# 1 Generating replicates of survival data

#######################################  
## sURVIVAL DATA SIMULATION FUNCTION ##  
## FOR MC MODEL 2 WITH 5 COVARIATES ##  
#######################################  
  
  
  
Surv.gen.cure.subtype.5var <-function(n, alfa, beta, shape, scale, simu, rate, hypo){  
 require(psych)  
 require(plyr)  
 require(MASS)  
 require(survival)  
   
 setwd('YOUR/FILE/LOCATION')  
   
 survtime<-matrix(rep(0,simu\*n),ncol=simu)  
 fstatus<-matrix(rep(0,simu\*n),ncol=simu)  
 fHer2<-matrix(rep(0,simu\*n),ncol=simu)  
 ftn<-matrix(rep(0,simu\*n),ncol=simu)  
 fluma<-matrix(rep(0,simu\*n),ncol=simu)  
 ftumct<-matrix(rep(0,simu\*n),ncol=simu)  
 fmens0<-matrix(rep(0,simu\*n),ncol=simu)  
  
 set.seed(2882488)  
   
  
 for (i in 1:simu) {   
   
 ## Generate covariates ##  
 ## and set coefficients ##  
 ##########################  
   
 set.seed(1000+i)  
  
pi<-runif(n,0,1)  
  
# Generate the binary covariates as categories of a multinomial master variable  
# based on their distributions in ANN complete case data  
  
Her2<-ifelse((pi<=0.034542314|(pi<=0.324697755 & pi>0.307426598)|(pi<=0.523316062 & pi>0.492227979)|(pi<=0.801381693 & pi>0.766839378)),1,0)   
LumA<-ifelse(((pi<=0.188255613 & pi>0.034542314)|(pi<=0.388601036 & pi>0.324697755)|(pi<=0.592400691 & pi>0.523316062)|(pi<=0.846286701 & pi>0.801381693)),1,0)  
TN<-ifelse(((pi<=0.227979275 & pi>0.188255613)|(pi<=0.419689119 & pi>0.388601036)|(pi<=0.649395509 & pi>0.592400691)|(pi<=0.924006908 & pi>0.846286701)),1,0)  
LumB<-ifelse(((pi<=0.307426598 & pi>0.227979275)|(pi<=0.492227979 & pi>0.419689119)|(pi<=0.766839378 & pi>0.649395509)|pi>0.924006908),1,0)  
mens0<-ifelse(((pi<=0.324697755 & pi>0.307426598)|(pi<=0.388601036 & pi>0.324697755)|(pi<=0.419689119 & pi>0.388601036)|(pi<=0.492227979 & pi>0.419689119)|(pi<=0.801381693 & pi>0.766839378)|(pi<=0.846286701 & pi>0.801381693)|(pi<=0.924006908 & pi>0.846286701)|pi>0.924006908),1,0)  
tumct<-ifelse(((pi<=0.523316062 & pi>0.492227979)|(pi<=0.801381693 & pi>0.766839378)|(pi<=0.592400691 & pi>0.523316062)|(pi<=0.846286701 & pi>0.801381693)|(pi<=0.649395509 & pi>0.592400691)|(pi<=0.924006908 & pi>0.846286701)|(pi<=0.766839378 & pi>0.649395509)|pi>0.924006908),1,0)  
  
  
x<-cbind(1,Her2,LumA,TN,mens0,tumct)  
  
 ## Cure/susceptible indicator ##  
 linpred <-x %\*% alfa  
 pb <-exp(linpred)/(1+exp(linpred))  
 ypb <-rbinom(n=n, size=1, prob=pb)  
 n1 <-sum(ypb)  
 n2 <-n-n1  
  
 covgroup <-cbind(ypb,Her2,LumA,TN,mens0,tumct)  
 covgroup <-as.data.frame(covgroup)  
 x.new<-covgroup[!(covgroup$ypb==1),]  
 x.new<-as.matrix(x.new)  
 x.new[,1]<-1   
   
 x.new1<-covgroup[!(covgroup$ypb==0),]  
 x.new1<-as.matrix(x.new1)  
   
