

Web-Appendix

Time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks: An application to peritoneal dialysis program

Laetitia Teixeira^{1, 2}, Carmen Cadarso-Suárez³,

Anabela Rodrigues^{1, 4}, and Denisa Mendonça^{1, 2}

¹ Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

² EPIUnit, Public Health Institute, University of Porto, Porto, Portugal

³ Unit of Biostatistics, Department of Statistics and Operations Research, School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

⁴ Nephrology Department, CHP-Hospital Geral de Santo António, Porto, Portugal

Address for correspondence: Laetitia Teixeira, Institute of Biomedical Sciences Abel Salazar, University of Porto, Rua de Jorge Viterbo Ferreira, 228 4050-313 Porto, Portugal.

E-mail: lcteixeira@icbas.up.pt.

Phone: (+351) 220 428 167.

Fax: (+351) 220 428 000.

Abstract: We offer script named `NestedCompRisks` to perform a similar analysis to that carried out in the manuscript "Time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks. An application to peritoneal dialysis program". In this web-appendix, `NestedCompRisks` is presented after a brief introduction section on the estimation procedure. This appendix has as main objective to explain the use of `NestedCompRisks` considering a hypothetical/simulated data, different from the data used in the main manuscript but with similar characteristics (structure of the multi-state model, number of transitions and number and type of covariates used). This script is composed by three main functions: (i) `Expl_NCR`: to perform the exploratory analysis, including e.g. Nelson-Aalen and Aalen-Johansen estimators, (ii) `Mult_NCR`: to present the results obtained from fitting semiparametric multi-state models in the presence of competing risks. The output includes flexible hazard rate curves and their confidence bands; and (iii) `AUC_NCR`: to calculate the time-dependent predictive accuracy of the fitted model through time-dependent ROC curves, and the temporal AUCs, together with their corresponding bootstrap confidence intervals.

Key words: time-dependent ROC curve; competing risks; multi-state models; peritoneal dialysis; STAR model; survival analysis

I - Estimation Procedure

1. Script_BayesX.txt

1.1. Data

The software BayesX is used to estimate the parameters of the STAR model. This software requires that the datafile (with extension .raw) has a specific structure, as presented in Figure 1:

- Each line of the data corresponds to one transition. Patients with two different transitions appear twice in the datafile;
- Data is composed by the following variables: 'id', 'st', 'beg', 'end', 'trans1', 'trans2', 'trans3', 'trans4', 'trans5', 'trans6', 'trans7', ..., 'covariates';
- 'id' represents the code used to identify each subject;
- 'st' represents the order of transition that was observed;
- 'beg' and 'end' represent the entrance and the exit times of the transition, respectively;
- 'trans1', 'trans2', 'trans3', 'trans4', 'trans5', 'trans6', 'trans7', ..., are dummy variables that represent the transitions that composed the multi-state model;
- Covariates are included at the end of the data.

id	st	beg	end	trans1	trans2	trans3	trans4	trans5	trans6	trans7	sex	...
1	1	0	7	0	0	0	0	1	0	0	1	
3	1	0	2.83	1	0	0	0	0	0	0	1	
3	2	2.83	22	0	0	0	1	0	0	0	1	
13	1	0	13	0	0	0	0	1	0	0	0	

Figure 1: Example of data structure required to estimate the parameters of STAR model using **BayesX**.

1.2. Script

Script_BayesX.txt is an example of the type of script used to estimate the parameters of a multi-state STAR model (Figure 2). For each transition that composed the multi-state model, information about reference values of continuous covariates needs to be included.

Analyzing for example the description of 'trans1' as represented in Figure 2, the variables 'sex', 'diab' and 'first' are considered as fixed effects and 'baseline' and 'age' are considered as flexible effects with reference values of 8 and 55, respectively.

