

“Analyzing Survival in “Evolucio_FG” data selecting diabetic patients.”

Investigators:

Pau Codina Verdaguer, `pau.codi@gmail.com`

Josep Lupón Roses, `jluponroses@gmail.com`

Version 1.0

- Joan Vila -

`joanviladomenech@gmail.com`

May 29, 2022

1 Version History

Version	Effective Date	Changes
1	29-May-2022	Primera proposta d'anàlisi

Contents

1	Version History	1
2	Càrrega de packages, dades, funcions	3
3	Introduction	3
3.1	Total individus i registres utilitzats al analitzar la supervivència	5
4	Survival analysis	6
4.1	Outcome = Cardiovascular Death (competing with other causes)	6
4.2	Outcome = Non-Cardiovascular Death (competing Cardiovascular Death)	7

2 Càrrega de packages, dades, funcions

```
rm(list=ls())
library(compareGroups)
library(tidyverse)
library(magrittr)
library(emmeans)
library(lme4)
library(multcomp)
library(lmerTest)
library(gdata)
library(Hmisc)
library(riskRegression)
library(prodlim)

# def.chunk.hook <- knitr::knit_hooks$get("chunk")
# knitr::knit_hooks$set(chunk = function(x, options) {
#   x <- def.chunk.hook(x, options)
#   ifelse(options$size == "scriptsize", paste0("\n \\", options$size, "\n\n", x, "\n\n \\", normalsize"),
#   # })

setwd("/Users/jvila/Dropbox/JLupon/FGdiabet/")

load("./dat/dat.rda")
load("./dat/datok.rda")

## Primary for CV
datok$crCV <- with(datok, ifelse(EXITUS==0, 0,
                                ifelse(CAUSA_EXITUS%in%c(1,2,3,4,5,7), 1,
                                ifelse(CAUSA_EXITUS%in%c(0,6), 2, NA))))
subset(datok, is.na(crCV))[, c("EXITUS", "CAUSA_EXITUS")]
attr(datok$crCV, "vari.label") <- "CV death for Competing risk"
attr(datok$crCV, "value.labels") <- c("Alive"=0, "crCV"=1, "No-CV death"= 2)
```

3 Introduction

En els resultats que em va ensenyar en Pau (Evolucio FG.html) s'havien eliminat els individus de ≥ 15 anys de seguiment.

Por poder analitzar la supervivencia, amb les dades que es disposen, cal tenir en compte:

- Que hi ha variables canviants en el temps
- Que hi ha *outcomes* que competeixen entre ells, morir de causa no cardiovascular 'competeix' amb morir de causa cardiovascular.

Per poder analitzar aquest tipus de dades cal manipular-les i re-ordenar-les:

Original:

```
options(width=300)
subset(dat, id==54)[, c("VISITMONTH", "DATA_0", "DATA_EXITUS", "CAUSA_EXITUS", "epi", "Edat", "SEX", "IMCcat", "EFcat", "etiology", "ACEI_ARB_s")]

##      VISITMONTH  DATA_0 DATA_EXITUS CAUSA_EXITUS      epi      Edat  SEX IMCcat EFcat      etiology ACEI_ARB_s
## 537           0 2-Oct-01      8-Feb-03           1 47.23547 62.32991 Women 25-<30  50+ Non-Ischemic      No
## 538           1   #NULL!      #NULL!          NA 40.67905 62.42574 Women 25-<30  50+ Non-Ischemic      No
## 539           3   #NULL!      #NULL!          NA 57.80133 62.67762 Women 25-<30  50+ Non-Ischemic      No
## 540           6   #NULL!      #NULL!          NA 48.48030 62.90486 Women 25-<30  50+ Non-Ischemic      No
## 541           9   #NULL!      #NULL!          NA 26.97315 63.25530 Women 25-<30  50+ Non-Ischemic      No
```

Nota que aquest pacient:

- es va visitar fins a 9 mesos més tard de la primera visita (aprox. juliol) i es va morir aprox 16 mesos més tard (al Febrer del 2003) de la primera visita
- els valors de 'epi' 'Edat' etc. sabem com li han anat canviant al llarg del temps
- el 'SEX' i la 'etiology', lògicament no li han canviat
- el 'IMCcat', la 'EFcat' i el tractament amb 'ACEI_ARB_s' segurament li ha canviat al llarg del temps de seguiment, però com que no sabem aquesta informació, assumim que ha estat constant al llarg del seguiment

