BIG DATA ANALYSIS OF EMERGENCY MEDICAL SERVICE APPLIED TO DETERMINE THE SURVIVAL RATE EFFECTIVE FACTORS AND PREDICT THE AMBULANCE TIME VARIABLES

[1]JIANN-SHINGSHIEH, [2]YU-TING YEH, [3]WEI-ZEN SUN, [4]MATTHEW HUEI-MING MA, [5]CHUN-YI DAI, [6]MUAMMAR SADRAWI, [7]MAYSAM F. ABBOD

[1,2,6]Department of Mechanical Engineering and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Chung-Li 32003, Taiwan.
[3,5]Department of Anesthesiology, College of Medicine, National Taiwan University, Taipei, 100, Taiwan
[4]Department of Emergency Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[7]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UB8 3PH, UK
E-mail: [1]rul3xji4@gmail.com, [2]jsshieh@saturn.yzu.edu.tw, [3]wzsun@ntu.edu.tw, [4]matthew@ntu.edu.tw, [5]chunyi.dai@gmail.com, [6]muammarsadrawi@yahoo.com, [7]maysam.abbod@brunel.ac.uk,

Abstract - Emergency medical service (EMS) takesan important part in out-of-hospital cases,and it takes decisively effect to patients' mortality rate. There are five factors have been scrutinized in this paper with large database to determine the correlation and effectiveness to survival rate, and also the difference between urban and suburban area. Seven years from 2007 to 2013 emergency record have conducted in study. Via applying analysis of variance (ANOVA), age, gender, response time, on-scene time and transportation time were used to be the analysis condition in survival rate and urban/suburban difference. Furthermore, age, gender, population density and total ambulance number were used as inputs to predict time outputs of response time, on-scene time, transportation time with artificial neural network (ANN). There are significant differences in all five factors of 7 years analysis, with age have the highest correlation (Pearson = -0.059), and on-scene time second highest (Pearson = -0.033) to survival rate. For urban and suburban comparison, each city has the highest correlation with time factors, and transportation time has the highest among other time factors. For time prediction, the best model performsmean absolute error (MAE) of3.2675 minutes, and response time has the lowest error of2.2498 minutes. Observing the result, it is suggested that patient with male around or higher 65 years old should be more concern and urgency. Urban and suburban do affect the out-of-hospital internal time in the study because urban patient spend less time on transportation time but more on on-scene time, while suburban has the opposite trend. In experimental prediction, model built with 4 years database could made the prediction within 3.2 minutes in training city but unable to apply to different city as well.

Keywords - Emergency medical service, Response time, On-scene time, Transportation time, Artificial neural network.

I. INTRODUCTION

With the technology improvement, our medical system also makes progress. But for all purpose is to improve the quality of our medical procedure and enhance the survival rate. In urban accident, Emergency Medical Service (EMS) plays the important role in such high population density like Taiwan. Emergency Medical Service's paramedic usually had the initial contact and treatment to patient, and it takes huge effect to patients' survival rate. In emergency medical, there are a lot of factors that will affect the patients'survival rate, such as patients' age, gender, pre-hospital time, first aid process, injury type or bystander CPR etc.Due to Taiwan is a high population density country with multiple topographies, in addition to the globalization population aging trend, diversity topographies have made different effect to the result of emergency medical. The emergency medical system usually attaches to the fire department in Taiwan so ambulancesdepart from the fire station, arrive accident scene, then transfer to hospital as appropriate as possible.During the attendance, all relevant information will be recorded, including patient age, gender, departure time, reach scene time, leave scene time, arrival hospital time, location, trauma type, call reason etc. [1-3]. By building and evaluating the huge dataset, the medical care quality can be improved without additional resource and make appropriate adjustment based on current deployment.

Big data analysis and machine learning are popular and non-negligible subjects nowadays. Big data allow us to have enough resource for mining the pattern and convincing result and build up a prediction model. Via data mining and analysis, the data distribution, composition structure, and conclude a pattern can be figured out. In this research, 7 years ofTaiwan EMS data is used for the application database.