 ## Time to event/censoring for the susceptible ##  
   
 ## Length of time to event under weibull ##  
 u<- runif(n2,0,1)  
 time1 <- ((-log(u))/(scale\*exp(x.new %\*% beta)))^(1/shape)  
   
 ## Recruitment time point during the accrual 5 yrs and follow-up period of 10 yrs ##  
 time.entry <- runif(n2,0,60)  
   
 ## Divide censoring to event by fixed follow-up of 10yrs ##  
 time1x <- time1+time.entry  
 delta2 <- ifelse(time1x>=180,0,1)  
 time2 <- (1-delta2)\*(60-time.entry)+delta2\*time1  
   
   
 ## Time to censoring for the cured ##  
 time.cen.cure <- c(rep(180,n1))-runif(n1,0,60)  
   
 ## Combine two groups ##  
 status <-c(delta2,rep(0,n1))  
 time <-c(time2, time.cen.cure)  
   
 dataset <- cbind.data.frame(status, time, rbind(x.new,x.new1))  
  
 survtime[,i]<-c(dataset$time)  
 fstatus[,i]<-c(dataset$status)  
 fHer2[,i]<-c(dataset$Her2)  
 ftumct[,i]<-c(dataset$tumct)  
 ftn[,i]<-c(dataset$TN)  
 fluma[,i]<-c(dataset$LumA)  
 fmens0[,i]<-c(dataset$mens0)  
  
 }  
   
 # Write/store output datasets by replicates of individual variables  
   
 if (rate == "high" & hypo == "null") {  
 write.csv(survtime, "timevarsubt.null5v\_hr.csv")  
 write.csv(fstatus, "statusvarsubt.null5v\_hr.csv")  
 write.csv(ftn, "TNvarsubt.null5v\_hr.csv")  
 write.csv(fluma, "LumAvarsubt.null5v\_hr.csv")  
 write.csv(fHer2, "Her2varsubt.null5v\_hr.csv")  
 write.csv(fmens0, "mens0varsubt.null5v\_hr.csv")  
 write.csv(ftumct, "tumctvarsubt.null5v\_hr.csv")  
 } else if (rate == "low" & hypo =="null") {  
 write.csv(survtime, "timevarsubt.null5v\_lr.csv")  
 write.csv(fstatus, "statusvarsubt.null5v\_lr.csv")  
 write.csv(ftn, "TNvarsubt.null5v\_lr.csv")  
 write.csv(fluma, "LumAvarsubt.null5v\_lr.csv")  
 write.csv(fHer2, "Her2varsubt.null5v\_lr.csv")  
 write.csv(fmens0, "mens0varsubt.null5v\_lr.csv")  
 write.csv(ftumct, "tumctvarsubt.null5v\_lr.csv")  
 } else if (rate == "medlow" & hypo =="null") {  
 write.csv(survtime, "timevarsubt.null5v\_medlr.csv")  
 write.csv(fstatus, "statusvarsubt.null5v\_medlr.csv")  
 write.csv(ftn, "TNvarsubt.null5v\_medlr.csv")  
 write.csv(fluma, "LumAvarsubt.null5v\_medlr.csv")  
 write.csv(fHer2, "Her2varsubt.null5v\_medlr.csv")  
 write.csv(fmens0, "mens0varsubt.null5v\_medlr.csv")  
 write.csv(ftumct, "tumctvarsubt.null5v\_medlr.csv")  
 } else if (rate == "exlow" & hypo =="null") {  
 write.csv(survtime, "timevarsubt.null5v\_exlr.csv")  
 write.csv(fstatus, "statusvarsubt.null5v\_exlr.csv")  
 write.csv(ftn, "TNvarsubt.null5v\_exlr.csv")  
 write.csv(fluma, "LumAvarsubt.null5v\_exlr.csv")  
 write.csv(fHer2, "Her2varsubt.null5v\_exlr.csv")  
 write.csv(fmens0, "mens0varsubt.null5v\_exlr.csv")  
 write.csv(ftumct, "tumctvarsubt.null5v\_exlr.csv")  
 } else if (rate == "medlow" & hypo =="al") {  
 write.csv(survtime, "timevarsubt.5v\_medlr.csv")  
 write.csv(fstatus, "statusvarsubt.5v\_medlr.csv")  
 write.csv(ftn, "TNvarsubt.5v\_medlr.csv")  
 write.csv(fluma, "LumAvarsubt.5v\_medlr.csv")  
 write.csv(fHer2, "Her2varsubt.5v\_medlr.csv")  
 write.csv(fmens0, "mens0varsubt.5v\_medlr.csv")  
 write.csv(ftumct, "tumctvarsubt.5v\_medlr.csv")   
   
