Is Right Heart Catheterization Really Beneficial?

Bin Fang 05/10/2006

Executive Summary

The effectiveness of Right Heart Catheterization (RHC) in the initial care of critically ill patients was believed by many physicians and cardiologists. In this project, we uses stratified cox model to study the relative risk of critically ill patients receiving and not receiving RHC. The data comes from a prospective study involving 5735 adult patients receiving care in the intensive care unit (ICU) in five US teaching hospitals between 1989 and 1994. The time of entry to the hospitals is regarded as start time of events and the censoring time is the last time of contract for those patients. The outcome variable is time until death (or the last time of contract for censored data) and our primary interest of covariate is receiving RHC within 24 hours after study entry or not. Other covariates includes demographic information, admission diagnosis and categories of comorbidities illness. By model model building, model diagnostics and model validation, we reach similar conclusion with previous research: when controlling other covariates, the risk of death for patients receiving RHC is 1.35 times the risk of those not receiving RHC. This finding implies necessity of randomized clinical trial of RHC for confirmation and may stimulate improvement of corresponding technology.

Introduction

In RHC, a special catheter is inserted from the neck or chest to the right side of the heart and then enters the pulmonary artery. In order to monitor the blood pressures inside the heart and lungs, the doctor may perform RHC in the diagnosis and/or management of Heart failure, Shock, Congenital heart disease, Valvular heart disease, Cardiomyopathy, Pulmonary hypertension. Well-known possible risks associated with a RHC includes bruising of the skin at the site where the catheter is inserted, excessive bleeding because of puncture of the vein during catheter insertion and Pneumothorax (partial collapse of the lung). These side-effects seem trivial and probably does not affect the survival. Thus, it is widespread accepted that the direct measurement of cardiac function provided by RHC is necessary to guide therapy for certain critically ill patients and that such management leads to better patient outcomes.

However, Connors *et al.* (1996) questioned this common belief by analyzing the data of a prospective study involving 5735 adult patients receiving care in the intensive care unit (ICU) for in five US teaching hospitals between 1989 and 1994. The fitted logistical regression model by Connors *et al.* indicated that RHC was associated with increased mortality and increased utilization of resources. In this project, we uses cox regression model to answer two questions: (1) Is use of RHC associated with increasing risk of death for critically ill patients? (2) Is there any effect modifier for the use of RHC?

Methods

The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) was a 5-center study of decision making and outcomes of seriously ill, hospitalized adult patients. Five medical centers participated in data collection and the statistical center at Duke University participated in data analysis. 5735 SUPPORT patients admitted or transferred to an ICU within the first 24 hours after entering the study were included in the data analysis conducted by Connors *et al.*, and we use the same data set as their research.

The outcome for this study is patient survival time. The primary covariate is RHC, which is coded present of performed within the first 24 hours after entering the study. The other covariates can be summarized as demographic information, admission diagnosis and categories of comorbidities illness. Demographic information (8 covariates such as age, gender, race) comes from interviews; admission diagnosis (33 covariates such as blood pressure, heart rate) and categories of comorbidities illness (12 covariates) come from chart abstraction.

There are missing values for two covariates: urine output (3028 missing values) and activities of daily living (4296 missing values). We impute missing values with column means before survival analysis. Approximately two thirds of total data (3823 observations) are randomly selected for model building and model diagnostics, and the remaining data (1912 observations) are used for model validation.

The statistical model that are used is cox regression model and the software is R studio (R 3.2.5).

Results

1. Summary Table

The means and standard deviations of continuous variables and numbers and percentages of categorical variables are summarized in Table 1a and Table 1b, respectively.

Since SUPPORT is a prospective study rather than a randomized clinical trial, the decision to use or withhold RHC depends on the physicians. Therefore, treatment selection (RHC or not) is confounded with other covariates. As it shows in table 1a, patients in No RHC Group has mean blood pressure of 84.87, whereas patients in RHC Group has mean blood pressure of 68.2. That is to say, mean blood pressure might affect physicians' decision on treatment selection. Other examples similar to this can be found in Table 1a and Table 1b.

