

Case-study Right heart catheterization

Hanchao Zhang, M.S.

Contents

Motivation	2
Data preparation	2
Analysis	6

Motivation

Right heart catheterization (RHC), also known as pulmonary artery catheterization, is an invasive test that mainly checks the working state of the heart by guiding a pulmonary artery (PA) catheter (a small and hollow tube) through the pulmonary artery and into the right chambers of the heart. By this way, the catheter can measure the functions of the heart such as blood pressure, cardiac output, etc., including other measures in connection to heart problems such as diurnal respiratory instability according to one study by Kumagai, et al. (2018). In turn, these measurements are used to treat and manage heart conditions such as heart failure, congenital heart disease, cardiomyopathy, pulmonary hypertension, etc. Thus, many cardiologists and critical care physicians believe that the direct measurements of the cardiac functions by RHC is necessary to the management of treating critically ill patients and that such management will theoretically lead to better health outcomes according to Connors, et al. (1996). However, due to the severity of the cardiac conditions of the patients as well as the invasive procedure of RHC with no guarantee of any beneficial outcomes, data measuring the benefits of the RHC procedure cannot easily be collected in a randomized control trial (RCT). To make up for this, observational studies were used to evaluate the effectiveness of the RHC procedure. Although, these studies are

susceptible to treatment selection bias as physicians can make the decision to use or withhold RHC in the treatment of their patients.

The purpose of this study is to reproduce the statistical analysis of the objective of the research, The Effectiveness of Right Heart Catheterization in the Initial Care, using the data it collected from a large group of critically ill patients and in predefined patient subgroups in order to determine how valid is the association between the use of the right heart catheterization during the first 24 hours of an ICU stay with subsequent survival, length of stay, intensity of care, and cost of care.

You can find the description of the data from <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html>

Data preparation

By reading the description of the data, we need to exclude several variables in the dataset. The variables including all the enrollment time, discharge time and so on, and the variables with percentage of missing value larger than 20% are also excluded.

