IS4250 Healthcare Analytics Project

Logistic regression model for identification of right ventricular dysfunction in patients with acute pulmonary embolism by means of computed tomography

Group 07

Wang Zhe	A0111785E
Wen Di	A0112491N

Link for the paper:

http://www.sciencedirect.com/science/article/pii/S0720048X13000715

Table of Contents

Summary of Paper

The paper we choose to analyze is titled as “Logistic regression model for identification of right ventricular dysfunction in patients with acute pulmonary embolism by means of computed tomography”. The main purpose of this paper is to increase the accuracy of identifying and predicting right ventricular dysfunction (RVD) experienced by patients with the use of computed tomography pulmonary angiography (CTPA). Most importantly, a logistic regression model constituted of several measurements would be calculated to theoretically support the prediction.

To facilitate the study, 97 sequential patients with acute pulmonary embolism were grouped according to with and without RVD using echocardiographic measurement of pulmonary artery systolic pressure (PASP). The PASP (showed in Fig. 1) that is less than or equal to 30 mmHg would be grouped under RVD (-), and the PASP greater than 30 mmHg would be grouped under RVD (+).

	RVD(−) PASP $\leq$ 30 mmHg; n = 46	RVD(+) PASP > 30 mmHg; n = 51	p
Age [years]	71.5 (27–91)	67 (22–92)	NS
BMI	26.79 (23.23–38.51)	26.57 (19.15–34.29)	NS
Systolic BP [mmHg]	119 (85–168)	124 (70–163)	NS
Heart rate [bpm]	94 (60–125)	96 (48–144)	NS
PASP [mmHg]	25 (18–30)	45 (32–95)	<0.001
Length of hospitalization [days]	9 (1–31)	12 (1–44)	0.03
ICU admission (*)	21 (46%)	34 (67%)	0.04
Mortality (*)	4 (8.7%)	7 (13.7%)	NS

Figure 1. Patients Characteristics

PE severity was graded with the pulmonary obstruction score. The following factors were taken into account: CT measurements of heart chambers and mediastinal vessels, position of interventricular septum and presence of contrast reflux into the inferior vena cava. The logistic regression model was prepared by means of backward conditional stepwise logistic regression to delete the weakest predictors. The final features (showed in Fig. 2) selected are obstruction score, maximum short axis of the Right Ventricle, and diameter of the inferior vena cava.

Results of logistic regression analysis.

		Estimate	SE	T test	P level	OR	95.0% C.I.	
							Lower	Upper
b_0	Intercept log. reg.	−8.361	1.92	18.85	<0.001	0.001	–	–
b_1	Obstruction score	0.132	0.02	29.75	<0.001	1.141	1.088	1.196
b_2	RV	0.078	0.03	8.48	0.004	1.082	1.025	1.140
b_3	IVC	0.094	0.05	3.61	0.012	1.098	0.997	1.211

Figure 2. Result of Logistic Regression Analysis

After running logistic regression, the obtained model:

$$Z = -8.361 + 0.132 \times \text{Obstruction Score} + 0.078 \times \text{RV} + 0.094 \times \text{IVC}$$

The calculated model, which is characterizing 79% sensitivity and 81% specificity, performed much better than single CT-based measurements. The area under the curve (showed in Fig. 3) is 86%, within 95% confidence interval. Therefore, it leads to the conclusion that using logistic regression model is able to identify RVD drastically better than solely CT-based measurements.