Table 1a: Descriptive Statistics for RHC Group and No RHC Group (Continuous Variables)

Covariates	No RHC (3551)		RHC (2184)		Covariates	No RHC (3551)		RHC (2184)	
	Mean	SD	Mean	SD		Mean	SD	Mean	SD
Age	61.8	15.6	60.8	17.3	PaCo2	40.0	14.2	36.8	11.0
Education	11.6	3.1	11.9	3.2	PH	7.4	0.1	7.4	0.1
Weight	15.3	11.4	16.3	12.6	Weight	65.0	29.5	72.4	27.7
Heart rate	112.9	40.9	118.9	41.4	Activities of Daily Living	1.2	1.9	1.0	1.7
Temperature	37.6	1.7	37.6	1.8	Duke Activity Status Index	20.4	5.5	20.7	5.0
Albumin	3.2	0.7	3.0	0.9	APACHE score	50.9	18.8	60.7	20.3
Hematocrit	32.7	8.8	30.5	7.4	Glasgow Coma Score	22.3	31.4	19.0	28.3
Bilirubin	2.0	4.4	2.7	5.3	Mean blood pressure	84.9	38.9	68.2	34.2
Creatinine	1.9	2.0	2.5	2.1	PaO2/FIO2 ratio	240.6	116.7	192.4	105.5
Sodium	137.0	7.7	136.3	7.6	Respiratory rate	29.0	14.0	26.7	14.2
Potassium	4.1	1.0	4.1	1.0	Urine output	2199.2	1445.2	2182.6	1635.0

Table 1b: Descriptive Statistics for RHC Group and No RHC Group (Categorical Variables)

Covarites	No RHC (3551)		RHC (2184)		Covarites	No RHC (3551)		RHC (2184)	
	No.	Pct.	No.	Pct.		No.	Pct.	No.	Pct.
ARF	1581	45%	909	41%	Medicaid	454	13%	193	9%
CHF	247	7%	209	9%	Medicare	947	27%	511	24%
Cirrhosis	175	5%	49	2%	Medicare & Medicaid	251	7%	123	6%
Colon Cancer	6	1%	1	1%	No insurance	186	5%	135	6%
Coma	341	10%	95	4%	Private	967	27%	731	33%
COPD	399	11%	58	3%	Private $ Medicare	746	21%	490	22%
Lung Cancer	34	1%	5	1%					
MOSF w/Malignancy	241	7%	158	7%					
MOSF w/Sepsis	527	15%	700	32%					
Cancer (Metastatic)	261	7%	123	6%	Black	585	16%	335	15%
No Cancer	2652	75%	1727	79%	Other	213	6%	142	7%
Cancer (No Metastatic)	638	18%	334	15%	white	2753	78%	1707	78%
No Acute MI	2984	84%	1738	80%	No DNR	3052	86%	2029	93%
Acute MI	567	16%	446	20%	DNR	499	14%	155	7%
No Heart Failure	2955	83%	1759	81%	No Respiratory Diagnosis	2070	58%	1552	71%
Heart Failure	596	17%	425	19%	Respiratory Diagnosis	1481	42%	632	29%
No Dementia	3138	88%	2033	93%	No Cardiovascular Diagnosis	2544	72%	1260	58%
Dementia	413	12%	151	7%	Cardiovascular Diagnosis	1007	28%	924	42%
No Psychiatry	3265	92%	2084	95%	No Neurological Diagnosis	2976	84%	2066	95%
Psychiatry	286	8%	100	5%	Neurological Diagnosis	575	16%	118	5%
No Pulmonary Disease	2777	78%	1869	86%	No Gastrointestinal Diagnosis	3029	85%	1764	81%
Pulmonary Disease	774	22%	315	14%	Gastrointestinal Diagnosis	522	15%	420	19%
No Renal Disease	3402	96%	2078	95%	No Renal Diagnosis	3404	96%	2036	93%
Renal Disease	149	4%	106	5%	Renal Diagnosis	147	4%	148	7%

No Cirrhosis	3286	93%	2048	94%		No Metabolic Diagnosis	3379	95%	2091	96%
Cirrhosis	265	7%	136	6%		Metabolic Diagnosis	172	5%	93	4%
No Upper GI Bleeding	3420	96%	2130	98%		No Hematologic Diagnosis	3312	93%	2069	95%
Upper GI Bleeding	131	4%	54	2%		Hematologic Diagnosis	239	7%	115	5%
No Metastatic Disease	2679	75%	1740	80%		No Sepsis Diagnosis	3036	86%	1668	76%
Metastatic Disease	872	25%	444	20%		Sepsis Diagnosis	515	14%	516	24%
No Immunosupperssion	2644	74%	1548	71%		No Trauma Diagnosis	3533	99%	2150	98%
Immunosupperssion	907	26%	636	29%		Trauma Diagnosis	18	1%	34	2%
No Transfer	3216	91%	1857	85%		No Orthopedic Diagnosis	3548	99%	2180	99%
Transfer	335	9%	327	15%		Orthopedic Diagnosis	3	1%	4	1%
No Myocardial Infarction	3446	97%	2089	96%		$11-$25k	713	20%	452	21%
Myocardial Infarction	105	3%	95	4%		$25-$50k	500	14%	393	18%
						>$50k	257	7%	194	9%
						Under$11k	2081	59%	1145	52%