Data includingpatients' basic characteristics and duty information, several factors are selected as inputs based on previous research [4-8]. For prediction, data composition being more irregular non-linear characteristics, and due to the complex relationship of variables, it is difficult to use traditional mathematical equation. Hence, the prediction model is developed using artificial neural network (ANN). ANN is one of the most popular methods to deal with nonlinear and non-stationary problems. As the previous studies [9-10], it can be modeled the complex relationships

between input and output. Prediction model could be estimated the time that ambulance takes during duty.
In this study, thefocus is on patients' characteristic and response time, on- scene time, transportation time, discussing the data contribution and the statistical result, compare difference between Taiwan's each city and difference between urban and suburbanin three major cities. Furthermore, we are also trying to build a model to predict response time, on-scene time and transportation time. By estimating these times, the medical resource can be deployed appropriatelywhich can contribute to the ambulance driver's safety.

II. PATIENTS AND METHODS

2.1. Experimental data

The database had recorded 5,825,595 patients from emergency data according to emergency medical service in Taiwan. The data contain patients' characteristic information, case number, date, call reason, call received time, ambulance departure time, reach scene time, leave scene time, arrived hospital time, location, past history, andpossible OHCA cause. For data analysis, statistics software is used to analyze cluster median and correlation coefficient, in order to show effective factors.

2.2. Data pre-processing

In this study, total 5,825,595 cases of the year 2007 to 2013 have been collected. Although there are around 5.8 million cases in the database, each parameter need to be evaluated and cleaned,±3 standard deviation (99%CI) is used to filter the unreasonable and abnormaldata. So, the final database size is 4,366,971 cases.

After filtration, the age has the range between 1 to 113 years old, and gender has been labeled in 1 and 2 as male and female.Pre-hospital survival rate has been labeled in 1 as survival and 0 as non-survival. Time factors have been evaluated and transferred into three parameters: response time, on-scene time, transportation time. According to the data recorded, three parameters are calculated by ambulance departure time, reaching scene time, leaving scene time, and arrivinghospital time.

Response time is the interval between ambulance departure time and reaching scene time. On-scene time is the interval between reaching scene time and leaving time. Transportation time is the interval of leaving scene time and arriving hospital time. Total 4,332,870 survivals and 34,101 non-survival cases from database are analyzed. Table 1 shows the distribution of the gender proportion of the database.

Table 1: Database Distribution by Gender and Survival.

EMS Database After Filter (817,090 cases)		
	Survival	Non-survival
Male	2,498,538 (57.66%)	21,785 (63.9%)
Female	1,834,332 (42.33%)	12,316(36.1%)
Total	4,332,870 (100%)	34,101 (100%)

For urban and suburban comparison, three municipalitiesof Taiwan were selected: New Taipei city, Taoyuan city and Kaohsiung city based on population density and location. There are two municipalities located in north Taiwan and one in south Taiwan that can be comparedto north and south Taiwan cities.Regarding the population density in the urban and suburban, the area with population density that are higher than the mean +1 standard deviation for urban and the rest are considered to be suburban.Tables2-4 show the distribution of the gender proportion of the municipality data.

Table 2: New Taipei Database Distribution by Gender and Survival.

New Taipei City Database		
	Survival	Non-survival
Male	394,371 (57.42%)	2,629 (62.97%)
Female	292,351 (42.57%)	1,546(37.02%)
Total	686,722 (100%)	7336(100%)

Table 3: Taoyuan Database Distribution by Gender and Survival.

Taoyuan City Database		
	Survival	Non-survival
Male	233,214 (58.04%)	3,048 (65.15%)
Female	168,556 (41.95%)	1,630 (34.84%)
Total	401,770 (100%)	4,678 (100%)

Table 4: Kaohsiung Database Distribution by Gender and Survival

Kaohsiung City Database		
	Survival	Non-survival
Male	352,892 (56.72%)	4,568 (63.71%)
Female	269,951 (43.27%)	2,601 (36.28%)
Total	623,843 (100%)	7,169 (100%)

To train the ANN, Taoyuan cityis selected based on 4 years data from 2008 to 2011 of total 223,823 cases for building model and 2012's 70,753 data for external testing. However, in order to test the model, Kaohsiung city was selected based on 7 years data which includes 193,551 cases. The mean absolute error (MAE) is used to measure the prediction accuracy.

2.3. Methodology

To analyze big database, correlation coefficient analysis, t-test and analysis of variance (ANOVA) [11] have been applied. Via correlation coefficient test, the result can be observed each factors' correlation between the survival rates.By observing analysis of variance result as well as the significancebetween survival and non-survival data in

each parameter input, the extent of the various factors effect can be evaluated.