We choose 20% as an arbitrary number since we do not want to lose too much information from the whole data set.

```
rhc <- read.csv("http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.csv")
rhc <- rhc[,-1]
var.vector <- c(
  "cat2",
  "sadmte",
  "dschdte",
  "dthdte",
  "lstctdte",
  "death",
  "surv2md1",
  "ptid"
)
var.which <- NULL
for (i in var.vector) {
  var.which[i] <- which(names(rhc) == i)
}
```

```
var.which2 <- NULL
for (i in names(rhc)) {
  if( sum(is.na(rhc[,i]))/nrow(rhc) > 0.2 ){
    var.which2[i] <- which(names(rhc) == i)
  }
}
```

```
dat.rhc <- rhc[,-c(var.which, var.which2)]
```

```
summary(dat.rhc)
```

```
##                cat1                ca                cardiohx
## ARF                :2490  Metastatic: 384  Min.      :0.0000
## MOSF w/Sepsis      :1227  No           :4379  1st Qu.:0.0000
## COPD               : 457  Yes          : 972  Median :0.0000
## CHF                : 456                      Mean  :0.1766
## Coma               : 436                      3rd Qu.:0.0000
## MOSF w/Malignancy: 399                      Max.    :1.0000
## (Other)            : 270
##      chfhx      dementhx      psychhx      chrpulhx
## Min.      :0.000  Min.      :0.00000  Min.      :0.00000  Min.      :0.0000
## 1st Qu.:0.000  1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.0000
## Median :0.000  Median :0.00000  Median :0.00000  Median :0.0000
## Mean    :0.178  Mean    :0.09834  Mean    :0.06731  Mean    :0.1899
## 3rd Qu.:0.000  3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:0.0000
## Max.    :1.000  Max.    :1.00000  Max.    :1.00000  Max.    :1.0000
##
##      renalhx      liverhx      gibledhx      malighx
## Min.      :0.00000  Min.      :0.00000  Min.      :0.00000  Min.      :0.0000
## 1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.00000  1st Qu.:0.0000
## Median :0.00000  Median :0.00000  Median :0.00000  Median :0.0000
## Mean    :0.04446  Mean    :0.06992  Mean    :0.03226  Mean    :0.2295
## 3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:0.00000  3rd Qu.:0.0000
## Max.    :1.00000  Max.    :1.00000  Max.    :1.00000  Max.    :1.0000
##
##      immunhx      transhx      amihx      age
## Min.      :0.000  Min.      :0.0000  Min.      :0.00000  Min.      : 18.04
## 1st Qu.:0.000  1st Qu.:0.0000  1st Qu.:0.00000  1st Qu.: 50.15
## Median :0.000  Median :0.0000  Median :0.00000  Median : 64.05
## Mean    :0.269  Mean    :0.1154  Mean    :0.03487  Mean    : 61.38
## 3rd Qu.:1.000  3rd Qu.:0.0000  3rd Qu.:0.00000  3rd Qu.: 73.93
## Max.    :1.000  Max.    :1.0000  Max.    :1.00000  Max.    :101.85
##
##      sex      edu      das2d3pc      t3d30      dth30
## Female:2543  Min.    : 0.00  Min.    :11.00  Min.    : 2.00  No :3817
## Male :3192   1st Qu.:10.00  1st Qu.:16.06  1st Qu.:16.00  Yes:1918
##              Median :12.00  Median :19.75  Median :30.00
##              Mean   :11.68  Mean   :20.50  Mean   :23.61
##              3rd Qu.:13.00  3rd Qu.:23.43  3rd Qu.:30.00
##              Max.   :30.00  Max.   :33.00  Max.   :30.00
##
##      aps1      scoma1      meanbp1      wblc1
## Min.    : 3.00  Min.    : 0  Min.    : 0.00  Min.    : 0.000
## 1st Qu.: 41.00  1st Qu.: 0  1st Qu.: 50.00  1st Qu.: 8.398
```

```

## Median : 54.00   Median : 0   Median : 63.00   Median : 14.100
## Mean   : 54.67   Mean   : 21   Mean   : 78.52   Mean   : 15.645
## 3rd Qu.: 67.00   3rd Qu.: 41   3rd Qu.:115.00   3rd Qu.: 20.049
## Max.   :147.00   Max.   :100   Max.   :259.00   Max.   :192.000
##
##      hrt1      resp1      temp1      pafi1
## Min.   : 0.0   Min.   : 0.00   Min.   :27.00   Min.   : 11.6
## 1st Qu.: 97.0   1st Qu.: 14.00   1st Qu.:36.09   1st Qu.:133.3
## Median :124.0   Median : 30.00   Median :38.09   Median :202.5
## Mean   :115.2   Mean   : 28.09   Mean   :37.62   Mean   :222.3
## 3rd Qu.:141.0   3rd Qu.: 38.00   3rd Qu.:39.00   3rd Qu.:316.6
## Max.   :250.0   Max.   :100.00   Max.   :43.00   Max.   :937.5
##
##      alb1      hema1      bili1      crea1
## Min.   : 0.300   Min.   : 2.00   Min.   : 0.09999   Min.   : 0.09999
## 1st Qu.: 2.600   1st Qu.:26.10   1st Qu.: 0.79993   1st Qu.: 1.00000
## Median : 3.500   Median :30.00   Median : 1.00977   Median : 1.50000
## Mean   : 3.093   Mean   :31.87   Mean   : 2.26707   Mean   : 2.13302
## 3rd Qu.: 3.500   3rd Qu.:36.30   3rd Qu.: 1.39990   3rd Qu.: 2.39990
## Max.   :29.000   Max.   :66.19   Max.   :58.19531   Max.   :25.09766
##
##      sod1      pot1      paco21      ph1
## Min.   :101.0   Min.   : 1.100   Min.   : 1.00   Min.   :6.579
## 1st Qu.:132.0   1st Qu.: 3.400   1st Qu.: 31.00   1st Qu.:7.340
## Median :136.0   Median : 3.800   Median : 37.00   Median :7.400
## Mean   :136.8   Mean   : 4.067   Mean   : 38.75   Mean   :7.388
## 3rd Qu.:142.0   3rd Qu.: 4.600   3rd Qu.: 42.00   3rd Qu.:7.460
## Max.   :178.0   Max.   :11.898   Max.   :156.00   Max.   :7.770
##
##      swang1      wtkilo1      dnr1      ninsclas
## No RHC:3551   Min.   : 0.00   No :5081   Medicaid      : 647
## RHC   :2184   1st Qu.: 56.30   Yes: 654   Medicare      :1458
##              Median : 70.00              Medicare & Medicaid: 374
##              Mean   : 67.83              No insurance      : 322
##              3rd Qu.: 83.70              Private           :1698
##              Max.   :244.00              Private & Medicare :1236
##
##      resp      card      neuro      gastr      renal      meta
## No :3622   No :3804   No :5042   No :4793   No :5440   No :5470
## Yes:2113   Yes:1931   Yes: 693   Yes: 942   Yes: 295   Yes: 265
##
##
##
##
##
##      hema      seps      trauma      ortho      race
## No :5381   No :4704   No :5683   No :5728   black: 920
## Yes: 354   Yes:1031   Yes: 52   Yes: 7   other: 355
##              white:4460
##
##
##
##
##      income