2. Univariate Analysis

Summary of cox model with RHC as the only covariate is shown in Table 2. It indicates that without adjusting for other covariates, the risk of death for patients receiving RHC is 1.16 times the risk of those not receiving RHC. Unadjusted Kaplan-Meier survival curves for treatment groups are shown in Figure 1. It seems that without adjusting for other covariates, RHC might lead to higher mortality between start of the study and 800 days, but slightly lower mortality 800 days after start of the study.

Table 2: summary of cox model with RHC as the only covariate

	coefficient	**exp(coefficient)**	**se(coefficient)**	**p-value**
RHC	0.15	1.16	0.033	6.89e-06

Figure 1: Unadjusted Kaplan-Meier Survival Curves

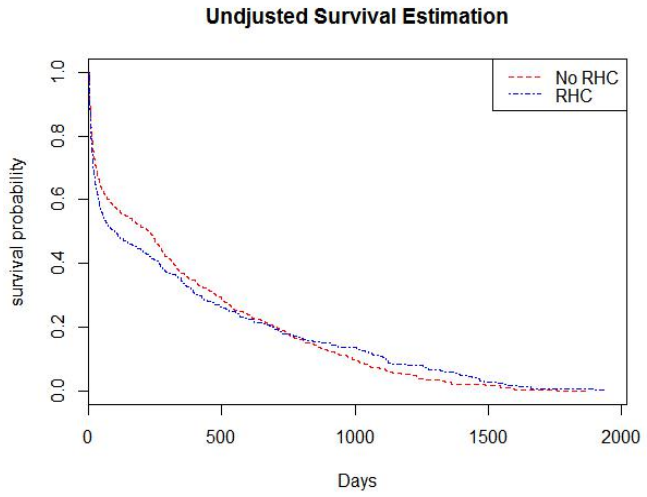
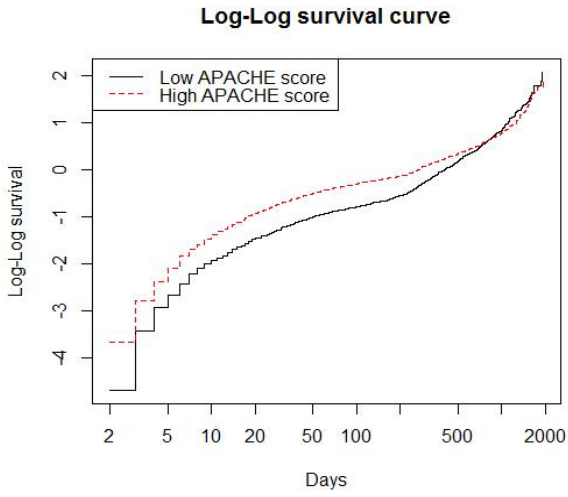


Figure 2: Log-Log Survival Curve for Two Strata



3. Model Building

We use stepwise strategy to select the model with lowest BIC value, which is the model fit1a (see Appendix for details), which includes 13 covariates. We use martingale residuals, Wald test and Likelihood Ratio Test to compare models which have linear forms of covariates with those which have square forms of covariates, and find that linear forms of covariates are enough to explain patient survival time. Interaction between RHC and other covariates are condidered afterwards, and according to Wald test and Likelihood Ratio Test, there are no significant interaction between RHC and other covariates. In other words, RHC affects patient survival time only through itself.

4. Model Diagnosis

4.1 Proportional Hazard Assumption

Schoenfeld residuals and log-log survival curves are used to check proportional hazard assumption. The effect of APACHE score seems time-varying (shown in Figure 2 last page), thus we fit a stratified cox model (APACHE score as strata) with 12 other covariates. We repeat model building process again and confirm this is the best model. The summary of the final model is listed in Table 3. We find that RHC increases the risk of death even more than it shows in the unadjusted model.

