# Using Bayesian Ordinal Regression Model to Identify Suicide Risk Factors

UIC BioStats, Pei-Shan Yen

May 1 2020

Abstract

Suicide hotlines is a vital prevention tool in reaching people with suicide ideation. The Taiwanese government established National Suicide Prevention Hotlines (TNSPH) in 2009 to reduce the country's suicide rate. Each year, TNSPH received around 70,000 phone calls. Volunteers must evaluate the suicide risk of the callers, but about 60% of callers cannot be assessed due to limited background information and call time duration. To efficiently identify the suicide risk during the phone call, exploring the essential predictors is the objective of this research. We use Bayesian Ordinal Regression Model to identify the critical predictors of known suicide risk groups. TNSPH volunteers may employ these crucial predictors to improve their consulting skills while interacting with the callers, saving countless lives.

keywords: Suicide Risk, Bayesian Ordinal Regression Model, MCMC

# 1 Introduction

## 1.1 Research Background

Suicide is recognized as a severe public health concern with nearly 800,000 deaths worldwide annually (World Health Organization, 2020). In other words, one person dies from suicide every 40 seconds. Suicide is not only a significant cause of premature mortality, but it also impacts seven to ten people close to each deceased (Christopher Lukas, 2007). In Taiwan, the issue of suicide has worsened in the last three decades, and the standardized suicide rate had risen from 7.4 per 100,000 in 1990 to 12.5 per 100,000 in 2017. In Taiwan, suicide not only leads to approximately 3,700 deaths per year but it is also the 11th leading cause of all deaths (Taiwan, Ministry of Health and Welfare, 2019).

Telephone crisis hotlines lead to significant decreases in suicides, and are an important tool in reaching individuals with suicidal ideation (Covington and Hovan, 2011). A major goal of hotlines is to respond to callers' short-term emotional volatility by providing immediate support regardless of time or location (Woodward and Wyllie, 2016). Those with suicidal ideation frequently experience psychological and physical disturbances, including depression, post- traumatic stress, social stigma, physical disorders, and heightened suicide risk (Christopher Lukas, 2007).

Taiwanese government established National Suicide Prevention Hotlines (TNSPH) in 2009 to reduce the country's suicide rate.TNSPH provide telephone counseling and crisis intervention with toll-free 24/7 services. Inside the TNSPH, there are approximately two hundred trained volunteers to answer the phone calls. The number of people in mental distress who have the will to contact TNSPH keeps growing. TNSPH responds each year to approximately 70,000 callers; around 15% of these callers are determined to be at risk for suicide (Shaw and Chiang, 2019).

## 1.2 Research Objective

For TNSPH, the accurate identification of suicide risk for the callers is critical because we must prevent any possible suicide attempts from happening by providing them with immediate help. However, several reasons lead to assessing suicide risk a problematic task. First, this situation is made worse by the anonymity of callers. Second, although the volunteers possibly identify repeated callers from the caller's voice and primary concern, there are still around 20% are first-time callers who often lack necessary information for volunteers to make an informed assessment. Third, the consulting time is often too short (Mean = 10 minutes) to finish the comprehensive evaluation.

Even though volunteers receive at least 200 training hours to help identify the signs of suicide, the volunteers still have little background information to evaluate suicide risk if the caller is unable or unwilling to express suicidal intent (Hunt el. 2018). As a result, usually, more than 60% of phone calls fail to assess suicide risk (TNSPH, 2020). Yen (2019) applied machine learning classification methods to help predict the unknown suicide risk. The Oversampling SMOTE Random Forest Model is constructed to identify high suicide risk among unknown risk callers with a false positive rate of 4.60%. The caller name list with predicted risk can be immediately used to track the caller's mental status and suicidal thoughts.

To efficiently identify the suicide risk during the phone call, exploring the essential predictors is the objective of this research. The Ordinal Regression Model with Frequentist/Bayesian approaches are reasonable method because the outcome suicide risk is the ordinal variable. The Decision Tree model will be used to verify the significant primary predictor of the risk model. The following sections will demonstrate the dataset, methodology, and analytic result.

## 2 Dataset

### 2.1 Data Structure

In this study, we use the caller data which was collected by TNSPH in 2013. These variables are grouped by demographic, socioeconomic, and clinical characteristics, suicidal factors, and call information. Figure 1 demonstrates the structure of all the variables.

