# **APPLYING STATISTICAL METHODS IN A REGISTRY DATASET OF CARDIOPULMONARY RESUSCITATION TO PREDICT PROBABILITY OF SURVIVAL BY CHEST COMPRESSION TIME IN CHILDREN**

by

### **Hsin-Hui Huang**

MPH, Yale University, 2002

MD, Kaohsiung Medical College, Taiwan, 1986

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

#### UNIVERSITY OF PITTSBURGH

### GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

### **Hsin-Hui Huang**

It was defended on

**July 28, 2014**

and approved by

### **Thesis Advisor:**

Lisa Weissfeld, PhD, MA Professor **Biostatistics** Graduate School of Public Health University of Pittsburgh

#### **Committee Member:**

(Joyce) Chung-Chou H. Chang, PhD Professor **Biostatistics** Graduate School of Public Health University of Pittsburgh

#### **Committee Member:**

Marnie Bertolet, PhD Assistant Professor Epidemiology Graduate School of Public Health University of Pittsburgh

Copyright © by Hsin-Hui Huang

2014

# **APPLYING STATISTICAL METHODS IN A REGISTRY DATASET OF CARDIOPULMONARY RESUSCITATION TO PREDICT PROBABILITY OF SURVIVAL BY CHEST COMPRESSION TIME IN CHILDREN**

Hsin-Hui Huang, MS

University of Pittsburgh, 2014

#### **ABSTRACT**

The focus of this thesis was to apply advanced statistical methods to the American Heart Association Get With The Guidelines-Resuscitation (AHA GWTG-R) registry, a registry data set derived from a prospective multi-sites observational study, the American Heart Association's National Registry of Cardiopulmonary Resuscitation (NRCPR). The data comprise comprehensive information related to the cardiopulmonary resuscitation (CPR) process, patients' outcome, and characteristics of both the patients and the hospitals. The purpose of the registry data is to provide information that can be used to improve the outcomes of sudden cardiac arrest (SCA) patients and updates protocol of CPR.

This thesis has two purposes. The first one is to investigate the relationship between the patients' disease and survival for SCA patients receiving different durations of chest compression. The second one is to establish a model for predicting the probability of survival according to the duration of CPR. In the clinical setting, a categorized variable may provide more meaningful inferences. To explore this option, a Generalized additive model (GAM) was used to identify cutoff points for the categorization of chest compression duration. This categorized variable was then used for the development of prediction models for survival and the Net reclassification index (NRI) was used to select the appropriate predictors for this model.

iv

Logistic regression, generalized estimating equations (GEE), and a generalized linear mixed model (GLMM) were performed to obtain the estimates of parameters. Thereafter, the probability of survival was estimated based on the results of the regression model.

Comprehensive registry data have been established for many healthcare problems, which include many observations and variables. A systematic process to analyze registry data is necessary. This thesis used multiple statistical techniques to create meaningful variables, select appropriate predictors, fit regression models, and predict the probabilities of outcome. The public health significance of this thesis is the identification of subgroups of SCA patients who may benefit from prolonged CPR duration and to assess significance of cluster effects in the registry data.







### **LIST OF TABLES**



### **LIST OF FIGURES**



### **1.0 INTRODUCTION**

<span id="page-9-0"></span>Sudden cardiac arrest (SCA) is defined as sudden, unexpected loss of heart function, breath and consciousness. If appropriate treatment is not provided immediately, these patients cannot survive. The common causes of SCA include arrhythmia, valvular heart disease, coronary heart disease, and so on. The Centers for Disease Control and Prevention (CDC) estimated that 2,000 SCA deaths occur in the population of individuals that were younger than 25 years old [\(Kung,](#page-55-1)  [Hoyert, Xu and Murphy 2008\)](#page-55-1). The general incidence of SCA is hard to estimate. The reported incidence of cardiovascular-related, out-of-hospital cardiac arrest (OHCA) in children and adolescents in the US ranges between 0.61 and 1.44 per 100,000 pediatric person-years [\(Meyer,](#page-56-0)  [et al. 2012\)](#page-56-0). The age-adjusted risk of SCA was higher in athletic young adults [\(Corrado, Basso,](#page-55-2)  [Rizzoli, Schiavon and Thiene 2003\)](#page-55-2). Furthermore, SCA accounted for 0.7% to 3% of pediatric hospital admission and up to 5.5% of pediatric intensive care unit (ICU) admissions [\(Reis,](#page-56-1)  [Nadkarni, Perondi, Grisi and Berg 2002\)](#page-56-1). Thus the costs of healthcare and rehabilitation following SCA lead to significant family, social and medical burden.

