

Socioeconomic factors affecting suicides in Sri Lanka

Abstract

Suicide is a serious public health problem in Sri Lanka with strong emotional repercussions for its survivors and for the families of its victims. More than 4000 people in Sri Lanka killed themselves in 2016 and were listed as having the third highest suicide rate. Hence, suicides have emphasized the importance of various factors in the prediction of suicide. Therefore, the main objectives of this study are to determine the major factors associated with suicide and predict the risk of suicide by age and gender separately. We analyzed data based on social and environmental surroundings from 2005 to 2011, obtained from the Sri Lanka Police. Additionally, the data from 2014 - 2016 were used for validation purposes. Factor analysis method was used to determine the variability among correlated variables. Moreover, LASSO regression model, Poisson regression model and Negative binomial regression model were used to predict the total suicides. AIC values were used to identify the best model in the Poisson regression and Negative binomial models. Besides, finest lambda value was used to identify the best LASSO regression model. All the predictions are made assuming that the prevailing conditions in the country affecting crime rates remain unchanged during the period. Further, the mean absolute percentage error (MAPE), is calculated, in order to find the prediction accuracy of the proposed model. Model accuracies for proposed LASSO regression model, Poisson regression model and Negative Binomial model are approximately 89.43%, 87.67% and 80.65% for males, and 76.22%, 92.81% and 90.58% for females respectively. Based on MAPE values, LASSO Regression model was selected as the best male model and the Poisson regression model was selected as the best female model.

Keywords: factor analysis, least absolute shrinkage and selection operator, akaike information criteria, mean absolute percentage error.

Volume 7 Issue 5 - 2018

UNJ Bandara, LS Nawarathna

Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka

Correspondence: UNJ Bandara, Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka, Email navodajayanath@gmail.com

Received: August 08, 2018 | **Published:** September 19, 2018

Introduction

Most of the people think the meaning of suicide is some person chooses to kill themselves, but the real explanation of suicide is some person acts of intentionally causing one's death (Suicide terminology, Wikipedia). Other than that, they do not want to live furthermore in the world. Most of suicidal people have some types of mental conditions at that moment. Moreover, suicide is a sign of serious depression, and hence suicide and depression have strong interrelated connection.¹⁻⁴ Approximately 800,000 people die due to suicide annually around the world, furthermore one person die due to suicide every 40 seconds in the world. Therefore, suicide is one of the most important global health problems in the world and it is a complex human behavior.⁵ Suicide occurs many countries, however suicide rate is relatively low in high income countries.⁶ Among total suicide around the world, 78% of suicides occur in low and middle-income countries. In Sri Lanka, more than 4000 people die due to suicide annually. According to the report conducted by the,⁷ Sri Lanka was ranked in the 3rd position among the 172 countries in terms of most suicide prone countries in the world. Since final few decades, Sri Lanka's suicide rate was very high, in 1995/96, suicide rate in Sri Lanka was 47.0 per 100,000 populations, that was more than two times global suicide rate.⁸ In 2016 suicide rate of Sri Lanka was 28.8 per 100,000 populations, while the global rate was 16 per 100,000 population considered by WHO, that was approximately two times global suicide rate.^{9,10}

When we consider district wise, most suicide were reported in rural areas. Maximum suicide is reported in Jaffna, Vavuniya, Monaragala,

Polonnaruwa districts and Colombo; Galle districts have recorded minimum suicides. Besides the leading suicide age range is 16 to 24 ages and suicide rate decrease after 30 year of age. After 50 years of age again increases the rate of suicide.¹¹ Teenagers have higher suicide rates because of their parents' pressure to fit in socially, to perform academic, their love affairs and spending more time mobile phone, laptop, social media become major causes for teenage suicide. Also, men suicide rate is greater than around 3.8 times women rate. Main objectives of this study are to provide the most current suicide statistics and examine the risk factors associated with suicide. Suicide is a statistically rare event and it is difficult to predict individual which die by suicide. Poisson regression model, negative binomial regression model and lasso regression were used as analysis methods.

