Pràctica 3 Modelització Avançada

Clara Albert - Joel Baltasar - Enric Sintes - Sergi Orozco

20/5/2021

Importació de les dades

```
library(readxl)
datos <- read_excel("~/GEA/3er/Modelització Avançada/Practica3.xlsx")
View(datos)
```

Qualitat de les dades

Veiem l'estructura de les dades.

```
str(datos)
```

```
## tibble [1,127 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ID                     : num [1:1127] 5 6 7 8 9 10 11 12 13 14 ...
##  $ study_num              : num [1:1127] 1 1 1 1 1 1 1 1 1 1 ...
##  $ treatment              : num [1:1127] 1 1 2 2 1 1 2 1 2 1 ...
##  $ decompensation_inclu1on: num [1:1127] 0 0 0 0 1 1 0 0 0 0 ...
##  $ Gender                 : num [1:1127] 1 2 2 1 1 1 1 2 1 1 ...
##  $ edad                   : num [1:1127] 65 55 45 22 70 57 60 50 65 50
...
##  $ ChildPughClass         : num [1:1127] 2 2 2 2 2 3 2 1 2 2 ...
##  $ inr_ing                : num [1:1127] 1.91 2 1.36 1.73 1.73 3.27
1.64 1.27 1.18 1.27 ...
##  $ hb_ing                 : num [1:1127] 110 120 110 90 130 110 120
100 100 130 ...
##  $ EP_HDAvariceal         : num [1:1127] 0 0 1 0 0 0 1 0 0 0 ...
##  $ EP_HRVaricesTime       : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
##  $ OLT_THYN               : num [1:1127] 0 0 0 0 0 0 0 0 0 0 ...
##  $ OLT_Time               : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
##  $ EP_DeathYN             : num [1:1127] 0 0 0 0 1 0 1 0 0 0 ...
##  $ EP_DeathTime           : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
```

Factorització de les dades

```
cols.to.factor <- sapply(datos, function(col) length(unique(col)) < 4);
head(cols.to.factor)
```

```
##                        ID                 study_num
treatment
##                     FALSE                     FALSE
TRUE
## decompensation_inclu1on                    Gender
edad
```

```
##                     TRUE                     TRUE
FALSE

datos[cols.to.factor] <- lapply(datos[ cols.to.factor] , factor)

str(datos)

## tibble [1,127 x 15] (S3: tbl_df/tbl/data.frame)
##  $ ID                     : num [1:1127] 5 6 7 8 9 10 11 12 13 14 ...
##  $ study_num              : num [1:1127] 1 1 1 1 1 1 1 1 1 1 ...
##  $ treatment              : Factor w/ 2 levels "1","2": 1 1 2 2 1 1 2
1 2 1 ...
##  $ decompensation_inclu1on: Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1
1 1 1 ...
##  $ Gender                 : Factor w/ 2 levels "1","2": 1 2 2 1 1 1 1
2 1 1 ...
##  $ edad                   : num [1:1127] 65 55 45 22 70 57 60 50 65 50
...
##  $ ChildPughClass         : Factor w/ 3 levels "1","2","3": 2 2 2 2 2
3 2 1 2 2 ...
##  $ inr_ing                : num [1:1127] 1.91 2 1.36 1.73 1.73 3.27
1.64 1.27 1.18 1.27 ...
##  $ hb_ing                 : num [1:1127] 110 120 110 90 130 110 120
100 100 130 ...
##  $ EP_HDAvariceal         : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 2
1 1 1 ...
##  $ EP_HRVaricesTime       : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
##  $ OLT_THYN               : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1
1 1 1 ...
##  $ OLT_Time               : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
##  $ EP_DeathYN             : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 2
1 1 1 ...
##  $ EP_DeathTime           : num [1:1127] 9 8 15 10 6 8 10 6 14 6 ...
```

