Statistical Methods in Cancer Epidemiology using R

Karri Seppä

Finnish Cancer Registry

Lecture 9

karri.seppa@cancer.fi

Mar 23, 2020

Points to be covered

1. Relative survival estimators
2. Age-standardised relative survival
3. Period approach
4. Regression modelling of relative survival.

- Packages `popEpi` and `rstpm2`.
- Functions `survtab()`, `glm()`, `relpois_ag()`, `stpm2()`.

Population-based cancer survival

- Relative (net) survival is the standard measure in population-based cancer survival
- Relative survival summarizes patients' excess mortality, i.e. additional mortality patients have compared with mortality in a relevant reference population.
- Does not rely on cause of death information
  - The **quality of cause-of-death information varies** over time, between types of cancer and between regions and countries.
  - Many cancer registries **do not record cause of death**.
  - Cause of death is rarely a simple dichotomy.

Excess mortality and relative survival

- Total mortality among patients $\lambda(t)$ is the sum of population mortality $\lambda_{\mathrm{P}}(t)$ and excess mortality $\lambda_{\mathrm{E}}(t)$ due to cancer:

$$\lambda(t) = \lambda_{\mathrm{P}}(t) + \lambda_{\mathrm{E}}(t)$$

- Relative survival $S_E(t)$ is the survival function related to the excess mortality:

$$S_{\mathrm{E}}(t) = \exp\left[-\int_0^t \lambda_{\mathrm{E}}(u)\ du\right]$$

  - Often interpreted as the survival **probability** in a hypothetical situation where the cancer would be the only possible cause of death with all other causes being eliminated.
  - **N.B.** Independence of competing risks is required but this assumption cannot be assessed based on the data!

Non-parametric relative survival estimators (hazard-based)

- Ederer II (Ederer and Heise 1958)

$$S_{\mathrm{E,EII}}(t_k) = \exp\left(-\Delta_j \frac{d_j - d_j^*}{y_j}\right)$$

  - Δ_j = length of follow-up time interval j
  - d_j = number of deaths
  - d_j^* = expected number of deaths (same as in SMR estimation)
  - y_j = number person-years

- Pohar Perme estimator (Pohar Perme et al. 2012)

$$S_{\mathrm{E,PP}}(t_k) \exp\left(-\Delta_j \frac{d_j^w - d_j^{*w}}{y_j^w}\right)$$

  - d_j^w, d_j^{*w} and y_j^w are based on individual-level observations of d_j, d_j^* and y_j weighted by the patient-specific cumulative expected survival probability (see Seppä et al. 2016).

Properties of the estimators

- Non-standardised Ederer II estimator overestimates relative survival because of informative censoring of old patients with high other-cause mortality
  - use age-standardised Ederer II
- Pohar Perme estimator corrects for this bias by giving larger weights for observations of older patients
- Both the Pohar Perme and the age-standardised Ederer II estimator are valid for 5-year net survival of cancer patients.

Age-standardised relative survival

- Weighted average of age-specific relative survival estimates

  $S_E(t) = \sum_{a=1}^{K} w_a S_{E,a}(t)$ where $\sum_{a=1}^{K} w_a = 1$

  - weight w_a is a standard for the proportion of patients in age group a at the beginning of follow-up

- E.g. International Cancer Survival Standards
  - ICSS1: elderly (most cancers)
  - ICSS2: little age dependence (eg. melanoma, brain, cervix)
  - ICSS3: young adults (eg. testis, hodgkin lymphoma); starify 15-44 into narrower groups

Age group (a)	w_{ICSS1}	w_{ICSS2}	w_{ICSS3}
15-44	0.07	0.28	0.60
45-54	0.12	0.17	0.10
55-64	0.23	0.21	0.10
65-74	0.29	0.20	0.10
75+	0.29	0.14	0.10

Ex. Oral cancer data

Relative survival analysis

- expected mortality is derived from the mortality rates in the population of Finland
- assume the date on diagnosis would a random date between 1985 and 2005.

```
orca <- read.table(file="oralca2.txt")
library(Epi)
library(popEpi)
set.seed(23032020)
orca$dg_date <- runif(nrow(orca), min=1985, max=2005)
orca$sex <- ifelse(orca$sex=="Male",0,1)
names(orca)[2] <- "dg_age"
Lex <- Lexis(entry = list(FUT = 0, AGE = dg_age, CAL = dg_date),
             exit = list(FUT = time),
             data = orca,
             exit.status =
               factor(event, levels = 0:2,
                      labels = c("alive", "canD", "othD")),
             merge = TRUE)
```