```

```
## > $50k      : 451
## $11-$25k    :1165
## $25-$50k    : 893
## Under $11k:3226
##
##
##
```

```
str(dat.rhc)
```

```
## 'data.frame':   5735 obs. of  52 variables:
## $ cat1      : Factor w/ 9 levels "ARF","CHF","Cirrhosis",...: 6 9 8 1 9 6 8 1 8 1 ...
## $ ca        : Factor w/ 3 levels "Metastatic","No",...: 3 2 3 2 2 2 1 2 3 3 ...
## $ cardiohx  : int  0 1 0 0 0 0 0 0 0 0 ...
## $ chfhx     : int  0 1 0 0 0 1 0 0 0 0 ...
## $ dementhx  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ psychhx   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ chrpulhx  : int  1 0 0 0 0 1 0 0 0 0 ...
## $ renalhx   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ liverhx   : int  0 0 0 0 0 0 0 0 0 0 ...
## $ gibledhx  : int  0 0 0 0 0 0 0 0 0 0 ...
## $ malighx   : int  1 0 1 0 0 0 1 0 0 1 ...
## $ immunhx   : int  0 1 1 1 0 0 0 0 0 0 ...
## $ transhx   : int  0 1 0 0 0 0 0 1 0 0 ...
## $ amihx     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ age       : num  70.3 78.2 46.1 75.3 67.9 ...
## $ sex       : Factor w/ 2 levels "Female","Male": 2 1 1 1 2 1 2 2 1 1 ...
## $ edu       : num  12 12 14.07 9 9.95 ...
## $ das2d3pc  : num  23.5 14.8 18.1 22.9 21.1 ...
## $ t3d30     : int  30 30 30 30 2 30 30 30 30 30 ...
## $ dth30     : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 1 1 1 1 ...
## $ aps1      : int  46 50 82 48 72 38 29 25 47 48 ...
## $ scoma1    : int  0 0 0 0 41 0 26 100 0 0 ...
## $ meanbp1   : num  41 63 57 55 65 115 67 128 53 73 ...
## $ wblc1     : num  22.1 28.9 0.05 23.3 29.7 ...
## $ hrt1      : int  124 137 130 58 125 134 135 102 118 141 ...
## $ resp1     : num  10 38 40 26 27 36 10 34 30 40 ...
## $ temp1     : num  38.7 38.9 36.4 35.8 34.8 ...
## $ paf1      : num  68 218 276 157 478 ...
## $ alb1      : num  3.5 2.6 3.5 3.5 3.5 ...
## $ hema1     : num  58 32.5 21.1 26.3 24 ...
## $ bili1     : num  1.01 0.7 1.01 0.4 1.01 ...
## $ crea1     : num  1.2 0.6 2.6 1.7 3.6 ...
## $ sod1      : int  145 137 146 117 126 138 136 136 136 146 ...
## $ pot1      : num  4 3.3 2.9 5.8 5.8 ...
## $ paco21    : num  40 34 16 30 17 68 45 26 40 30 ...
## $ ph1       : num  7.36 7.33 7.36 7.46 7.23 ...
## $ swang1    : Factor w/ 2 levels "No RHC","RHC": 1 2 2 1 2 1 1 1 1 2 ...
## $ wtkilo1   : num  64.7 45.7 0 54.6 78.4 ...
## $ dnr1      : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 1 1 1 1 ...
## $ ninsclas  : Factor w/ 6 levels "Medicaid","Medicare",...: 2 6 5 6 2 2 5 5 5 1 ...
## $ resp      : Factor w/ 2 levels "No","Yes": 2 1 1 2 1 2 1 2 1 1 ...
## $ card      : Factor w/ 2 levels "No","Yes": 2 1 2 1 2 1 1 1 1 1 ...
## $ neuro     : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 2 1 2 ...
## $ gastr     : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 2 ...
```

```
## $ renal      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 1 ...
## $ meta       : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ hema       : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...
## $ seps       : Factor w/ 2 levels "No","Yes": 1 2 1 1 1 1 1 2 1 1 ...
## $ trauma     : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ ortho      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ race       : Factor w/ 3 levels "black","other",...: 3 3 3 3 3 3 3 3 3 ...
## $ income     : Factor w/ 4 levels "> $50k","$11-$25k",...: 4 4 3 2 4 4 3 3 4 4 ...
```