Fem un resum estadístic per veure si hi ha alguna dada que no està preparada per l'anàlisi.

```
summary(datos)

##        ID           study_num      treatment decompensation_inclu1on
Gender
##  Min.   :   5.0   Min.   : 1.000   1:602     0:522
1:807
##  1st Qu.: 290.5   1st Qu.: 2.000   2:525     1:605
2:320
##  Median : 658.0   Median : 6.000
##  Mean   : 671.4   Mean   : 5.755
##  3rd Qu.:1021.5   3rd Qu.: 9.000
##  Max.   :1400.0   Max.   :11.000
##
##       edad       ChildPughClass    inr_ing          hb_ing
```

```
EP_HDAvariceal
##  Min.   :18.00   1:399          Min.   :0.000   Min.   : 50.0   0:897
##  1st Qu.:49.00   2:481          1st Qu.:1.100   1st Qu.:101.2   1:230
##  Median :56.00   3:247          Median :1.300   Median :120.0
##  Mean   :55.65                  Mean   :1.364   Mean   :117.5
##  3rd Qu.:64.00                  3rd Qu.:1.500   3rd Qu.:134.0
##  Max.   :81.00                  Max.   :9.900   Max.   :173.0
##                                 NA's   :37      NA's   :417
##  EP_HRVaricesTime OLT_THYN    OLT_Time      EP_DeathYN  EP_DeathTime
##  Min.   : 0.93    0:1078   Min.   : 1.00   0:862      Min.   : 1.00
##  1st Qu.: 8.00    1:  49   1st Qu.: 9.00   1:265      1st Qu.: 9.00
##  Median :14.10             Median :16.00              Median :16.23
##  Mean   :19.22             Mean   :20.37              Mean   :20.61
##  3rd Qu.:24.00             3rd Qu.:25.00              3rd Qu.:25.54
##  Max.   :81.37             Max.   :81.37              Max.   :81.37
##
```

Missing values

```
apply(is.na(datos),2,sum)
```

```
##                       ID                study_num
treatment
##                        0                        0
0
## decompensation_inclu1on                   Gender
edad
##                        0                        0
0
##            ChildPughClass                  inr_ing
hb_ing
##                        0                       37
417
##            EP_HDAvariceal         EP_HRVaricesTime
OLT_THYN
##                        0                        0
0
##                  OLT_Time               EP_DeathYN
EP_DeathTime
##                        0                        0
0
```

La dada inr_ing té 37 missings i la variable hb_ing en té 417.

Filtració dades

```
table(datos$study_num, datos$treatment)
```

```
##
##        1  2
##    1  79 85
##    2  75 73
```

```
##   3  75 73
##   4  32  0
##   5   0 28
##   6  30 28
##   7  50 46
##   8  66  0
##   9  81 80
##   10 83 86
##   11 31 26
```

Veiem que els estudis 4,5 i 8 no tenen dades pels 2 gups de tractament i per tant els podem esborrar.

```
datos = datos[!(datos$study_num == 4 | datos$study_num == 5 |
datos$study_num == 8), ]
dim(datos)
```

```
## [1] 1001   15
```

```
table(datos$study_num, datos$treatment)
```

```
##
##       1  2
##   1  79 85
##   2  75 73
##   3  75 73
##   6  30 28
##   7  50 46
##   9  81 80
##   10 83 86
##   11 31 26
```

Anàlisi comparativa dades basals

Només realitzarem l'anàlisi de les dades basals, per tant les seleccionem i fem una taula per comparar aquestes dades a partir dels dos grups de tractament.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tableone)
```

```r
vars = colnames(datos); vars = vars[c(4,5,6,7,8,9)]
continues=names(datos[cols.to.factor==FALSE])

## Construct a table
tabUnmatched <- CreateTableOne(vars = vars, strata = "treatment", data =
datos, test = TRUE)

## Show table with SMD
print(tabUnmatched, nonnormal = continues,smd = TRUE)
```

```
##                                    Stratified by treatment
##                                     1
##   n                                    504
##   decompensation_inclu1on = 1 (%)      262 (52.0)
##   Gender = 2 (%)                       134 (26.6)
##   edad (median [IQR])                 58.00 [51.00, 67.25]
##   ChildPughClass (%)
##      1                                 167 (33.1)
##      2                                 214 (42.5)
##      3                                 123 (24.4)
##   inr_ing (median [IQR])              1.30 [1.10, 1.50]
##   hb_ing (median [IQR])              120.00 [101.00, 134.00]
##                                    Stratified by treatment
##                                     2                       p      test
SMD
##   n                                    497
##   decompensation_inclu1on = 1 (%)      283 (56.9)            0.131
0.100
##   Gender = 2 (%)                       142 (28.6)            0.528
0.044
##   edad (median [IQR])                 53.00 [45.00, 61.00]   <0.001
nonnorm  0.593
##   ChildPughClass (%)                                        0.127
0.129
##      1                                 184 (37.0)
##      2                                 217 (43.7)
##      3                                  96 (19.3)
##   inr_ing (median [IQR])              1.30 [1.10, 1.50]     0.707
nonnorm  0.087
##   hb_ing (median [IQR])              119.00 [100.00, 133.00]  0.345
nonnorm  0.070
```

