

# Demo-Data analysis

Li-Pang Chen and Bangxu Qiu

2023-06-16

## R Markdown

In this R Markdown, we demonstrate analysis of the Signal Tandmobiel Study in the main text entitled “Analysis of Length-Biased and Partly Interval-Censored Survival Data with Mismeasured Covariates”. To illustrate our analysis, we primarily focus on tooth label 11 as shown in the main text. If one would like to examine other teeth as demonstrated in the supporting information, then the command in “tandmob2” can be adjusted manually.

In the following code, we aim to reproduce the result and separately examine several estimation methods in the main text. All methods' names can be referred to the main text. To simply the illustration and focus on showing different methods, we fix the misclassification probability 0.05 when correcting for measurement error (Methods 1 and 5); one can revise the code manually to demonstrate 0.005 misclassification probability.

Method 1: SIMEXboost(0.05)

```
#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset
```

```
#required packages
library(mice)
```

```
##
## 载入程辑包：'mice'
```

```
## The following object is masked from 'package:stats':
##
##      filter
```

```
## The following objects are masked from 'package:base':
##
##      cbind, rbind
```

```
library(DMwR2)
```

```
## Registered S3 method overwritten by 'quantmod':
##      method      from
##      as.zoo.data.frame zoo
```

```
library(openxlsx)
library(survival)
library(survminer)
```

```
## 载入需要的程辑包：ggplot2
```

```
## 载入需要的程辑包：ggpubr
```

```
##  
## 载入程辑包：'survminer'
```

```
## The following object is masked from 'package:survival':  
##  
##      myeloma
```

```
#Data setup
```

```
## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.
```

```
## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses  
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]  
A<-AA[-c(which(is.na(AA[,22]))),]
```

```
## remove the first three non-informative columns and other teeth labels  
data<-A[,c(-1,-2,-4,-(23:32),-(33:132))]  
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)
```

[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1[-c(1741:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])
data2<-data1[,c(-(18:19))]

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=0.03
  dataF2=dataF1[-c(((kfold-1)*(n/5))+1):(((kfold)*(n/5)))),]

  LR<-cbind(dataF2[,1],dataF2[,2])

  betarbind<-NULL;eta<-c(0,0.25,0.5,0.75,1);B=5
  for(cc in 1:length(eta)){
    betaihat<-NULL
    for(i in 1:B){
      X1<-t(dataF2[,c(3:43)])
      Xchange1<-X1[c(1,(5:41)),]
      orthogonal<-matrix(c(sqrt(0.5),sqrt(0.5),sqrt(0.5),-sqrt(0.5)),nrow=2,ncol=2,byrow=TRUE)
      orthogonalT<-t(orthogonal)
      piepsilon<-orthogonalT%*(diag(c(1^(eta[cc]),0.9^(eta[cc]))))%*orthogonal
      # 0.9 here is for misclassification probability 0.05; 0.99 can be used to represent misclassification probability 0.005. It is due to the decomposition of a matrix. See the main text for details.
      Xchange1pSet<-NULL
      for(s in 1:length(Xchange1[,1])){
        exchange1<-runif(length(X1[,1]),0,1)
        change1=which(exchange1<piepsilon[2])
        Xchange1p<-Xchange1[s,]
        if(length(change1)!=0) for(o in 1:length(change1)){
          if(Xchange1[s,][change1[o]]==(0)) Xchange1p[change1[o]]=1
          else Xchange1p[change1[o]]=(0)
        }
        Xchange1pSet<-rbind(Xchange1pSet,Xchange1p)
      }

      WetaB<-rbind(Xchange1pSet[1,],X1[c(2:4),],Xchange1pSet[c(2:38),])

      X1censor<-WetaB
      TT1censor<-log(LR[,1]);TT1<-TT1censor
      functionx<-function(x){
        b=1/(exp(1)*(exp(1)-1))
        a=(log(x)/x)*(b/((pi)*(((log(x)/x)^2)+b^2)))
        return(a)
      }
      function2<-function(x){
        b=1/(exp(1)*(exp(1)-1))

```