Survival rates have increased in the decades after an improved standard resuscitation protocol was implemented [\(Girotra, et al. 2013,](#page-55-3) [Sutton, et al. 2014\)](#page-56-2). Among the patients with SCA, children with cardiac disease had better survival, but those who with trauma had the worst outcome [\(Meert, et al. 2009\)](#page-55-4). The potential factors that are associated with survival of pediatric SCA patients have been investigated. A Japanese nationwide population-based study indicated that pediatric patients with OHCA can benefit from bystander cardiopulmonary resuscitation (CPR) and public access-automated external defibrillator [\(Akahane, et](#page-55-5) al. 2013). A prospective study based on the National Registry of Cardiopulmonary Resuscitation showed that patients

with bradycardia, which is arrhythmia with slow heart rate, were more likely to survive after CPR [\(Donoghue, et al.](#page-55-6) 2009). A prospective, multinational, observational study investigated factors that impact survival on hospital discharge for 502 in-hospital pediatric SCA patients. The study concluded that low developmental index, underlying diseases, such as cancer, longer compression duration, and more inotropic drug use were associated with mortality of these patients [\(Lopez-Herce, et al. 2013\)](#page-55-7).

Though longer chest compression duration is considered an indicator of a poor survival outcome for SCA patients, some pediatric patients may benefit from prolonged chest compression and their characteristics were not thoroughly studied. More medical staffs and drugs are required for prolonged CPR. In order to balance the medical cost and survival benefits of prolonged CPR for SCA patients, it is important to identify specific characteristics of SCA patients whose chance of survival is increased with prolonged chest compressions. This study aimed to identify the disease categories of the SCA patients who can benefit from prolonged CPR duration and predict the probability of survival based on CPR duration for each disease category. I will explore the relationship of the illness categories, chest compression duration, and their survival outcomes by using the AHA GWTG-R data (see below). The following process illustrates the statistical methods used for analyzing a registry data set to establish a model to predict survival of pediatric SCA patients.

2

#### **1.1 RESUSCITATION DATA**

<span id="page-11-0"></span>The American Heart Association Get With The Guidelines-Resuscitation (AHA GWTG-R) is derived from the American Heart Association's National Registry of Cardiopulmonary Resuscitation (NRCPR). NRCPR is a prospective, multi-site, observational study starting in 1999 and was incorporated into GWTG in 2010. This is an ongoing study. The primary aim of the study is quality improvement of the CPR protocol, so that more lives can be saved through appropriate CPR procedure. The program collects resuscitation data from participating hospitals and then provides these hospitals with feedback on their resuscitation practice and outcomes of patients. Furthermore, new evidence-based guidelines can be developed from the data (www.heart.org/resuscitation) [\(Peberdy, et al. 2003\)](#page-56-3).

#### **1.2 OBJECTIVES**

#### <span id="page-11-2"></span><span id="page-11-1"></span>**1.2.1 Survival outcome**

The study aimed to investigate the relationship between illness categories of pediatric SCA patients and their outcomes on hospital discharge in each chest compression duration group. Furthermore, this study also determined the characteristics of patients and CPR factors that can predict the probability of survival.

### **2.0 METHODOLOGY**

### **2.1 DESIGN**

<span id="page-12-1"></span><span id="page-12-0"></span>The AHA GWTG-R is a prospective, multicenter registry of in-hospital cardiac arrest (IHCA) and resuscitation events using Utstein-style data reporting [\(Cummins, et al. 1997,](#page-55-8) [Jacobs,](#page-55-9)  [et al. 2004\)](#page-55-9). This study included subjects registered in 328 US and Canadian hospitals from January  $1<sup>st</sup>$ , 2000 through December 31 $<sup>st</sup>$ , 2009.</sup>

#### **2.2 SUBJECTS (INCLUSION/EXCLUSION CRITERIA)**

<span id="page-12-2"></span>All subjects <18 years of age with pulseless IHCA events were included in this study. The subjects must accept at least 1 minute of chest compression. Patients experiencing the events in hospitals and in the other locations (outpatient clinics within the hospital, visitors, and inpatients of rehabilitation, skilled nursing, and mental health facilities attached to study hospitals) were included. The subjects with events that happened outside of the hospital or in the neonatal intensive care unit (NICU), delivery room, or nursery were excluded. The subjects with the variable illness categories of newborn, obstetric, or other illnesses were excluded, too. If the subjects received more than 180 minutes of chest compression, the chest compression duration was winsorized at a pre-determined maximum of 180 minutes to reduce the effects of the possible extreme outliers.

#### **2.3 MEASURES**

<span id="page-13-0"></span>Index events indicated the first cardiopulmonary arrest event during the patient's hospitalization. The illness categories were defined by the characteristics of the patient at the time of cardiopulmonary arrest. General medical condition indicated a non-cardiovascular medical illness. Medical cardiac patients had a primary diagnosis of a cardiovascular medical illness. General surgical patients were enrolled at preoperative status with a general surgical illness or at a postoperative status after non-cardiovascular surgery. Surgical cardiac condition indicated a postoperative status after cardiovascular surgery. Trauma patients were subjects experiencing single or multiple injuries. Patients with "do not attempt resuscitation" before their first IHCA were excluded.