```r
## Construct a table
tabUnmatched <- CreateTableOne(vars = vars, strata =
c("study_num","treatment"), data = datos, test = TRUE)
```

```
## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf
```

```
## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf

## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf

## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf

## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf

## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in min(x, na.rm = TRUE): ningún argumento finito para min;
retornando
## Inf

## Warning in max(x, na.rm = TRUE): ningun argumento finito para max;
retornando -
## Inf

## Warning in StdDiff(variable = var, group = strataVar): Variable has
only NA's in
## at least one stratum. na.rm turned off.
```

```
## Show table with SMD
print(tabUnmatched, nonnormal = continues,smd = TRUE)
```

```
##                                    Stratified by study_num:treatment
##                                     1:1                     2:1
```

```
##   n                                      79                      75
##   decompensation_inclu1on = 1 (%)     32 (40.5)               40 (53.3)
##   Gender = 2 (%)                      21 (26.6)               22 (29.3)
##   edad (median [IQR])              55.00 [47.50, 61.00]   59.00
[51.00, 67.00]
##   ChildPughClass (%)
##      1                                27 (34.2)               28 (37.3)
##      2                                34 (43.0)               16 (21.3)
##      3                                18 (22.8)               31 (41.3)
##   inr_ing (median [IQR])            1.36 [1.22, 1.61]      1.10 [1.00,
1.50]
##   hb_ing (median [IQR])           120.00 [100.00, 130.00]    NA [NA,
NA]
##                                   Stratified by study_num:treatment
##                                    3:1
##   n                                   75
##   decompensation_inclu1on = 1 (%)     25 (33.3)
##   Gender = 2 (%)                      22 (29.3)
##   edad (median [IQR])              64.00 [55.00, 70.50]
##   ChildPughClass (%)
##      1                                31 (41.3)
##      2                                31 (41.3)
##      3                                13 (17.3)
##   inr_ing (median [IQR])            1.26 [1.17, 1.40]
##   hb_ing (median [IQR])           125.00 [114.25, 139.75]
##                                   Stratified by study_num:treatment
##                                    6:1                     7:1
##   n                                   30                      50
##   decompensation_inclu1on = 1 (%)     20 (66.7)               19 (38.0)
##   Gender = 2 (%)                       7 (23.3)               10 (20.0)
##   edad (median [IQR])              56.00 [52.25, 68.00]   61.00
[52.00, 70.75]
##   ChildPughClass (%)
##      1                                 5 (16.7)               22 (44.0)
##      2                                20 (66.7)               17 (34.0)
##      3                                 5 (16.7)               11 (22.0)
##   inr_ing (median [IQR])            1.20 [1.20, 1.30]      1.10 [1.00,
1.48]
##   hb_ing (median [IQR])           130.00 [120.00, 142.50]    NA [NA,
NA]
##                                   Stratified by study_num:treatment
##                                    9:1
##   n                                   81
##   decompensation_inclu1on = 1 (%)     81 (100.0)
##   Gender = 2 (%)                      18 ( 22.2)
##   edad (median [IQR])              55.00 [49.00, 64.00]
##   ChildPughClass (%)
##      1                                 0 (  0.0)
##      2                                52 ( 64.2)
##      3                                29 ( 35.8)
```