```

a=(b/((pi)*(((log(x)/x)^2)+b^2)))
return(a)
}
Iter=50;time=0
for(j in 1:Iter){
  betazero1=betazero
  Lbeta1<-NULL
  for(i in 1:length(TT1censor)){
    Lbeta<-(LR[,1][i]*exp(-sum(X1censor[,i]*betazero1)))
    Lbeta1<-cbind(Lbeta1,Lbeta)
  }
  Rbeta1<-NULL
  for(i in 1:length(TT1censor)){
    Rbeta<-(LR[,2][i]*exp(-sum(X1censor[,i]*betazero1)))
    Rbeta1<-cbind(Rbeta1,Rbeta)
  }
  dFtgeneration<-NULL
  for(i in 1:length(TT1censor)){
    a<-runif(1000,Lbeta1[i],Rbeta1[i])
    a1<-functionx(a)
    dFt=sum((Rbeta1[i]-Lbeta1[i])*a1)/length(a1)
    b<-runif(1000,Lbeta1[i],Rbeta1[i])
    b1<-function2(b)
    dFt1<-sum((Rbeta1[i]-Lbeta1[i])*b1)/length(b1)
    censorYi<-dFt/dFt1
    dFtgeneration<-rbind(dFtgeneration,censorYi)
  }
  for(i in 1:length(dFtgeneration)){
    if(dFtgeneration[i]=="Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="-Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="NaN") dFtgeneration[i]=0
  }
  ubetastep2<-NULL
  for(i in 1:length(dFtgeneration)){
    a<-X1censor[,i]*dFtgeneration[i]-namita*betazero1
    ubetastep2<-rbind(ubetastep2,a)
  }
  #compute u
  ubetastep<-ubetastep2
  u<-NULL
  for(i in 1:p){
    u<-c(u,sum(ubetastep[,i])/length(TT1))
  }
  for(i in 1:p){
    Delta=u
    #if(abs(Delta[i])>0.6*max(abs(Delta))) betazero[i]=betazero[i]+0.01*sign(Delta[i])#Delta[i]=Delta[i]+0.05*sign(Delta[i])
    if(abs(Delta[i])>0*max(abs(Delta))) betazero[i]=betazero[i]+0.02*sign(Delta[i])
    #if(abs(Delta[i])==max(abs(Delta))) betazero[i]=betazero[i]+0.005*Delta[i]
    #if(time==Iter/2) betazero[which(abs(betazero)<0.05)]=0
    if(time==Iter-1) betazero[which(abs(betazero)<0.02)]=0
  }
}

```

```

        time=time+1
    }
    betaihat<-rbind(betaihat,betazero)
}
betarbind<-rbind(betarbind,betaihat)
}
meanbeta0hat<-NULL;meanbeta0.25hat<-NULL;meanbeta0.5hat<-NULL;meanbeta0.75hat<-NULL;meanbeta1h
at<-NULL
beta0hat<-betarbind[1:B,];beta0.25hat<-betarbind[(B+1):(2*B),];
beta0.5hat<-betarbind[((2*B)+1):(3*B),];beta0.75hat<-betarbind[((3*B)+1):(4*B),];
beta1hat<-betarbind[((4*B)+1):(5*B),]
for (i in 1:p){
    a<-mean(beta0hat[,i])
    b<-mean(beta0.25hat[,i])
    c<-mean(beta0.5hat[,i])
    d<-mean(beta0.75hat[,i])
    e<-mean(beta1hat[,i])
    meanbeta0hat<-cbind(meanbeta0hat,a)
    meanbeta0.25hat<-cbind(meanbeta0.25hat,b)
    meanbeta0.5hat<-cbind(meanbeta0.5hat,c)
    meanbeta0.75hat<-cbind(meanbeta0.75hat,d)
    meanbeta1hat<-cbind(meanbeta1hat,e)
}

gamma1hat<-NULL;gamma0hat<-NULL;
for(i in 1:p){
    x<-c(0,0.25,0.5,0.75,1)
    y<-c(meanbeta0hat[i],meanbeta0.25hat[i],
        meanbeta0.5hat[i],meanbeta0.75hat[i],meanbeta1hat[i])
    c<-lm(y~x)

    gamma0<-as.numeric(c$coefficients[1])
    gamma1<-as.numeric(c$coefficients[2])

    gamma0hat<-c(gamma0hat,gamma0)
    gamma1hat<-c(gamma1hat,gamma1)
}
betacorrect<-gamma0hat-gamma1hat

for(i in 1:p){
    if(abs(betacorrect[i])<0.005) betacorrect[i]=0
}
kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):((kfold*(n/5))))],][3:41]
for(kk in 1:length(kfoldprec[,1])){
    Ttuta<-sum(kfoldprec[kk,]*betacorrect)
    TtutaSet<-c(TtutaSet,Ttuta)
}
}

```

