Patient Survival Prediction Based on Physiological Measurements Using a Logistic Regression Model

January 21, 2020

Executive Summary

Of the millions of hospital patients who are admitted, only some are discharged while others die during their stay. Some physiological measurements can indicate whether or not a patient is in a more critical condition than other patients. The assessment of how likely a patient is to die could lead to a more effective allocation of hospital resources so as to prevent patient death. The purpose of this study is to create a predictive model that would use physiological measurements to predict whether a patient survives or dies. The dataset used in the analysis was collected from 112 critically ill patients in Southern California. Welch's t test and stepwise model selection were used to create the final predictive model. The sensitivity value of 0.767 and specificity value of 0.753 indicated that the model performed well in predicting patient survival. It was determined that shock type, mean arterial pressure, mean central venous pressure and body surface index were associated with patient survival.

1 Introduction

There are numerous hospitals located throughout Southern California, with millions of patients admitted every year. A difficult task these hospitals face is deciding which patients should be prioritized when it comes to medical attention. Those with a higher risk of dying would ideally be treated first, but this can be difficult to determine. Upon admission and discharge, or in some cases death, many physiological measurements are recorded. Such measurements could potentially give insight into the seriousness of a patient's condition.

If the hospital staff is able to ascertain a patient's risk of dying based on these measurements, then medical resources can be more effectively distributed so that those more at risk of dying are given priority, potentially reducing the number of deaths. The purpose of this investigation is to make such a model that uses physiological measurements to accurately predict whether a patient survives or dies.

Methods

The dataset utilized in this study contains physiological measurements for 112 critically ill patients in Southern California upon admission and before discharge or death. A basic description of each variable in the dataset is listed in Table 1. There were no missing values found. A likely typo

Table 1: Dataset variable descriptions.

Variable	Description	Units	Comment
ID	ID	none	
AGE	Age	yr	
HT	Height	cm	
Sex		none	1=Male, 2=Female
SURVIVE	Survival	none	1=Survived, 3=Died
SHOCK_TYPE	Shock Type	none	2=Non-shock 3=Hypovolemic shock 4=Cardiogenic shock 5=Bacterial shock 6=Neurogenic shock 7=Other
SBP	Systolic pressure	mmHg	
MAP	Mean arterial pressure	mmHg	
HR	Heart rate	beats/min	
DBP	Diastolic pressure	mmHg	
MCVP	Mean central venous pressure	cm H_2O	
BSI	Body surface index	m^2	
CI	Cardiac index	liters/min*m^2	
AT	Appearance time	sec	
MCT	Mean circulation time	sec	
UO	Urinary output	ml/hr	
PVI	Plasma volume index	ml/kg	
RCI	Red cell index	ml/kg	
HG	Hemoglobin	gm/100 ml	
HCT	Hematocrit	percent	
RECORD	Card sequence	none	1=Initial, 2=Final

recorded an initial height of 70 cm and final height of 170 cm for the patient with ID 539, so the initial height was changed to 170 cm. The hematocrit values listed were greater than 100%, so the implied decimal point was added. The sex, survive, shock type and record variables were changed to factors and renamed. Specifically, the sex variable was changed such that 1 was represented by "M" and 2 by "F". The survive variable was changed such that 1 was represented by "Survived" and 3 by "Died". The shock type variables were represented by shortened versions of their names: "Non", "Hypovol", "Cardio", "Bacterial", "Neuro" and "Other". The record variable was changed such that 1 was represented by "Initial" and 2 by "Final". The final measurements were filtered out of the original dataset and not used in the study since they were collected prior to discharge or death, where the survival outcome was evident. The variable ID was dropped because it was only used to identify the patients. The variable RECORD was dropped after subsetting the data into initial and final datasets.

The response variable indicated whether a patient survived or not. The two categorical predictor variables were sex and shock type. The continuous predictor variables were age, height, systolic blood pressure, mean arterial pressure, heart rate, diastolic blood pressure, mean central venous pressure, body surface index, cardiac index, appearance time of symptoms, mean circulation time, urinary output, plasma volume index, red cell index, hemoglobin levels and hematocrit levels.

The response variable states whether a patient survived or died, indicating that a logistic

regression model would be appropriate in predicting the binary response of survival. Exploratory data analysis was performed as a means of discovering the effects of the predictor variables on the response variable. An initial logistic regression model was made using all the variables, which was reduced to an optimal final model with fewer variables by considering both variable significance and stepwise model selection using AIC selection criterion. The fitness of the model was assessed by performing diagnostics. Throughout the study, R version 3.5.3 was used.

