Simulación

Yanink Neried Caro Vega y Lizbeth Naranjo Albarrán

Proyecto I: Métodos Estadísticos para la Investigación en VIH/SIDA

Increased Mortality After Tuberculosis Treatment Completion in Persons Living With Human Immunodeficiency Virus in Latin America

We assessed the association between cured tuberculosis (TB) and mortality among persons living with human immunodeficiency virus (HIV) in Latin America. We compared survival among persons with and without TB at enrollment in HIV care, starting 9 months after clinic enrollment. In multivariable analysis, TB was associated with higher long-term mortality (hazard ratio, 1.57; 95% confidence interval, 1.25–1.99).

Usando base de datos CCASAnet

Usar datos de CCASAnet:

```
art <- read.csv(paste0(dir,"art_sim.csv"))
basic <- read.csv(paste0(dir,"basic_sim.csv"))
follow <- read.csv(paste0(dir,"follow_sim.csv"))
lab_cd4 <- read.csv(paste0(dir,"lab_cd4_sim.csv"))
lab_rna <- read.csv(paste0(dir,"lab_rna_sim.csv"))
visit <- read.csv(paste0(dir,"visit_sim.csv"))
## #dim(art); #dim(basic); #dim(follow); #dim(lab_cd4); #dim(lab_rna); #dim(visit)
head(art,3)
```

```
##   patient      site     art_id      art_sd pi nnrti1 nnrti2 nnrti nrti t20 ccr5
## 1    ar.1 argentina 3TC,AZT,NVP 2007-05-16  0      1      0     1    2   0    0
## 2    ar.1 argentina 3TC,AZT,EFV 2007-05-30  0      1      0     1    2   0    0
## 3    ar.1 argentina 3TC,ABC,AZT 2007-08-03  0      0      0     0    3   0    0
##   ii1 ii2 rtv_drug numdrugs art_class      art_ed              art_rs
## 1   0   0       No        3     HAART 2007-05-28 Toxicity Dermatologic
## 2   0   0       No        3     HAART 2007-07-04             Unknown
## 3   0   0       No        3     HAART
```

```
head(basic,3)
```

```
##   patient      site baseline_d baseline_d_num male      age    birth_d
## 1    ar.1 argentina 2007-04-13          754.5    1 34.16329 1973-02-12
## 2    ar.2 argentina 2010-07-06         1934.5    0 45.89359 1964-08-13
## 3    ar.3 argentina 2011-03-28         2199.5    1 47.23421 1964-01-02
```

```
##   hivdiagnosis_d_num hivdiagnosis_d aids_y aids.miss                mode
## 1           3030.601     2007-04-13      9        0   Homosexual contact
## 2          -1270.448     1999-01-07      0        0  Injecting drug user
## 3          -1052.261     1999-08-13      0        0 Heterosexual contact
##   clinicaltrial_y mode_oth recart_y aids_d birth_d_a aids_cl_y aids_cl_d
## 1               0       NA        0        D              9
## 2               0       NA        0        D              0
## 3               0       NA        9        D              0
```

```r
head(follow,3)
```

```
##   patient     site  l_alive_d death_y death_d death_d_a drop_rs drop_rs_oth
## 1    ar.1 argentina 2013-02-04       0                D
## 2   ar.10 argentina 2013-02-13       0                D
## 3  ar.100 argentina 2013-07-12       0                D
```

```r
head(lab_cd4,3)
```

```
##   patient     site      cd4_d cd4_v time
## 1    ar.1 argentina 2007-01-19   405    0
## 2    ar.1 argentina 2008-05-07   490  474
## 3    ar.1 argentina 2009-08-26   238  950
```

```r
head(lab_rna,3)
```

```
##   patient     site      rna_d rna_v
## 1    ar.1 argentina 2007-01-19 74724
## 2    ar.1 argentina 2013-02-04   -40
## 3    ar.1 argentina 2011-11-03   399
```

```r
head(visit,3)
```

```
##   patient     site    visit_d cdcstage whostage
## 1    ar.1 argentina 2007-04-13       NA       NA
## 2    ar.2 argentina 2010-07-06       NA       NA
## 3    ar.3 argentina 2011-03-28       NA       NA
```

Cambiar formato de las fechas.

```r
class(basic$birth_d)
```

```
## [1] "factor"
```

```r
basic$birth_d <- as.Date(basic$birth_d,"%Y-%m-%d")
class(basic$birth_d)
```

```
## [1] "Date"
```

```r
basic$baseline_d <- as.Date(basic$baseline_d,"%Y-%m-%d")
follow$death_d <- as.Date(follow$death_d,"%Y-%m-%d")
follow$l_alive_d <- as.Date(follow$l_alive_d,"%Y-%m-%d")
lab_cd4$cd4_d <- as.Date(lab_cd4$cd4_d,"%Y-%m-%d")
```

Simular otras variables necesarias:

```r
lab_cd4$cd4_v_m227 = ifelse(lab_cd4$cd4_v<227,1,0)
basic$education = sample(c(1,0),size=nrow(basic),replace=TRUE,prob=c(0.15,85)) ### patients (15%) had n
basic$age_M35 = ifelse(basic$age>35,1,0)
```