```

TT<-data.frame(exp(TtutaSet))

Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)

pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.5128924
```

```
## [1] 0.9689655
```

## Method 2: Naive

```

#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)

#Data setup

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:32),-(33:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)

```



[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1[-c(1741:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])
data2<-data1[,c(-(18:19))]

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  Iter=100;time=0
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=1.02
  dataF2=dataF1[-c((((kfold-1)*(n/5))+1):(((kfold)*(n/5))))),]
  LR<-cbind(dataF2[,1],dataF2[,2])

  X1<-t(dataF2[,c(3:43)])
  X1censor<-X1
  TT1censor<-log(LR[,1]);TT1<-TT1censor

  functionx<-function(x){
    b=1/(exp(1)*(exp(1)-1))
    a=(log(x)/x)*(b/((pi)*(((log(x)/x)^2)+b^2)))
    return(a)
  }
  function2<-function(x){
    b=1/(exp(1)*(exp(1)-1))
    a=(b/((pi)*(((log(x)/x)^2)+b^2)))
    return(a)
  }
  for(j in 1:Iter){
    betazero1=betazero
    Lbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Lbeta<- (LR[,1][i]*exp(-sum(X1censor[,i]*betazero1)))
      Lbeta1<-cbind(Lbeta1,Lbeta)
    }
    Rbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Rbeta<- (LR[,2][i]*exp(-sum(X1censor[,i]*betazero1)))
      Rbeta1<-cbind(Rbeta1,Rbeta)
    }
    dFtgeneration<-NULL
    for(i in 1:length(TT1censor)){
      a<-runif(2000,Lbeta1[i],Rbeta1[i])
      a1<-functionx(a)
      dFt=sum((Rbeta1[i]-Lbeta1[i])*a1)/length(a1)
      b<-runif(2000,Lbeta1[i],Rbeta1[i])
      b1<-function2(b)
      dFt1<-sum((Rbeta1[i]-Lbeta1[i])*b1)/length(b1)
    }
  }
}

```

```

    censorYi<-dFt/dFt1
    dFtgeneration<-rbind(dFtgeneration,censorYi)
  }
  for(i in 1:length(dFtgeneration)){
    if(dFtgeneration[i]=="Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="-Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="NaN") dFtgeneration[i]=0
  }
  ubetastep2<-NULL
  for(i in 1:length(dFtgeneration)){
    a<-X1censor[,i]*dFtgeneration[i]-namita*betazero1
    ubetastep2<-rbind(ubetastep2,a)
  }
  #compute u
  ubetastep<-ubetastep2
  u<-NULL
  for(i in 1:p){
    u<-c(u,sum(ubetastep[,i])/length(TT1))
  }
  for(i in 1:p){
    Delta=u

    if(abs(Delta[i])>0.9*max(abs(Delta))) betazero[i]=betazero[i]+0.02*sign(Delta[i])
    if(time==Iter-1) betazero[which(abs(betazero)<0.02)]=0
  }
  time=time+1
}
kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):((kfold*(n/5)))),][3:41]
for(kk in 1:length(kfoldprec[,1])){
  Ttuta<-sum(kfoldprec[kk,]*betazero)
  TtutaSet<-c(TtutaSet,Ttuta)
}
}

TT<-data.frame(exp(TtutaSet))

Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)

pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])

```

```
doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out
```

```
## [1] 0.4493023
```

```
## [1] 0.9517241
```

### Method 3: PIC(Gao et al.)

```
#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)

#Data setup

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:32),-(33:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)
```

[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1[-c(1741:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])
data2<-data1[,c(-(18:19))]

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  Iter=100;time=0
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=0.6
  dataF2=dataF1[-c((((kfold-1)*(n/5))+1):(((kfold)*(n/5)))),]
  LR<-cbind(dataF2[,1],dataF2[,2])