Artificial neural network have been applied to build the training model and testingusingMATLAB. In this study, our ANN training model was built in 4 inputs that contain age, gender, population density and total ambulance number in each area and the three outputs of response time, on-scene time and transportation time. The topology of ANN was 2 hidden layer, 10 and 15 nodes, and 1500 epochs training iteration. Training data of proportion for building model were set up in 80% training, 10% validation, and 10% testing.

III. RESULTS AND DISCUSSION

3.1. Taiwan EMS 7 year data analysis

Total 4,366,971 cases have been analyzed via IBM SPSS Statistic software. Based on age, gender, response time, on-scene time and transportation time, EMS data evaluated each factors and produce statistical result and correlation comparison table. The result shows median value with standard deviation and p-value for survival and non-survival. There is the significant division between survival and non-survival in age gender. Median age value is 49.79±23.1years old for survival and 65.26± 19.58 years old for non-survival ($p<0.01$). Gender result show that male has the higher proportion in non-survival cluster with median gender level value is 1.42 ± 0.49 for survival and 1.36 ± 0.48 for non-survival (<0.01). For time factors, response time and on-scene time both have the trend with lower time spent in survival cluster, but opposite in transportation time. The median response time (RT) value is 5.83± 3.7 minutes for survival and 6.12± 3.82 minutes for non-survival ($p<0.01$). Median on-scene time (OT) is 7.22 ± 5.24 minutes and 9.16± 5.4 for survival and non-survival($p<0.01$). Median transportation time (TT) value is 8.56 ± 6.94 and 7.773 ± 6.5 minutes for survival and non-survival ($p<0.01$).

Figs. 1-4 show the average of each factors associated with survival proportion. The relation curve is shown in Fig.1. It can be seen there is the significant trend in age, that survival rate has negative correlation with age. Survival ratio decrease with ageing, and the slope is getting relatively stable around 5 years old to 83 years old. Figs. 2 and 3 show the response time and on-scenetime average curve associated with survival proportion. Steady slope can be observed from first minute to around 17 minutes in response time, while on-scene slope has more rapidly decreased compared with the first 17 minutes. Transportation time shows the opposite trend curve to response time and on-scene time, survival ratio is increasing with time parameter.

Correlation coefficientshave shown in Table 5. Five factors with survival rate were cross-compared to each other in correlation table, and it shows the Pearson correlation coefficient, p-value and data size. Age has the highest correlation with -0.059 among 5 factors. The negative correlation coefficient represents the decrease slope of age associated with survival rate that opposite trend was shown. The second highest factor is on-scene time with -0.33, and response time is the last factor with correlation coefficient -0.007.

3.2. Three Municipalities data analysis associated with urban and suburban comparison

Five factors have been analyzed based on previous research and 7 years result. Age in New Taipei city with urban and suburbanis 48.84 ± 22.72 and 48.19 ± 23.04 years old ($p<0.01$). Median gender value ratio are 1.44 ± 0.496 and 1.41 ± 0.492 to urban and suburban ($p<0.01$). For time factors in urban and suburban analysis, response time are 5.61 ± 3.189 and 6.46 ± 4.464 minutes ($p<0.01$). On-scene time are 9.87 ± 6.92 and 8.83 ± 7.1 minutes ($p>0.05$). Transportation time are 7.01 ± 4.8 and 10 ± 7.45 ($p<0.01$).

Population age in Taoyuan city forurban and suburbanis46.76 ± 22.54 and 47.55 ± 23.43 years old ($p<0.01$). Median gender value ratio are 1.45 ± 0.497 and 1.41 ± 0.492 with urban and suburban ($p<0.01$). For the time factors in urban and suburban, response time are 6.13 ± 3.74 and 6.04 ± 4.32 minutes ($p<0.01$). On-scene time are 9.02 ± 5.6 and 7.96 ± 5.58 minutes ($p<0.01$). Transportation time are 5.58 ± 4.385 and 9.51 ± 8 minutes ($p<0.01$).

Population age in Kaohsiung city forurban and suburbanis 50.62 ± 22.73 and 50.52 ± 23 years old ($p<0.01$). Median gender value ratio are 1.45± 0.49 and 1.42 ± 0.49 ($p<0.01$). For time factors in urban and suburban, response time are 4.32 ± 3.41 and 5.36 ± 4.24 minutes ($p<0.01$). On-scene time are 8.08 ± 5.7 and 7.08 ± 5.7 minutes ($p<0.01$). Transportation time are 6.54 ± 6.159 and 9.58 ± 8.85 minutes ($p<0.01$).