Table 3: Summary of the Final Model

Variables	Coefficient	Exp(Coefficient)	Standard Error	P-value
RHC	0.22	1.25	0.043	3.5e-07 ***
Coma	0.85	2.34	0.086	< 2e-16 **
Lung Cancer	0.57	1.77	0.200	4.2e-3 **
MOSF w/Malignancy	0.40	1.49	1.423	1.64e-06 ***
No Cancer	-0.28	0.76	0.053	1.7e-07 ***
Cirrhosis	0.32	1.38	0.089	2.9e-04 ***
Age	0.01	1.01	0.001	2.0e-11 ***
Duke Activity Status Index	-0.03	0.97	0.004	1.5e-11 ***
Glasgow Coma Score	0.00	1.00	0.001	2.4e-09 ***
Bilirubin	0.03	1.03	0.004	4.5e-10 ***
DNR status	0.71	2.04	0.059	< 2e-16 ***
Hematologic Diagnosis	0.29	1.33	0.080	3.6e-04 ***

4.2 Outliers and Influential Points with respect to RHC

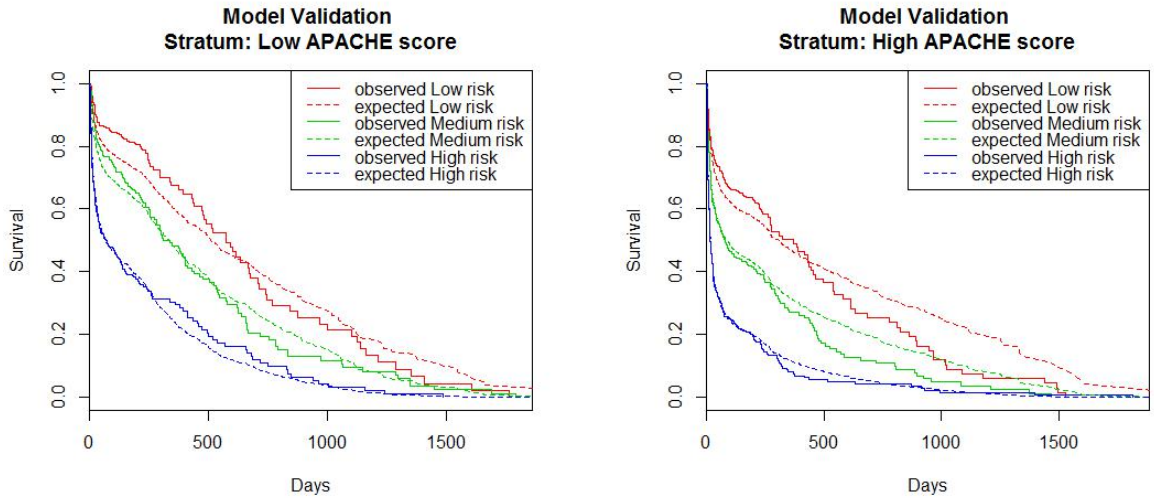
Table 4 lists two outliers and the most influential observation with respect to RHC. Patient 5653 receives RHC within 24 hours after entry to the study, and his other covariates imply early death, whereas he lives much longer than expected. Patient 4624 and 1317 do not receive RHC within 24 hours after entry to the study, and their other covariates imply later death, whereas they die much earlier than expected. Patient 1317 is the most influential observation with respect to RHC in the sense that if he was removed from the data set, the coefficient of RHC would dramatically change.

Table 4: Some outliers and Influential Points with respect to RHC

Observation	RHC	MOSF w/Malignancy	Cancer	DNR status
5653 (outlying)	RHC	Yes	Yes	DNR
4624 (outlying)	No RHC	No	No	No DNR
1317 (influential)	No RHC	No	No	No DNR

5. Model validation

Figure 3: Observed and Expected Survival Curves for Validation Data



As it shows in Figure 3, the observed survival curves (Kaplan-Meier) and expected survival curves (adjusted survival curves) for both Low APACHE score group and high APACHE score group are in similar shapes. Also, the estimated coefficients with use of validation data are similar to those with use of training data. Thus, the stratified cox model is a good predictive model for survival time.

Conclusion

Based on the model, we can conclude when controlling other covariates, the risk of death for patients receiving RHC is 1.35 times the risk of those not receiving RHC.