  X1<-t(dataF2[,c(3:43)])
  X1censor<-X1
  TT1censor<-log(LR[,1]);TT1<-TT1censor
  functionx<-function(x){
    a=1/x
    return(a)
  }
  function2<-function(x){
    a=1/(x^2)
    return(a)
  }
  for(j in 1:Iter){
    betazero1=betazero
    Lbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Lbeta<-log(LR[,1])[i]-sum(X1censor[,i]*betazero1)
      Lbeta1<-cbind(Lbeta1,Lbeta)
    }
    Rbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Rbeta<-log(LR[,2])[i]-sum(X1censor[,i]*betazero1)
      Rbeta1<-cbind(Rbeta1,Rbeta)
    }
    dFtgeneration<-NULL
    for(i in 1:length(TT1censor)){
      a<-runif(300,Lbeta1[i],Rbeta1[i])
      a1<-functionx(a)
      dFt=sum((Rbeta1[i]-Lbeta1[i])*a1)/length(a1)
      b<-runif(300,Lbeta1[i],Rbeta1[i])
      b1<-function2(b)
      dFt1<-sum((Rbeta1[i]-Lbeta1[i])*b1)/length(b1)
      censorYi<-dFt/dFt1
      dFtgeneration<-rbind(dFtgeneration,censorYi)
    }
    for(i in 1:length(dFtgeneration)){

```



```

    if(dFtgeneration[i]=="Inf") dFtgeneration[i]=0
    if(dFtgeneration[i=="-Inf") dFtgeneration[i]=0
    if(dFtgeneration[i=="NaN") dFtgeneration[i]=0
  }
  ubetastep2<-NULL
  for(i in 1:length(dFtgeneration)){
    a<-X1censor[,i]*dFtgeneration[i]-namita*betazero1
    ubetastep2<-rbind(ubetastep2,a)
  }
  #compute u
  ubetastep<-ubetastep2
  u<-NULL
  for(i in 1:p){
    u<-c(u,sum(ubetastep[,i])/length(TT1))
  }
  for(i in 1:p){
    Delta=u
    if(abs(Delta[i])>0*max(abs(Delta))) betazero[i]=betazero[i]+0.05*sign(Delta[i])
    if(time==Iter-1) betazero[which(abs(betazero)<0.05)]=0
  }
  time=time+1
}
kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):((kfold*(n/5)))),][3:41]
for(kk in 1:length(kfoldprec[,1])){
  Ttuta<-sum(kfoldprec[kk,]*betazero)
  TtutaSet<-c(TtutaSet,Ttuta)
}
}

TT<-data.frame(exp(TtutaSet))
Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)
pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.4659965
```

```
## [1] 0.9373563
```

#### Method 4: LBPIC

```
#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)

#Data setup

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:32),-(33:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)
```

[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1[-c(1741:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])
data2<-data1[,c(-(18:19))]

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  Iter=100;time=0
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=1.02
  dataF2=dataF1[-c(((kfold-1)*(n/5))+1):((kfold)*(n/5))),]
  LR<-cbind(dataF2[,1],dataF2[,2])
  X1<-t(dataF2[,c(3:43)])
  X1censor<-X1
  TT1censor<-log(LR[,1]);TT1<-TT1censor

  functionx<-function(x){
    b=1/(exp(1)*(exp(1)-1))
    a=(log(x)/x)*(b/((pi)*(((log(x)/x)^2)+b^2)))
    return(a)
  }
  function2<-function(x){
    b=1/(exp(1)*(exp(1)-1))
    a=(b/((pi)*(((log(x)/x)^2)+b^2)))
    return(a)
  }
  for(j in 1:Iter){
    betazero1=betazero
    Lbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Lbeta<-(LR[,1][i]*exp(-sum(X1censor[,i]*betazero1)))
      Lbeta1<-cbind(Lbeta1,Lbeta)
    }
    Rbeta1<-NULL
    for(i in 1:length(TT1censor)){
      Rbeta<-(LR[,2][i]*exp(-sum(X1censor[,i]*betazero1)))
      Rbeta1<-cbind(Rbeta1,Rbeta)
    }
    dFtgeneration<-NULL
    for(i in 1:length(TT1censor)){
      a<-runif(2000,Lbeta1[i],Rbeta1[i])
      a1<-functionx(a)
      dFt=sum((Rbeta1[i]-Lbeta1[i])*a1)/length(a1)
      b<-runif(2000,Lbeta1[i],Rbeta1[i])
      b1<-function2(b)
      dFt1<-sum((Rbeta1[i]-Lbeta1[i])*b1)/length(b1)
      censorYi<-dFt/dFt1
    }
  }
}