2 Exploratory Data Analysis

There were 69 patients in the study who survived and 43 who died, with 62% of patients surviving and 38% dying. The relationship between the response variable and the categorical predictor variables was investigated using contingency tables. The relationship between the shock type and response variable was shown in Table 2. The relationship between the sex and response variable was shown in Table 5b in the Appendix B. The χ^2-test confirmed that there was a significant association between patient survival and shock type at the 0.05 significance level, with a p-value of 6.373e-02. The results of the tests are included in Table 6 in the Appendix B.

Table 2: Contingency table for survive and shock type.

	Non	Hypovolemic	Cardiogenic	Bacterial	Neurogenic	Other
Survived	31	7	10	9	9	3
Died	3	10	10	6	7	7

The relationship between the response variable and the continuous predictor variables was investigated. The distributions of the continuous predictor variables hematocrit and mean arterial pressure with respect to patient survival were illustrated using boxplots, as seen in Figure 1.

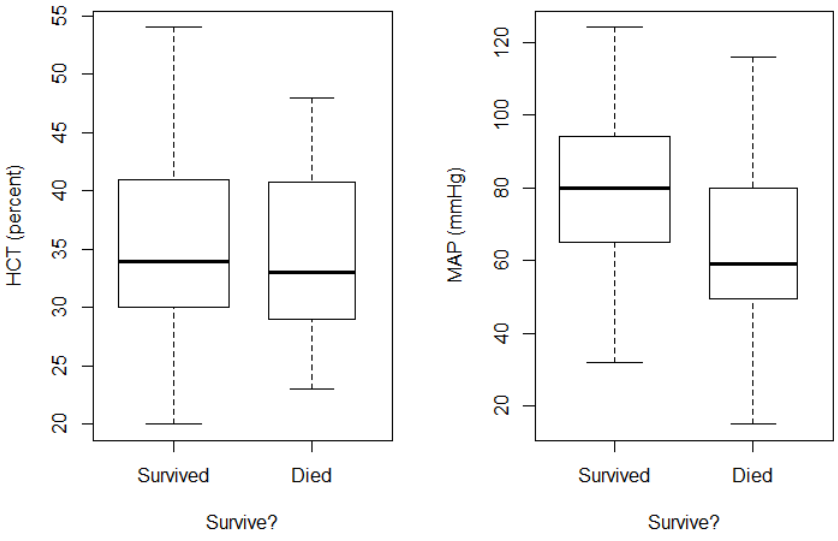

Figure 1: Boxplots for hematocrit (left) and mean arterial pressure (right).

The difference in means between the response variable and the continuous predictor variables was investigated using the Welch's two-sample t-test. The mean arterial pressure variable was found to have a significant difference in means between patients who survived and those who died, with a p-value of 9.99e-05. The hematocrit variable showed no significant difference in means, and thus hematocrit had no association with the response variable. The results of the tests are included in Table 7 in the Appendix B.

The correlation plot in Figure 2 was used to visualize the associations between the continuous categorical variables. The following notable associations were found: systolic blood pressure and mean arterial pressure, systolic pressure and diastolic pressure, body surface index and height, mean circulation time and cardiac index.

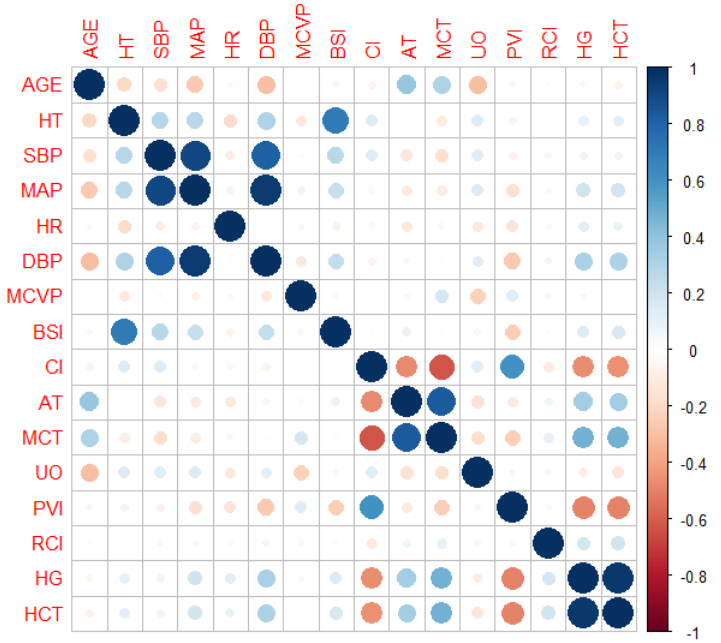

Figure 2: Correlation plot with variables from the dataset.