The primary outcome was survival on hospital discharge. Compared to continuous variables, categorized variables may be more practical in the clinical setting. Therefore, the chest compression duration was categorized based on the results of generalized additive model (GAM).

#### **2.4 DATA ANALYSIS**

#### <span id="page-14-1"></span><span id="page-14-0"></span>**2.4.1 Generalized Additive Model (GAM)**

A generalized additive model (GAM) was used to determine transformation of the chest compression duration variable. This method extends the usual likelihood-based regression models and develops its estimation. GAM assumes that the mean of the dependent variable depends on an additive predictor through a nonlinear link function. For a linear model with covariates,  $X_1, X_2, \cdots, X_n$ , the linear function,  $\sum_{i=1}^{p} \beta_i$  $_1^p$   $\beta_i$ X<sub>i</sub>, of the likelihood-based regression model is replaced by an smooth function,  $\sum_1^p s_i(X_i)$  $\frac{p}{1} s_i(X_i)$ , that defines the additive component.

For a generalized linear model (GLM),  $g\{\mu\} = \beta_0 + \sum_{i=1}^p \beta_i(X_i)$  $_{i=1}^{p} \beta_i(X_i)$ , where  $\mu$  is the conditional expectation of Y given  $x_1, \dots, x_n$ . Therefore, GAM is  $g\{\mu\} = \eta = s_0 + \sum_{i=1}^p s_i(X_i)$  $_{i=1}^{p} s_i(X_i)$ .

The *local scoring algorithm* and the *weighted backfitting algorithm* are used to estimate the  $s_i(\cdot)$ 's. The algorithms find new estimates of the functions by smoothing the partial residuals till the partial functions converge. The local scoring algorithm is used when the dependent variables are categorical, and the backfitting algorithm is used for the model with continuous dependent variables. Any nonparametric smoothing method, such as lowess and B-spline, can be used to estimate the  $s_i(\cdot)$ 's. This procedure can reduce a multiple regression to a series of twodimensional partial regression problems. The results can be plotted on a two dimensional graph to show the partial effects of each  $X_i$  on Y [\(Hastie and Tibshirani 1986\)](#page-55-10).

The GAM provides a nonparametric method to see the relationship between the predictors and the outcome. In a clinical setting, categorizing some continuous variables may be more

applicable, especially for predicting clinical outcomes. Therefore, information obtained from GAM can be used to determine appropriate cutoff points based on the data. In GAM, the relationship between  $X$ 's, the horizontal axis, and  $s(X)$ , the vertical axis, can be plotted. The line for  $s(X) = 0$  indicates the average value of the covariate. The average-risk cutoff points are those  $x$ 's  $\in X$ , such that  $s(x)=0$ . Furthermore, the points where the slopes change are the extra cut-off points. The selection of the extra cut-off points are based on the graphical visualization of the slopes and the clinical significance [\(Barrio, Arostegui, Quintana and Group 2013\)](#page-55-11).

In this analysis, only one predictor, a continuous variable of chest compression duration, was used in the GAM. And the survival status on discharge was the dependent variable. Categorization of the continuous chest compression duration variable was determined by the GAM method and expert opinions.

#### <span id="page-16-0"></span>**2.4.2 Net Reclassification Index (NRI)**

The AHA GWTG-R includes variables recording clinical and administrative data in each resuscitation event. The purpose of this study is to predict the probability of survival according to the characteristics of subjects. In order to determine the appropriate covariates for the models, the net reclassification index (NRI) was used. This method quantifies whether a new independent variable can provide a clinically relevant improvement in the prediction of the dependent variable. The model with established predictors is indicated as the "old" model. The model with one additional new predictor is denoted as the "new" model. The NRI is estimated by the following equations:

Event NRI: *NRI*<sub>*e*</sub> = *P* (*up*|*event*) – *P* (*down*|*event*), where

{ "up": the new risk model places a person into a higher risk category than the old model "down": the new model places a person into a lower risk category "event":  $cases - persons who survive$ "noevent": controls – persons who die

Non-event NRI: *NRIne= P (up|noevent) – P (down|noevent)*

An NRI is the sum of event NRI and non-event NRI, yielding

*NRI= P (up|event) - P (down|event) + P (down|nonevent) - P (up|nonevent).* 

If the new predictor increases the predicted risk for an event and decreases the predicted risk for a non-event, *P (up|event)* and *P (down|event)* provide the positive components of the NRI. However, the risk of the event moving down and the risk of the non-event moving up indicate that the new predictor compromises the prediction ability of the model [\(Pencina, D'Agostino,](#page-56-4)  [D'Agostino and Vasan 2008\)](#page-56-4).