1.3. Results

For each transition that composed the multi-state model, the results obtained for fixed and flexible effects using **Script_BayesX.txt** are saved in different files. Considering the model presented in Figure 2, several files are generated for each transition: one for the fixed covariates; two for the baseline flexible effects; two for each continuous covariates considered as flexible effects. All these results referring to the fixed and flexible effects together with a detailed explanation are summarized in a pdf file.

```

dataset d

d.infile using 'directory of the database'

remlreg m1

m1.outfile= 'directory where results are to be saved'

m1.mregress

trans1 = end(baseline, reference=8) + sex + diab + first + age (psplinerw2,
reference=55):

trans2 = end(baseline, reference=17) + sex + diab + first + age
(psplinerw2, reference=55):

trans3= end(baseline, reference=18) + sex + diab + first + age (psplinerw2,
reference=55):

trans4= end(baseline, reference=20) + sex + diab + first + age (psplinerw2,
reference=55):

trans5= end(baseline, reference=13) + sex + diab + first + age (psplinerw2,
reference=55):

trans6= end(baseline, reference=12) + sex + diab + first + age (psplinerw2,
reference=55):

trans7= end(baseline, reference=12) + sex + diab + first + age (psplinerw2,
reference=55),

family=multistate state=st lefttrunc=beg eps = 0.0001 maxit =30000 using d

```

Figure 2: Script_BayesX.txt used to estimate the parameters of the STAR model

II - NestedCompRisks

The script named `NestedCompRisks` is composed by three main functions: (1) `Expl_NCR`: to perform the exploratory analysis, (2) `Mult_NCR`: to present the results obtained from fitting semiparametric multi-state models in the presence of competing risks; and (3) `AUC_NCR`: to calculate the time-dependent predictive accuracy of the fitted model through time-dependent ROC curves, and the temporal AUCs, together with their corresponding bootstrap confidence intervals.

1. `Expl_NCR(data)`

The function `Expl_NCR(data)` is used to perform the exploratory analysis for a multi-state model in the presence of competing risks. This function allows the estimation and representation of Nelson-Aalen and Aalen-Johansen estimators.

1.1. Data

The datafile required in order to use the function `Expl_NCR(data)` has to present a specific structure (with extension `.txt`), as shown in Figure 3:

- Each line of the data corresponds to one transition. Patients with two different transition appears twice in the datafile;
- Data is composed by the following variables: 'id', 'entry', 'exit', 'from', 'to';
- 'id' represents the code used to identify each subject;
- 'entry' and 'exit' represent the entrance time and the exit time of the transition, respectively;

- 'from' and 'to' represent the states associated to the entrance time and the exit time, respectively.
- If patient was censored, code 'cens' needs to be considered.

id	entry	exit	from	to
1	0	7	0	5
3	0	2.83	0	1
3	2.83	22	1	4
...
395	0	34	0	'cens'

Figure 3: Example of data structure required by the function `Expl_NCR(data)`.

1.2. Script

The function `Expl_NCR(data)` requires only the argument 'data', that must comply the structure defined in the previous section. Figure 4 represents some extracts of this function.

```
Expl_NCR=function (data) {
tra.prog=matrix(FALSE, 8, 8, dimnames=list(as.character(0:7),
as.character(0:7)))
tra.prog[1, c(2, 6:8)]=TRUE
tra.prog[2, 3:5]=TRUE
tra.prog
```

```

m11=mvna(data, c("0", "1", "2", "3", "4", "5", "6", "7"), tra.prog, "cens")
m12=etm(data, c("0", "1", "2", "3", "4", "5", "6", "7"), tra.prog,"cens", s
= 0)
...
}

```

Figure 4: Some extracts of the function `Expl_NCR(data)`.

1.3. Results

Using the function `Expl_NCR(data)`, results of exploratory analysis are summarized in two graphs, one that represents Nelson-Aalen estimator and one that represents Aalen-Johansen estimator (not shown).

2. `Mult_NCR(dir)`

Considering the results obtained by `Script_BayesX.txt` using the BayesX software, the function `Mult_NCR(dir)` allows the graphic representation of the results of the flexible effects considered in the multi-state model. Additionally, for each transition the 95% pointwise confidence bands and the reference values considered are included.