Changing the type of the data is also very important for survival analysis, and by checking the data, we also sometimes need to deal with variables with characters. In our data set, the status of having RHC or not is denoted with character. A way to adjust it is to directly force it into factor. However, this kind of factor format sometimes does not work in some of the packages in R. Therefore, we forced the character into a numeric number and deducted it by 1.

```
for (i in names(dat.rhc)) {
  if(is.numeric(dat.rhc[,i])){
    if(mean(dat.rhc[,i], na.rm = T) < 5){
      dat.rhc[,i] <- as.factor((dat.rhc[,i]))
    }else{
      dat.rhc[,i] <- ((dat.rhc[,i]))
    }
  }
}
dat.rhc$dth30 <- as.numeric(dat.rhc$dth30) - 1
dat.rhc$swang1 <- as.numeric(dat.rhc$swang1) - 1
table(dat.rhc$swang1)

##
##      0      1
## 3551 2184
```

Analysis

Cox model and Survival Random Forest

After cleaning the data, we can fit the cox model and survival random forest with our data

ADD SOME MORE INFORMATION HERE ABOUT COX MODEL AND SURVIVAL ANALYSIS

```
fit1 <- coxph(Surv(t3d30,dth30) ~ . ,data = dat.rhc)

## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :
## Ran out of iterations and did not converge

fit2 <- rfsrc(Surv(t3d30,dth30) ~ ., data = dat.rhc, nsplit = 10, importance = TRUE, ntree = 100)
```

Cross validation

ADD SOME MORE INFORMATION HERE ABOUT CROSS VALIDATION

We can definitely compute cross-validation test error by our own code. However, for the survival analysis, the ‘pec’ package save us lots of effort to do the cross-validation and also the function is faster.

```
#fitpec1 <- pec(list("Cox" = fit1), formula = Surv(t3d30,dth30) ~ 1, data = dat.rhc, cens.model = 'cox')
```

Restricted Mean Survival time (Causal Inference)

ADD SOME MORE INFORMATION HERE ABOUT RMST

Causal inference is the process of drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect. The main difference between causal inference and inference of association is that the former analyzes the response of the effect variable when the cause is changed.[1][2] The science of why things occur is called etiology. Causal inference is an example of causal reasoning (https://en.wikipedia.org/wiki/Causal_inference)