This study provides the most current suicide statistics and examines the risk reasons associated with suicide. Suicide is a statistically rare event and it is difficult to predict individual which die by suicide.^{12,13} Poisson regression model, multiple linear regression model and lasso regression were used as analysis methods. The main objectives of this study are to determine major significance socioeconomic factors associated with suicide, predict the future risk of suicide by gender and selecting the best model for suicide predictions.

This article is organized as follows. In Section 2, we describe the materials and methods used to analyze the suicide data from 2005-2011. Section 4 describes the results obtained from the study and Section 5 concludes with a discussion. All the computations in this article have been performed using the statistical programming language R.

Materials and methods

This study is based on data from 2005 to 2011 and 2014 – 2016 total suicides in Sri Lanka which collected from Statistics Department of Sri Lanka Police. The dependent variable is risk factor of total suicide and nine independent numerical variables that are considered in this study are listed in Table 1. This data set is a count data set and it does not contain any negative integers. Moreover, explanatory variables are discrete variables.

Line graph was plotted to identify variation pattern of total suicide in both genders from 2005 to 2011-time period. Then the factor analysis was used to explain whether the relationships of variables observed, correlated variables have similar patterns regarding to total deaths of suicide. We predict the total suicides using the LASSO (Least Absolute Shrinkage and Selection Operator) regression, Poisson regression and Negative binomial models.

LASSO regression model

LASSO regression model is used for attribute selection, shrinkage and improving the accuracy of predictive models. It is a parametric model, implies that it can be applicable to a specific model that has to be a postulated. This has the effect of shrinking coefficient values (and the complexity of the model), allowing some with a minor effect to the response to become zero. When increasing the value of lambda parameter, it will reduce the magnitude of the coefficients, but it will not result in the exclusion of any of the variables. LASSO coefficients, $\hat{\beta}_\epsilon^L$ minimize the quantity,

$$\sum_{i=1}^n (y_i - \beta_o - \sum_{j=1}^p \beta_j * x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| = RSS + \lambda \sum_{j=1}^p |\beta_j|$$

To fit the lasso model, alpha value is specified as one. But alpha value equals to one is not a default. Then deviance graph was plotted, which explains deviance percentage of how much deviance explained, also it is equal to the R-squared in the regression model.

Poisson regression model

Poisson regression is also generalized linear model format of regression analysis. Sometimes Poisson regression model known as log linear model. Characteristic of Poisson distribution is that the mean and variance are same. But in certain circumstances observed variance is greater than mean value. This suicide data set is a count data set and it does not contain any negative integers. Moreover, explanatory variables are discrete variables and hence, can apply Poisson regression model.

$$\log(E(Y|x)) = \alpha + \beta *$$

We use only significant variables from training data set and then predicted the number of suicides.

Negative binomial regression

Negative binomial regression is a type of generalized linear model in which the dependent variable is a count of the number of times an event occurs. This model can be use under, dependent variable is an observed count that follows negative binomial distribution and the possible values of response variable are nonnegative. Negative

binomial regression is implemented using maximum likelihood estimation. This model is generalization of Poisson regression and based on Poisson-gamma mixture distribution (reference).

Table 1 Description of variables

Variable	Variable name
Y	Total suicides
X ₁	Family disputes
X ₂	Physical disabilities
X ₃	Mental disorders
X ₄	Loss love affairs
X ₅	Economic problems
X ₆	Addiction drugs
X ₇	Problem with elders
X ₈	Loss relationships
X ₉	Employment problem

Model validation

The data from 2005 – 2011 were used for model fitting and the data from 2014 - 2016 were used for model validation purposes. AIC (Akaike Information Criterion) values were used to identify best model. AIC values measure relative quality of the statistical models and the best model has lowest AIC value. Then deviance graph was plotted, which explains deviance percentage of how much deviance explained. Moreover, the mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is measured, in order to find the prediction accuracy of the proposed models.

$$MAPE = \frac{100}{n} * \sum_{i=1}^n \frac{|A_i - F_i|}{|A_i|}$$

Where, A_i is the Actual value and F_i is the Predict value. MAPE value measures the absolute error percentage. After constructing three models, accuracy of these three models were compared using MAPE value and model with the maximum accuracy was selected as the best model.

Results

Data from 2005 to 2011 was considered for the preliminary analysis of male and female suicides and line graph was obtained using the training data set.