Cohen's D statistic was calculated so as to assess the magnitude of the effect of each continuous predictor variable on the response variable. The mean arterial pressure variable was found to have Cohen's D value of 0.817, signifying a large-level effect on patient survival, as expected from two variables found to have an association. In contrast, the variable hematocrit was found to have a Cohen's D value of 0.0926, indicating a negligible effect on patient survival, as expected from two variables found to have no association.

3 Statistical Analyses

The p-value from the Welch's t test was used along with the stepwise AIC algorithm to create an optimal logistic regression model. All variables were initially included in the full model. Select variables were omitted depending on the agreement between the Welch's t test p-value and the stepwise AIC model selection, while others were added based on previous findings.

After the first model change, both Welch's t test and the AIC algorithm found shock type,

mean central venous, body surface index and urinary output to be significant variables. All other variables in the full model were removed, while mean arterial pressure was added based on the previously determined significance in the mean difference; the resulting AIC score was 119.79. After the second model change, both Welch's t test and the AIC algorithm found shock type, mean central venous, body surface index and mean arterial pressure to be significant. The urinary output variable was omitted based on the defined procedure; the resulting AIC score was 120.81. The final model included shock type, mean central venous, body surface index and mean arterial pressure variables. Although the AIC score was slightly larger when urinary output was removed, the slight correlation with mean central venous and current lack of significance supported the choice of omission. Interaction terms were not included since AIC only slightly decreased while the model became more complex.

Diagnostics

The continuous covariates in the final model were transformed into binomial explanatory variable patterns (EVPs) so that diagnostics could be performed later. By binning all of the covariates into 2 intervals, the aggregation of the data into sets of similar covariates resulted in 39 EVPs and 109 observations.

Diagnostic plots made from the EVPs are shown in Figure 3, located in the Appendix C. In the standardized Pearson residual plot, there were three points marked as possible outliers. The same three points were marked as possible outliers in the Cook's distance versus leverage plot. Upon further examination, the measures of influence for the potential outliers shown in Table 3 indicated that it was not necessary to omit the sets of values, since none of the values had a studentized residual greater than 3 or less than -3.

Table 3: Potential EVP outliers/influential points are listed.

	SHOCK_TYP	MAP_int	MCVP_int	BSI_int	SURVIVE	std.res	cookd	h
16	Neuro	(15,72.5]	(80,302]	(109,169]	1	0.66	0.70	0.43
32	Non	(15,72.5]	(80,302]	(80, 302]	2	0.14	0.28	0.30
38	Neuro	(72.5,124]	(80,302]	(169, 224]	2	0.32	0.14	0.34

The ability of the final model to correctly classify a patient survival or death was assessed using sensitivity and specificity, with values of 0.767 and 0.753, respectively; the cutoff point of 0.38 was based on the proportion of patients who died in the dataset. Such high values indicate that the final model does a decent job of classifying whether a patient survives or dies.

The relationship between patient survival and each predictor variable can be described using odds ratios. For a proper interpretation of a percent increase or decrease in the odds of patient survival, it is assumed that all other variable unit changes are held constant. There was a 2,904% increase in the odds of the admitted patient dying if the experienced shock type was "other". A 1,450% increase in the odds of the admitted patient dying occurred if the experienced shock type was "hypovolemic". There was an 827% increase in the odds of the admitted patient dying if the experienced shock type was "cardiogenic". A 598% increase in the odds of the admitted patient dying occurred if the experienced shock type was "neurogenic". There was a 279% increase in the odds of the admitted patient dying if the experienced shock type was "bacterial". For every cm H_20 increase in mean central venous pressure, the odds of the admitted patient dying increased by 2%. For every m2 increase in body surface index, the odds of the admitted patient dying decreased

by 2.8%. For every mmHg increase in mean arterial pressure, the odds of the admitted patient dying decreased by 2.9%. A summary of the coefficients and odds ratios for each of the final model variables are listed in Table 4.

Table 4: Summary of the coefficients and odds ratios for each of the final model variables.