The major limitation of the study is it is an observational study rather than a clinical trial. As discussed before, physicians will choose different treatments according to patients' physical status. It is likely that patients with worse status received RHC and without RHC they might die even earlier. Thus, to completely rule out this selection bias, a randomized clinical trial is necessary. Another limitation lies in the assumption of the model. We assumes that the effect of RHC is time-independent. However, as it shows in Figure 1 and Schoenfeld residuals (data not shown), the effect of RHC are probably time-dependent. It seems possible that RHC gives rise to higher mortality in the short term but are beneficial in the long run. A cox model with time-varying effects might help answer this question. Thirdly, the data set does not include the information which center the data are collected, and multivariate model might be a better choice if there are some correlation within the patients in a specific medical center.

Anyway, this project is helpful for the following research and discussion about the benefits of RHC for critically ill patients.

References

1. http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html
2. http://www.hopkinsmedicine.org/healthlibrary
3. Klein and Moeschberger (2003): Survival Analysis-Techniques for Censored and Truncated Data (2nd Ed).
4. Connors *et al.* (1996): The effectiveness of RHC in the initial care of critically ill patients. *J American Medical Association* 276:889-897.

Appendix

Data set can be downloaded from the following link:
http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.csv

R codes are as follows:

```
##################################
rhc <- read.csv("rhc.csv")
summary(rhc)
attach(rhc)
library(survival)

##################################
# data cleaning
##################################

colSums( is.na(rhc) )
#In dthdte, replace NA with value from lstctdte to find days

rhc$dthdte1<-ifelse(is.na(rhc$dthdte), rhc$lstctdte, rhc$dthdte)
rhc$time<-rhc$dthdte1-rhc$sadmdte

rhc$cat12=ifelse(cat1=="CHF",1,0)
rhc$cat13=ifelse(cat1=="Cirrhosis",1,0)
rhc$cat14=ifelse(cat1=="Colon Cancer",1,0)
rhc$cat15=ifelse(cat1=="Coma",1,0)
rhc$cat16=ifelse(cat1=="COPD",1,0)
rhc$cat17=ifelse(cat1=="Lung Cancer",1,0)
rhc$cat18=ifelse(cat1=="MOSF w/Malignancy",1,0)
rhc$cat19=ifelse(cat1=="MOSF w/Sepsis",1,0)

rhc$caN=ifelse(ca=="No",1,0)
rhc$caY=ifelse(ca=="Yes",1,0)


rhc$ninsclas2=ifelse(ninsclas=="Medicare",1,0)
rhc$ninsclas3=ifelse(ninsclas=="Medicare & Medicaid",1,0)
rhc$ninsclas4=ifelse(ninsclas=="No insurance",1,0)
rhc$ninsclas5=ifelse(ninsclas=="Private",1,0)
rhc$ninsclas6=ifelse(ninsclas=="Private & Medicare",1,0)

rhc$black=ifelse(race=="black",1,0)
rhc$white=ifelse(race=="white",1,0)

rhc$income2=ifelse(income=="$25-$50k",1,0)
rhc$income3=ifelse(income=="> $50k",1,0)
rhc$income4=ifelse(income=="Under $11k",1,0)

rhc$death1<-ifelse(death=="Yes", 1, 0)

#mean(adld3p,na.rm=TRUE)
#rhc$adld3p1=ifelse(is.na(adld3p),1.182071,adld3p)
#mean(urin1,na.rm=TRUE)
```

```
#rhc$urin11=ifelse(is.na(urin1),2192.454,urin1)

quantile(rhc$aps1, 0.5)
rhc$aps1_grp <-ifelse(rhc$aps1<54, 1, 2)
table(rhc$aps1_grp)


# Randomly sample 3823 data as training set,
# the remaining 1912 data are used as validation set.

set.seed(1)
train <- sample(1:5735,3823)
rhc.training <- rhc[train,]
dim(rhc.training)
validation <- setdiff(1:5735,train)
rhc.validation <- rhc[validation,]
dim(rhc.validation)


############################
# Unajusted survival curve
############################

fitKM <- survfit( Surv(time,death1)~swang1, type='kaplan',
                          conf.type='log-log', data=rhc.training)
summary(fitKM)
plot(fitKM, xlab='Days', ylab='survival probability', col=c(2,4),lty=c(2,4),
        main="Undjusted Survival Estimation")
legend("topright",c("No RHC","RHC"), col=c(2,4),lty=c(2,4))

fit.rhc <- coxph(Surv(time,death1)~swang1,
                       method='breslow', data=rhc)
summary(fit.rhc)


######################
# Model Selection
######################


fit1 <- coxph(Surv(time,death1)~cat12+cat13+cat14+cat15+cat16+cat17+cat18+cat19
                       +caN+caY+cardiohx+chfhx+dementhx
                       +psychhx+chrpulhx+renalhx+liverhx+gibledhx+malighx
                       +immunhx+transhx+amihx+age+sex+edu
                       +das2d3pc+aps1+scoma1