2.1. Script

The argument 'dir' required by this function represents the directory where results are to be saved, as defined in the third line of the `Script_BayesX.txt` represented in Figure 2. Figure 5 shown some extracts of the function `Mult_NCR(dir)`.


```

Mult_NCR = function(dir) {

data_bas1=read.csv(paste(dir,"/","m1__f_end_logbaseline.res",sep=""),
sep="")

data_age1=read.csv(paste(dir,"/","m1__f_age_pspline.res",sep=""), sep="")

...

windows()

layout(matrix(c(1, 1, 1, 1, 2, 1, 3, 4, 5, 6, 7, 8), 2, 6, byrow = TRUE))

plot(0, type="n", xlab=NULL, ylab=NULL, col.axis="white", col.lab="white",
tck=0)

box(col = "white")

plot(data_age1$age,data_age1$ci95lowerref, type="l", xlab="Age",
ylab="Log Hazard Ratio",xlim=c(10,90), ylim=c(-8,4), col = "white",
main="Peritonitis")

polygon(c(data_age1$age,rev(data_age1$age)), c(data_age1$ci95lowerref,
rev(data_age1$ci95 upperref)),col=8, lty=0)

lines(data_age1$age, data_age1$pmoderef)

...

```

Figure 5: Some extracts of the function `Mult_NCR(dir)`.

2.2. Results

As an example of the type of results provided by this function, Figure 6 represents adjusted smooth log hazard ratio with 95% pointwise confidence bands for age with 55 years as reference value.

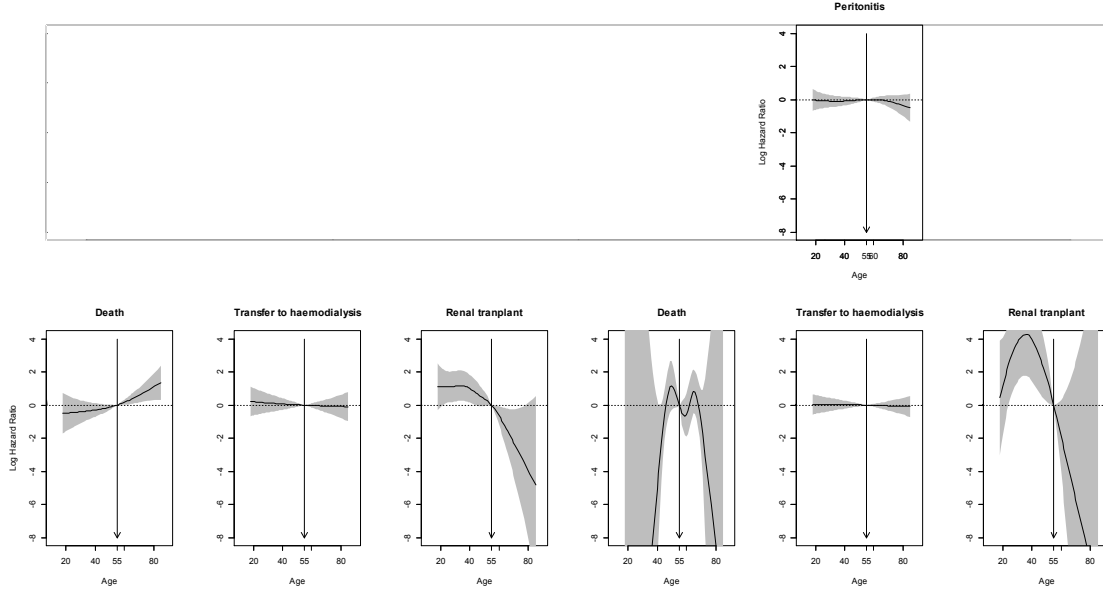


Figure 6: Adjusted smooth log hazard ratio estimates with 95% pointwise confidence bands for age (value 55 as reference) for STAR model.

3. AUC_NCR(data, dir)

The function `AUC_NCR(data,dir)` calculates and represents the time-dependent predictive accuracy of the fitted model through time-dependent ROC curves, and the temporal AUCs, together with their corresponding bootstrap confidence intervals.

3.1. Data

The data structure required by this function is similar to the structure considered in the script `Script_BayesX.txt` (see Figure 1).

3.2. Script

The argument 'data' required by the function `AUC_NCR(data,dir)` refers to the data file and the argument 'dir' represents the directory where results are to be saved, as defined in the third line of the `Script_BayesX.txt` represented in Figure 2.