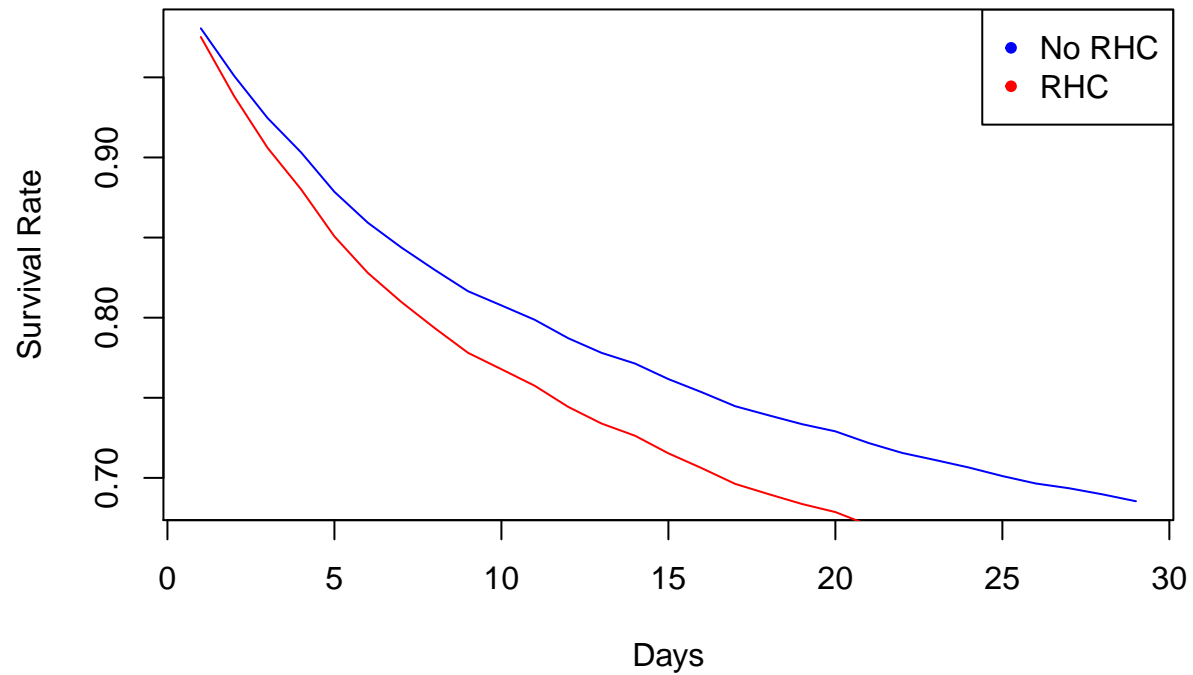
We will use the idea of it to compute the difference of survival time between having RHC and not having RHC by both model.

```
rhc0 <- rhc1 <- dat.rhc
rhc0$swang1 <- 0
rhc1$swang1 <- 1

time1 <- survfit(fit1, newdata = rhc0)$surv
time2 <- survfit(fit1, newdata = rhc1)$surv

rowMeans(time1) %>%
  plot(type = "l", col = "blue", main = 'RMST-Cox model', xlab = 'Days', ylab = 'Survival Rate')
rowMeans(time2) %>%
  lines(col = "red")
legend('topright', c('No RHC', 'RHC'), col = c('blue', 'red'), pch = 20)
```

RMST-Cox model



```
time3 <- predict(fit2, newdata = rhc0)$survival
time4 <- predict(fit2, newdata = rhc1)$survival
```

```
colMeans(time3) %>%
```

```
  plot(type = "l", col = "blue", main = 'RMST-Survival random forest', xlab = 'Days', ylab = 'Survival Rate')
```

```
colMeans(time4) %>%
```

```
  lines(col = "red")
```

```
legend('topright', c('No RHC', 'RHC'), col = c('blue', 'red'), pch = 20)
```


RMST–Survival random forest