                       +meanbp1+wblc1+hrt1+resp1+temp1+pafi1
                       +alb1+hema1+bili1+crea1+sod1+pot1
                       +paco21+ph1+swang1+wtkilo1+dnr1
                       +resp+card+neuro+gastr+renal+meta+hema
                       +seps+trauma+ortho+ninsclas2+ninsclas3
                       +ninsclas4+ninsclas5+ninsclas6+black+white
                       +income2+income3+income4,
                       method='breslow', data=rhc.training)
```

```
summary(fit1)


step(fit1)    # Stepwise AIC
step(fit1, k = log(3823))    # Stepwise BIC


# backward BIC
fit1a <- coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                        caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                        swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a)


###################################
# Transformation
###################################

#age
plot(rhc.training$age, resid(fit1a), xlab='age', ylab='Residual')
lines(lowess(rhc.training$age, resid(fit1a), f=.5), col="red")

fit1a.age <- coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                             caN + liverhx + age +I(age^2) + das2d3pc + aps1 + scoma1 + bili1 +
                             swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a.age)
# Reduced Model is preferred.

#das2d3pc
plot(rhc.training$das2d3pc, resid(fit1a), xlab='das2d3pc', ylab='Residual')
lines(lowess(rhc.training$das2d3pc, resid(fit1a), f=.5), col="red")

fit1a.das2d3pc <- coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                   caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                                   swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a.das2d3pc)
anova(fit1a.das2d3pc, fit1a)
AIC(fit1a.das2d3pc, fit1a)
BIC(fit1a.das2d3pc, fit1a)
# Reduced Model is preferred.

#aps1
plot(rhc.training$aps1, resid(fit1a), xlab='aps1', ylab='Residual')
lines(lowess(rhc.training$aps1, resid(fit1a), f=.5), col="red")

fit1a.aps1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                             caN + liverhx + age + das2d3pc + aps1 +I(aps1^2) + scoma1 + bili1 +
                             swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a.aps1)
# Reduced Model is preferred.

#scoma1
plot(rhc.training$scoma1, resid(fit1a), xlab='scoma1', ylab='Residual')
```

```
lines(lowess(rhc.training$scoma1, resid(fit1a), f=.5), col="red")

fit1a.scoma1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + aps1 + scoma1 +I(scoma1^2) + bili1 +
                                swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a.scoma1)
anova(fit1a.scoma1, fit1a)
AIC(fit1a.scoma1, fit1a)
BIC(fit1a.scoma1, fit1a)
# Reduced Model is preferred.

#bili1
plot(rhc.training$bili1, resid(fit1a), xlab='bili1', ylab='Residual')
lines(lowess(rhc.training$bili1, resid(fit1a), f=.5), col="red")

fit1a.bili1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                               caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +I(bili1^2)+
                               swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit1a.bili1)
anova(fit1a.bili1, fit1a)
AIC(fit1a.bili1, fit1a)
BIC(fit1a.bili1, fit1a)
# Reduced Model is preferred.

# Conclusion:    Model with linear forms is preferred.


##############################
# Interaction with RHC
##############################

# Interaction between cat1 and RHC
fit.with.cat1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                  caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                                  swang1 + dnr1 + hema + cat15*swang1 + cat17*swang1 + cat18*swang1,
                               data = rhc.training, method = "breslow")
summary(fit.with.cat1)
#No interaction

# Interaction between caN and RHC
fit.with.caN<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                                swang1 + dnr1 + hema + caN*swang1,
                              data = rhc.training, method = "breslow")
summary(fit.with.caN)
anova(fit.with.caN, fit1a.das2d3pc)
AIC(fit.with.caN, fit1a.das2d3pc)
BIC(fit.with.caN, fit1a.das2d3pc)
#No interaction

# Interaction between age and RHC
fit.with.age<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
```

```
                              swang1 + dnr1 + hema + age*swang1,
