an estimate of net survival = the proportion of patients alive in the absence of other causes of death than cancer

names(pm) <- c("sex","CAL","AGE","haz")  
w <- table(Lex$agegr)  
w <- list(agegr = as.numeric(w))  
st.rel <- survtab(  
 Surv(time = FUT, event = lex.Xst%in%c("canD", "othD"))   
 ~ adjust(agegr), data = Lex, pophaz = pm,  
 surv.type = "surv.rel",breaks = BL, weights=w)  
#summary(st.rel,t=c(1,5,10,15))

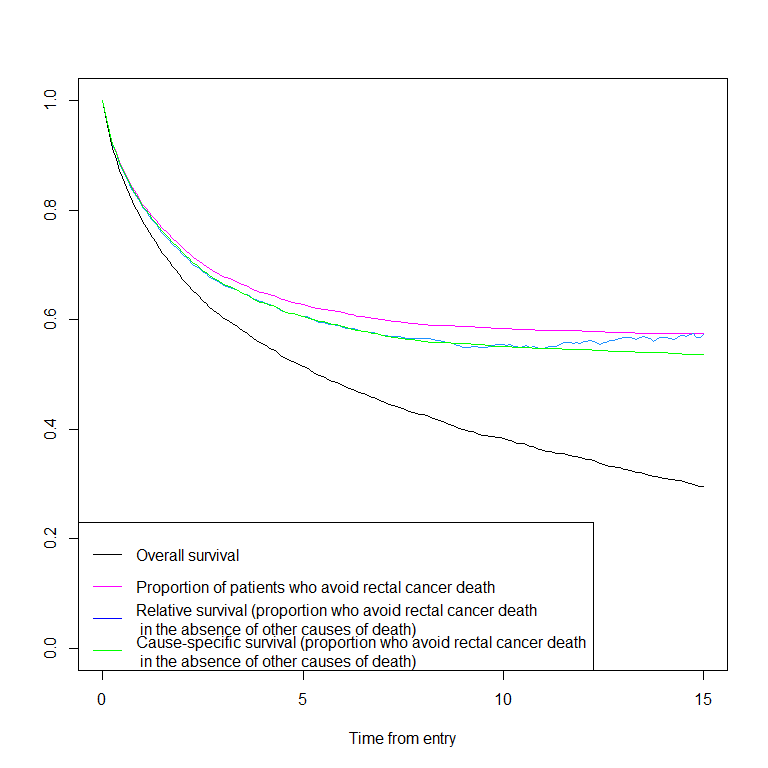
By default, survtab function uses the Ederer II estimator for relative survival.

## Cause-specific survival (event = death from rectal cancer)

In cause-specific survival (normal Kaplan-Meier), outcome event is death from rectal cancer. It gives an estimate of net survival = the proportion of patients alive in the absence of other causes of death than cancer

st.cause<- survtab(  
 Surv(time = FUT, event = lex.Xst) ~ adjust(agegr),   
 data = Lex, pophaz = pm,  
 surv.type = "surv.cause",breaks = BL, weights=w)  
#summary(st.cause,t=c(1,5,10,15))

plot(st.obs,conf.int=F,ylim=c(0,1),ylab="")  
lines(st.cif$Tstop,1-st.cif$CIF\_canD,col="magenta")  
lines(st.rel,y="r.e2.as",conf.int=F,col="dodgerblue")  
lines(st.cause,y="surv.obs.canD.as",conf.int=F,col="green")  
legend("bottomleft",c("Overall survival",  
 "Proportion of patients who avoid rectal cancer death",  
 "Relative survival (proportion who avoid rectal cancer death\n in the absence of other causes of death)",  
 "Cause-specific survival (proportion who avoid rectal cancer death\n in the absence of other causes of death)"),  
 col=c("black","magenta","blue","green"),lty=1)



Compare the estimates and their interpretation. Remember that the probability interpretation of relative survival and cause-specific survival relies on the assumption that competing risks are independent, and you never know whether it is true or not.

Relative and cause-specific survival are, however, useful summary measures of excess mortality and cause-specific mortality, respectively, when the aim is to compare cancer survival between calendar periods (or e.g. countries) between which mortality due to other causes of death may also differ.

# Age-standardised relative survival

Has rectal cancer survival in improved from 1993-2002 to 2003-2012. Estimate the age-standardised relative survival using ICSS weights and the Pohar Perme estimator.