	Coefficient	Odds ratio	SE	p-value	95% CI on OR
Intercept	3.140	23.000	2.500	0.210	(0.195 , 3950.000)
Hypovolemic Shock	2.740	15.500	0.935	0.003	(2.750 , 115.000)
Cardiogenic Shock	2.230	9.270	0.835	0.008	(1.980 , 56.100)
Bacterial Shock	1.330	3.790	0.897	0.137	(0.683 , 24.900)
Neurogenic Shock	1.940	6.980	0.871	0.026	(1.360 , 44.400)
Other	3.410	30.400	1.070	0.001	(4.270 , 308.000)
MAP	-0.029	0.971	0.013	0.023	(0.945 , 0.995)
MCVP	0.015	1.020	0.005	0.002	(1.010 , 1.030)
BSI	-0.028	0.972	0.014	0.048	(0.944 , 0.999)

4 Conclusion

The finding that the physiological measurements shock type, mean arterial pressure, mean central venous pressure and body surface index were useful in ascertaining whether a patient survived or died made intuitive sense. An overweight or tall patient with a large body surface index experiencing shock would be likely to experience drastic changes in arterial and central venous pressures, making it more likely that the patient will be in critical condition or even die. Understanding the importance of these variables could potentially lead to fewer patients deaths.

The high sensitivity and specificity indicated that the model performed reasonably well when identifying patients who were going to die and patients who were not going to die, respectively.

Discussion

The non-shock variable was quite vague and limited the accuracy of the model, since this could include life-threatening or non-life-threatening conditions. The small size of the dataset limited what analyses and model assessments could be successfully carried out. The small dataset also could have led to exclusion of variables in the final model.

It is evident that some physiological measurements are more useful than others in predicting patient death. A potentially useful factor that could predict the survival of a patient is their medical history. This could be further investigated in future studies.