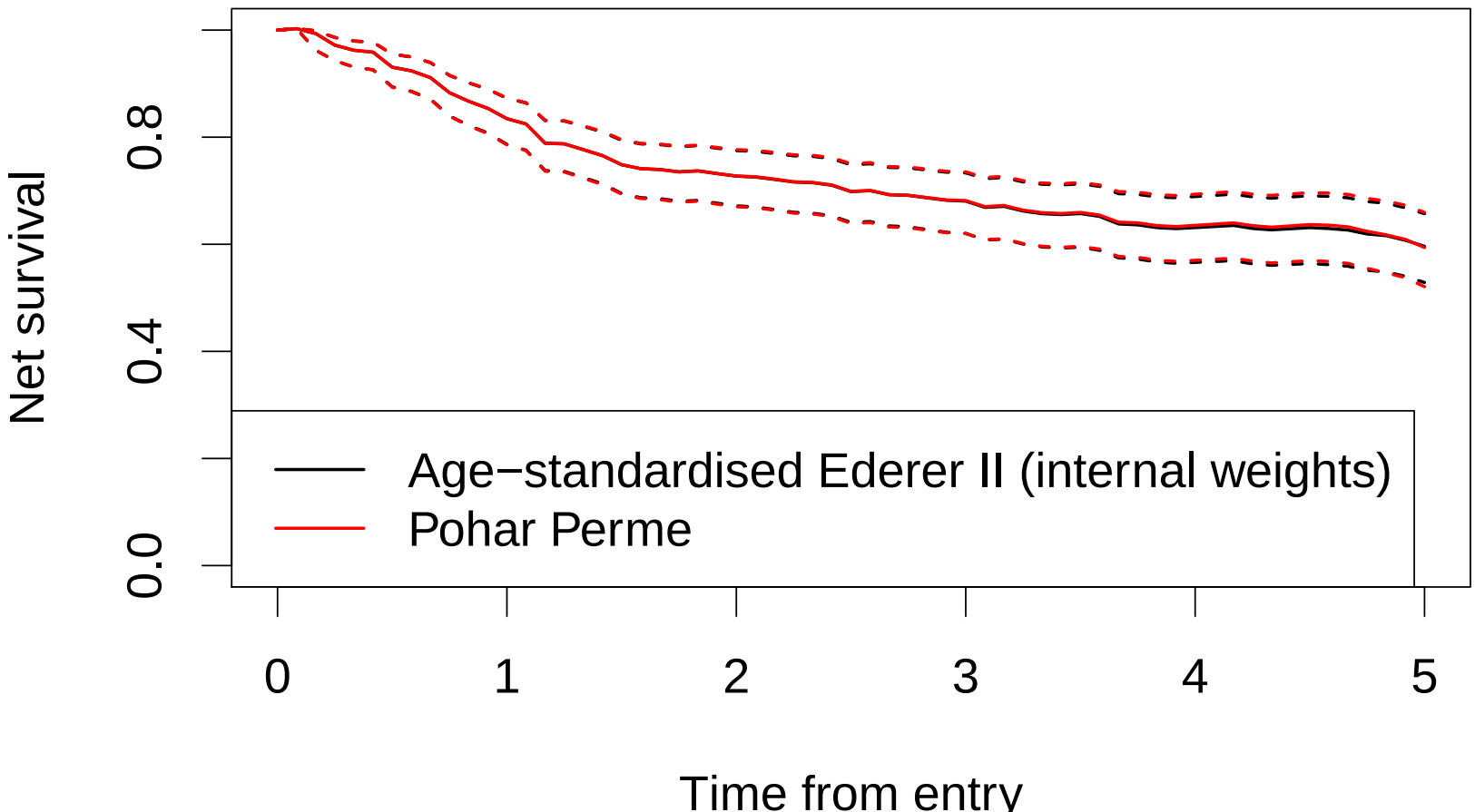
Ex. Oral cancer data, Ederer II and Pohar Perme

```
pm <- popmort
names(pm) <- c("sex", "CAL", "AGE", "haz")
cutpoints <- c(0,55,65,75,Inf)
Lex$agegr <- cut(Lex$dg_age, cutpoints, right=F)

#Age-standardised Ederer II estimator
surv.EII <- survtab(FUT ~ 1 + adjust(agegr), data = Lex,
                    surv.type = "surv.rel",
                    pophaz = pm, weights = "internal",
                    breaks = list(FUT = seq(0, 5, 1/12)))
#Pohar Perme estimator
surv.PP <- survtab(FUT ~ 1, data = Lex,
                    surv.type = "surv.rel", relsurv.method = "pp",
                    pophaz = pm,
                    breaks = list(FUT = seq(0, 5, 1/12)))
```