Unir tablas, usando variables en común "patient" y "site":

```r
datos_bf <- merge(basic,follow, by=c("patient","site"),all=TRUE)
lab_cd4_base <- subset(lab_cd4,lab_cd4$time==0) ### CD4 baseline
datos_bf4 <- merge(datos_bf,lab_cd4_base, by=c("patient","site"), all.x=FALSE,all.y=FALSE,nu.dups=TRUE)
```

Generar una tabla para enfermedad:

```r
diseases = data.frame(patient=basic$patient, site=basic$site)
diseases$tuberculosis_y = 0
diseases$tuberculosis_d = as.Date(NA)
patients = unique(lab_cd4$patient) ### pacientes con laboratorio CD4
Ntb = n_distinct(patients)


tuberculosis = sample(c(1,0),size=Ntb,replace=TRUE,prob=c(6.8,93.2))


# Ntb_a = n_distinct(follow$patient[follow$death_y==0])
# Ntb_d = n_distinct(follow$patient[follow$death_y==1])
# prob_a = -log(0.932)/log(1.57)
# prob_d = -log(0.068)/log(1.57)
# tb_alive = sample(c(1,0),size=Ntb_a,replace=TRUE,prob=c(6.8,93.2))
# tb_death = sample(c(1,0),size=Ntb_d,replace=TRUE,prob=c(6.8,93.2))
# tuberculosis = sample(c(1,0),size=Ntb,replace=TRUE,prob=c(6.8,93.2))
```

Simular la fecha en la que se diagnostica, tomando como base las fechas de CD4:

```r
for(i in 1:Ntb){
  if(tuberculosis[i]==1){
    id = patients[i]
    datos_id = subset(lab_cd4,lab_cd4$patient==id)
    id.num = which(diseases$patient==as.character(id))
    diseases$tuberculosis_y[id.num] = 1
    diseases$tuberculosis_d[id.num] = sample(datos_id$cd4_d,size=1)
  }
}
head(diseases)


##   patient       site tuberculosis_y tuberculosis_d
```

```
## 1    ar.1 argentina              0          <NA>
## 2    ar.2 argentina              0          <NA>
## 3    ar.3 argentina              0          <NA>
## 4    ar.4 argentina              0          <NA>
## 5    ar.5 argentina              0          <NA>
## 6    ar.6 argentina              0          <NA>
```

```r
datos_bf4d <- merge(datos_bf4,diseases, by=c("patient","site"), all.x=FALSE,all.y=FALSE,nu.dups=TRUE)
```

Estimar modelo de Cox usando tabla "follow"

```r
library(survival)
datos_bf4d$tiempo = ifelse(datos_bf4d$death_y==1,
                           as.numeric(difftime(datos_bf4d$death_d,datos_bf4d$birth_d, units="auto"))/365
                           as.numeric(difftime(datos_bf4d$l_alive_d,datos_bf4d$birth_d, units="auto"))/3

datos_bf4d$muerte = datos_bf4d$death_y
```