The potential predictors were selected based on the results of descriptive statistics and opinions of experts, the emergent physicians. The most important predictors, determined by physicians, formed the basic model. A new predictor was added into the basic model, and NRI was performed for the new predictor. If the new predictor significantly increased the prediction power of the survival outcome, the new predictor was incorporated into the old model and formed the new basic model. If the new predictor could not improve prediction of the survival outcome, the predictor was dropped. The chosen predictors were then used for further analyses.

#### <span id="page-18-0"></span>**2.4.3 Logistic Regression Model**

The outcome is a binary variable and logistic regression is the approach of choice in this setting. One goal of the study was to model the conditional probability  $P(Y=1 | X=x)$  as a function of *x*. One assumption of a logistic regression model is the independence of every subject or observation. The unknown parameters in the function were estimated by maximum likelihood.

The general form of logistic regression model is

$$
logit(P[Y_i = 1 | X_i]) = ln\left(\frac{p(X_i)}{1 - p(X_i)}\right) = X_i \beta, \text{ where } \begin{cases} \beta(parameter): \beta_0, \beta_1, \beta_2, \cdots, \beta_p \\ X(variable): 1, x_1, x_2, \cdots, x_p \\ t(subject): 1, 2, \cdots, n \end{cases}
$$

.

Solving for  $p(X_i)$ , the probability of the event occurring, the result is

$$
p(X_i) = \frac{e^{X_i \beta}}{1 + e^{X_i \beta}} = \frac{1}{1 + e^{-(X_i \beta)}}.
$$

In this analysis, the major outcome was survival of the subjects. The first purpose was to test whether patients in some illness categories had higher odds of survival in a specific compression duration category. The hypothesis was that the odds of survival are not the same for patients with different illness categories and in different chest compression duration categories. In other words, compression duration categories were the effect modifier of the relationship between illness categories and survival. The interaction of illness and compression duration categories was included in the regression model.

The second purpose was to predict the probability of survival for patients with different illness categories based on the chest compression duration. This predictive model was established based on the results of the NRI. After the values of the parameters were estimated, the probability of survival for chest compression duration from 0 minute through 180 minutes was calculated. The average values of the other covariates were used in calculating the probability of survival. The prediction indicated the probability of survival for an average patient receiving chest compression from 0 through 180 minutes.

#### <span id="page-20-0"></span>**2.4.4 Generalized Estimating Equation Model (GEE)**

Though the procedure of CPR has been established and standardized, the survival of SCA subjects may depend on the quality of critical care within the hospitals. Therefore, the survival of subjects within the same hospital may be correlated. To obtain the population average estimates by considering the correlated data within each hospital, a generalized estimating equation (GEE) model was used [\(Zeger and Liang 1986\)](#page-56-5). The outcome and covariates were the same as those in the logistic model.

The GEE model: 
$$
logit(P[Y_{ij} = 1 | X_{ij}]) = X_{ij} \beta, \{i: 1 - N \text{ clusters}, j: 1 - J \text{ subjects}\}
$$

Marginal mean (population-average mean):  $\mu_{ij} = E(Y_{ij} | X_{ij}) = P(Y_{ij} = 1 | X_{ij})$ 

Variance-covariance matrix for correlated data (N clusters and J observation per cluster):

$$
V = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_N \end{pmatrix}
$$
  
= 
$$
\begin{pmatrix} Cov(Y_{1j}, Y_{1k}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Cov(Y_{Nj}, Y_{Nk}) \end{pmatrix}, j, k = 1, ..., J
$$
  
= 
$$
\begin{pmatrix} \emptyset A_1^{1/2} R_1 A_1^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \emptyset A_N^{1/2} R_N A_N^{1/2} \end{pmatrix}
$$

 $V=V(Y)$ : a  $N \times N$  variance-covariance matrix of the dependent variable

*A*<sup>*i*</sup>: a *J*×*J* diagonal matrix with *V*( $\mu$ <sup>*ij*</sup>) as the *j*th diagonal element

*Ri*: a *J*×*J* working correlation matrix

: a overdispersion parameter

Working variance-covariance matrix for  $Y_i$  is equal to  $V_i = \phi A_i^{1/2} R_i A_i^1$ 

Common working correlation structure:

Independence, where  $R_{jk}=0$ 

$$
R_i = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}
$$

Exchangeable, where *Rjk*=ρ

$$
R_i = \begin{bmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{bmatrix}
$$

Autoregressive, AR(1), where  $R_{jk} = \rho^{|j-k|}$ 

$$
R_i = \begin{bmatrix} 1 & \cdots & \rho^{J-1} \\ \vdots & \ddots & \vdots \\ \rho^{J-1} & \cdots & 1 \end{bmatrix}
$$

Unstructured, where  $R_{jk} = \rho_{jk}$ 

$$
R_i = \begin{bmatrix} 1 & \cdots & \rho_{1,j} \\ \vdots & \ddots & \vdots \\ \rho_{J,1} & \cdots & 1 \end{bmatrix}
$$

The major assumption of the GEE model is that the data satisfy the missing completely at random (MCAR) assumption. The registry data were established by the trained coordinators who record all required variables during resuscitation and the following clinical results. The data collection procedure was consistent within each hospital. Therefore, the missing values were not

likely to be associated with covariates and outcomes. For the purposes of this analysis, we assume that the MCAR assumption was not violated.