```
##   inr_ing (median [IQR])           1.45 [1.20, 1.70]
##   hb_ing (median [IQR])          106.00 [90.00, 120.00]
##                                 Stratified by study_num:treatment
##                                  10:1                    11:1
##   n                                  83                      31
##   decompensation_inclu1on = 1 (%)    36 (43.4)                9 (29.0)
##   Gender = 2 (%)                     20 (24.1)               14 (45.2)
##   edad (median [IQR])             58.00 [52.00, 66.50]   69.00
[60.00, 77.00]
##   ChildPughClass (%)
##      1                               41 (49.4)               13 (41.9)
##      2                               33 (39.8)               11 (35.5)
##      3                                9 (10.8)                7 (22.6)
##   inr_ing (median [IQR])           1.20 [1.10, 1.30]      1.60 [1.35,
2.10]
##   hb_ing (median [IQR])          125.00 [108.50, 141.00]    NA [NA,
NA]
##                                 Stratified by study_num:treatment
##                                  1:2                     2:2
##   n                                  85                      73
##   decompensation_inclu1on = 1 (%)    34 (40.0)               48 (65.8)
##   Gender = 2 (%)                     22 (25.9)               19 (26.0)
##   edad (median [IQR])             50.00 [37.00, 57.00]   54.00
[46.00, 64.00]
##   ChildPughClass (%)
##      1                               42 (49.4)               23 (31.5)
##      2                               36 (42.4)               18 (24.7)
##      3                                7 ( 8.2)               32 (43.8)
##   inr_ing (median [IQR])           1.30 [1.18, 1.55]      1.20 [1.10,
1.40]
##   hb_ing (median [IQR])          110.00 [100.00, 127.50]    NA [NA,
NA]
##                                 Stratified by study_num:treatment
##                                  3:2
##   n                                  73
##   decompensation_inclu1on = 1 (%)    38 (52.1)
##   Gender = 2 (%)                     25 (34.2)
##   edad (median [IQR])             56.00 [47.00, 61.00]
##   ChildPughClass (%)
##      1                               33 (45.2)
##      2                               30 (41.1)
##      3                               10 (13.7)
##   inr_ing (median [IQR])           1.32 [1.13, 1.50]
##   hb_ing (median [IQR])          131.00 [119.00, 144.00]
##                                 Stratified by study_num:treatment
##                                  6:2                     7:2
##   n                                  28                      46
##   decompensation_inclu1on = 1 (%)    18 (64.3)               20 (43.5)
##   Gender = 2 (%)                      9 (32.1)               13 (28.3)
##   edad (median [IQR])             56.00 [46.75, 62.00]   55.00
```

```
[49.00, 63.50]
##   ChildPughClass (%)
##      1                                  10 (35.7)              20 (43.5)
##      2                                  14 (50.0)              20 (43.5)
##      3                                   4 (14.3)               6 (13.0)
##   inr_ing (median [IQR])              1.20 [1.10, 1.40]      1.10 [1.00,
1.37]
##   hb_ing (median [IQR])            130.00 [120.00, 140.00]    NA [NA,
NA]
##                                    Stratified by study_num:treatment
##                                     9:2
##   n                                    80
##   decompensation_inclu1on = 1 (%)      80 (100.0)
##   Gender = 2 (%)                       18 ( 22.5)
##   edad (median [IQR])               46.00 [40.00, 55.00]
##   ChildPughClass (%)
##      1                                  0 (  0.0)
##      2                                 49 ( 61.3)
##      3                                 31 ( 38.8)
##   inr_ing (median [IQR])              1.50 [1.28, 1.72]
##   hb_ing (median [IQR])             100.00 [80.00, 110.50]
##                                    Stratified by study_num:treatment
##                                     10:2                    11:2
##   n                                    86                      26
##   decompensation_inclu1on = 1 (%)      34 (39.5)              11 (42.3)
##   Gender = 2 (%)                       27 (31.4)               9 (34.6)
##   edad (median [IQR])               53.00 [48.00, 58.00]   62.50
[51.00, 66.75]
##   ChildPughClass (%)
##      1                                 45 (52.3)              11 (42.3)
##      2                                 37 (43.0)              13 (50.0)
##      3                                  4 ( 4.7)               2 ( 7.7)
##   inr_ing (median [IQR])              1.20 [1.10, 1.40]      1.35 [1.10,
1.50]
##   hb_ing (median [IQR])            119.00 [106.00, 135.50]    NA [NA,
NA]
##                                    Stratified by study_num:treatment
##                                     p      test    SMD
##   n
##   decompensation_inclu1on = 1 (%) <0.001          0.580
##   Gender = 2 (%)                   0.611          0.153
##   edad (median [IQR])             <0.001 nonnorm  0.559
##   ChildPughClass (%)              <0.001          0.621
##      1
##      2
##      3
##   inr_ing (median [IQR])          <0.001 nonnorm  0.380
##   hb_ing (median [IQR])           <0.001 nonnorm     NA
```

Quan comparem els dos grups globalment, les dades basals són gairebé totes comparables entre elles, ja que la seva diferència mitjana estadaritzada és més petita que 0.2 en la majoria dels casos. Quan afegim el número d'estudi per comparar-les, el SMD creix en general i moltes variables passen a ser no comparables entre grups i estudis.

Supòsit A

Sota el supòosit A tots els estudis del meta-anàlisi són assaigs clínics. Això vol dir que els estudis tenen un mostreig aleatori i, per tant, les possibles variables confusores estan neutralitzades pel factor aleatòri. Al ser un assaigs clínics podem fer l'anàlisi del HR (hazard ratio) més fàcil, ja que en el model de Cox no hem d'incloure covariables, només el tractament.

Supòsit B

Meta-analisis sobre el supòsit A

Primer pas Analizem cada estudi per separat, obtenim com a resultat l'estimador (ln (HR)) i l'error estàndard (ee(ln (HR))).

Pasem la variable tractament a factor