Per analitzar les dades cal re-ordenar-les:

```
subset(datok, id==54)[, c("VISITMONTH", "end", "EXITUS", "crCV", "epi", "Edat", "SEX", "IMCcat", "EFcat", "etiology", "ACEI_ARB_s")]

##      VISITMONTH end EXITUS crCV      epi      Edat  SEX IMCcat EFcat      etiology ACEI_ARB_s
## 532           0   1      0   0 47.23547 62.32991 Women 25-<30  50+ Non-Ischemic      No
## 533           1   3      0   0 40.67905 62.42574 Women 25-<30  50+ Non-Ischemic      No
## 534           3   6      0   0 57.80133 62.67762 Women 25-<30  50+ Non-Ischemic      No
## 535           6   9      0   0 48.48030 62.90486 Women 25-<30  50+ Non-Ischemic      No
## 536           9  16      1   1 26.97315 63.25530 Women 25-<30  50+ Non-Ischemic      No
```

Nota que p.e.

- des del mes 6 fins al mes 9 s'assumeix que el valor d'epi ha estat de 48.48030
- en estat vital és de 'viu' en tots els seguiment, fins el mes 16 en que mort de causa 1 (Cardiovascular)

Nota aquest pacient:

```
subset(dat, id==2454)[, c("VISITMONTH", "DATA_O", "DATA_EXITUS", "CAUSA_EXITUS", "epi", "Edat", "SEX", "IMCcat", "EFcat", "etiology", "ACEI_ARB_s")]

##      VISITMONTH  DATA_O DATA_EXITUS CAUSA_EXITUS      epi      Edat  SEX IMCcat EFcat      etiology ACEI_ARB_s
## 9457          0 27-Dec-18      #NULL!      NA 99.05472 57.73854 Women 25-<30 <=40% Non-Ischemic      Yes
## 9458          1      #NULL!      #NULL!      NA 96.53552 57.82888 Women 25-<30 <=40% Non-Ischemic      Yes
## 9459          3      #NULL!      #NULL!      NA 95.51449 57.98494 Women 25-<30 <=40% Non-Ischemic      Yes
## 9460          6      #NULL!      #NULL!      NA 92.15475 58.23409 Women 25-<30 <=40% Non-Ischemic      Yes
## 9461         12      #NULL!      #NULL!      NA 78.49559 58.68857 Women 25-<30 <=40% Non-Ischemic      Yes
```

- 12 mesos més tard del reclutament (la darrera visita que se li va fer) encara estava viu
- a partir del mes 12 el valor de 'epi' era de 78.49559, però a efectes de supervivència, no fem servir aquesta informació
- nota que un cop re-ordenat, tenim una fila menys d'aquest pacient

```
subset(datok, id==2454)[, c("VISITMONTH", "end", "EXITUS", "crCV", "epi", "Edat", "SEX", "IMCcat", "EFcat", "etiology", "ACEI_ARB_s")]

##      VISITMONTH end EXITUS crCV      epi      Edat  SEX IMCcat EFcat      etiology ACEI_ARB_s
## 9059          0   1      0   0 99.05472 57.73854 Women 25-<30 <=40% Non-Ischemic      Yes
## 9060          1   3      0   0 96.53552 57.82888 Women 25-<30 <=40% Non-Ischemic      Yes
## 9061          3   6      0   0 95.51449 57.98494 Women 25-<30 <=40% Non-Ischemic      Yes
## 9062          6  12      0   0 92.15475 58.23409 Women 25-<30 <=40% Non-Ischemic      Yes
```

3.1 Total individus i registres utilitzats al analitzar la supervivència

Per poder fer la base de dades adequada per analitzar supervivència m'ha calgut re-ordenar-la com he explicat més amunt i a més eliminar visites repetides fetes en el mateix mes, p.e.

```
subset(dat, id==434)[1:5, c("VISITMONTH", "DATA_O", "DATA_EXITUS", "CAUSA_EXITUS", "epi", "Edat", "SEX", "IMCcat", "EFcat", "etiology", "ACEI_ARB_s")]

##      VISITMONTH  DATA_O DATA_EXITUS CAUSA_EXITUS      epi      Edat  SEX IMCcat EFcat      etiology ACEI_ARB_s
## 2214          0 22-Apr-04      #NULL!      NA 96.89872 54.57906 Men      30+ <=40% Non-Ischemic      Yes
## 2215          3      #NULL!      #NULL!      NA 90.85616 54.71321 Men      30+ <=40% Non-Ischemic      Yes
## 2216          3      #NULL!      #NULL!      NA 54.72842 54.78987 Men      30+ <=40% Non-Ischemic      Yes
## 2217          6      #NULL!      #NULL!      NA 54.63272 55.03901 Men      30+ <=40% Non-Ischemic      Yes
## 2218          9      #NULL!      #NULL!      NA 68.06016 55.32649 Men      30+ <=40% Non-Ischemic      Yes
```

Nota que aquest individu té dos visites al mes 3. El que he fet és eliminar la segona.