Ex. Oral cancer data, Ederer II and Pohar Perme

```
par(mar=c(4,4,0.5,1))
plot(surv.EII, lwd=2, ylim=c(0,1))
lines(surv.PP, lwd=2, col="red")
legend("bottomleft",c("Age-standardised Ederer II (internal weights)",
                      "Pohar Perme"),
       lty=1, lwd=2 ,col=c("black","red"))
```



Ex. Oral cancer data, ICSS weights

Use external weights (ICSS1) in age standardisation:

```
ICSS$agegr <- cut(ICSS$age, cutpoints, right = FALSE)
w <- aggregate(ICSS1~agegr, data = ICSS, FUN = sum)
w
```

```
     agegr ICSS1
1    [0,55) 19000
2   [55,65) 23000
3   [65,75) 29000
4 [75,Inf) 29000
```

Ex. Oral cancer data, ICSS weights

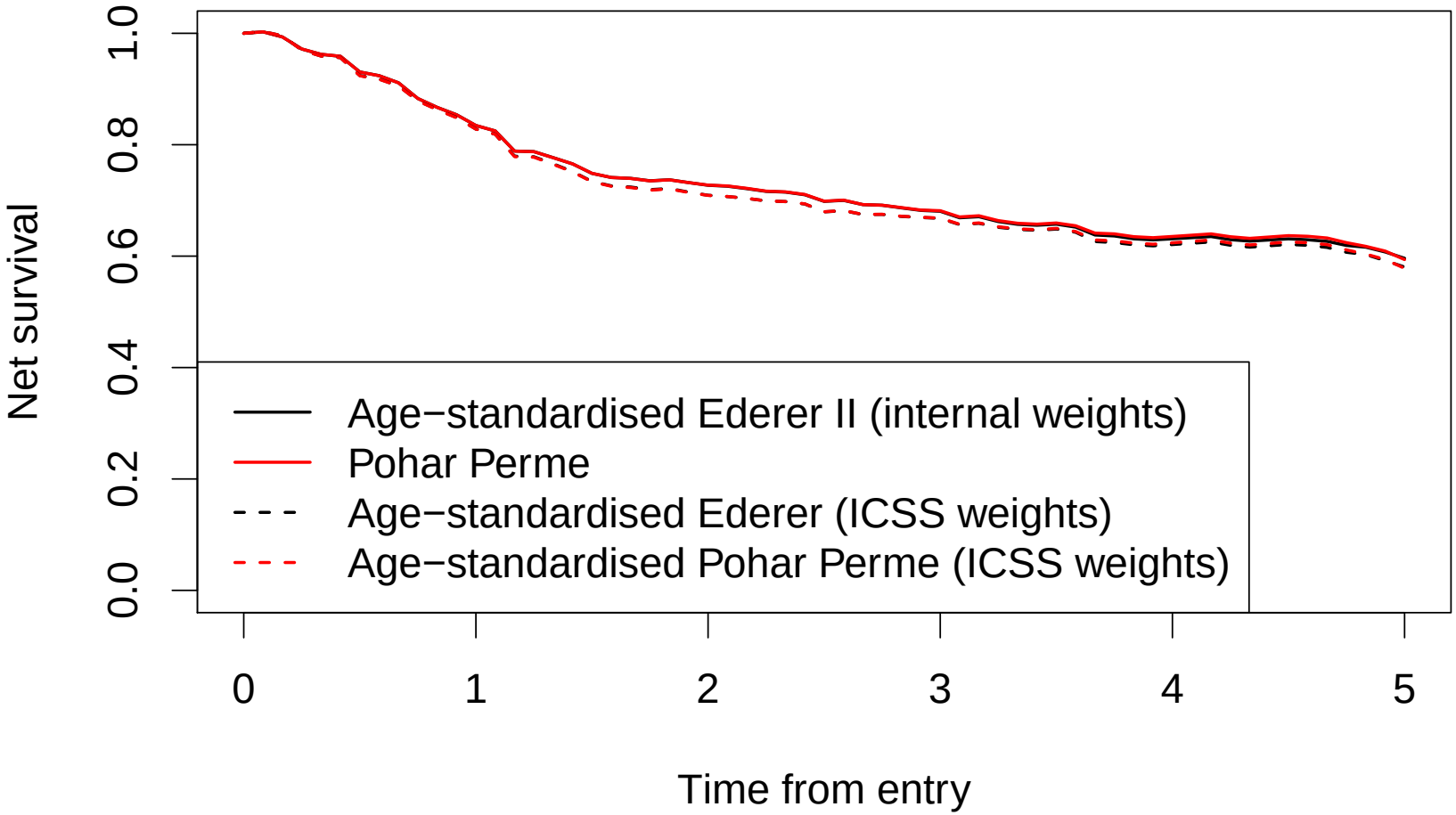
```
Lex$agegr <- cut(Lex$dg_age, cutpoints, right=F)
surv.EII.ICSS <- survtab(FUT ~ 1+adjust(agegr), data = Lex,
                surv.type = "surv.rel",
                pophaz = pm, weights = w$ICSS1,
                breaks = list(FUT = seq(0, 5, 1/12)))
surv.PP.ICSS <- survtab(FUT ~ 1+adjust(agegr), data = Lex,
                surv.type = "surv.rel",relsurv.method = "pp",
                pophaz = pm, weights = w$ICSS1,
                breaks = list(FUT = seq(0, 5, 1/12)))
```