According to Figure 1, number of male suicides has always high propensity than female suicides in every year. But male suicides gradually decrease annually. There is no significant difference in female suicides during this period. Male suicide percentage is approximately equal to 3.5 times higher tendency than female suicide percentage, which implies that there is a huge impact on the male suicide for Sri Lanka suicide rate. Factor analysis was used to investigate same pattern variables in the data set. In this study permanently there are nine factors included. Latent small number of variables were found using factor analysis and highest positive factor

loading were selected regarding the relevant factor. Further number of factors based on observed variables were identified and factors were named accordingly (Table 2).

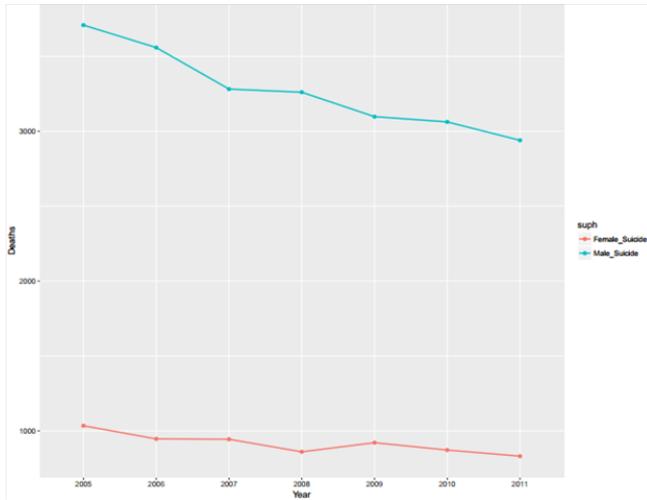


Figure 1 Line graph of total number of suicide male and female separately for the period (2005-2011).

“Harassment of family”, “Physical disabilities”, “Mental disorders” variables strongly associated with Factor1, and hence these variables are belonging to Factor1. Therefore, factor 1 can be renamed as Social Problems. “Economic problems”, “Addiction to drugs”, “Problems with elders” variables are strongly associated with Factor 2 and hence can be renamed as “low personal income”. Moreover, Factor 3 can be renamed based on losing personal life goals as “Employment

problems”, “Loss love affairs” and” Loss relationships” variables strongly associated with Factor 3. Under LASSO regression, total suicide of male and total suicide of female was considered from 2005 to 2011 time period. There are no missing values in the data base. Nine variables are the independent variables and response variable is number of complete suicides of this study. Figures 2–4 shows the actual and predicted male and female suicides respectively using the LASSO regression model.

Table 3 shows the AIC values for models fitted using Poisson regression and negative binomial models according backward selection method.

According to above AIC values table, we can clearly see that the corresponding AIC values for the different models. Under the procedure the minimum AIC value is 254.44, that is the full model. Under AIC value condition, best model includes lowest AIC value. Therefore full model is the best model for total male suicide model. Selected models with minimum AIC value for each category are highlighted. The plot of male and female suicides for fitted Poisson Regression and Negative binomial Regression and are shown in Figures 4–7.

Table 4 shows the comparison of model accuracies of the fitted models.

LASSO regression model, Poisson regression model and negative Binomial model has the model accuracies 89.43%, 87.67% and 80.65% respectively for males and 76.22%, 92.81% and 90.58% respectively for females. Based on the model accuracies, LASSO regression model is selected as the best suicide model for males, whereas Poisson regression model was selected as the best suicide model for females.