Modelo de Cox:

```r
mod_cox <- coxph(Surv(tiempo, muerte) ~ tuberculosis_y + cd4_v_m227 + education + age_M35 , data=datos_
summary(mod_cox)
```

```
## Call:
## coxph(formula = Surv(tiempo, muerte) ~ tuberculosis_y + cd4_v_m227 +
##     education + age_M35, data = datos_bf4d)
##
##   n= 23409, number of events= 955
##
##                    coef exp(coef) se(coef)        z Pr(>|z|)
## tuberculosis_y  0.13577   1.14541  0.12182    1.114    0.265
## cd4_v_m227      1.46025   4.30703  0.09128   15.997   <2e-16 ***
## education       0.46452   1.59125  0.70815    0.656    0.512
## age_M35        -1.59398   0.20312  0.08051  -19.800   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## tuberculosis_y    1.1454     0.8730    0.9021    1.4543
## cd4_v_m227        4.3070     0.2322    3.6014    5.1509
## education         1.5913     0.6284    0.3972    6.3755
## age_M35           0.2031     4.9233    0.1735    0.2378
##
## Concordance= 0.757  (se = 0.007 )
## Likelihood ratio test= 695.2  on 4 df,   p=<2e-16
## Wald test            = 639.9  on 4 df,   p=<2e-16
## Score (logrank) test = 736.8  on 4 df,   p=<2e-16
```

```r
km <- survfit(Surv(tiempo, muerte) ~ tuberculosis_y, type = "kaplan-meier",conf.type="log-log", data=da
plot(km, conf.int=T,xlab="time", ylab="Survival", lty=c(1,4),
```

```
      lwd=2, xlim=c(0,100), col=c("red","red","red","blue","blue","blue"))
legend("topright", c("No Baseline TB","Baseline TB"), lty=c(1,4),col=c("red","blue"))

plot(km, conf.int=T,xlab="time", fun="event", ylab="Probability of Death", lty=c(1,4),
      lwd=2, xlim=c(0,100), col=c("blue","blue","blue","red","red","red"))
legend("topright", c("No Baseline TB","Baseline TB"), lty=c(1,4),col=c("red","blue"))
```
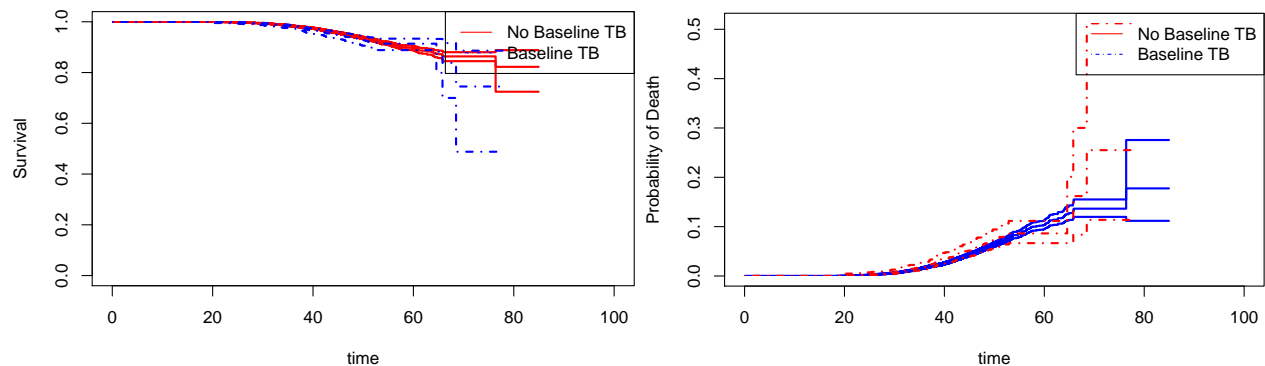

Simular variable respuesta para modelo de Cox

Valores de los parametros y simulacion del tiempo:

```
attach(datos_bf4d)
N = n_distinct(patient)
beta_HR_TB = log(1.57)
beta_HR_CD4 = log(1.57)
beta_HR_age = log(1.56)
beta_HR_edu = log(1.24)


XBeta =  tuberculosis_y*beta_HR_TB + cd4_v_m227*beta_HR_CD4 + education*beta_HR_edu + age_M35*beta_HR_ag

sigma = 3   ### parametro para weibull
tiempo = rweibull(n=N, shape=1/sigma, scale=exp(XBeta))

muerte = rep(0,N)
muerte[tuberculosis_y==1] = sample(c(1,0),size=sum(tuberculosis_y==1),replace=TRUE,
                                   prob=c(10.2,89.8))
muerte[tuberculosis_y==0] = sample(c(1,0),size=N-sum(tuberculosis_y==1),replace=TRUE,
                                   prob=c(5.6,94.4))
```

Modelo de Cox:

```
mod_cox <- coxph(Surv(tiempo, muerte) ~ tuberculosis_y + cd4_v_m227 + education + age_M35 )
summary(mod_cox)


## Call:
## coxph(formula = Surv(tiempo, muerte) ~ tuberculosis_y + cd4_v_m227 +
##     education + age_M35)
##
##   n= 23409, number of events= 1369
```

```
## 
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## tuberculosis_y  0.53403   1.70579  0.07991  6.683 2.34e-11 ***
## cd4_v_m227     -0.21758   0.80446  0.05454 -3.989 6.63e-05 ***
## education       0.45425   1.57500  0.44825  1.013    0.311
## age_M35        -0.07183   0.93068  0.05522 -1.301    0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
##                exp(coef) exp(-coef) lower .95 upper .95
## tuberculosis_y    1.7058     0.5862    1.4585    1.9950
## cd4_v_m227        0.8045     1.2431    0.7229    0.8952
## education         1.5750     0.6349    0.6542    3.7917
## age_M35           0.9307     1.0745    0.8352    1.0371
## 
## Concordance= 0.552  (se = 0.009 )
## Likelihood ratio test= 59.03  on 4 df,   p=5e-12
## Wald test            = 65.06  on 4 df,   p=2e-13
## Score (logrank) test = 66.27  on 4 df,   p=1e-13
```

```r
km <- survfit(Surv(tiempo, muerte) ~ tuberculosis_y, type = "kaplan-meier",conf.type="log-log")
plot(km, conf.int=T,xlab="time", ylab="Survival", lty=c(1,4),
     lwd=2, xlim=c(0,100), col=c("red","red","red","blue","blue","blue"))
legend("topright", c("No Baseline TB","Baseline TB"), lty=c(1,4),col=c("red","blue"))

plot(km, conf.int=T,xlab="time", fun="event", ylab="Probability of Death", lty=c(1,4),
     lwd=2, xlim=c(0,100), col=c("blue","blue","blue","red","red","red"))
legend("topright", c("No Baseline TB","Baseline TB"), lty=c(1,4),col=c("red","blue"))
```