 } else if (rate == "exlow" & hypo =="al") {  
 write.csv(survtime, "timevarsubt.5v\_exlr.csv")  
 write.csv(fstatus, "statusvarsubt.5v\_exlr.csv")  
 write.csv(ftn, "TNvarsubt.5v\_exlr.csv")  
 write.csv(fluma, "LumAvarsubt.5v\_exlr.csv")  
 write.csv(fHer2, "Her2varsubt.5v\_exlr.csv")  
 write.csv(fmens0, "mens0varsubt.5v\_exlr.csv")  
 write.csv(ftumct, "tumctvarsubt.5v\_exlr.csv")   
   
 }else if (rate == "low" & hypo == "al"){  
 write.csv(survtime, "timevarsubt.5v\_lr.csv")  
 write.csv(fstatus, "statusvarsubt.5v\_lr.csv")  
 write.csv(ftn, "TNvarsubt.5v\_lr.csv")  
 write.csv(fluma, "LumAvarsubt.5v\_lr.csv")  
 write.csv(fHer2, "Her2varsubt.5v\_lr.csv")  
 write.csv(fmens0, "mens0varsubt.5v\_lr.csv")  
 write.csv(ftumct, "tumctvarsubt.5v\_lr.csv")   
 } else if (rate == "high" & hypo =="al"){  
 write.csv(survtime, "timevarsubt.5v\_hr.csv")  
 write.csv(fstatus, "statusvarsubt.5v\_hr.csv")  
 write.csv(ftn, "TNvarsubt.5v\_hr.csv")  
 write.csv(fluma, "LumAvarsubt.5v\_hr.csv")  
 write.csv(fHer2, "Her2varsubt.5v\_hr.csv")  
 write.csv(fmens0, "mens0varsubt.5v\_hr.csv")  
 write.csv(ftumct, "tumctvarsubt.5v\_hr.csv")   
 }  
}  
  
#### DESCRIPTION ##################################################################  
# INPUT  
# n : sample size in each simulation  
# alpha : a vector of coefficients of linear predictor for logistic part  
# beta : a vector of coefficients of linear predictor for survival part  
# shape : value of the shape parameter for survival time in weibull distribution  
# scale : value of the scale parameter for survival time in weibull distribution  
# simu : number of simulations  
# mixval : a vector of initial values for nlm on coefficients in mixture model  
# coefn : total number of coefficients for the mixture model  
# OUTPUT  
# NULL.