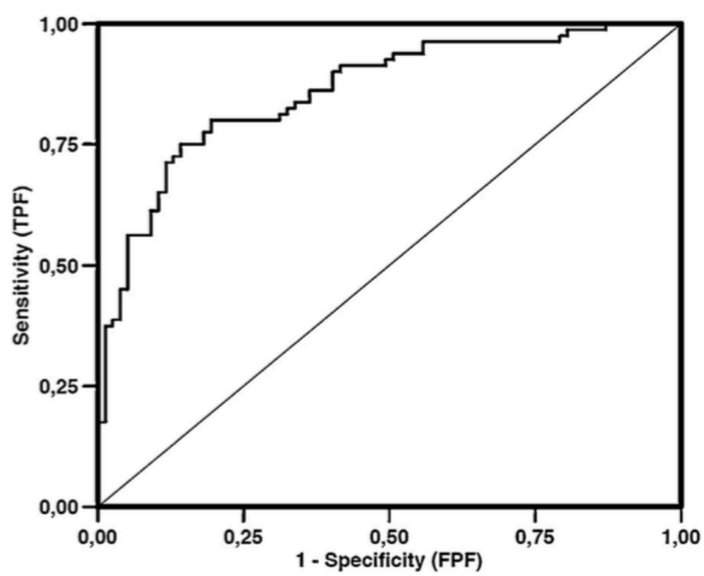

Figure 3. ROC Curve

Relevance of Paper

Comparing to left ventricle, lesser attention has been paid to how right ventricular dysfunction may be best detected and measured historically. Although the information regarding right ventricular function is relatively limited, it is essential to gain deeper understanding about right ventricular dysfunction as it is a critical contributor of various diseases (American Heart Association, 2006).

The study of investigating a logistic regression model to identify right ventricular dysfunction which is diagnosed using computed tomography pulmonary angiography denoting higher sensitivity and specificity would help improve doctors' diagnosis of the related symptoms. Consequently, this study would improve the quality of healthcare in the area of treating and healing right ventricular dysfunction for patients with acute pulmonary embolism as well as increasing its reliability.

Issues and Challenges

Acute pulmonary embolism (PE) refers to obstruction of the pulmonary artery or one of its branches by material (e.g. thrombus, tumor, air, or fat). It is a form of venous thromboembolism (VTE) that is commonly exists and can be deadly. The clinical presentation of PE is dynamic and usually nonspecific which in term make the diagnosis challenging.

Evaluation of heart failure patients with suspected acute PE is challenging because of the substantial overlap in symptoms and signs of both illnesses (G. P., 2008). Although dyspnea (breathing difficulty) is the most commonly reported symptom in both PE and heart failure, severe dyspnea out of proportion to objective findings of pulmonary vascular congestion suggests that another process such as PE is compromising gas exchange. Systemic arterial hypotension, cardiogenic shock, or cardiac arrest may be observed in decompensated heart failure or massive PE. Submassive PE should be considered in normotensive heart failure patients with evidence of right heart failure such as elevated jugular venous pressure, tricuspid regurgitation, and an accentuated sound of pulmonic closure. However, heart failure patients often show signs of right ventricular dysfunction without PE.

Laboratory testing of heart failure patients with suspected acute PE is complicated because elevated D-dimer levels may be present without other symptoms or illness. Similarly, cardiac biomarkers, including troponin, brain-type natriuretic peptide, and pro-brain–type natriuretic peptide, may be difficult to interpret because they often are abnormally elevated in heart failure. Nevertheless, heart failure patients with elevated cardiac biomarkers and findings of new or worsened Right Ventricular Dysfunction (RVD) should undergo further evaluation for PE.

Contrast-enhanced chest CT is the preferred imaging modality to evaluate suspected PE in heart failure patients. However, use of intravenous contrast-enhanced chest CT may be difficult for heart failure patients with chronic kidney disease who will have increased vulnerability to contrast nephropathy. In addition, the rapid bolus administration of intravenous contrast may cause a sudden increase in intra-cardiac pressures and pulmonary edema. Heart failure patients presenting with pulmonary vascular congestion or systemic hypertension should be stabilized before undergoing chest CT, in order to have accurate result.

Limitations of the Study

Due to the retrospective character of the study, which is designed to analyze pre-existing data that are collected from medical records of patients, the level of evidence could be inferior compared with prospective studies. The result may be prone to selection bias and not representative of the general population if convenience sampling influences control. Therefore, further evaluation of applicability of the computed logistic regression model should be carried out on a new and larger group of patients.

Replications

We are going to simulate the multivariate logistic regression and generate a simple ROC curve that is similar in this paper. We cannot use pure random generated data to plot the ROC curve as it will only show around 50-60% area under the curve. Hence, we have crafted some data in project.csv and create the logistic regression model. After the predicted value is calculated, it will plot the ROC curve. Eventually, it will be tested by the Hosmer-Lemeshow test to determine the goodness of fit of the model.

```
#Author: Wang Zhe, Wen Di
#Project: IS4250 Healthcare Analytics

library(pROC)
library(ResourceSelection)

#you may need change the path to read the csv file
setwd('/Users/wang/Desktop')
data = read.csv('project.csv')

#running logistic regression
mod <- glm(data$Z~data$Obstruction+data$RV+data$IVC, family="binomial")
summary(mod)
predpr <- predict(mod)

#plot roc curve
roccurve <- roc(data$Z ~ predpr)
plot(roccurve,legacy.axes=TRUE)

#running hosmer-lemeshow test
hoslem.test(mod$y, fitted(mod), g=10)
```