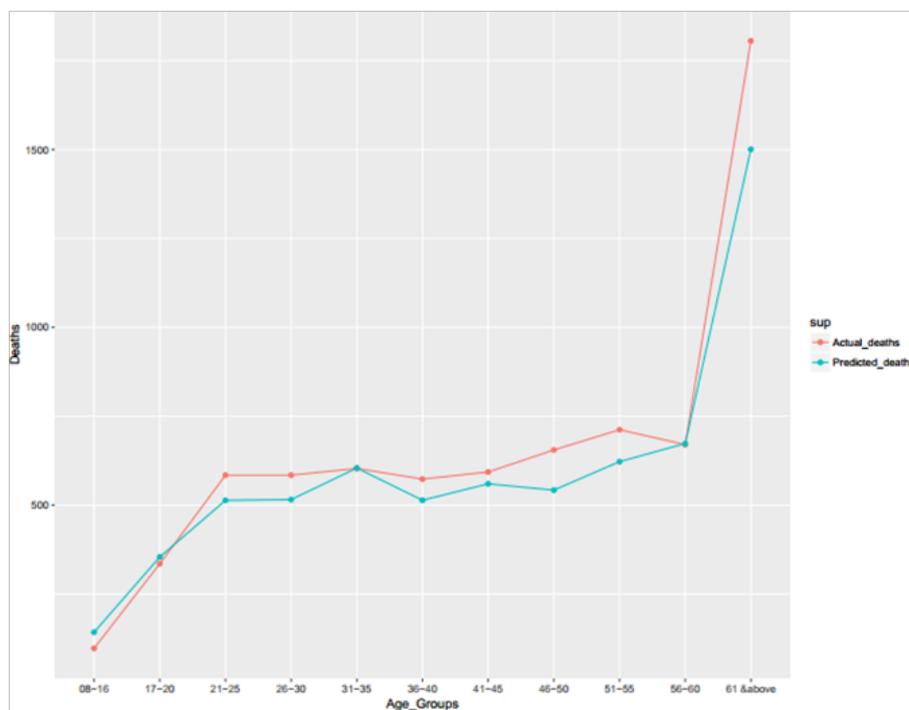


Figure 2 Actual and Predicted male suicides using the LASSO model.

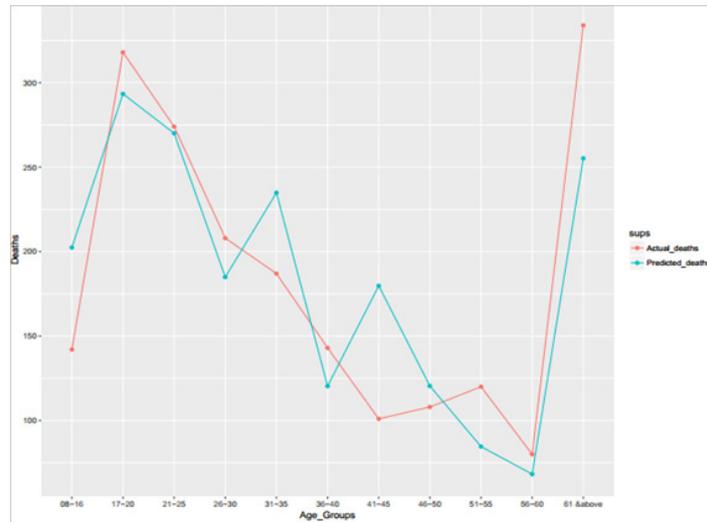


Figure 3 Actual and Predicted female suicide in LASSO model.

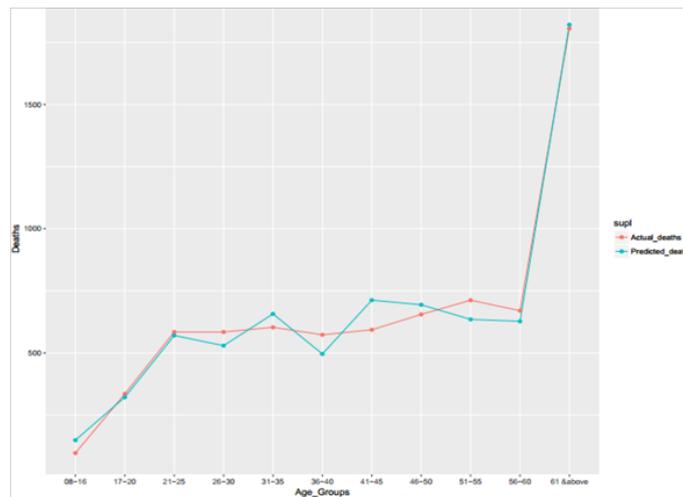


Figure 4 Actual and Predicted total male suicides using Poisson Regression model.