Lex$period <- cut( cal.yr(Lex$dg\_date), c(1993,2003,2013), right=F,dig.lab = 4)  
ICSS1 <- aggregate(ICSS1~cut(ICSS$age,   
 c(0,45,55,65,75,Inf), right = FALSE),   
 data = ICSS, FUN = sum)  
w <- list(agegr = ICSS1$ICSS1)  
st.rel <- survtab(  
 Surv(time = FUT, event = lex.Xst%in%c("canD", "othD"))   
 ~ sex + period + adjust(agegr),   
 data = Lex,pophaz = pm,surv.type = "surv.rel",  
 relsurv.method = "pp",breaks = BL, weights=w)  
summary(st.rel,t=c(1,5,10,15))

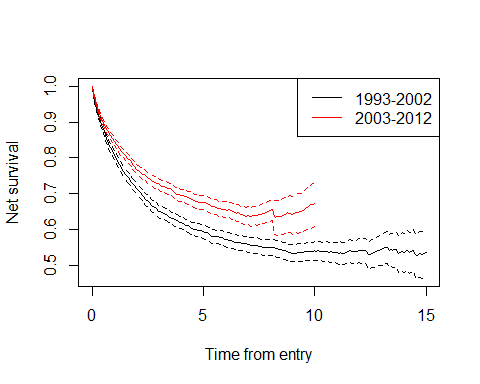
sex period surv.int Tstart Tstop delta d pyrs d.pp  
1: 1 [1993,2003) 12 0.9166667 1 0.08333333 60 248.824778 63.650766  
2: 1 [1993,2003) 60 4.9166667 5 0.08333333 13 154.472450 22.833276  
3: 1 [1993,2003) 120 9.9166667 10 0.08333333 5 112.798768 10.656958  
4: 1 [1993,2003) 180 14.9166667 15 0.08333333 1 40.066222 2.495137  
5: 1 [2003,2013) 12 0.9166667 1 0.08333333 47 260.312286 49.862711  
6: 1 [2003,2013) 60 4.9166667 5 0.08333333 4 98.659993 5.202564  
7: 1 [2003,2013) 120 9.9166667 10 0.08333333 0 0.365332 0.000000  
8: 1 [2003,2013) NA NA NA NA NA NA NA  
 d.exp.pp d.pp.2 pyrs.pp surv.obs.as.lo surv.obs.as surv.obs.as.hi  
1: 9.95612421 67.830645 256.7741010 0.7721363 0.7849992 0.7972344  
2: 10.44526079 67.354206 185.2233079 0.5018975 0.5177134 0.5332825  
3: 15.08950748 27.344593 168.0066044 0.3856067 0.4010478 0.4164331  
4: 12.89305701 6.225711 82.2248335 0.3077360 0.3235001 0.3393524  
5: 7.90154463 53.215723 266.6109419 0.8165912 0.8277087 0.8382208  
6: 4.98799512 6.953770 113.2821802 0.5879067 0.6043383 0.6203412  
7: 0.01376493 0.000000 0.4333571 0.4804040 0.5017128 0.5226223  
8: NA NA NA NA NA NA  
 SE.surv.obs.as r.pp.as.lo r.pp.as r.pp.as.hi SE.r.pp.as  
1: 0.006401154 0.7930900 0.8064752 0.8190961 0.006631840  
2: 0.008008062 0.5741027 0.5930722 0.6115029 0.009542744  
3: 0.007866129 0.5127324 0.5394885 0.5654465 0.013454343  
4: 0.008068372 0.4653960 0.5360451 0.6015042 0.034839317  
5: 0.005515817 0.8342032 0.8457952 0.8566479 0.005722654  
6: 0.008275263 0.6549621 0.6749420 0.6940489 0.009971536  
7: 0.010774136 0.6069758 0.6725246 0.7296206 0.031297710  
8: NA NA NA NA NA

summary prints estimates at given time points: at 1, 5, 10 and 15 years after diagnosis (column Tstop). Column surv.int shows the follow-up time interval (monthly intervals were defined in breaks): at the end of 12th follow-up time interval Tstop=1 (1-year survival). The length of the intervals is delta in years. Notice that d, pyrs and d.exp are the number of deaths, person-years and the expected number of deaths within a monthly interval only, e.g. from 11 to <12 months (surv.int=12). The Pohar Perme estimator is based on the weighted observations d.pp, pyrs.pp and d.exp.pp.