Ex. Oral cancer data, internal vs ICSS weights

```
par(mar=c(4,4,0.5,1))
plot(surv.EII, lwd=2, conf.int=F,ylim=c(0,1))
lines(surv.PP, lwd=2,col="red", conf.int=F)
lines(surv.EII.ICSS, lwd=2, lty=2, conf.int=F)
lines(surv.PP.ICSS, lwd=2,col="red",lty=2, conf.int=F)

legend("bottomleft",c("Age-standardised Ederer II (internal weights)",
                      "Pohar Perme",
                      "Age-standardised Ederer (ICSS weights)",
                      "Age-standardised Pohar Perme (ICSS weights)"),
       lty=c(1,1,2,2), lwd=2,col=c("black","red","black","red"))
```

Ex. Oral cancer data, ICSS weights



Period approach

- Survival estimates in a cohort of patients tend to be outdated
  - 5-year survival estimate requires follow-up data on patients diagnosed $\geq$ 5 years ago.
- Use period approach (Brenner et al. 2004) to produce maximally updated estimates of patient survival
  - follow-up of patients is restricted to a calendar time window.
  - e.g. patients diagnosed in 2011-2018 and followed-up in 2016-2018.
  - patients diagnosed earlier than 2016 do not contribute to survival estimate before coming to the follow-up window 2016-2018 (late entry).
  - interpretation: survival in patients, who would have the same follow-up time specific excess mortality rates as those observed in 2016-2018. (A prediction for patients diagnosed in 2016-2018.)

Period approach

Example: 5-year survival based on patients diagnosed in 2011-2018 and followed in calendar time window 2016-2018.