Appendices

Appendix A: Annotated R Code

```
library(dplyr) # for filtering
library(corrplot) # for correlation plot
library(MASS) # for stepAIC
library(effsize) # for Cohen's d
library(caret) # for sensitivity and specificity

#
#

## Read in data
ca<-read.csv("DAE_2020.csv",header=F); dim(ca)  #  224  21
colnames(ca)<-c("ID", "AGE", "HT", "Sex", "SURVIVE", "SHOCK_TYP", "SBP",
"MAP", "HR", "DBP", "MCVP", "BSI", "CI", "AT", "MCT", "UO",
"PVI", "RCI", "HG", "HCT", "RECORD")
# Data summary
str(ca)
head(ca, n = 5)
summary(ca)

# Search for missing values
sum(is.na(ca)) # 0 missing values

# Investigate unique values based on questionable summary mins, maxes
ca[ca$HT == 70, "ID"] # ID 539 has really small height
ca[ca$ID == 539, "HT"] # 70 and 170 (should be similar or the same for initial and final)
ca[ca$ID == 540, "HT"] # 160 and 160 (example of how it should be)
ca[ca$ID == 539, "HT"] <- 170 # changed inital and final HT to 170

ca$HCT # all values are over 100%, so likely variable w/implied decimal
ca$HCT <- (ca$HCT)/10 # now correct percent

# Set and rename factors
ca$Sex <- as.factor(ca$Sex)
ca$SURVIVE <- as.factor(ca$SURVIVE)
ca$SHOCK_TYP <- as.factor(ca$SHOCK_TYP)
ca$RECORD <- as.factor(ca$RECORD)
levels(ca$Sex) <- c("M", "F")
```

```
levels(ca$SURVIVE) <- c("Survived", "Died")
levels(ca$SHOCK_TYP) <- c("Non", "Hypovol","Cardio", "Bacterial", "Neuro", "Other")
levels(ca$RECORD) <- c("Initial", "Final")

## Filter the data into initial and final data sets (NEED dplyr package)
ca_in <- filter(ca, ca$RECORD == "Initial"); dim(ca_in) # 112 21; NOTE: Using this dataset
ca_fi <- filter(ca, ca$RECORD == "Final"); dim(ca_fi) # 112 21
attach(ca_in)

# Remove ID and Record column for ca_in dataset
ca_in$ID <- NULL
ca_in$RECORD <- NULL
head(ca_in)
SURVIVE

##################################
# EDA: (Response variabe SURVIVE)
##################################

## EDA: Contingency tables for categorical variables

# frequency table for response variable (SURVIVE)
table(SURVIVE)
# SURVIVE
# Survived     Died
# 69       43
signif(prop.table(table(SURVIVE)), digits=3)
# SURVIVE
# Survived     Died
# 0.616    0.384

# Contingency tables for categorical covariates
table(SURVIVE, SHOCK_TYP)
# SHOCK_TYP
# SURVIVE    Non Hypovol Cardio Bacterial Neuro Other
# Survived  31       7     10         9     9     3
# Died       3      10     10         6     7     7
signif(prop.table(table(SURVIVE, SHOCK_TYP)), digits=3)

table(SURVIVE, Sex)
# Sex
# SURVIVE     M  F
# Survived 41 28
# Died     17 26
```

```
signif(prop.table(table(SURVIVE, Sex)), digits=3)

## EDA: Chi-square tests for SURVIVE vs. categorical variables
chisq.test(SHOCK_TYP, SURVIVE) ##
chisq.test(Sex, SURVIVE) ##
chisq.test(Sex, SHOCK_TYP)
# pvalues: 0.0007616, 0.06376, 0.4092

## EDA: Boxplots for SURVIVE vs. continuous covariates
plot(SURVIVE,AGE, ylab = "AGE (year)", xlab="Survive?", varwidth=T)
plot(SURVIVE,HT, ylab = "HT (cm)", xlab="Survive?", varwidth=T) ##
   dead had lower HT
plot(SURVIVE,SBP, ylab = "SBP (mmHg)", xlab="Survive?", varwidth=T)
    ## dead had lower SBP
plot(SURVIVE,MAP, ylab = "MAP (mmHg)", xlab="Survive?", varwidth=T)
    ## dead had lower MAP
plot(SURVIVE,HR, ylab = "HR (beats/min)", xlab="Survive?", varwidth
   =T)
plot(SURVIVE,DBP, ylab = "DBP (mmHg)", xlab="Survive?", varwidth=T)
    ## dead had lower DBP
plot(SURVIVE,MCVP, ylab = "MCVP (cmH2O)", xlab="Survive?", varwidth
   =T) ## dead had higher MCVP
plot(SURVIVE,BSI, ylab = "BSI (m2)", xlab="Survive?", varwidth=T)
   ## dead slightly lower BSI
plot(SURVIVE,CI, ylab = "CI (liters/min m2)", xlab="Survive?",
   varwidth=T)
plot(SURVIVE,AT, ylab = "AT (sec)", xlab="Survive?", varwidth=T)
plot(SURVIVE,MCT, ylab = "MCT (sec)", xlab="Survive?", varwidth=T)
   ## dead had higher MCT
plot(SURVIVE,UO, ylab = "UO (ml/hr)", xlab="Survive?", varwidth=T)
   ## similar means, but different IQR
plot(SURVIVE,PVI, ylab = "PVI (ml/kg)", xlab="Survive?", varwidth=T
   )
plot(SURVIVE,RCI, ylab = "RCI (ml/kg)", xlab="Survive?", varwidth=T
   )
plot(SURVIVE,HG, ylab = "HG (gm/100 ml)", xlab="Survive?", varwidth
   =T)
plot(SURVIVE,HCT, ylab = "HCT (percent)", xlab="Survive?", varwidth
   =T) # example of same
# possible variables of interest: SBP, MAP, DBP, MCVP, BSI, MCT, UO
par(mfrow=c(1,2))
plot(SURVIVE,HCT, ylab = "HCT (percent)", xlab="Survive?", varwidth
   =T)
plot(SURVIVE,MAP, ylab = "MAP (mmHg)", xlab="Survive?", varwidth=T)
par(mfrow=c(1,1))

# EDA: Correlation plot
```

```
corrplot(cor(ca_in[sapply(ca_in, is.numeric)]))
# associations of variables of interest:
# SBP: MAP, SBP: DBP, BSI: HT, MCT:AT, MCT:CI

# SURVIVE factor changed to numeric for later analyses
SURVIVE <- ifelse(SURVIVE=="Survived",0,1)