                            data = rhc.training, method = "breslow")
summary(fit.with.age)
#No interaction

# Interaction between liverhx and RHC
fit.with.liverhx<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                              caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                              swang1 + dnr1 + hema + liverhx*swang1,
                            data = rhc.training, method = "breslow")
summary(fit.with.liverhx)
#No interaction

# Interaction between das2d3pc and RHC
fit.with.das2d3pc<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                   caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                                   swang1 + dnr1 + hema + das2d3pc*swang1,
                                 data = rhc.training, method = "breslow")
summary(fit.with.das2d3pc)
#No interaction

# Interaction between aps1 and RHC
fit.with.aps1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                              caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                              swang1 + dnr1 + hema + aps1*swang1,
                            data = rhc.training, method = "breslow")
summary(fit.with.aps1)
#No interaction

# Interaction between scoma1 and RHC
fit.with.scoma1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                                swang1 + dnr1 + hema + scoma1*swang1,
                              data = rhc.training, method = "breslow")
summary(fit.with.scoma1)
#No interaction

# Interaction between bili1 and RHC
fit.with.bili1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                               caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                               swang1 + dnr1 + hema + bili1*swang1,
                             data = rhc.training, method = "breslow")
summary(fit.with.bili1)
#No interaction

# Interaction between hema and RHC
fit.with.hema<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                              caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                              swang1 + dnr1 + hema + hema*swang1,
                            data = rhc.training, method = "breslow")
summary(fit.with.hema)
#No interaction
```

```
#No interaction terms are included in the model.
#Model with linear forms and no interaction terms is the best model after model building.


############################
# Cox PH Assumption
############################

## using schoenfeld residuals

temp<-cox.zph(fit1a)
temp

## log-log survival curves
par(mfrow=c(1,1))

# dnr1
fit_dnr1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                            caN + liverhx + age + das2d3pc + aps1 + scoma1 + bili1 +
                            swang1 + strata(dnr1) + hema, data = rhc.training, method = "breslow")
fit_dnr11<-survfit(fit_dnr1)
plot(fit_dnr11,fun='cloglog',lty=1:2,col=1:2)
legend("topleft",c("No DNR", "DNR"),lty=1:2,col=1:2, cex=0.8)

temp<-cox.zph(fit_dnr1)
temp
# do not stratify on dnr1


# aps1
fit_aps1<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                            caN + liverhx + age + das2d3pc + strata(aps1_grp) + scoma1 + bili1 +
                            swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
fit_aps11<-survfit(fit_aps1)
plot(fit_aps11,fun='cloglog',lty=1:4,col=1:4)
legend("topleft",c("Low", "Medium", "High"),lty=1:4,col=1:4)

temp<-cox.zph(fit_aps1)
temp


# use stratified model (aps1 as strata)


############################
# Outliers
############################

fit_training <- coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + strata(aps1_grp) + scoma1 + bili1 +
                                swang1 + dnr1 + hema, data = rhc.training, method = "breslow")
summary(fit_training)
```

```
dresid<-resid(fit_training,'dev')
length(dresid)

par(mfrow=c(1,1))
plot(dresid, ylim=c(-3.5, 3.5), main="Outliers",
        xlab="observations", ylab="Deviance Residuals")

head(sort(dresid))
tail(sort(dresid))


###################################
##Influential Points
###################################

sresid<-resid(fit_training,'score')
dim(sresid)

apply(sresid, 2, which.max)

data.frame(rhc$cat15, rhc$cat17, rhc$cat18, rhc$caN, rhc$liverhx,
                rhc$age, rhc$das2d3pc, rhc$scoma1, rhc$bili1, rhc$swang1,
                rhc$dnr1, rhc$hema)[c(5653, 4624, 1317),]
dresid[1317]

###################################