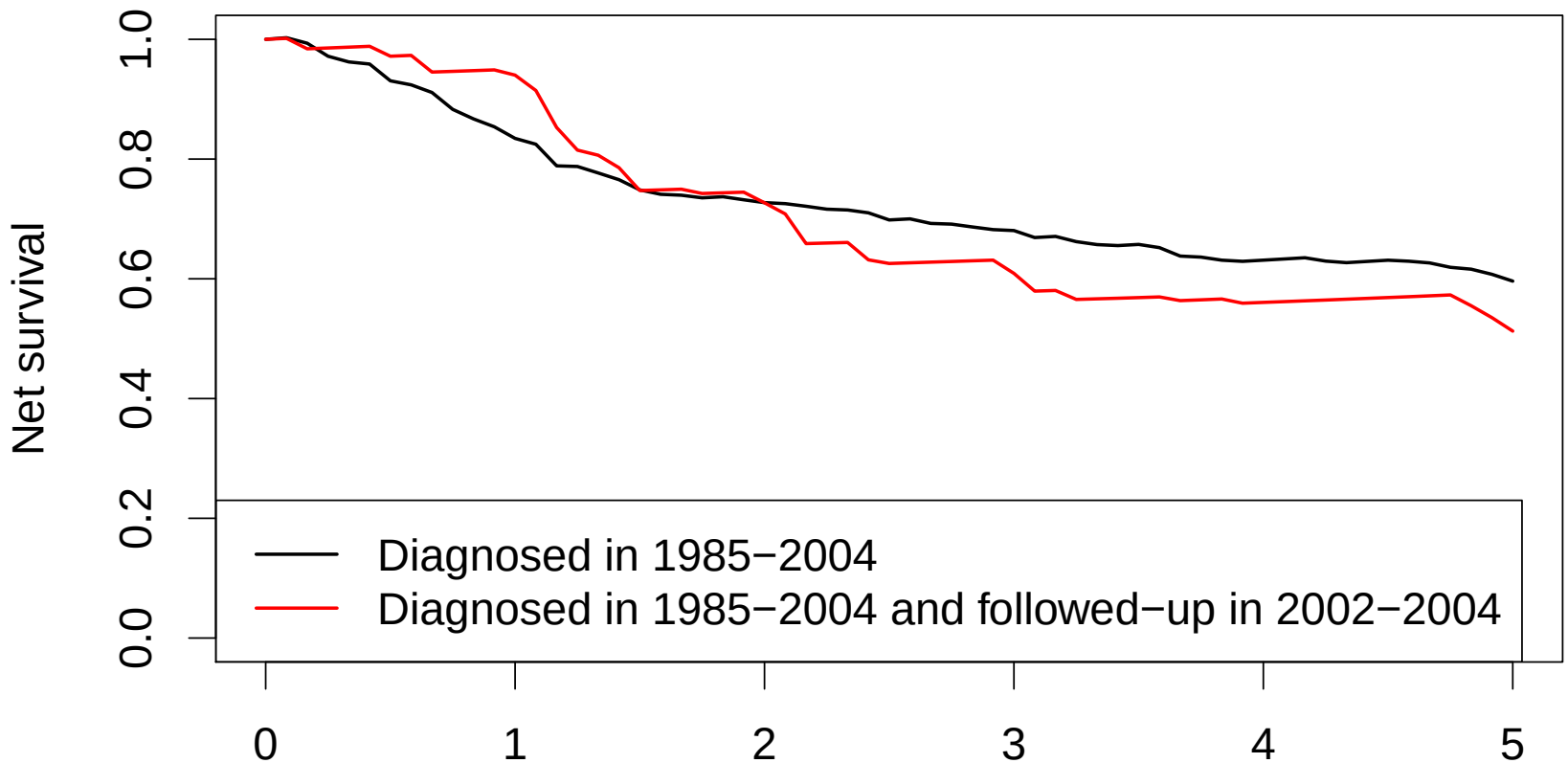
Year of diagnosis	Year of follow-up							
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
2011	0-1	1-2	2-3	3-4	4-5	5		
2012		0-1	1-2	2-3	3-4	4-5	5	
2013			0-1	1-2	2-3	3-4	4-5	5
2014				0-1	1-2	2-3	3-4	4-5
2015					0-1	1-2	2-3	3-4
2016						0-1	1-2	2-3
2017							0-1	1-2
2018								0-1

Ex. Oral cancer data, period approach

```
#Age-standardised Ederer II estimator
surv.EII.period <- survtab(FUT ~ 1 + adjust(agegr), data = Lex,
                surv.type = "surv.rel",
                pophaz = pm, weights = "internal",
                breaks = list(FUT = seq(0, 5, 1/12), CAL=c(2002,2005)))
```

Ex. Oral cancer data, period approach

```
#Age-standardised Ederer II estimator
par(mar=c(4,4,0.5,1))
plot(surv.EII, lwd=2, conf.int=F,ylim=c(0,1))
lines(surv.EII.period, lwd=2,col="red", conf.int=F)
legend("bottomleft",c("Diagnosed in 1985-2004",
         "Diagnosed in 1985-2004 and followed-up in 2002-2004"),
       lty=c(1,1), lwd=2,col=c("black","red"))
```



Regression models for relative survival

1. Logarithm of excess mortality

$$\log[\lambda_E(t)] = \alpha + \beta X$$

- can be modelled by using Poisson regression with a user-specified link function (Dickman et al. 2004).