SURVIVE<-as.numeric(as.character(SURVIVE))
ca_in$SURVIVE <- SURVIVE

## EDA: Welch's two-sample t-test for continuous covariates
t.test(AGE[SURVIVE==0], AGE[SURVIVE==1])
t.test(HT[SURVIVE==0], HT[SURVIVE==1]) ## 7
t.test(SBP[SURVIVE==0], SBP[SURVIVE==1]) ## 2
t.test(MAP[SURVIVE==0], MAP[SURVIVE==1]) ## 1
t.test(HR[SURVIVE==0], HR[SURVIVE==1])
t.test(DBP[SURVIVE==0], DBP[SURVIVE==1]) ## 3
t.test(MCVP[SURVIVE==0], MCVP[SURVIVE==1]) ## 5
t.test(BSI[SURVIVE==0], BSI[SURVIVE==1]) ## 6
t.test(CI[SURVIVE==0], CI[SURVIVE==1])
t.test(AT[SURVIVE==0], AT[SURVIVE==1])
t.test(MCT[SURVIVE==0], MCT[SURVIVE==1]) ## 8
t.test(UO[SURVIVE==0], UO[SURVIVE==1]) ## 4
t.test(PVI[SURVIVE==0], PVI[SURVIVE==1])
t.test(RCI[SURVIVE==0], RCI[SURVIVE==1])
t.test(HG[SURVIVE==0], HG[SURVIVE==1])
t.test(HCT[SURVIVE==0], HCT[SURVIVE==1])
# pvalues: 0.1882, 0.06965, 0.000116, 9.991e-05, 0.3006, 0.0003469,
    0.004267,
# 0.02435, 0.2308, 0.3418, 0.0888, 0.0004551, 0.5683, 0.2342,
   0.5729, 0.6216

## EDA: Cohen's d (standardized mean difference)
# note: 0.2 (small), 0.5 (medium), and 0.8 (large)
cohen.d(AGE, factor(SURVIVE))  #
cohen.d(HT, factor(SURVIVE))  #
cohen.d(SBP, factor(SURVIVE)) ##
cohen.d(MAP, factor(SURVIVE)) ###
cohen.d(HR, factor(SURVIVE))
cohen.d(DBP, factor(SURVIVE)) ##
cohen.d(MCVP, factor(SURVIVE)) ##
cohen.d(BSI, factor(SURVIVE)) #
cohen.d(CI, factor(SURVIVE)) #
cohen.d(AT, factor(SURVIVE))
cohen.d(MCT, factor(SURVIVE)) #
cohen.d(UO, factor(SURVIVE)) ##
cohen.d(PVI, factor(SURVIVE)) ##
cohen.d(RCI, factor(SURVIVE)) #
```

```
cohen.d(HG, factor(SURVIVE))
cohen.d(HCT, factor(SURVIVE))
# d: -0.259, 0.331, 0.79, 0.817, -0.199, 0.7417, -0.564, 0.466,
# 0.227, -0.182, -0.349, 0.593, 0.495, 0.2094, 0.107, 0.0926

##############################################
# MODEL BUILDING: (Response variabe SURVIVE)
##############################################

## Logistic Models
# (Full logistic regressiong model)
fit.full <- glm(SURVIVE~., family=binomial(link=logit), data=ca_in)
summary(fit.full)
# Signigicant variables: SHOCK_TYP, MCVP, BSI, UO, PVI

full_aic <- stepAIC(fit.full)
anova(full_aic) # Chisqr: RCI insignif
# Variables: SHOCK_TYP, MAP, MCVP, BSI, UO, RCI; AIC = 119.09

# (Remove PVI, RCI; Add MAP)
fit.1 <- glm(SURVIVE~SHOCK_TYP+ MAP + MCVP+ BSI+ UO,
family=binomial(link=logit), data=ca_in)
summary(fit.1)
# Signigicant variables: SHOCK_TYP, MCVP, BSI

first_aic <- stepAIC(fit.1)
anova(first_aic)
# Variables: SHOCK_TYP, MAP, MCVP, BSI, UO

# (Remove UO)
fit.final <- glm(SURVIVE~SHOCK_TYP+ MAP + MCVP+ BSI,
family=binomial(link=logit), data=ca_in)
summary(fit.final)
# Signigicant variables: SHOCK_TYP, MAP, MCVP, BSI

final_aic <- stepAIC(fit.final) # variables
anova(final_aic)
# Variables: SHOCK_TYP, MAP, MCVP, BSI

# (Add interaction terms)
fit.3 <- glm(SURVIVE~SHOCK_TYP+ MAP + MCVP+ BSI+
MAP:MCVP+ MAP:BSI+ MCVP:BSI,
family=binomial(link=logit), data=ca_in)
summary(fit.3)
# Signigicant variables: SHOCK_TYP, MCVP, MCVP:BSI; BSI

third_aic <- stepAIC(fit.3) # variables
```

```
anova(third_aic)
# Variables: SHOCK_TYP, MAP, MCVP, BSI, MCVP:BSI

# (Remove interaction terms except MCVP:BSI)
fit.4 <- glm(SURVIVE~SHOCK_TYP + MAP+ MCVP+ BSI+ MCVP:BSI,
family=binomial(link=logit), data=ca_in)
summary(fit.4)
# Signigicant variables: SHOCK_TYP, BSI, MCVP:BSI; MCVP

fourth_aic <- stepAIC(fit.4) # variables
anova(fourth_aic)
# Variables: SHOCK_TYP, MAP, MCVP, BSI, MCVP:BSI

###############
# Diagnostics
###############

## First, transform data into binomial explanatory variable
   patterns (EVPs)
g <- 2 # number of categories
MAP_int <- cut(MAP, quantile(MAP,0:g/g), include.lower=TRUE)
MCVP_int <- cut(MCVP, quantile(MCVP,0:g/g), include.lower=TRUE)
BSI_int <- cut(BSI, quantile(BSI,0:g/g), include.lower=TRUE)

# Compute for each covariate pattern the number of deaths (w) and
   the number of patients (n)
w <- aggregate(formula = SURVIVE~SHOCK_TYP+MAP_int+MCVP_int+BSI_int
   , data = ca_in, FUN = sum)
n <- aggregate(formula= SURVIVE~SHOCK_TYP+MAP_int+MCVP_int+BSI_int,
   data= ca_in, FUN=length)
w.n <- data.frame(w, trials=n$SURVIVE, prop=round(w$SURVIVE/
   n$SURVIVE,2))
w.n
head(w.n) # View the EVPs
nrow(w.n) # Number of EVPs (covariate patterns) = 39
sum(w.n$trials)  # Number of observations = 109

# Recall final logistic model
fit.prelim1=glm(SURVIVE~SHOCK_TYP+MAP+MCVP+BSI,
family=binomial(link=logit), data=ca_in)

# preliminary model: weighted logistic regression of EVP
   proportions on explanatory vars in final model
mod.prelim1<- glm(formula=SURVIVE/trials~SHOCK_TYP+MAP_int+MCVP_int
   +BSI_int,
family=binomial(link=logit),data=w.n, weights=trials)