```
treatment_f=factor(datos$treatment,levels=c(1,2),labels =
c("cas","control"))
datos$study_num=as.factor(datos$study_num)

datos = cbind(datos,treatment_f)

library("survival")
library("survminer")

## Loading required package: ggplot2

## Loading required package: ggpubr
```

Variable mort

```
HR<-coxph(Surv(EP_DeathTime,as.numeric(EP_DeathYN)) ~
treatment_f*strata(study_num), data = datos)
sum=summary(HR)

#estimador
lnHR_mort<-unname(HR$coefficients)

#error estandard
EElnHR_mort<-unname(sum$coefficients[,3])
```

Variable OLT

```
#1)
HR<-coxph(Surv(OLT_Time,as.numeric(OLT_THYN)) ~
treatment_f*strata(study_num), data = datos)

## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights =
weights, :
## Loglik converged before variable 1,2,3,4,6 ; coefficient may be
infinite.

sum=summary(HR)

#estimador
lnHR_olt<-unname(HR$coefficients)

#error estandard
EElnHR_olt<-unname(sum$coefficients[,3])
```

Variable EP

```
#1)
HR<-coxph(Surv(EP_HRVaricesTime,as.numeric(EP_HDAvariceal)) ~
treatment_f*strata(study_num), data = datos)
sum=summary(HR)

#estimador
lnHR_ep<-unname(HR$coefficients)