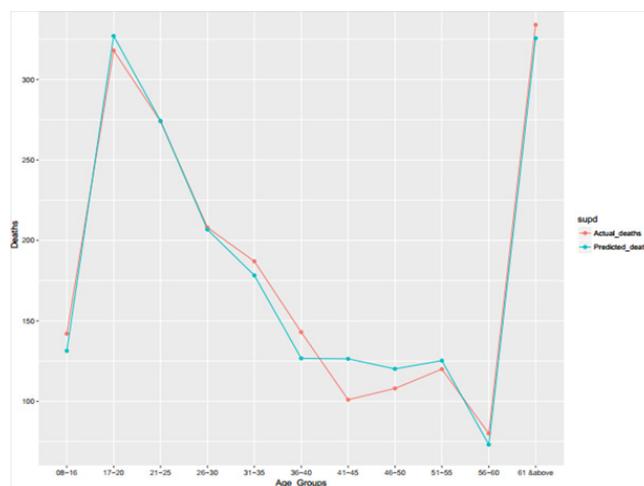


Figure 5 Actual and Predicted total female suicides using Poisson Regression model.

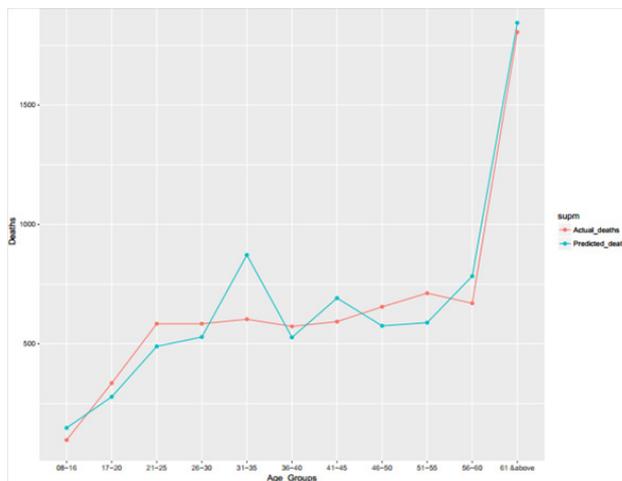


Figure 6 Actual and Predicted total male suicides using negative binomial model.

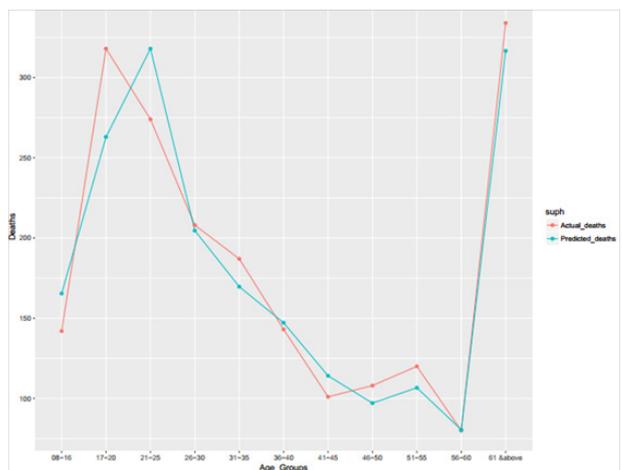


Figure 7 Actual and Predicted total female suicides using negative binomial model.

Table 2 Description of three factor model

Variable	Factor 1	Factor 2	Factor 3
X ₁	0.83	0	0.5
X ₂	0.88	0.32	0
X ₃	0.86	0	-0.25
X ₄	-0.8	-0.23	0.46
X ₅	0	0.89	-0.39
X ₆	0.57	0.79	0
X ₇	0	0.92	0
X ₈	-0.54	0	0.84
X ₉	0	0	0.96