The purpose of the analysis was to compare the GEE model to the logistic regression model, so the covariates in the models were the same. Quasilikelihood under the independence model criterion (QIC) was used to determine the appropriate working correlation structure [\(Pan 2001\)](#page-56-6).

#### <span id="page-23-0"></span>**2.4.5 Generalized Linear Mixed Model (GLMM)**

The third model is the generalized linear mixed model (GLMM). Contrary to GEE that estimates a population-average mean, GLMM estimates a conditional mean [\(Williams 1982\)](#page-56-7).

The general form of the model is

$$
\overline{\mathbf{\hat{y}}}^{N\times 1} = \overline{\mathbf{\hat{X}}\mathbf{\hat{B}}_{N\times p} \mathbf{\hat{p}}\mathbf{\hat{z}}} + \overline{\mathbf{\hat{Z}}\mathbf{\hat{y}}_{N\times q} \mathbf{\hat{q}}\mathbf{\hat{z}}}^{N\times 1} + \overline{\mathbf{\hat{z}}}^{N\times 1}.
$$

$$
logit(P[Y = 1|X]) = ln\left(\frac{p(X)}{1-p(X)}\right) = X\beta + Z\gamma,
$$

 $N=\sum_{i=1}^q n_i$  $y_{j=1}^{q} n_j$ , where *j*: *1-q* clusters

 $\gamma \sim N(0, G)$ , where  $G = \sigma_{int}^2$  for a random intercept.

GLMM uses a logistic link function for a binary outcome.

This registry data set was established through participating hospitals, so this study can be considered as a clustered study design. A random intercept effect was applied to this model. The covariates in GLMM were the same as those in the logistic regression model.

### **3.0 RESULTS**

### **3.1 DEMOGRAPHIC CHARACTERISTICS**

<span id="page-24-1"></span><span id="page-24-0"></span>A total of 2,564 subjects were eligible in this study. Figure 1 showed the distribution of chest compression duration. More than 50% of the subjects received less than 25 minutes of chest compression. Table 1 showed the demographic characteristics of the subjects. More than 40% of patients were in the general medical group. Only 194 subjects were in a general surgical group. Overall, 40% of the subjects received chest compression for less than 16 minutes. For subjects with general surgical condition or with trauma, almost half of them had less than 16 minutes of chest compression. Compared to the other patients, a higher proportion of subjects with cardiac diseases accepted longer chest compression duration. Nearly half of the patients with a surgical cardiac condition were younger than 1 month; however, more than 50% of trauma patients were 8 years of age and older. Gender was not significantly different across all illness groups. Most of the SCA patients were inpatients of healthcare facilities. Most SCA events happened in the intensive care unit (ICU). In all illness categories, patients mainly presented hypotension or hypoperfusion. Most patients required invasive airway establishment and mechanical ventilator support, especially for those patients with a surgical or traumatic condition. Twenty-three percent of patients with a surgical cardiac condition needed a pacemaker. Preexisting respiratory insufficiency and hypotension/hypoperfusion were common across all illness groups. Otherwise, incidence of other pre-existing health conditions was low. More than 40% of traumatic patients were sent to the emergency department (ED) during the weekend. For patients with a general medical or a general surgical condition, their first pulseless rhythm was mainly presented as asystole; however, the trauma and surgical cardiac patients mainly experienced pulseless electrical activity (PEA). More than 80% of all patients used epinephrine, but other

16

vasopressors were not administered for most patients. More patients with a surgical cardiac condition required invasive procedures, such as an invasive airway insertion, pacemaker, and continuous sedation.

More than 70% of SCA patients died after CPR. Table 2 showed the outcomes of patients with individual illness categories. Patients with surgical condition had higher survival rates, but only 10% of trauma patients survived. Table 3 showed the survival rate of SCA patients in each chest compression duration category. Forty-six percent of the patients receiving less than 16 minutes of chest compression duration survived. The survival rates were similar for patients receiving 16-35 minutes and longer than 35 minutes of chest compression duration.



<span id="page-26-0"></span>**Figure 1. Distribution of Chest Compression Duration**



<span id="page-27-0"></span>





















## <span id="page-37-0"></span>**Table 2. Outcome for each illness category**

### <span id="page-37-1"></span>**Table 3. Outcome for each chest compression category**



### **3.2 GAM ANALYSIS**

<span id="page-38-0"></span>A GAM analysis was performed with survival status as the dependent variable and continuous chest compression duration (minutes) as the independent variable. The model used the spline smoother. Figure 2 showed the graph of chest compression duration and the smooth function. The curve of the smooth function approached zero at about 15 minutes and 65 minutes of chest compression duration. The slope of the curve changed at about 35 minutes and 140 minutes of compression duration. Large variation of the smooth function estimate after 50 minutes of chest compression indicated that the outcome varied as the compression duration increased and that the number of subjects may be very small. Therefore, the cutoff points after 50 minutes of chest compression were ignored. The categories of the compression duration were determined as 0-15, 16-35, and more than 35 minutes.



