**Supplementary materials**

**S4 - R CODE FOR MULTIPLE IMPUTATIONS**

## 1. MULTIPLE IMPUTATIONS USING A UNIFORM DISTRIBUTION

## WITHIN CENSORING INTERVALS

##Times layout (a probability will be estimated for each time in MyTimes)

MyTimes <- seq(0, max(MyData$T2[is.finite(MyData$T2)]), by=0.01)

##number of imputations

M <- 200

##for the M imputations, for each ‘interval-censored’ patient,

##impute a value between T1 and T2 using a uniform distribution

imputation <- t(with(MyData[MyData$censoring%in%"interval",],

mapply(runif, n=M, min=T1, max=T2)))

##RESULTS

##for each imputation (+ data of patients not interval censored)

##use the following function to determine the survival probability

##at each time of MyTimes

imputation.res<-function(i)

{

stmp <- survfit(Surv(standard,status)~1,

data=rbind(data.frame(standard=imputation[,i],status=1),

MyData[!MyData$censoring%in%"interval",

c("standard","status")]),

conf.type="log-log",se.fit=TRUE,error="greenwood")

mysummary <- summary(stmp, times=MyTimes, extend=TRUE)

return( cbind ( surv=mysummary$surv,

se=mysummary$std.err,

low=mysummary$lower,

up=mysummary$upper))

}

##apply this function to each imputation

MI1<-sapply(1:M,FUN=imputation.res,simplify="array")

##each column of MI1 represents an imputation

##each row of MI1 represents a time from MyTimes

#Mean parameter of the survival probability at each time from MyTimes:

MyMeans1<-apply(MI1[,"surv",],MARGIN=1,FUN=mean)

## Mean parameter of the lower bound at each time from MyTimes:

MyLow1<-apply(MI1[,"low",],MARGIN=1,FUN=mean)

## Mean parameter of the upper bound at each time from MyTimes:

MyUp1<-apply(MI1[,"up",],MARGIN=1,FUN=mean)

## 2. MULTIPLE IMPUTATIONS USING THE NPMLE DISTRIBUTION

##0. here is the NPMLE of the data (once):

NPMLE <- icfit( Surv( T1, T2, type = "interval2") ~ 1,

data = MyData, conf.int=FALSE)

##0. detect all intervals and bounds with dedicated probabilities:

tables <- data.frame(IntTimes=c(0,as.vector(NPMLE$intmap)),

pf=c(0,rep(NPMLE$pf,each=2)))

tables<-tables[with(tables, is.finite(IntTimes)),]

tables<-unique(tables)

##0. define the time layout

AllTimes<-seq(0,max(tables$IntTimes),by=0.01)

##define the NPMLE Probability density function for each time from AllTimes

MyPDF <- data.frame(TIMES=AllTimes,

PDF=with(tables,

approx(IntTimes,pf,AllTimes,method="linear"))$y)

##0. define the NPMLE Cumulative density function (CDF)

## for each time from AllTimes

MyCDF<-data.frame(TIMES=MyPDF$TIMES,

CDF=cumsum(MyPDF$PDF/sum(MyPDF$PDF)))

##0.Function to impute a value between T1 and T2, knowing the CDF:

NPMLEI<-function(N,t1,t2)

{

##1/ simulation with a uniform distribution between min and max value

##reached by the CDF between T1 and T2

##2/ using the approx() function, impute the simulated time,

##knowing the CDF by a linear approximation

return( with( MyCDF[MyCDF$TIMES>t1 & MyCDF$TIMES<t2,],

approx( CDF, TIMES, runif(N,min(CDF),max(CDF)),

method="linear")$y))

}

##number of imputations

M <- 200

##for the M imputations, for each ‘interval-censored’ patient,

##impute a value between T1 and T2 using the NPMLE distribution

Imputation <- t(with(MyData[MyData$censoring%in%"interval",],

mapply(NPMLEI,N=M,t1=T1,t2=T2)))

##apply this function to each imputation

MI2<-sapply(1:M,FUN=imputation.res,simplify="array")

##As MI1, each column of MI2 represents an imputation

##each row of MI2 represents a time from MyTimes

#Mean parameter of the survival probability at each time from AllTimes:

MyMeans2<-apply(MI2[,"surv",],MARGIN=1,FUN=mean)

## Mean parameter of the lower bound at each time from AllTimes:

MyLow2<-apply(MI2[,"low",],MARGIN=1,FUN=mean)

## Mean parameter of the upper bound at each time from AllTimes:

MyUp2<-apply(MI2[,"up",],MARGIN=1,FUN=mean)