Observing the correlation coefficient table in Tables 5-7, each city has the highest correlation with transportation time. For New Taipei City with 0.231, Taoyuan city with 0.198, and Kaohsiung city with 0.16. According to the coefficient, transportation time has the positive value with urban and suburban cluster. With the label that urban as 1 andsuburban as 2, it represents the transportation time increases in county cluster. Most of the time factors have higher correlation with age and gender in urban and suburban comparison.

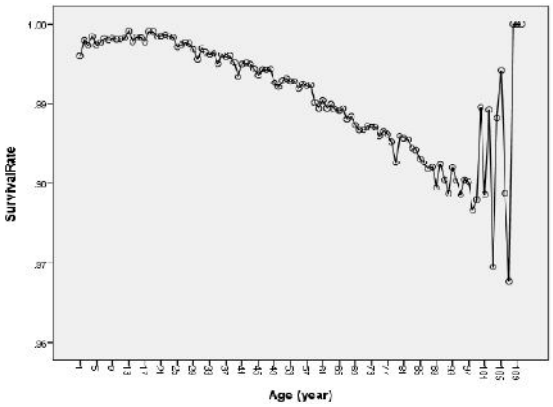

Fig.1. Age associated with survival rate.

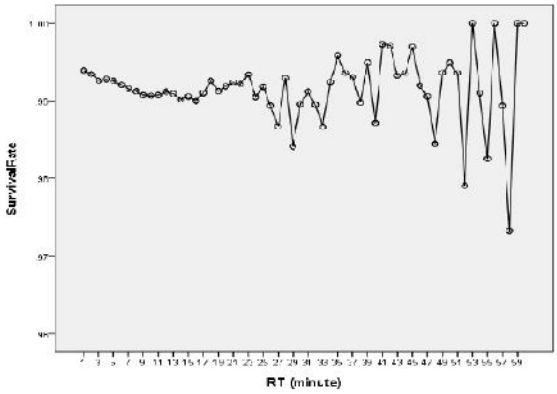

Fig.3. On-scene time associated with survival rate.

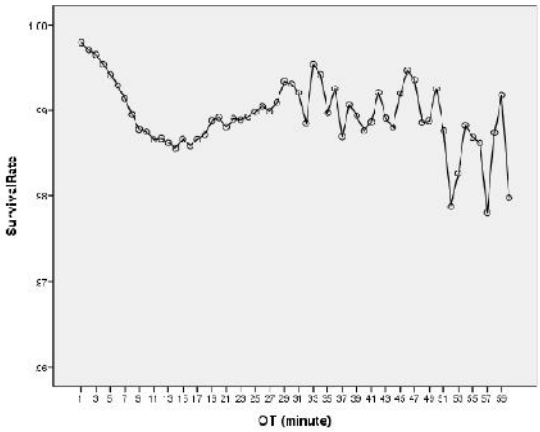

Fig.2. Response time associated with survival rate.

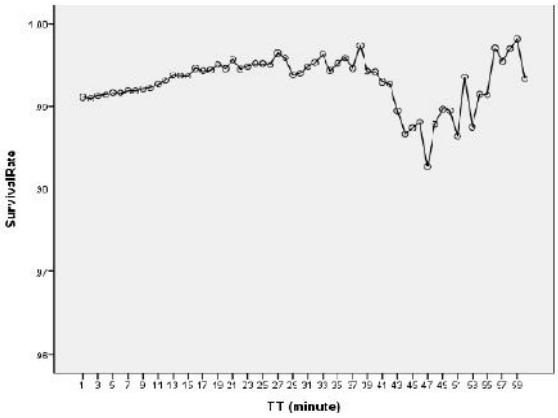

Fig.4. Transportation time associated with survival rate.

Table 5:Seven yearscorrelation coefficient between survival and 5 factors.

		Age	Gender	RT	OT	TT
Survival and Non-survival	PearsonCorrelation	-0.059*	0.011*	-0.07*	-0.033*	0.01*
	p-value	0.000	0.000	0.000	0.000	0.000
	Patient number	4,366,971	4,366,971	4,366,971	4,366,971	4,366,971
* The correlation is significant at the 0.01 level						

Table 6: New Taipei City 7 year correlation coefficient between urban and suburban and 5 factors.