<span id="page-39-0"></span>**Figure 2. Generalized Additive Model based on chest compression duration of all subjects**

#### **3.3 NRI ANALYSIS**

<span id="page-40-0"></span>The net reclassification index (NRI) was used to determine the independent covariates for the model to predict the survival outcome. The base model was determined by expert opinions. The chosen variables were illness categories, chest compression duration, pulse rhythm on visit, age groups, event location, weekend on visit, daytime on visit, bypass procedure, calcium gluconate injection, previous septicemia episode, previous renal insufficient condition, and use of continuous vasoactive agents. All the above variables must be in the predictive model.

Table 4 showed the NRI test results. The candidate variables were selected based on the expert opinions and the results of descriptive statistics in table 1. The variables being statistically different across the illness categories were tested by NRI. Only variables with p-value <0.05 and with positive NRI, indicating that they provided positive prediction toward survival, were included in the final model. The variable of prior major trauma was excluded from the final model because it was closely related to one of the illness categories, "Trauma". Excluding this variable also avoided the problem of collinearity. In addition to the base model, five more variables were added into the final model. These five variables were use of sodium bicarbonate, prior cardiopulmonary arrest, use of apnea monitor, use of pulse monitor, and pre-existing condition of hypotension/hypoperfusion.



## <span id="page-41-0"></span>**Table 4. NRI and test results for all possible covariates**



#### <span id="page-43-0"></span>**3.4 REGRESSION COEFFICIENTS FROM THE THREE MODELS**

This study aimed to evaluate the association of illness categories with survival across different compression duration groups. The hypothesis was that the relationship of illness and survival was modified by compression duration. Interaction of compression duration groups and illness categories was included in the regression model. Though the type 3 analysis of effects for the interaction term was not statistically significant, several individual levels of the interaction term were statistically significant. Therefore, the interaction term was kept in the model.

Logistic regression, generalized estimating equation (GEE), and generalized linear mixed model (GLMM) were used for the analysis of all eligible subjects. Four different correlation structures, including independent, exchangeable, autoregressive, and unstructured, working correlation structures were used in the GEE model. Quasilikelihood under the independence model criterion (QIC) was used to select the most appropriate working correlation structure. Table 5 showed the QIC values of the models with the four working correlation structures. Though the model with the first order of autoregressive working correlation structure had the smallest QIC, the QIC values for all correlation structures were similar. The exchangeable correlation structure is more appropriate for the characteristics of the data. Therefore, the coefficients in table 6 and odds ratios in table 7 were obtained from the GEE model with exchangeable working correlation structure.



#### <span id="page-44-0"></span>**Table 5. QIC for different working correlation structure**

Table 6 summarized the estimates of parameters from the logistic regression, GEE model, and GLMM model. In general, the estimates of parameters were similar among the three models. Though the standard deviation of most estimates was slightly higher in the GEE and GLMM, the conclusion derived from the three models are similar. In the regression model, the type 3 analysis of effects showed significant effects for most covariates, except for time and weekend of events. However, the expert opinion preferred keeping them in the final model. The two major interests of the study, compression duration groups and illness categories, were significantly related to survival of the subjects. When inspecting the individual covariates, the odds of survival were not significantly different between the age group of 1 -8 years and the age group of younger than 1 month for the logistic model and GLMM. However, GEE showed significant different survival odds between the two groups. On the contrary, the odds of survival for the subjects experiencing SCA in the procedure room was significantly different from the odds of survival for the subjects experiencing SCA in the ICU for the logistic model and GLMM, but not for the GEE model.

<span id="page-45-0"></span>

**Table 6. Coefficients of covariates in logistic regression, GEE, and Mixed model (random intercept)**





### **3.5 ODDS RATIOS**

<span id="page-48-0"></span>Table 7 showed the odds ratios of the other illness categories vs. the general medical condition in each compression duration group. In the group of chest compression less than 16 minutes, patients with medical cardiac condition and with surgical cardiac condition had higher odds of survival compared to subjects with general medical condition. On the contrary, subjects with trauma were less likely to survive. In the group of chest compression duration between 16 and 35 minutes, only subjects with a surgical cardiac condition had a better odds to survive compared to the subjects with a general medical condition. And the subjects with trauma still had the worse outcome. In the group of chest compression longer than 35 minutes, the subjects with a medical cardiac or with a surgical cardiac condition had a better outcome. Similar results can be obtained from GEE and GLMM models. In the group of chest compression less than 16 minutes, the odds of survival for the subjects with medical cardiac condition and for the subjects with general medical condition were not statistically different in the GLMM model; however, the odds ratios and their 95% confidence intervals were greater than 1 in the logistic model and GEE model. Comparing the estimates and the 95% confidence intervals of the results in the three models, they were still very similar and the lower limit of the confidence interval derived from GLMM was slightly smaller than 1. Therefore, we still concluded that the results from the three models were not significantly different.