En resum:

- la base de dades original tenia 9469 i 935 individus
- un cop eliminades les visites de 15 anys o més de seguiment, la base de dades té 9374 i 935 individus
- un cop re-ordenat (eliminat darrer registre en el viu, i les visites repetides) la base de dades té 8985 i 935 individus

4 Survival analysis

4.1 Outcome = Cardiovascular Death (competing with other causes)

```
# funcio per arreglar el resultat dels models
arreglaMod <- function(mymodel){
  resu <- summary(mymodel)
  resu <- data.frame(cbind(resu$conf.int[, c("exp(coef)", "2.5%", "97.5%")], resu$coef[, "p-value"]))
  names(resu) <- c("HR", "Lower95%CI", "Upper95%CI", "pvalue")
  resu$pvalue <- round(resu$pvalue, 5)
  return(resu)
}

# funcio per comparar models
compareModels <- function(oldmodel, newmodel){
  compare <- data.frame(ordre= seq(1, nrow(oldmodel)), xvari= rownames(oldmodel), coefOld=log(oldmodel$HR), pvalOld = oldmodel$pvalue)
  theNew <- data.frame( xvari= rownames(newmodel), coefNew=log(newmodel$HR), pvalNew = newmodel$pvalue)
  compare <- merge(compare, theNew, by = "xvari", all.x=TRUE)
  compare <- compare[order(compare$ordre), ]
  compare$change <- with(compare, round((coefOld - coefNew)/ coefOld * 100, 2))
  return(compare)
}

## Primary for CV
datok$crCV <- with(datok, ifelse(EXITUS==0, 0,
                                ifelse(CAUSA_EXITUS%in%c(1,2,3,4,5,7), 1,
                                ifelse(CAUSA_EXITUS%in%c(0,6),2, NA))))
attr(datok$crCV, "vari.label") <- "CV death for Competing risk"
attr(datok$crCV, "value.labels") <- c("Alive" =0, "crCV"=1, "No-CV death"= 2)
```

Initial Model

```
modia <- FGR(Hist(entry = VISITMONTH, time= end, event = crCV) ~ epi + SEX + Edat + IMCcat + etiology +
              HTA + EFcat + ACEI_ARB_s + ARNI_s + MRA_s, datok, cause=1)
resuCV1 <- arreglaMod(modia)
resuCV1
```

	HR	Lower95%CI	Upper95%CI	pvalue
## epi	0.9902678	0.9851827	0.9953792	0.00020
## SEXMen	0.9521854	0.7451562	1.2167341	0.70000
## Edat	1.0103359	0.9957994	1.0250845	0.16000
## IMCcat18.5-<25	1.8210727	0.2254491	14.7097735	0.57000
## IMCcat25-<30	1.4817126	0.1856688	11.8246673	0.71000
## IMCcat30+	1.6633234	0.2076943	13.3207520	0.63000
## etiologyIschemic	1.9191764	1.5145838	2.4318482	0.00000
## HTAYes	1.2875125	0.9879084	1.6779779	0.06100
## EFcat41-49%	0.7926932	0.5282824	1.1894444	0.26000
## EFcat50+	1.6507692	1.2144808	2.2437894	0.00140
## ACEI_ARB_sYes	0.3701754	0.2718439	0.5040752	0.00000
## ARNI_sYes	0.3241129	0.1884894	0.5573214	0.00005
## MRA_sYes	0.7986086	0.6284279	1.0148749	0.06600

Provo de treure algunes variables:

```
modib <- update(modia, . ~. - SEX - IMCcat)
resuCV1b <- arreglaMod(modib)

compareModels(resuCV1, resuCV1b)
```

	xvari	ordre	coefOld	pvalOld	coefNew	pvalNew	change
## 6	epi	1	-0.00977985	0.00020	-0.00955021	0.00024	2.35
## 13	SEXMen	2	-0.04899556	0.70000	NA	NA	NA
## 3	Edat	3	0.01028283	0.16000	0.01211920	0.08700	-17.86
## 9	IMCcat18.5-<25	4	0.59942574	0.57000	NA	NA	NA
## 10	IMCcat25-<30	5	0.39319858	0.71000	NA	NA	NA
## 11	IMCcat30+	6	0.50881765	0.63000	NA	NA	NA
## 7	etiologyIschemic	7	0.65189612	0.00000	0.63735196	0.00000	2.23
## 8	HTAYes	8	0.25271209	0.06100	0.28791542	0.03100	-13.93
## 4	EFcat41-49%	9	-0.23231902	0.26000	-0.22348448	0.27000	3.80
## 5	EFcat50+	10	0.50124139	0.00140	0.50634023	0.00083	-1.02
## 1	ACEI_ARB_sYes	11	-0.99377843	0.00000	-0.95375632	0.00000	4.03
## 2	ARNI_sYes	12	-1.12666343	0.00005	-1.15249506	0.00002	-2.29
## 12	MRA_sYes	13	-0.22488428	0.06600	-0.22317658	0.06100	0.76

Veig que l'Edat i la HTA canvia més d'un 10%. Crec que és millor deixar el primer model.