The age-standardised survival estimates (both overall survival surv.obs.as and relative survival r.pp.as (based on the Pohar Perme estimator) with 95% confidence intervals and standard errors were calculated for the two periods.

Plot the age-standardised relative survival curves in patients diagnosed in 1993-2002 and in 2003-2012:

plot(st.rel[st.rel$period=="[1993,2003)",])  
lines(st.rel[st.rel$period=="[2003,2013)",],col="red")  
legend("topright",c("1993-2002","2003-2012"),col=c("black","red"),lty=1)



# Regression model for relative survival

Estimate the relative excess risk (ratio of excess hazards) for the difference in rectal cancer survival between 1993-2002 and 2003-2012. First, tabulate the observations by follow-up time interval, age group, sex and period.

BL <- list(fot=c(0,3/12,1:5))  
names(pm) <- c("sex","year","agegroup","haz")  
sire$period <- cut( cal.yr(sire$dg\_date), c(1993,2003,2013), right=F,dig.lab = 4)  
Lex <- lexpand(sire, birth = bi\_date,  
 entry = dg\_date, exit = ex\_date,  
 status = status%in%1:2,  
 breaks = BL, pophaz = pm,pp=F,  
 aggre = list(fot,agegr,sex,period)  
)

The piecewise constant excess hazard model can be fitted using relpois\_ag function in popEpi package:

m1 <- relpois\_ag(formula = from0to1 ~ period+fot\*agegr,   
 data = Lex, d.exp = d.exp, offset = log(pyrs))  
ci.exp(m1)

exp(Est.) 2.5% 97.5%  
(Intercept) 0.04737401 0.01488776 0.1507478  
period[2003,2013) 0.77927020 0.71773516 0.8460810  
fot[0.25, 1) 2.71244453 0.79186945 9.2911216  
fot[1, 2) 2.21222063 0.64337392 7.6066498  
fot[2, 3) 1.17643066 0.31455346 4.3998534  
fot[3, 4) 0.23069625 0.03301250 1.6121394  
fot[4, 5) 0.25656241 0.03714556 1.7720630  
agegr[45,55) 1.61632499 0.45018845 5.8031398  
agegr[55,65) 3.38092135 1.02971915 11.1007251  
agegr[65,75) 4.01745353 1.23918760 13.0246081  
agegr[75,Inf) 15.27985442 4.78676930 48.7748492  
fot[0.25, 1):agegr[45,55) 0.65757436 0.16742843 2.5826201  
fot[1, 2):agegr[45,55) 0.47746790 0.11999239 1.8999172  
fot[2, 3):agegr[45,55) 0.70913424 0.16303899 3.0843627  
fot[3, 4):agegr[45,55) 3.70825310 0.47610636 28.8824982  
fot[4, 5):agegr[45,55) 2.53271577 0.32296122 19.8619799  
fot[0.25, 1):agegr[55,65) 0.25892520 0.07220726 0.9284698  
fot[1, 2):agegr[55,65) 0.28557714 0.07938867 1.0272788  
fot[2, 3):agegr[55,65) 0.43614525 0.11105614 1.7128515  
fot[3, 4):agegr[55,65) 1.34896798 0.18410639 9.8840383  
fot[4, 5):agegr[55,65) 0.47238841 0.06087296 3.6658446  
fot[0.25, 1):agegr[65,75) 0.34662599 0.09864939 1.2179455  
fot[1, 2):agegr[65,75) 0.31294291 0.08856101 1.1058282  
fot[2, 3):agegr[65,75) 0.39687364 0.10265032 1.5344198  
fot[3, 4):agegr[65,75) 1.48450863 0.20602409 10.6966417  
fot[4, 5):agegr[65,75) 1.34872879 0.18871692 9.6391428  
fot[0.25, 1):agegr[75,Inf) 0.15399934 0.04460816 0.5316471  
fot[1, 2):agegr[75,Inf) 0.09915330 0.02846077 0.3454361  
fot[2, 3):agegr[75,Inf) 0.11761547 0.03064471 0.4514123  
fot[3, 4):agegr[75,Inf) 0.32724123 0.04441284 2.4111680  
fot[4, 5):agegr[75,Inf) 0.27202182 0.03644168 2.0305287

Age-adjusted excess mortality among patients diagnosed in 2003-2012 is 0.78-fold compared to patients diagnosed in 1993-2002.