### **Table 7. Odds ratios of illness categories for survival within each compression duration category for logistic regression, GEE, and mixed model**

<span id="page-49-0"></span>

#### **3.6 PREDICTION MODEL**

<span id="page-50-0"></span>The estimates of parameters were similar in logistic regression, GEE, and GLMM. To simplify the process of prediction, the predicted probability of survival was estimated based on the results of logistic regression. All variables selected by NRI were included in the model. Interaction of continuous chest compression duration and illness categories was included in the model. The average value of each covariate was used in the model for the prediction of the probability of survival.

Figure 3 showed the predicted probability of survival for each illness category from 0 minutes through 180 minutes of chest compression. Patients with a surgical cardiac condition had the best probability of survival. Within 10 minutes of chest compression, the survival probability for this group of patients was up to 50%. The probabilities of survival were similar in the beginning for subjects with a medical cardiac and a general medical condition. However, the probability dropped faster for the subjects with a general medical condition as the duration of chest compression increased. Though patients with a general surgical condition had a higher probability of survival in the beginning compared to those who with a general medical condition, the probabilities were tied after 70 minutes of chest compression. Both groups of patients had a survival probability of less than 10%.



<span id="page-51-0"></span>**Figure 3. Predicated probability of survival for each illness category**

#### **4.0 DISCUSSION**

### **4.1 CONCLUSION**

#### <span id="page-52-2"></span><span id="page-52-1"></span><span id="page-52-0"></span>**4.1.1 Generalized Additive Model (GAM) analysis**

Though continuous variables may provide more information, they may not be practical in the clinical setting. A GAM can appropriately categorize continuous predictors based on the relationship of the predictors and the outcome through a nonparametric smoothing function. This method determined several cutoff points of chest compression duration for this data set. However, not all cutoff points should be used. The variance increased dramatically for longer chest compression duration because fewer cases received such a long duration of chest compression. The slope of the curve did not change dramatically after 35 minutes of chest compression duration. Therefore, the cutoff points after 35 minutes were not used in this study.

#### <span id="page-52-3"></span>**4.1.2 Net Reclassification Index (NRI) analysis**

Many characteristics of the patients were significantly different across the disease groups (Table 1). Though they can be included in the predictive model, the problems of over-fitting and computational demand may compromise the feasibility of the models. Also, the final model with too many predictors is not practical in the clinical setting. Using the NRI to determine the most appropriate variables for prediction has become popular recently. This method can specifically select the variables predicting the desired outcome, but not predicting the opposite outcome. Therefore, the model determined by NRI can be more precise.

# <span id="page-53-0"></span>**4.1.3 Logistic regression, Generalized Estimating Equation (GEE), and Generalized Linear Mixed Model (GLMM) analyses**

This study showed that estimates of parameters obtained from a logistic regression model, a GEE model, and a GLMM were similar and the conclusion derived from the three modeling methods was the same. The hospital based registry data were considered as a clustered observational study design. The subjects within the same cluster were correlated because the clinical practice and patients' characteristics may be similar within a specific cluster, but different from the other clusters. Both GEE and GLMM can deal with correlated data. However, the benefits of the two complicated methods were not confirmed in this study.

Several factors may undermine the performance of GEE and GLMM. First, the sizes of clusters were diverse. Some clusters had hundreds of patients, but small sized clusters were more common in the data set. Therefore, subjects may be independent. Second, survival of SCA may mainly depend on appropriate CPR procedure and characteristics of subjects, but not on characteristics related to the hospitals. Therefore, the hospital effect is limited. Random effects may not be necessary to analyze data related to SCA patients.

The methods for correlated data may not be necessary for problems that are mainly related to characteristics of patients but not depend on features of clusters.

45

### **4.2 PUBLIC HEALTH ASPECT**

<span id="page-54-0"></span>The registry data that include comprehensive variables have been established for many healthcare problems. In order to explore the data appropriately and to derive clinically applicable inferences, a systematic process is necessary. This study demonstrated the process used to analyze hospital based registry data of SCA episodes.

The process involves the categorization of observations for clinical application, the establishment of an appropriate model to predict outcome, and the computation of estimates through methods that take correlated data into consideration. The similar results from the three methods in this study may be caused by the specific entity of the health problem. Therefore, using estimates from logistic regression to predict probability of survival is appropriate. However, for the other health problems, it may be still necessary to consider the random effect related to the different hospitals.