```

```

    dFtgeneration<-rbind(dFtgeneration,censorYi)
  }
  for(i in 1:length(dFtgeneration)){
    if(dFtgeneration[i]=="Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="-Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="NaN") dFtgeneration[i]=0
  }
  ubetastep2<-NULL
  for(i in 1:length(dFtgeneration)){
    a<-X1censor[,i]*dFtgeneration[i]-namita*betazero1
    ubetastep2<-rbind(ubetastep2,a)
  }
  #compute u
  ubetastep<-ubetastep2
  u<-NULL
  for(i in 1:p){
    u<-c(u,sum(ubetastep[,i])/length(TT1))
  }
  for(i in 1:p){
    Delta=u
    if(abs(Delta[i])>0*max(abs(Delta))) betazero[i]=betazero[i]+0.02*sign(Delta[i])
    if(time==Iter-1) betazero[which(abs(betazero)<0.02)]=0
  }
  time=time+1
}
kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):((kfold*(n/5)))),][3:41]
for(kk in 1:length(kfoldprec[,1])){
  Ttuta<-sum(kfoldprec[kk,]*betazero)
  TtutaSet<-c(TtutaSet,Ttuta)
}
}

TT<-data.frame(exp(TtutaSet))
Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)
pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.5348843
```

```
## [1] 0.9810345
```

#### Method 5: SIMEX(0.05)

```
#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)

#Data setup

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:32),-(33:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)
```

[illegible]



[illegible]

```

data1<-complete(C)
data1<-data1[-c(1741:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])
data2<-data1[,c(-(18:19))]

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=0.03
  dataF2=dataF1[-c(((kfold-1)*(n/5))+1):(((kfold)*(n/5)))),]
  LR<-cbind(dataF2[,1],dataF2[,2])
  betarbind<-NULL;eta<-c(0,0.25,0.5,0.75,1);B=5
  for(cc in 1:length(eta)){
    betaihat<-NULL
    for(i in 1:B){
      X1<-t(dataF2[,c(3:43)])
      Xchange1<-X1[c(1,(5:41)),]
      orthogonal<-matrix(c(sqrt(0.5),sqrt(0.5),sqrt(0.5),-sqrt(0.5)),nrow=2,ncol=2,byrow=TRUE)
      orthogonalT<-t(orthogonal)
      piepsilon<-orthogonalT%*(diag(c(1^(eta[cc]),0.9^(eta[cc]))))%*orthogonal
      # 0.9 here is for misclassification probability 0.05; 0.99 can be used to represent
t misclassification probability 0.005. It is due to the decomposition of a matrix. See the main
text for details.
      Xchange1pSet<-NULL
      for(s in 1:length(Xchange1[,1])){
        exchange1<-runif(length(X1[1,]),0,1)
        change1<-which(exchange1<piepsilon[2])
        Xchange1p<-Xchange1[s,]
        if(length(change1)!=0) for(o in 1:length(change1)){
          if(Xchange1[s,][change1[o]]==(0)) Xchange1p[change1[o]]=1
          else Xchange1p[change1[o]]=(0)
        }
        Xchange1pSet<-rbind(Xchange1pSet,Xchange1p)
      }
    }

    WetaB<-rbind(Xchange1pSet[1,],X1[c(2:4),],Xchange1pSet[c(2:38),])
    X1censor<-WetaB
    TT1censor<-log(LR[,1]);TT1<-TT1censor
    functionx<-function(x){
      b=1/(exp(1)*(exp(1)-1))
      a=(log(x)/x)*(b/((pi)*(((log(x)/x)^2)+b^2)))
      return(a)
    }
    function2<-function(x){
      b=1/(exp(1)*(exp(1)-1))
      a=(b/((pi)*(((log(x)/x)^2)+b^2)))
      return(a)
    }
  }
}