		Age	Gender	RT	OT	TT
Urban and suburban	PearsonCorrelation	-0.014*	-0.026*	-0.108*	-0.073*	0.231*
	p-value	0.000	0.000	0.000	0.000	0.000
	Patient number	690,897	690,897	690,897	690,897	690,897
* The correlation is significant at the 0.01 level						

Table 7:Taoyuan City 7 year correlation coefficient between urban and suburban and 5 factors

		Age	Gender	RT	OT	TT
Urban and suburban	PearsonCorrelation	0.013*	-0.028*	-0.006*	-0.072*	0.198*
	p-value	0.000	0.000	0.000	0.000	0.000
	Patient number	406,448	406,448	406,448	406,448	406,448
* The correlation is significant at the 0.01 level						

Table 8:Kaohsiung City 7 year correlation coefficient between urban and suburban and 5 factors.

		Age	Gender	RT	OT	TT
Urban and suburban	PearsonCorrelation	-0.002	-0.027*	-0.095*	-0.08*	0.168*
	p-value	0.114	0.000	0.000	0.000	0.000
	Patient number	631,012	631,012	631,012	631,012	631,012
* The correlation is significant at the 0.01 level						

3.3. Artificial neural network model training

Total 4 years of 223,823 cases have been used in a training model and 70,753 cases for external testing. In addition to Taoyuan urban and suburban, the model is tested on Kaohsiung cityfor193,551 cases. Training for the best performance, the model mean MAE is 3.2675 minutes, response time MAE is2.2498 minutes, on-scene time MAE is 3.6654 minutes, and transportation time MAE is 3.9051 minutes. According to Taoyuan city's good performance results, the same is being applied to Kaohsiung city. For Kaohsiung city, the best model performance is mean MAE of 55.2911 min, response time MAE is44.0799, on-scene time MAE is 56.7244 minutes, and transportation time MAE is 65.069 minutes. The results show that the model may not be applied to another city.

CONCLUSIONS

In this study, an attempt is made to get more precisely on each factor actually effect to the patient, improving EMS quality with currently resource and building model to predict proper time to reach the efficient mission and care about drivers' safety. Getting further with each result, based on seven years statistical results with each parameter, age is significant divided into different cluster that mortality mean value is 65.26 ± 19.58 years old. Age around 65 years old or further should pay more attention.Observing the result, male has the higher proportion to female cluster in mortality rate [12].For time parameters, on-scene time takes highly correlation among other study parameters. Although response time does not show the higher correlation to mortality rate, but there is the study show that it would be a survival advantage to the unselected patient within 4 minutes [13].The transportation time result shows the opposite trend compared to other time factors. The relationship between survival and time are complex and diverse.

It isconsidered that time factor will also affect by urban or suburban, especially transportation time. To figure out the effect factors, the difference between urban and suburbanis investigated in three municipalities. Time factors have shown the higher correlation between urban and suburban in each city. Transportation time has the most highly correlation with the clustering, and it also shows that shorter on-scene time associated with longer transportation time[14]. Response time is lower in the urban result. It may consider by fire station density in urban area that the density is higher than suburban city and it reduces the time for response. For suburban area, on-scene time is much lower than urban and hashigher transportation time. Observed previous research[15], trauma type has the higher percentage call in suburban cities, and for trauma type patient, there is the research [16] to reduce on-scene time to improve survival rate.

Time factors are the only variables that can be adjusted in the study. The statistical results are utilized to measure the out-of-hospital internal time. Result shows that the experimental model can predict time parameters with around mean absolute error 3.2675 minutes but is not available to apply on Kaohsiung city with Taoyuan city's model. Different city might not be able to apply the same model, but need to establish their own.

Applying the big data, the critical factors can be analyzed to find the patients' survival rate and correlation between each parameter, and also help assigning resources properly and create maximum benefit on critical factors. In addition to assign resources, applying large data set for training model to predict out-of-hospital EMS internal time help us to evaluate proper time and can be considered to driver's safety. In future work, more variable condition could be measured by additional inputlike trauma type [17], bystander CPR [18] and also improve the performance of prediction.

REFERENCES

[1] Feero, S., Hedges, J. R., Simmons, E., & Irwin, L. (1995). Does out-of-hospital EMS time affect trauma survival?. The American journal of emergency medicine, 13(2), 133-135.

[2] Böttiger, B. W., Grabner, C., Bauer, H., Bode, C., Weber, T., Motsch, J., & Martin, E. (1999). Long term outcome after out-of-hospital cardiac arrest with physician staffed emergency medical services: the Utstein style applied to a midsized urban/suburban area. Heart, 82(6), 674-679.