## Model Validation
###################################


#    Comparing coef. estimates

fit_validation<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + strata(aps1_grp) + scoma1 + bili1 +
                                swang1 + dnr1 + hema, data = rhc.validation, method = "breslow")
summary(fit_validation)

fit_all<-coxph(Surv(time, death1) ~ cat15 + cat17 + cat18 +
                                caN + liverhx + age + das2d3pc + strata(aps1_grp) + scoma1 + bili1 +
                                swang1 + dnr1 + hema, data = rhc, method = "breslow")
summary(fit_all)

tablecomp<-round( cbind(fit_training$coef, confint(fit_training),
                                fit_validation$coef, confint(fit_validation) ), 2)
tablecomp


# Comparing observed vs. expected survival

rs <- cbind(rhc$cat15, rhc$cat17, rhc$cat18, rhc$caN, rhc$liverhx,
                rhc$age, rhc$das2d3pc, rhc$scoma1, rhc$bili1, rhc$swang1,
                rhc$dnr1, rhc$hema)[validation,] %*% coef(fit_all)
dim(rs)
```

```
## stratum: Low APACHE score

rs_lowaps <- rs[rhc.validation$aps1_grp==1]

quantile(rs_lowaps, c(.33,.67))                    # 33% and 67% quantiles of risk scores

riskgrp_lowaps<-ifelse(rs_lowaps<=1.217248, 1, 2)
riskgrp_lowaps<-ifelse(rs_lowaps>=1.569997, 3, riskgrp_lowaps)
table(riskgrp_lowaps)

rhc.validation.lowaps <- rhc.validation[rhc.validation$aps1_grp==1,]
dim(rhc.validation.lowaps)

# observed

fit_lowaps <- survfit(Surv(time,death1)~riskgrp_lowaps, data=rhc.validation.lowaps)


# expected
zvalues <- data.frame(rhc$cat15, rhc$cat17, rhc$cat18, rhc$caN, rhc$liverhx,
                      rhc$age, rhc$das2d3pc, rhc$scoma1, rhc$bili1, rhc$swang1,
                      rhc$dnr1, rhc$hema)[validation,]
dim(zvalues)

zvalues_lowaps <- zvalues[rhc.validation$aps1_grp==1,]
dim(zvalues_lowaps)
names(zvalues_lowaps)

detach()
library(survival)
attach(rhc.validation.lowaps)
expect_lowaps <- survexp(~riskgrp_lowaps+ratetable(cat15=cat15, cat17=cat17, cat18=cat18,
                                                   caN=caN, liverhx=liverhx, age=age,
                                                   das2d3pc=das2d3pc, aps1_grp=1,
                                                   scoma1=scoma1, bili1=bili1, swang1=swang1,
                                                   dnr1=dnr1, hema=hema),
                         data=zvalues_lowaps, ratetable=fit_training)

par(mfrow=c(1,1))
plot(fit_lowaps,xlab="Days", ylab="Survival", main="Model Validation \nStratum: Low APACHE score",
       col=2:4, lty=1)
lines(expect_lowaps,col=2:4,lty=2)
legend("topright",paste( rep(c("observed", "expected"), 3),
                         rep(c("Low risk", "Medium risk", "High risk"), each=2) ),
          col=rep(2:4, each=2), lty=rep(1:2,3), cex=1 )



## stratum: High APACHE score

rs_highaps <- rs[rhc.validation$aps1_grp==2]
```

```
quantile(rs_highaps, c(.33,.67))                    # 33% and 67% quantiles of risk scores

riskgrp_highaps<-ifelse(rs_highaps<=1.359365, 1, 2)
riskgrp_highaps<-ifelse(rs_highaps>=1.812699, 3, riskgrp_highaps)
table(riskgrp_highaps)

rhc.validation.highaps <- rhc.validation[rhc.validation$aps1_grp==2,]
dim(rhc.validation.highaps)

# observed

fit_highaps<-survfit(Surv(time,death1)~riskgrp_highaps, data=rhc.validation.highaps)


# expected

zvalues_highaps <- zvalues[rhc.validation$aps1_grp==2,]
dim(zvalues_highaps)
names(zvalues_highaps)

detach()
library(survival)
attach(rhc.validation.highaps)
expect_highaps <- survexp(~riskgrp_highaps+ratetable(cat15=cat15, cat17=cat17, cat18=cat18,
                                                     caN=caN, liverhx=liverhx, age=age,
                                                     das2d3pc=das2d3pc, aps1_grp=2,
                                                     scoma1=scoma1, bili1=bili1, swang1=swang1,
                                                     dnr1=dnr1, hema=hema),
                          data=zvalues_highaps, ratetable=fit_training)

plot(fit_highaps,xlab="Days", ylab="Survival", main="Model Validation \nStratum: High APACHE score",
       col=2:4, lty=1)
lines(expect_highaps,col=2:4,lty=2)
legend("topright",paste( rep(c("observed", "expected"), 3),
                                  rep(c("Low risk", "Medium risk", "High risk"), each=2) ),
          col=rep(2:4, each=2), lty=rep(1:2,3), cex=1)


detach()
```