Table 3 AIC values for fitted models according backward selection method

Method	Gender	Model	AIC values
Poisson Regression	Male	Null	5110.1
		Total~X ¹	4114.2
		Total~X ¹ +X ²	1057.2
		Total~X ¹ +X ² +X ³	816.11
		Total~X ¹ +X ² +X ³ +X ⁴	590.63
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵	435.52
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶	389.37
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷	296.52
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸	283.84
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸ +X ⁹	254.44
	Female	Null	1793
		Total~X ¹	984.9
		Total~X ¹ +X ²	782
		Total~X ¹ +X ² +X ³	469.22
		Total~X ¹ +X ² +X ³ +X ⁴	278.18
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵	267.25
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶	206.6
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷	105.94
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸	107.76
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸ +X ⁹	109.4
Negative Binomial	Male	Null	186.38
		Total~X ¹	186.38
		Total~X ¹ +X ²	177.9
		Total~X ¹ +X ² +X ³	176.98
		Total~X ¹ +X ² +X ³ +X ⁴	173.29
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵	174.69
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶	171.68
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷	169.02
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸	165.41
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸ +X ⁹	164.08
	Female	Null	155.49
		Total~X ¹	155.49
		Total~X ¹ +X ²	155.11
		Total~X ¹ +X ² +X ³	151.31
		Total~X ¹ +X ² +X ³ +X ⁴	145.72
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵	146.39
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶	142.84
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷	107.94
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸	109.76
		Total~X ¹ +X ² +X ³ +X ⁴ +X ⁵ +X ⁶ +X ⁷ +X ⁸ +X ⁹	111.4

Table 4 Model comparison for both male and females

Model	Male		Female	
	MAPE	Model Accuracy	MAPE	Model Accuracy
LASSO Regression	10.57%	89.43%	23.78%	76.22%
Poisson Regression	12.33%	87.67%	7.19%	92.81%
Negative binomial Regression	19.35%	80.65%	9.42%	90.58%

Conclusion

In last few years risk factors of complete suicide in Sri Lanka have not been changed in significantly. First, we found that male suicide percentage approximately equal to the 3.5 times high propensity than that of female suicide percentage. In this study, it was determined that the most motivational risk factor for complete suicide is categorized into three factors, namely physical problems, low personal income problems, losing personal life goals. Furthermore, LASSO regression, Poisson regression and negative binomial regression models were used to identify the significance of variables associated with suicide, and total suicides were predicted. LASSO regression model was a better model for male complete suicide among these three fitted models. For female suicides, the lowest MAPE value is obtained in Poisson regression model and hence it is the better model for female complete suicide among three fitted models.

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Junsoo R, Johngwon P, Jinsuk L, et al. Factors That Affect Suicidal Attempt Risk Among Korean. *J Prev Med Public Health*. 2015;48(1):28–37.
2. Kim JW, Lee WH, Myung W, et al. Predicting national suicide numbers with social media data. *PLoS ONE*. 2013;8(4):e61809.
3. Rajapakse T, Griffiths KM, Christensen H, et al. A comparison of non-fatal self-poisoning among males and females, in Sri Lanka. *BMC Psychiatry*. 2014;14:221.
4. Samaraweera S, Sumathipala A, Siribaddana S, et al. Completed Suicide among Sinhalese in Sri Lanka: A Psychological Autopsy Study. *Suicide and Life-Threatening Behavior*. 2008;38(2):221–228.
5. Lemaire CM, Graham DP. Factors associated with suicidal ideation in OEF/OIF veterans. *Journal of Affective Disorders*. 2011;130:231–238.
6. Thalagala N. Suicide trends in Sri Lanka 1880–2006; Social Demographic and Geographical Variations. *Journal of the College of community physicians of Sri Lanka*. 2009;14(1):24–32.
7. Aslam H. Top 10 Countries with highest Suicide rate in 2016. 2016.
8. Gunnell D. The impact of pesticide regulations on suicide in Sri Lanka. *Int J Epidemiol*. 2007;36(6):1235–1242.
9. Hassamal S, Keyser-Marcus L, Breden EC, et al. A Brief Analysis of suicide Methods and Trends in virginia from 2003 to 2012. *BioMed Research International*. 2015.
10. Huang Y, Chen C, Wang L, et al. Methods of suicide predict the risks and. *Neuropsychiatric Disease and Treatment: Dovepress*. 2014; 711–718.
11. Widger T. Suicide in Sri Lanka: Understanding the Crisis. 2014.
12. Balasooriya NN, Kularathne MG. Trends of risk factors of complete suicide by gender and age. *Journal of Public Health and Epidemiology*. 2016;115–120.
13. Brijesh S, Asis De, Edwin VT, et al. Time Trend of the Suicide Incidence in India: a Statistical Modeling. *American Journal of Public Health Research*. 2015;3:80–87.