2. Logarithm of cumulative excess mortality

$$\log[\Lambda_E(t)] = \text{spline}[log(t) \mid \alpha] + \beta X$$

- can be modelled using spline functions for the baseline cumulative excess hazard (`rstpm2` package)

Ex. Oral cancer data, piecewise constant excess hazard

```
#cutpoints for piecewise constant excess hazard
fb <- c(0,1,2,3,5)
pm2 <- pm
names(pm2) <- c("sex","year","agegroup","haz")
orca$agegr <- cut(orca$dg_age, cutpoints, right=F)
#combines orca and population mortality rates
Lex <- lexpand(orca, birth = dg_date-dg_age, entry = dg_date,
               exit = dg_date+time, status=event %in% 1:2,
               breaks = list(fot=fb),
               pophaz=pm2, pp = FALSE,
               aggre = list(agegr, fot))
```

Ex. Oral cancer data, piecewise constant excess hazard

Aggregated data

```
Lex
```

	agegr	fot	pyrs	at.risk	d.exp	from0to0	from0to1
1:	[0,55)	0	94.163	100	0.4176292	0	15
2:	[0,55)	1	81.649	85	0.3764739	1	4
3:	[0,55)	2	75.902	80	0.3727492	5	8
4:	[0,55)	3	118.216	67	0.5752607	8	6
5:	[55,65)	0	66.439	71	0.7529778	0	9
6:	[55,65)	1	53.859	62	0.6535970	3	10
7:	[55,65)	2	46.049	49	0.6024387	3	4
8:	[55,65)	3	72.744	42	0.9819390	3	7
9:	[65,75)	0	83.234	89	2.3992428	0	14
10:	[65,75)	1	68.080	75	2.0061784	0	12
11:	[65,75)	2	60.721	63	1.8970814	1	4
12:	[65,75)	3	91.164	58	3.0119818	8	10
13:	[75,Inf)	0	66.487	78	5.8403911	0	26
14:	[75,Inf)	1	42.927	52	3.9569316	0	15
15:	[75,Inf)	2	33.630	37	3.4714270	0	5
16:	[75,Inf)	3	55.426	32	6.5913538	2	10

Ex. Oral cancer data, piecewise constant excess hazard

Proportional hazards between age groups?

```
## fit model using aggregated data
rpm <- relpois_ag(formula = from0to1 ~ fot + agegr,
                  data = Lex, d.exp = d.exp, offset = log(pyrs))
ci.exp(rpm)
```

```
              exp(Est.)      2.5%     97.5%
(Intercept)   0.1299963 0.0845701 0.1998229
fot[1, 2)     0.7921478 0.4951471 1.2672964
fot[2, 3)     0.4500391 0.2392291 0.8466161
fot[3, 5)     0.3974890 0.2250555 0.7020377
agegr[55,65)  1.3009831 0.7600595 2.2268746
agegr[65,75)  1.1746621 0.6841282 2.0169189
agegr[75,Inf) 2.1077956 1.2363790 3.5933985
```

```
## non-proportional effect of age (=follow-up time age interaction)
rpm2 <- update(rpm, . ~ fot*agegr)
anova(rpm, rpm2, test="LRT")$`Pr(>Chi)`[2]
```

```
[1] 0.3147943
```

Ex. Oral cancer data, spline functions

Mortality from any cause and from cancer, and excess mortality

```
library(rstpm2)
orca$agegr <- cut(orca$dg_age, cutpoints, right=F)

#mortality from any cause
fit <- stpm2(Surv(time,event%in%c(1,2))~1,
             data=orca,df=3)

#cause (cancer)-specific mortality
fit2 <- stpm2(Surv(time,event%in%c(1))~1,
              data=orca,df=3)

#excess mortality
orca$ex_age <- floor(orca$dg_age+orca$time)
orca$ex_year <- floor(orca$dg_date+orca$time)
orca <- merge(orca,pm,
              by.x=c("sex","ex_age","ex_year"),
              by.y=c("sex","AGE","CAL"))
fit3 <- stpm2(Surv(time,event%in%c(1,2))~1,
              data=orca,df=3,
              bhazard=orca$haz)
```

Ex. Oral cancer data, hazard

```
plot(fit,newdata=data.frame(time=seq(0.1,5,0.1)),
     type="hazard",add=FALSE,ci=FALSE,line.col=1,
     ylim=c(0.01,0.3),rug=F)
plot(fit2,newdata=data.frame(time=seq(0.1,5,0.1)),
     type="hazard",add=TRUE,ci=FALSE,line.col=2)
plot(fit3,newdata=data.frame(time=seq(0.1,5,0.1)),
     type="hazard",add=TRUE,ci=FALSE,line.col=3)
legend("topright",
       legend=c("Total mortality",
                "Oral cancer mortality",
                "Excess mortality"),
       col=c("black","red","green"),lty=1)
```

Ex. Oral cancer data, hazard