```

```

}
Iter=50;time=0
for(j in 1:Iter){
  betazero1=betazero
  Lbeta1<-NULL
  for(i in 1:length(TT1censor)){
    Lbeta<-(LR[,1][i]*exp(-sum(X1censor[,i]*betazero1)))
    Lbeta1<-cbind(Lbeta1,Lbeta)
  }
  Rbeta1<-NULL
  for(i in 1:length(TT1censor)){
    Rbeta<-(LR[,2][i]*exp(-sum(X1censor[,i]*betazero1)))
    Rbeta1<-cbind(Rbeta1,Rbeta)
  }
  dFtgeneration<-NULL
  for(i in 1:length(TT1censor)){
    a<-runif(1000,Lbeta1[i],Rbeta1[i])
    a1<-functionx(a)
    dFt=sum((Rbeta1[i]-Lbeta1[i])*a1)/length(a1)
    b<-runif(1000,Lbeta1[i],Rbeta1[i])
    b1<-function2(b)
    dFt1<-sum((Rbeta1[i]-Lbeta1[i])*b1)/length(b1)
    censorYi<-dFt/dFt1
    dFtgeneration<-rbind(dFtgeneration,censorYi)
  }
  for(i in 1:length(dFtgeneration)){
    if(dFtgeneration[i]=="Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="-Inf") dFtgeneration[i]=0
    if(dFtgeneration[i]=="NaN") dFtgeneration[i]=0
  }
  ubetastep2<-NULL
  for(i in 1:length(dFtgeneration)){
    a<-X1censor[,i]*dFtgeneration[i]-namita*betazero1
    ubetastep2<-rbind(ubetastep2,a)
  }
  #compute u
  ubetastep<-ubetastep2
  u<-NULL
  for(i in 1:p){
    u<-c(u,sum(ubetastep[,i])/length(TT1))
  }
  for(i in 1:p){
    Delta=u
    if(abs(Delta[i])>0*max(abs(Delta))) betazero[i]=betazero[i]+0.02*sign(Delta[i])
    if(time==Iter-1) betazero[which(abs(betazero)<0.02)]=0
  }
  time=time+1
}
betaihat<-rbind(betaihat,betazero)
}
betarbind<-rbind(betarbind,betaihat)
}

```

```

meanbeta0hat<-NULL;meanbeta0.25hat<-NULL;meanbeta0.5hat<-NULL;meanbeta0.75hat<-NULL;meanbeta1h
at<-NULL
beta0hat<-betarbind[1:B,];beta0.25hat<-betarbind[(B+1):(2*B),];
beta0.5hat<-betarbind[((2*B)+1):(3*B),];beta0.75hat<-betarbind[((3*B)+1):(4*B),];
beta1hat<-betarbind[((4*B)+1):(5*B),]
for (i in 1:p){
  a<-mean(beta0hat[,i])
  b<-mean(beta0.25hat[,i])
  c<-mean(beta0.5hat[,i])
  d<-mean(beta0.75hat[,i])
  e<-mean(beta1hat[,i])
  meanbeta0hat<-cbind(meanbeta0hat,a)
  meanbeta0.25hat<-cbind(meanbeta0.25hat,b)
  meanbeta0.5hat<-cbind(meanbeta0.5hat,c)
  meanbeta0.75hat<-cbind(meanbeta0.75hat,d)
  meanbeta1hat<-cbind(meanbeta1hat,e)
}

gamma1hat<-NULL;gamma0hat<-NULL;
for(i in 1:p){
  x<-c(0,0.25,0.5,0.75,1)
  y<-c(meanbeta0hat[i],meanbeta0.25hat[i],
      meanbeta0.5hat[i],meanbeta0.75hat[i],meanbeta1hat[i])
  c<-lm(y~x)
  #c<-lm(y~x+I(x*x))
  gamma0<-as.numeric(c$coefficients[1])
  gamma1<-as.numeric(c$coefficients[2])

  gamma0hat<-c(gamma0hat,gamma0)
  gamma1hat<-c(gamma1hat,gamma1)
}
betacorrect<-gamma0hat-gamma1hat

for(i in 1:p){
  if(abs(betacorrect[i])<0.005) betacorrect[i]=0
}
kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):((kfold*(n/5))))],[3:41]
for(kk in 1:length(kfoldprec[,1])){
  Ttuta<-sum(kfoldprec[kk,]*betacorrect)
  TtutaSet<-c(TtutaSet,Ttuta)
}

}
TT<-data.frame(exp(TtutaSet))

Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)

```

```

pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.5128924
```

```
## [1] 0.9689655
```