4.2 Outcome = Non-Cardiovascular Death (competing Cardiovascular Death)

Initial Model

```
mod2a <- FGR(Hist(entry = VISITMONTH, time= end, event = crCV) ~ epi + SEX + Edat + IMCcat + etiology +  
  HTA + EFcat + ACEI_ARB_s + ARNI_s + MRA_s, datok, cause=2)  
resuCV2 <- arreglaMod(mod2a)  
resuCV2
```

	HR	Lower95%CI	Upper95%CI	pvalue
## epi	0.9941938	0.98799344	1.0004331	0.06800
## SEXMen	1.4950104	1.09532749	2.0405370	0.01100
## Edat	1.0325923	1.01580153	1.0496606	0.00013
## IMCcat18.5-<25	0.1369559	0.05421655	0.3459628	0.00003
## IMCcat25-<30	0.1426834	0.05836432	0.3488184	0.00002
## IMCcat30+	0.1446473	0.05874840	0.3561430	0.00003
## etiologyIschemic	1.1160523	0.86410427	1.4414611	0.40000
## HTAYes	1.6954900	1.23282669	2.3317846	0.00120
## EFcat41-49%	1.2617714	0.82966166	1.9189352	0.28000
## EFcat50+	1.3271501	0.91002831	1.9354642	0.14000
## ACEI_ARB_sYes	0.4186053	0.29325735	0.5975312	0.00000
## ARNI_sYes	0.4007386	0.23142242	0.6939320	0.00110
## MRA_sYes	0.5053750	0.38532246	0.6628315	0.00000

Provo de treure algunes variables:

```
mod2b <- update(mod2a, . ~. - etiology - EFcat)  
resuCV2b <- arreglaMod(mod2b)
```

```
compareModels(resuCV2, resuCV2b)
```

	xvari	ordre	coefOld	pvalOld	coefNew	pvalNew	change
## 6	epi	1	-0.005823084	0.06800	-0.006134561	0.05500	-5.35
## 13	SEXMen	2	0.402133194	0.01100	0.367655863	0.01200	8.57
## 3	Edat	3	0.032072433	0.00013	0.032893017	0.00007	-2.56
## 9	IMCcat18.5-<25	4	-1.988096556	0.00003	-1.983623617	0.00003	0.22
## 10	IMCcat25-<30	5	-1.947127179	0.00002	-1.927829439	0.00002	0.99
## 11	IMCcat30+	6	-1.933457134	0.00003	-1.906230254	0.00003	1.41
## 7	etiologyIschemic	7	0.109797714	0.40000	NA	NA	NA
## 8	HTAYes	8	0.527971778	0.00120	0.516466270	0.00150	2.18
## 4	EFcat41-49%	9	0.232516579	0.28000	NA	NA	NA
## 5	EFcat50+	10	0.283033824	0.14000	NA	NA	NA
## 1	ACEI_ARB_sYes	11	-0.870826768	0.00000	-0.922112657	0.00000	-5.89
## 2	ARNI_sYes	12	-0.914445951	0.00110	-0.937230965	0.00079	-2.49
## 12	MRA_sYes	13	-0.682454571	0.00000	-0.700990009	0.00000	-2.72

Es poden treure les variables (no cavien els coeficients més del 10%). El model final queda:

```
resuCV2b
```

	HR	Lower95%CI	Upper95%CI	pvalue
## epi	0.9938842	0.98766729	1.0001403	0.05500
## SEXMen	1.4443449	1.08544921	1.9219068	0.01200
## Edat	1.0334400	1.01681622	1.0503355	0.00007
## IMCcat18.5-<25	0.1375698	0.05435207	0.3482012	0.00003
## IMCcat25-<30	0.1454636	0.05950256	0.3556092	0.00002
## IMCcat30+	0.1486397	0.06040852	0.3657390	0.00003
## HTAYes	1.6760943	1.21842030	2.3056839	0.00150
## ACEI_ARB_sYes	0.3976780	0.28222099	0.5603686	0.00000
## ARNI_sYes	0.3917110	0.22664817	0.6769854	0.00079
## MRA_sYes	0.4960939	0.37916932	0.6490746	0.00000