In cancer survival, it is almost always necessary to include interactions between follow-up time and age into the model (i.e. excess hazards are not proportional between age groups):

m0 <- relpois\_ag(formula = from0to1 ~ period+fot + agegr,   
 data = Lex, d.exp = d.exp, offset = log(pyrs))  
anova(m0,m1,test="LRT")

Analysis of Deviance Table  
  
Model 1: from0to1 ~ period + fot + agegr  
Model 2: from0to1 ~ period + fot \* agegr  
 Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
1 49 217.260   
2 29 44.758 20 172.5 < 2.2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The same model can be fitted using glm function. The link function of relative survival Poisson regression can be specified as follows:

Lex$FUT <- cut( Lex$fot, c(0,3/12,1:5), right=F)  
d\_star <- Lex$d.exp  
  
#define link function for relative survival regression  
rspois <- poisson()  
rspois$link <- "glm relative survival model with Poisson error"  
rspois$linkfun <- function(mu) log(mu - d\_star)  
rspois$linkinv <- function(eta) d\_star + exp(eta)  
assign(".d\_star", Lex$d.exp, env = .GlobalEnv)  
rspois$initialize <- expression({  
 if (any(y < 0)) stop(paste("Negative values not allowed for",   
 "the Poisson family"))  
 n <- rep.int(1, nobs)  
 mustart <- pmax(y, .d\_star) + 0.1  
})  
  
#glm model  
m2 <- glm(formula = from0to1 ~ period+FUT\*agegr,   
 data = Lex, family=rspois, offset = log(pyrs))  
ci.exp(m2)

exp(Est.) 2.5% 97.5%  
(Intercept) 0.04737401 0.01488776 0.1507478  
period[2003,2013) 0.77927020 0.71773516 0.8460810  
FUT[0.25,1) 2.71244453 0.79186945 9.2911216  
FUT[1,2) 2.21222063 0.64337392 7.6066498  
FUT[2,3) 1.17643066 0.31455346 4.3998534  
FUT[3,4) 0.23069625 0.03301250 1.6121394  
FUT[4,5) 0.25656241 0.03714556 1.7720630  
agegr[45,55) 1.61632499 0.45018845 5.8031398  
agegr[55,65) 3.38092135 1.02971915 11.1007251  
agegr[65,75) 4.01745353 1.23918760 13.0246081  
agegr[75,Inf) 15.27985442 4.78676930 48.7748492  
FUT[0.25,1):agegr[45,55) 0.65757436 0.16742843 2.5826201  
FUT[1,2):agegr[45,55) 0.47746790 0.11999239 1.8999172  
FUT[2,3):agegr[45,55) 0.70913424 0.16303899 3.0843627  
FUT[3,4):agegr[45,55) 3.70825310 0.47610636 28.8824982  
FUT[4,5):agegr[45,55) 2.53271577 0.32296122 19.8619799  
FUT[0.25,1):agegr[55,65) 0.25892520 0.07220726 0.9284698  
FUT[1,2):agegr[55,65) 0.28557714 0.07938867 1.0272788  
FUT[2,3):agegr[55,65) 0.43614525 0.11105614 1.7128515  
FUT[3,4):agegr[55,65) 1.34896798 0.18410639 9.8840383  
FUT[4,5):agegr[55,65) 0.47238841 0.06087296 3.6658446  
FUT[0.25,1):agegr[65,75) 0.34662599 0.09864939 1.2179455  
FUT[1,2):agegr[65,75) 0.31294291 0.08856101 1.1058282  
FUT[2,3):agegr[65,75) 0.39687364 0.10265032 1.5344198  
FUT[3,4):agegr[65,75) 1.48450863 0.20602409 10.6966417  
FUT[4,5):agegr[65,75) 1.34872879 0.18871692 9.6391428  
FUT[0.25,1):agegr[75,Inf) 0.15399934 0.04460816 0.5316471  
FUT[1,2):agegr[75,Inf) 0.09915330 0.02846077 0.3454361  
FUT[2,3):agegr[75,Inf) 0.11761547 0.03064471 0.4514123  
FUT[3,4):agegr[75,Inf) 0.32724123 0.04441284 2.4111680  
FUT[4,5):agegr[75,Inf) 0.27202182 0.03644168 2.0305287