## Mehod 6: IC\_Par

```

#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)
library(icenReg)

```

```
## 载入需要的程辑包：Rcpp
```

```
## 载入需要的程辑包：coda
```

### *#Data setup*

```

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

```

```

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

```

```

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:76),-(77:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)#delete NA

```

[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1#[-c(1501:(Length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])#response
data2<-data1[,c(-(18:19))]#covariates
dataF<-centralImputation(cbind(Y,data2))
dataF1=dataF;TtutaSet<-NULL

dataF<-centralImputation(cbind(Y,data2))

TtutaSet<-NULL;dataF1=dataF;

## 5-fold cross validation
for(kfold in 1:5){
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=0
  dataF2=dataF1[-c((((kfold-1)*(n/5))+1):(((kfold)*(n/5))))),]
  LR<-cbind(dataF2[,1],dataF2[,2])
  X1<-as.data.frame(dataF2[,c(3:43)])
  IC_par = ic_par(cbind(LR[,1],LR[,2]) ~ .,data=X1, model="ph",dist="lnorm")
  betazero<-IC_par$coef[-c(1,2)]
  kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):(((kfold)*(n/5))))),][3:41]
  for(kk in 1:length(kfoldprec[,1])){
    Ttuta<-sum(kfoldprec[kk,]*betazero)
    TtutaSet<-c(TtutaSet,Ttuta)
  }
}

TT<-data.frame(exp(TtutaSet))
Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)
pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])
TtutaSet1<-TtutaSet#[-interval]
Y1<-Y#[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin

```

```
## [1] 0.8066281
```

```
## [1] 0.9982808
```



```

TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.8079655
```

```
## [1] 0.9982808
```

### Method 7: IC\_Bayesian

```

#upload a dataset from .RData
load("C:\\Users\\user\\Downloads\\tandmob2.RData") #tandmob2 is a dataset

#required packages
library(mice)
library(DMwR2)
library(openxlsx)
library(survival)
library(survminer)
library(icenReg)

#Data setup

## Note: From the datasets, responses for tooth label 11 are columns 21 and 22. Following an order in the main text, labels are 21, 31, 41, and so on, and the corresponding columns are (20,21),(22,23),(24,25), and so on.

## keep tooth label 11 from columns 21 and 22 and remove NA to get interval-censored responses
AA<-tandmob2[-c(which(is.na(tandmob2[,21]))),]
A<-AA[-c(which(is.na(AA[,22]))),]

## remove the first three non-informative columns and other teeth labels
data<-A[,c(-1,-2,-4,-(23:76),-(77:132))]
C<-mice(data,m=3,method = "pmm",maxit = 10,seed=163)#delete NA

```

[illegible]

[illegible]

```

data1<-complete(C)
data1<-data1#[-c(1501:(length(A[,1]))),]
Y<-cbind(data1[,18],data1[,19])#response
data2<-data1[,c(-(18:19))]#covariates
dataF<-centralImputation(cbind(Y,data2))
dataF1=dataF;TtutaSet<-NULL

## 5-fold cross validation
for(kfold in 1:5){
  n=length(dataF1[,1]);p=41;betazero=t(rep(0,p));namita=0
  dataF2=dataF1[-c((((kfold-1)*(n/5))+1):(((kfold)*(n/5))))),]
  LR<-cbind(dataF2[,1],dataF2[,2])
  X1<-as.data.frame(dataF2[,c(3:43)])
  IC_Bayes = ic_bayes(cbind(LR[,1],LR[,2]) ~ .,data=X1
                    ,model="ph",dist = "lnorm")
  betazero<-IC_Bayes$coef[-c(1,2)]
  kfoldprec<-dataF1[c((((kfold-1)*(n/5))+1):(((kfold)*(n/5))))),][3:41]
  for(kk in 1:length(kfoldprec[,1])){
    Ttuta<-sum(kfoldprec[kk,]*betazero)
    TtutaSet<-c(TtutaSet,Ttuta)
  }
}

TT<-data.frame(exp(TtutaSet))
Set<-NULL
for(i in 1:length(TT[,1])){
  if(Y[i,1]<TT[i,]&Y[i,2]>TT[i,]) Set=c(Set,TT[i,])
}
interval<-which(Y[,1]<TT&Y[,2]>TT)

pin<-length(which(Y[,1]<TT&Y[,2]>TT))/length(Y[,1])