#########################################  
## Example to generate 5000 replicates ##  
## of data sample size n=1000 with ##  
## event rate 25% under the null ##  
#########################################  
  
# By altering the intercept values of alfa and beta vectors, change to intended event rate   
Surv.gen.cure.subtype.5var(n=1000,   
 alfa=c(1.28, 0,0,0, 0.6,-0.732),   
 beta=c(-7.13,0,0,0,0.85,0.137),   
 rate="high", hypo="null",  
 shape=1.77, scale=1, simu=5000) ##  
  
# Check for final distribution of event rate  
simu.status<-read.csv("statusvarsubt.null5v\_hr.csv", header=TRUE)  
summary(apply(simu.status[,-1], 2, mean))

# 2 Mixture Cure model fitting for 5-covariate model

## 2.1 A single data example of high event rate and parameter values under the null

# Import 5000 replicates by individual variable files  
simu.Her2<-read.csv("Her2varsubt.null5v\_hr.csv", header=TRUE)   
simu.LumA<-read.csv("LumAvarsubt.null5v\_hr.csv", header=TRUE)  
simu.TN<-read.csv("TNvarsubt.null5v\_hr.csv", header=TRUE)  
simu.TUMCT<-read.csv("tumctvarsubt.null5v\_hr.csv", header=TRUE)  
simu.MENS0<-read.csv("mens0varsubt.null5v\_hr.csv", header=TRUE)  
simu.status<-read.csv("statusvarsubt.null5v\_hr.csv", header=TRUE)  
simu.time<-read.csv("timevarsubt.null5v\_hr.csv", header=TRUE)  
  
# install online package mixcuref  
  
devtools::install\_github("ChangchangXu-LTRI/mixcuref", build = TRUE, build\_opts = c())  
  
library(mixcuref)  
  
# Example single simulated dataset   
  
i=1  
 Her2 <- as.factor(simu.Her2[,i+1])  
 LumA <- as.factor(simu.LumA[,i+1])  
 TN <- as.factor(simu.TN[,i+1])  
 TUMCT <- as.factor(simu.TUMCT[,i+1])  
 MENS0 <- as.factor(simu.MENS0[,i+1])  
 primdata<-cbind.data.frame(Her2, TN, LumA, MENS0, TUMCT, simu.status[,i+1],simu.time[,i+1])  
 colnames(primdata)<-c("Her2", "TN", "LuminalA", "MENS0", "TUMCT", "CENS","Time")  
  
## Parameter estimation under Firth-type penalization via mixcuref::mixcure.penal.est  
 mix.pl.est <- mixcuref::mixcure.penal.est(Surv(Time, CENS == 1) ~ Her2 + LuminalA + TN + MENS0 + TUMCT,data=primdata,init=c(5,-0.1,-0.1,0.1,0.1,0.1,-10,-0.1,0.1,-0.1,-0.1,0.1,2.5), pl=T)   
 ## 2d LRT under FT-PL via mixcuref::mixcure.penal.2d.nested.lrt  
 mix.pl.2dlrt <- mixcuref::mixcure.penal.2d.nested.lrt(Surv(Time, CENS == 1) ~ Her2 + LuminalA + TN + MENS0 + TUMCT,data=primdata, loglik=mix.pl.est$coefficients$alpha[,7], init=c(2,-0.1,-0.1,-0.1,-0.1,-0.1,-10,-0.1,-0.1,-0.1,-0.1,0.1,1), pl=T)   
  
## Parameter estimation under normal likelihood  
 mix.est <- mixcuref::mixcure.penal.est(Surv(Time, CENS == 1) ~ Her2 + LuminalA + TN + MENS0 + TUMCT,data=primdata,init=c(5,-0.1,-0.1,0.1,0.1,0.1,-10,-0.1,0.1,-0.1,-0.1,0.1,2.5), pl=F)   
 ## 2d LRT under ML via mixcuref::mixcure.penal.2d.nested.lrt  
 mix.2dlrt <- mixcuref::mixcure.penal.2d.nested.lrt(Surv(Time, CENS == 1) ~ Her2 + LuminalA + TN + MENS0 + TUMCT,data=primdata, loglik=mix.est$coefficients$alpha[,7], init=c(2,-0.1,-0.1,-0.1,-0.1,-0.1,-10,-0.1,-0.1,-0.1,-0.1,0.1,1), pl=F)