#error estandard
EElnHR_ep<-unname(sum$coefficients[,3])
```

Segon pas

Ara hem de calcular l'estimador resum de tots els estudis, la significació estadística i resultats d'heterogeneïtat.

En aquest cas utilitzarem la ponderació per l'invers de la variància. 1/ee(ln (HRi))^2

Ara, tenint en compte els pesos, podem calcular, Hazard Ratio (HR) i error estàndard de l'estimador.

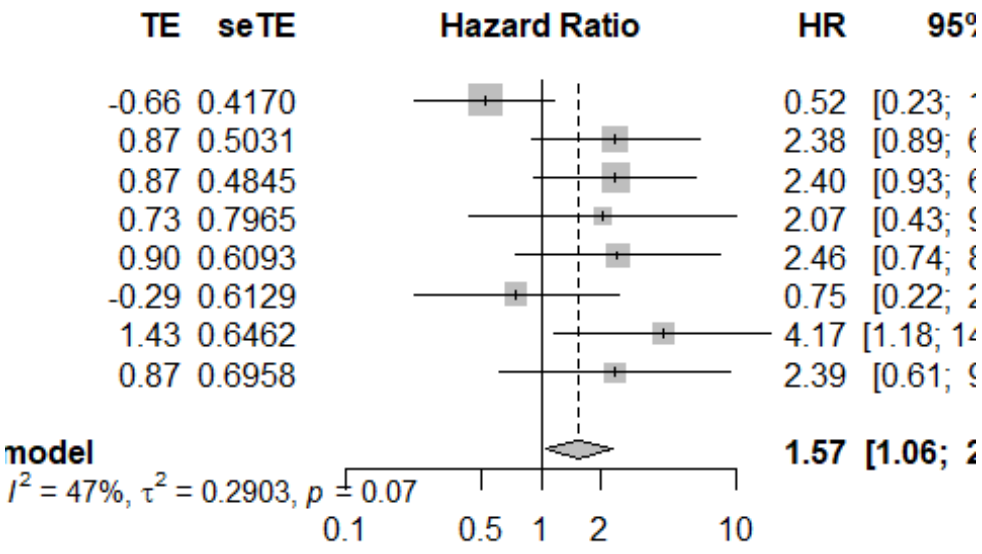
L'estimació dels pesos fet anteriorment és en el cas de efectes fixes. Si volem aplicar el mètode d'efectes aleatòris l'estimació dels pesos és una mica diferent.

Per fer aquest segon pas de mètodes aleatoris utilitzarem la funcio metagen de la llibreria meta de R.

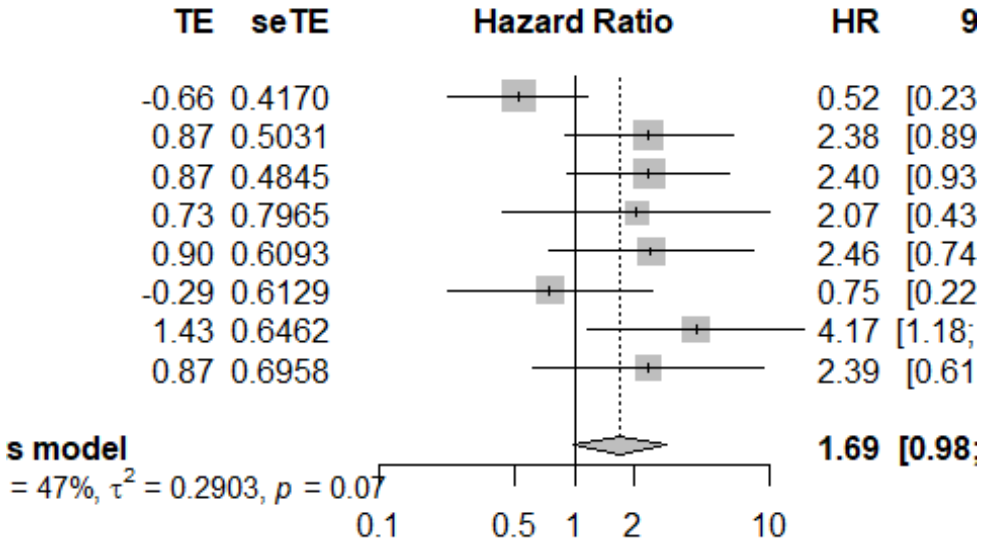
La millor manera de resumir el resultats és mitjançant el forest plot, on mirem com és l'estimador resum i si és significatiu o no.

Variable mort

```
library(meta)

## Registered S3 methods overwritten by 'lme4':
##   method                           from
##   cooks.distance.influence.merMod car
##   influence.merMod                 car
##   dfbeta.influence.merMod          car
##   dfbetas.influence.merMod         car

## Loading 'meta' package (version 4.18-1).
## Type 'help(meta)' for a brief overview.

studies = unique(datos$study_num)

# Efectes fixos
pooled1<-metagen(TE=lnHR_mort , seTE= EElnHR_mort, studlab= studies,
data=HR, sm="HR", comb.fixed = TRUE, comb.random = FALSE ,method.tau =
"REML")
forest(pooled1)
```

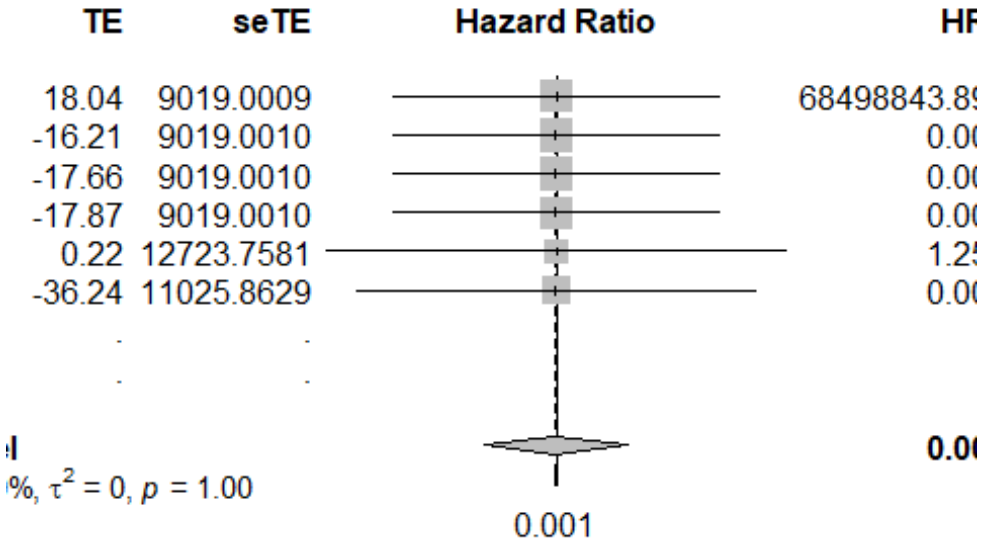