### **BIBLIOGRAPHY**

- <span id="page-55-5"></span><span id="page-55-0"></span>Akahane, M., et al. (2013), "Characteristics and Outcomes of Pediatric out-of-Hospital Cardiac Arrest by Scholastic Age Category," *Pediatr Crit Care Med*, 14, 130-136.
- <span id="page-55-11"></span>Barrio, I., Arostegui, I., Quintana, J. M., and Group, I. C. (2013), "Use of Generalised Additive Models to Categorise Continuous Variables in Clinical Prediction," *BMC Med Res Methodol*, 13, 83.
- <span id="page-55-2"></span>Corrado, D., Basso, C., Rizzoli, G., Schiavon, M., and Thiene, G. (2003), "Does Sports Activity Enhance the Risk of Sudden Death in Adolescents and Young Adults?," *J Am Coll Cardiol*, 42, 1959-1963.
- <span id="page-55-8"></span>Cummins, R. O., et al. (1997), "Recommended Guidelines for Reviewing, Reporting, and Conducting Research on in-Hospital Resuscitation: The in-Hospital 'Utstein Style'. A Statement for Healthcare Professionals from the American Heart Association, the European Resuscitation Council, the Heart and Stroke Foundation of Canada, the Australian Resuscitation Council, and the Resuscitation Councils of Southern Africa," *Resuscitation*, 34, 151-183.
- <span id="page-55-6"></span>Donoghue, A., et al. (2009), "Cardiopulmonary Resuscitation for Bradycardia with Poor Perfusion Versus Pulseless Cardiac Arrest," *Pediatrics*, 124, 1541-1548.
- <span id="page-55-3"></span>Girotra, S., et al. (2013), "Survival Trends in Pediatric in-Hospital Cardiac Arrests: An Analysis from Get with the Guidelines-Resuscitation," *Circ Cardiovasc Qual Outcomes*, 6, 42-49.
- <span id="page-55-10"></span>Hastie, T., and Tibshirani, R. (1986), "Generalized Additive Models," *Statistical Science*, 1, 297- 318.
- <span id="page-55-9"></span>Jacobs, I., et al. (2004), "Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update and Simplification of the Utstein Templates for Resuscitation Registries. A Statement for Healthcare Professionals from a Task Force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian Resuscitation Council, New Zealand Resuscitation Council, Heart and Stroke Foundation of Canada, Interamerican Heart Foundation, Resuscitation Council of Southern Africa)," *Resuscitation*, 63, 233-249.
- <span id="page-55-1"></span>Kung, H. C., Hoyert, D. L., Xu, J., and Murphy, S. L. (2008), "Deaths: Final Data for 2005," *Natl Vital Stat Rep*, 56, 1-120.
- <span id="page-55-7"></span>Lopez-Herce, J., et al. (2013), "Factors Associated with Mortality in Pediatric in-Hospital Cardiac Arrest: A Prospective Multicenter Multinational Observational Study," *Intensive Care Med*, 39, 309-318.
- <span id="page-55-4"></span>Meert, K. L., et al. (2009), "Multicenter Cohort Study of in-Hospital Pediatric Cardiac Arrest," *Pediatr Crit Care Med*, 10, 544-553.
- <span id="page-56-0"></span>Meyer, L., et al. (2012), "Incidence, Causes, and Survival Trends from Cardiovascular-Related Sudden Cardiac Arrest in Children and Young Adults 0 to 35 Years of Age: A 30-Year Review," *Circulation*, 126, 1363-1372.
- <span id="page-56-6"></span>Pan, W. (2001), "Akaike's Information Criterion in Generalized Estimating Equations," *Biometrics*, 57, 120-125.
- <span id="page-56-3"></span>Peberdy, M. A., et al. (2003), "Cardiopulmonary Resuscitation of Adults in the Hospital: A Report of 14720 Cardiac Arrests from the National Registry of Cardiopulmonary Resuscitation," *Resuscitation*, 58, 297-308.
- <span id="page-56-4"></span>Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., and Vasan, R. S. (2008), "Evaluating the Added Predictive Ability of a New Marker: From Area under the Roc Curve to Reclassification and Beyond," *Stat Med*, 27, 157-172; discussion 207-112.
- <span id="page-56-1"></span>Reis, A. G., Nadkarni, V., Perondi, M. B., Grisi, S., and Berg, R. A. (2002), "A Prospective Investigation into the Epidemiology of in-Hospital Pediatric Cardiopulmonary Resuscitation Using the International Utstein Reporting Style," *Pediatrics*, 109, 200-209.
- <span id="page-56-2"></span>Sutton, R. M., et al. (2014), "2010 American Heart Association Recommended Compression Depths During Pediatric in-Hospital Resuscitations Are Associated with Survival," *Resuscitation*.
- <span id="page-56-7"></span>Williams, D. A. (1982), "Extra-Binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144-148.
- <span id="page-56-5"></span>Zeger, S. L., and Liang, K. Y. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics*, 42, 121-130.