```

```
## Diagnostics: residual plots
one.fourth.root=function(x){
x^0.25
}
source("Examine.logistic.reg(1)(1).R")
save1 <- examine.logistic.reg(mod.prelim1,identify.points = T,
scale.n=one.fourth.root,scale.cookd=sqrt)

# Store prediction, residual, Cook's D, and leverage values
w.n.diag1 <- data.frame(w.n, pi.hat=round(save1$pi.hat, 2), std.res
   =round(save1$stand.resid, 2),
cookd=round(save1$cookd, 2), h=round(save1$h, 2))
p <- length(mod.prelim1$coef) # number of parameters in model (#
   coefficients = 9)

# Examine outliers/influential points idenitfied by plots
ck.out <- abs(w.n.diag1$std.res)>2.5 | w.n.diag1$cookd>4/nrow(w.n)
   | w.n.diag1$h > 3*p/nrow(w.n)
extract.EVPs <- w.n.diag1[ck.out, ]
extract.EVPs[order(extract.EVPs$SURVIVE),]

## Diagnostics: Sensitivity and Specificity
# Recall imbalanced classes:
# SURVIVE
# Survived     Died
# 0.616     0.384
threshold <- 0.38
predicted_values <- ifelse(predict(fit.final, type="response")>
   threshold,1,0)
actual_values <- fit.final$y
confmatrix <- table(predicted_values, actual_values)
confmatrix
sensitivity <- confmatrix[2,2]/(confmatrix[2,2] + confmatrix[1,2])
   # 0.767
specificity <- confmatrix[1,1]/(confmatrix[1,1] + confmatrix[2,1])
   # 0.753

## Diagnostics: Odds ratio
beta_hat <- formatC(signif(fit.final$coeff,digits=3),digits=3,
   format='f',flag='#')
OR <- formatC(signif(exp(fit.final$coeff),digits=3),digits=3,format
   ='f',flag='#')
SE <- formatC(signif(summary(fit.final)$coeff[,2],digits=3),digits
   =3,format='f',flag='#')
CI_bounds <- formatC(signif(exp(confint(fit.final)),digits=3),
   digits=3,format='f',flag='#')
p_val <- formatC(signif(summary(fit.final)$coeff[,4],digits=3),
```

```
   digits=3,format='f',flag='#')
OR_table <- cbind(beta_hat,OR,SE,p_val,matrix(paste("(",CI_bounds
   [,1],
",",CI_bounds[,2],")")))
colnames(OR_table)<-cbind("Coefficient","Odds Ratio","SE","p-value
   ",
"95% CI on OR")
OR_table
```