TtutaSet1<-TtutaSet[-interval]
Y1<-Y[-interval,]
doutSet<-NULL
for(i in 1:length(TtutaSet1)){
  dout<-min(abs(Y1[i,1]-exp(TtutaSet1[i]))/(Y1[i,1]),abs(Y1[i,2]-exp(TtutaSet1[i]))/Y1[i,2])
  doutSet<-c(doutSet,dout)
}
mean(doutSet);1-pin #d_out & p_out

```

```
## [1] 0.8065933
```

```
## [1] 0.9988539
```

We can repeat the above code to obtain numerical results for all teeth. In our early analysis, we have obtained predicted failure times for all teeth under all estimation methods, which have also been summarized in .csv files in the corresponding author's GitHub. We now use those .csv files to reproduce survivor curves as displayed in Figure 1 in the main text.

```

noSIMEXnoBoost<-read.xlsx("C:\\noSIMEXnoBoost.xlsx")
Naive<-read.xlsx("C:\\Naive.xlsx")
SIMEXBoost0.005<-read.xlsx("C:\\SIMEXBoost0.005.xlsx")
SIMEXBoost0.05<-read.xlsx("C:\\SIMEXBoost0.05.xlsx")
SIMEX0.005noBoost<-read.xlsx("C:\\SIMEX0.005noBoost.xlsx")
SIMEX0.05noBoost<-read.xlsx("C:\\SIMEX0.05noBoost.xlsx")
PICGao<-read.xlsx("C:\\PIC(Gao et al.).xlsx")
ICbayes<-read.xlsx("C:\\IC_Bayes.xlsx")
ICpar<-read.xlsx("C:\\IC_par.xlsx")
#View(Naive[,5])
data<-NULL
for(i in 1:28){
  aa<-cbind(noSIMEXnoBoost[,i],Naive[,i],SIMEXBoost0.005[,i]
            ,SIMEXBoost0.05[,i],SIMEX0.005noBoost[,i],SIMEX0.05noBoost[,i],PICGao[,i],ICbayes[,
i],ICpar[,i])
  data<-cbind(data,aa)
}
# tooth5<-data[,c(1:6)]
#View(data[,c(1:9)])

#max(as.numeric(aa[,1]));max(as.numeric(bb[,1]));max(as.numeric(cc[,1]));max(as.numeric(dd[,
1]));max(as.numeric(ff[,1]));max(as.numeric(jj[,1]))

#min(as.numeric(aa[,1]));min(as.numeric(bb[,1]));min(as.numeric(cc[,1]));min(as.numeric(dd[,
1]));min(as.numeric(ff[,1]));min(as.numeric(jj[,1]))
# tooth7<-da
aa<-na.omit(data[,244]);aa<-cbind(aa,rep("LBPIC",length(aa)))
bb<-na.omit(data[,245]);bb<-cbind(bb,rep("Naive",length(bb)))
cc<-na.omit(data[,246]);cc<-cbind(cc,rep("SIMEXBoost(0.005)",length(cc)))
dd<-na.omit(data[,247]);dd<-cbind(dd,rep("SIMEXBoost(0.05)",length(dd)))
ee<-na.omit(data[,248]);ee<-cbind(ee,rep("SIMEX(0.005)",length(ee)))
ff<-na.omit(data[,249]);ff<-cbind(ff,rep("SIMEX(0.05)",length(ff)))
jj<-na.omit(data[,250]);jj<-cbind(jj,rep("PIC(Gao et al.)",length(jj)))
kk<-na.omit(data[,251]);kk<-cbind(kk,rep("ICBayes",length(kk)))
ll<-na.omit(data[,252]);ll<-cbind(ll,rep("ICpar",length(ll)))

gg<-as.data.frame(rbind(aa,bb,cc,dd,ee,ff,jj,kk,ll))
colnames(gg)<-c("T","method")
hh<-Surv(as.numeric(gg[,1]))
fit <- survfit(hh~ method, data=gg)

```

Based on the programming code above, we can draw the following survivor curves for the tooth label 11 under several estimation methods:

IC\_Par    LBPIC    PIC(Gao et al.)    SIMEX(0.005)  
IC\_Bayesian    SIMEXBoost(0.005)    SIMEX(0.05)    SIMEXBoost(0.05)