Figure 4. The Simulated ROC Curve - R Code

```
Call:
glm(formula = data$Z ~ data$Obstruction + data$RV + data$IVC,
    family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1633  -0.9539   0.4518   0.8816   1.7692

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       6.24492    2.47470   2.524   0.0116 *
data$Obstruction -0.11981    0.04850  -2.470   0.0135 *
data$RV           0.04122    0.03287   1.254   0.2098
data$IVC         -0.05559    0.03041  -1.828   0.0675 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 61.29  on 44  degrees of freedom
Residual deviance: 50.32  on 41  degrees of freedom
AIC: 58.32

Number of Fisher Scoring iterations: 4
```

Figure 5. Summary of Logistic Regression

```
Call:
roc.formula(formula = data$Z ~ predpr)


Data: predpr in 19 controls (data$Z 0) < 26 cases (data$Z 1).
Area under the curve: 0.7713
```

Figure 6. Area Under the Curve

```
Hosmer and Lemeshow goodness of fit (GOF) test


data:  mod$y, fitted(mod)
X-squared = 8.0694, df = 8, p-value = 0.4267
```

Figure 7. Hosmer and Lemeshow Goodness of Fit

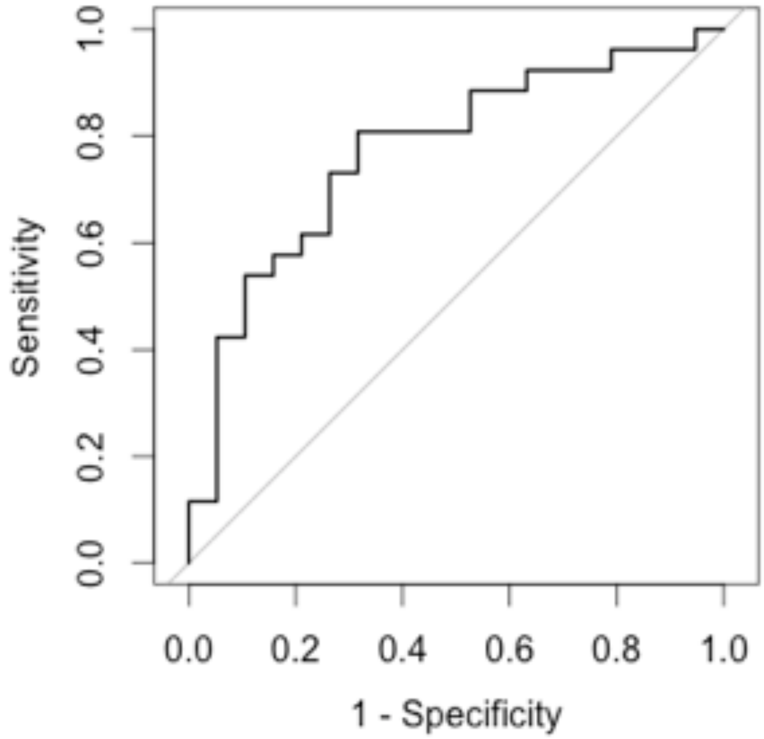

Figure 8. The Simulated ROC Curve Diagram

Conclusion

The paper reveals that the researchers make use of binary logistic regression analysis of CT-based measurements to establish the prognostic model. Although the study has the limitation of retrospection, the combination of CT-derived measurements is depicted to have greater specificity and sensitivity than single-based CT parameters. Based on the test on the group of 97 patients, the model indicates its reliability in assessing RVD based on the acute PE. The suggested model seems to be a promising indicator of RVD in this group of patients. The feasibility would be evaluated more precisely if a larger sample with PE could be obtained.

References

American Heart Association (2006). *Right Ventricular Function and Failure*. Retrieved March 25, 2016 from http://circ.ahajournals.org/content/114/17/1883.full

G. P., MD. (2008). Pulmonary Embolism in Heart Failure. Retrieved April 08, 2016, from http://circ.ahajournals.org/content/118/15/1598.full