Figure 7 shown some extracts of the function `AUC_NCR(data,dir)`.

```
AUC_NCR=function(data2, dir){
  data=as.data.frame(cbind(data2$end, data2$trans1, data2$trans2,
    data2$trans3, data2$trans4, data2$trans5, data2$trans6, data2$trans7,
    data2$sex, data2$diab, data2$first, data2$age))
  names(data)=c("end","trans1","trans2","trans3","trans4","trans5","trans6",
    "trans7", "sex","diab","first","age")
  smooth.f=as.data.frame(cbind(data2$age))
  names(smooth.f)=c("age")
  linear.predictor=function(model, data, smooth.f = NULL) {
    beta1=read.table(paste(model,"_FixedEffects.res",sep=""), header=TRUE)
    ...
    m1=data[, as.character(beta1$varname[-1])]
    ...
    lp1=as.matrix(m1)%*%beta1$pmode[-1]
    lp1=lp1 + beta1$pmode[1]
    ...
    if (!is.null(smooth.f))
    for (i in 1:length(smooth.f)) {
      cov1=read.table(paste(model,"_f_",smooth.f[i],"_pspline.res",sep=""),
        header=TRUE)
```

```

sf1=approxfun(cov1[,smooth.f[i]],cov1$pmode,yleft=min(cov1$pmode),
yright=max(cov1$pmode))(data[,smooth.f[i]])

lp1=lp1 + sf1

...

lp=data.frame(lp1, lp2, lp3, lp4, lp5, lp6, lp7)
}

setwd(dir)

times=seq(min(data2$end),max(data2$end),by=1)

aux1=function(d){

eta1=linear.predictor("m1", data = d, smooth.f = c("age"))

AUC1=NULL

for(t in times) {

AUC1=c(AUC1,survivalROC.C(Stime=d$end,status=d$trans1,
marker=eta1$lp1,predict.time=t,span=0.05)$AUC)

}

AUC1 ...

data1.mod=censboot(data,aux1,R=999)

data1.env=envelope(data1.mod,level=0.95)$point

...}

```

Figure 7: Some extracts of the function `AUC_NCR(data,dir)`.

3.3. Results

As an example of the type of results provided by this function, Figure 8 represents temporal AUC and bootstrap 95% confidence interval

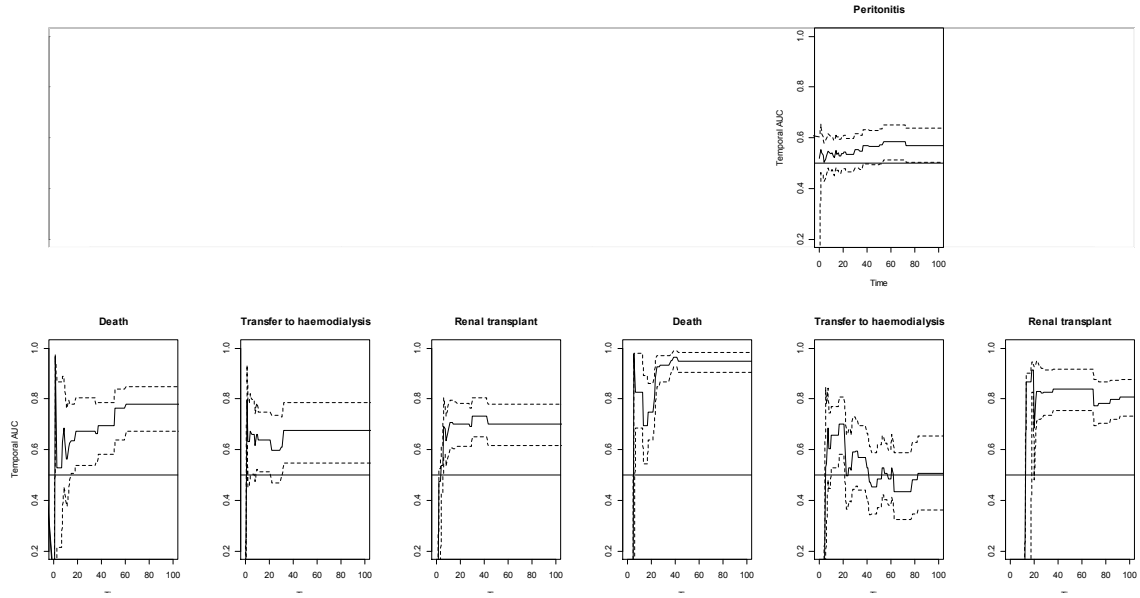


Figure 8: Temporal AUC for STAR model (solid line) and bootstrap 95% confidence interval (dotted line)