```
#Efectes aleatoris
pooled2<-metagen(TE=lnHR_mort , seTE= EElnHR_mort, studlab= studies,
data=HR, sm="HR",comb.fixed = FALSE, comb.random = TRUE,method.tau =
"REML")
forest(pooled2)
```

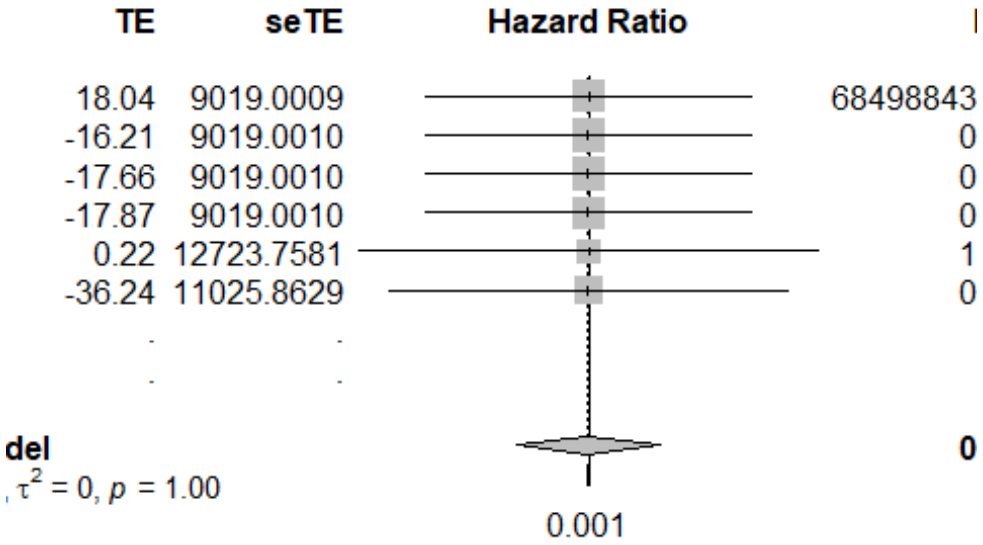


Variable OLT

```
library(meta)
studies = unique(datos$study_num)

# Efectes fixos
pooled1<-metagen(TE=lnHR_olt , seTE= EElnHR_olt, studlab= studies,
data=HR, sm="HR", comb.fixed = TRUE, comb.random = FALSE ,method.tau =
"REML")