Appendix B: Additional Contingency Table; χ^2-test and Welch's t test results

Table 5: Contingency table for survive and sex.

	M	F
Survived	41	28
Died	17	26

Table 6: Results of χ^2-test of association for survival and sex, as well as survival and shock type.

	χ^2	df	p-value
Sex	21.142	5	7.616e-04
Shock type	3.4369	1	6.376e-02

Table 7: Results from the Welch's t test.

	t	df	p value
Age	-1.32	87.44	0.1882
Height	3.12	70.97	2.57e-03
Diastolic pressure	3.73	82.73	3.46e-04
Systolic pressure	4.03	86.8	1.16e-04
Mean arterial pressure	4.09	81.34	9.99e-05
Heart rate	-1.04	92.61	3.00e-01
Diastolic pressure	3.73	82.73	3.46e-04
Mean central venous pressure	-2.93	91.75	4.26e-03
Body surface index	2.29	76.75	2.43e-02
Cardiac index	1.20	97.53	2.30e-01
Appearance time	-0.955	94.59	3.41e-01
Mean circulation time	-1.72	76.99	8.88e-02
Urinary output	3.63	96.32	4.55e-04
Plasma volume index	0.572	92.53	5.68e-01
Red cell index	1.19	110	2.34e-01
Hemoglobin	0.565	95.34	5.72e-01
Hematocrit	0.495	99.67	6.21e-01

Appendix C: Diagnostic plots

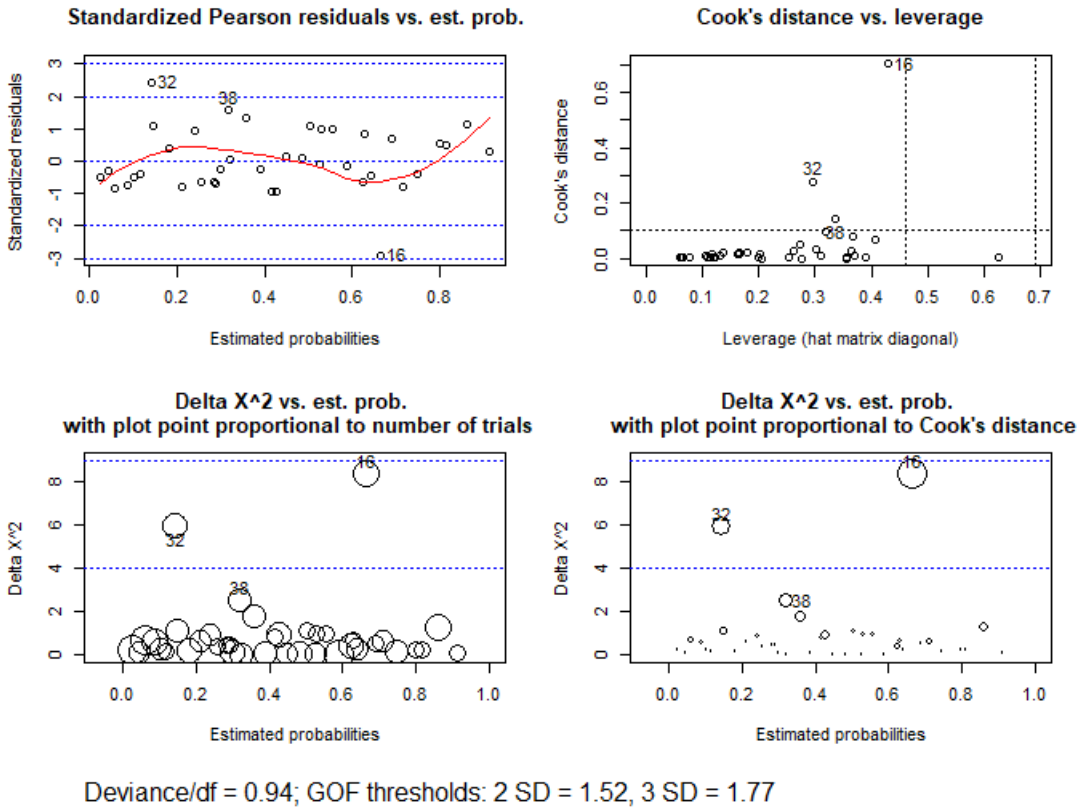

Figure 3: Diagnostic plots made from EVPs.