[3] Hsia, R. Y., Kanzaria, H. K., Srebotnjak, T., Maselli, J., McCulloch, C., &Auerbach, A. D. (2012). Is emergency department closure resulting in increased distance to the nearest emergency department associated with increased inpatient mortality?. Annals of emergency medicine, 60(6), 707-715.

[4] Blackwell, T. H., & Kaufman, J. S. (2002). Response time effectiveness: comparison of response time and survival in an urban emergency medical services system. Academic Emergency Medicine, 9(4), 288-295.

[5] Spaite, D. W., Valenzuela, T. D., Meislin, H. W., Criss, E. A., &Hinsberg, P. (1993). Prospective validation of a new model for evaluating emergency medical services systems by in-

field observation of specific time intervals in prehospital care. Annals of emergency medicine, 22(4), 638-645.

[6] Ramanujam, P., Castillo, E., Patel, E., Vilke, G., Wilson, M. P., &Dunford, J. V. (2009). Prehospital transport time intervals for acute stroke patients. The Journal of emergency medicine, 37(1), 40-45.

[7] Spaite, D. W., Stiell, I. G., Bobrow, B. J., de Boer, M., Maloney, J., Denninghoff, K., & Wells, G. A. (2009). Effect of transport interval on out-of-hospital cardiac arrest survival in the OPALS study: implications for triaging patients to specialized cardiac arrest centers. Annals of emergency medicine, 54(2), 248-255.

[8] Jiang, Y. J., Ma, M. H. M., Sun, W. Z., Chang, K. W., Abbod, M. F., & Shieh, J. S. (2012). Ensembled neural networks applied to modeling survival rate for the patients with out-of-hospital cardiac arrest. Artificial Life and Robotics, 17(2), 241-244.

[9] Abbod, M. F., Cheng, K. Y., Cui, X. R., Huang, S. J., Han, Y. Y., & Shieh, J. S. (2011). Ensembled neural networks for brain death prediction for patients with severe head injury. Biomedical Signal Processing and Control, 6(4), 414-421..

[10] Liu, Q., Cui, X., Abbod, M. F., Huang, S. J., Han, Y. Y., & Shieh, J. S. (2011). Brain death prediction based on ensembled artificial neural networks in neurosurgical intensive care unit. Journal of the Taiwan Institute of Chemical Engineers, 42(1), 97-107.

[11] Lindman, H. R. (2012). Analysis of variance in experimental design. Springer Science & Business Media.

[12] McCoy, C. E., Menchine, M., Sampson, S., Anderson, C., & Kahn, C. (2013). Emergency medical services out-of-hospital scene and transport times and their association with mortality in trauma patients presenting to an urban Level I trauma center. Annals of emergency medicine, 61(2), 167-174.

[13] Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A., &Markovchick, V. J. (2005). Paramedic response time: does it affect patient survival?. Academic Emergency Medicine, 12(7), 594-600.

[14] Roudsari, B. S., Nathens, A. B., Arreola-Risa, C., Cameron, P., Civil, I., Grigoriou, G., ...&Liberman, M. (2007). Emergency Medical Service (EMS) systems in developed and developing countries. Injury, 38(9), 1001-1013.

[15] Huang, C. H., Chen, W. J., Ma, M. H. M., Lai, C. L., Lin, F. Y., & Lee, Y. T. (2001). Ambulance utilization in metropolitan and rural areas in Taiwan. JOURNAL-FORMOSAN MEDICAL ASSOCIATION, 100(9), 581-586.

[16] Spaite, D. W., David, J. T., Valenzuela, T. D., Criss, E. A., Meislin, H. W., Mahoney, M., & Ross, J. (1991). The impact of injury severity and prehospital procedures on scene time in victims of major trauma. Annals of emergency medicine, 20(12), 1299-1305.

[17] Newgard, C. D., Schmicker, R. H., Hedges, J. R., Trickett, J. P., Davis, D. P., Bulger, E. M., ... & Brown, T. B. (2010). Emergency medical services intervals and survival in trauma: assessment of the “golden hour” in a North American prospective cohort. Annals of emergency medicine, 55(3), 235-246.

[18] Sasson, C., Keirns, C. C., Smith, D. M., Sayre, M. R., Macy, M. L., Meurer, W. J., McNally, B. F., Kellermann, A. L., Iwashyna, T. J. (2011). Examining the contextual effects of neighborhood on out-of-hospital cardiac arrest and the provision of bystander cardiopulmonary resuscitation. Resuscitation, 82(6), 674-679.

★★★