## Warning in metagen(TE = lnHR_olt, seTE = EElnHR_olt, studlab =
studies, : Zero
## values in seTE replaced by NAs.

forest(pooled1)
```

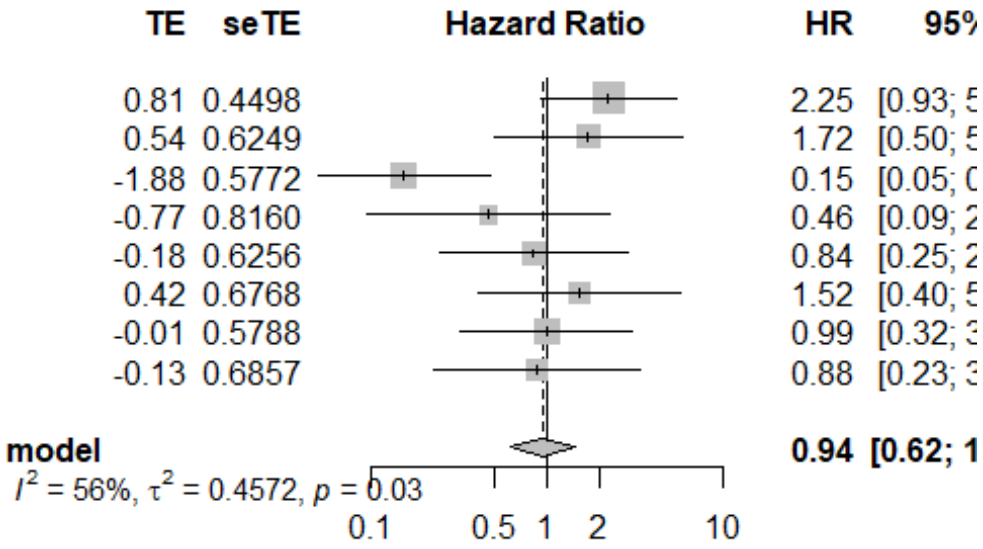



```
#Efectes aleatoris
pooled2<-metagen(TE=lnHR_olt , seTE= EElnHR_olt , studlab= studies,
data=HR, sm="HR",comb.fixed = FALSE, comb.random = TRUE,method.tau =
"REML")

## Warning in metagen(TE = lnHR_olt, seTE = EElnHR_olt, studlab =
studies, : Zero
## values in seTE replaced by NAs.

forest(pooled2)
```

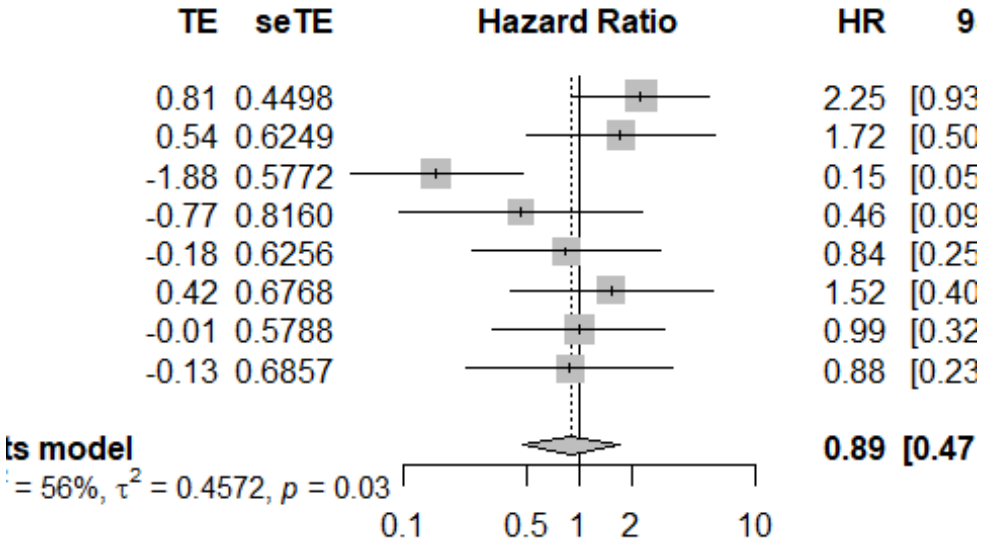


Variable EP

```
library(meta)
studies = unique(datos$study_num)

# Efectes fixos
pooled1<-metagen(TE=lnHR_ep , seTE= EElnHR_ep, studlab= studies, data=HR,
sm="HR", comb.fixed = TRUE, comb.random = FALSE ,method.tau = "REML")
forest(pooled1)
```




```
#Efectes aleatoris
pooled2<-metagen(TE=lnHR_ep , seTE= EElnHR_ep, studlab= studies, data=HR,
sm="HR",comb.fixed = FALSE, comb.random = TRUE,method.tau = "REML")
forest(pooled2)
```


TE
seTE
Hazard Ratio
HR
9
0.81 0.4498
0.54 0.6249
-1.88 0.5772
-0.77 0.8160
-0.18 0.6256
0.42 0.6768
-0.01 0.5788
-0.13 0.6857
2.25 [0.93
1.72 [0.50
0.15 [0.05
0.46 [0.09
0.84 [0.25
1.52 [0.40
0.99 [0.32
0.88 [0.23
ts model
0.89 [0.47
= 56%, τ² = 0.4572, p = 0.03
0.1
0.5
1
2
10