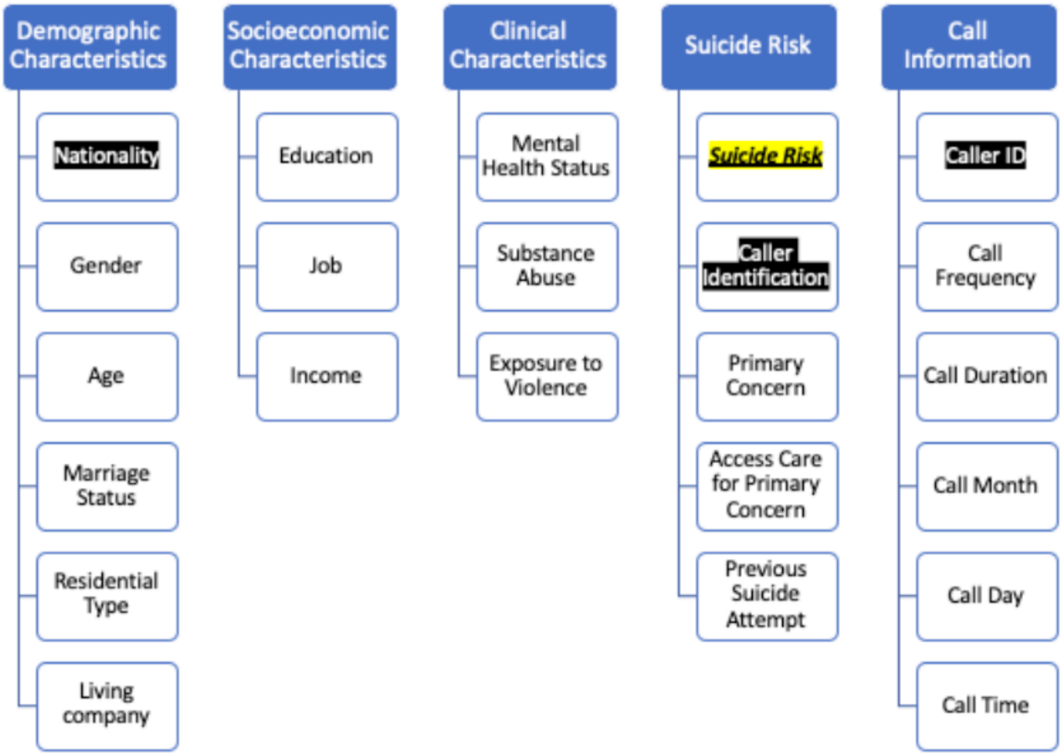



Figure 1. Data Structure

Based on the level of suicidal risk, TNSPH volunteers divides callers into six groups consisting of: (i) no suicide risk (NSR), (ii) low suicide risk (LSR), (iii) moderate suicide risk (MSR), (iv) high suicide risk (HSR), and (v) imminent suicide risk (ISR), and provide them with different strategies and resources. Consider the sample size and the ease of intepreation, we collapse the outcome into 3 levels, including no suicide risk callers (NSR) , low suicide risk callers (LSR), and high suicide risk callers (HSR).

### 2.2 Data Prepare Process

The total sample size of the data is 73,750 with 23 categorial variables. To narrow down the scope of the current investigation, we will not analyze the entire dataset; instead, complete analysis with final sample size 5,966 is performed for this study. We only focus on specific cases which are obtained by employing the following filtering criteria in Table 1.

Table 1. Data Refine Process

| | | Remove Sample Size | Final Sample Size |
|---|---|---|---|
| Step0 | Original Data Set | 0 | 73,050 |
| Step1 | Refine Research Population | 984 | 72,066 |
| Step2 | Data Clean: remove ineligible calls | 15,714 | 56,352 |
| Step3 | Complete Analysis | 40,583 | 15,769 |
| Step4 | Remove Suicide Risk Cannot be Evaluated Group | 9,803 | 5,966 |

Step 1. Refine Research Population: 1) remove non-Taiwanese nationals and those calling on behalf of someone else to report an ISR or a suspected high suicide risk case. After removing the two types of cases, we can remove the variables Nationality and Caller Identification in the dataset because these become only one level that lead to no analytic effect to our prediction model. 2) remove ID variable - this variable is not for analysis use. In summary, we remove three variables from the dataset. Now, the remaining number of predictors is $23 - 1 - 3 = 19$.

Step 2. Clean the Data: 1) remove ineligible phone calls, including informational calls, referral calls, obscene calls, and prank calls. 2) remove cases with call duration of less than 1 minute since these extremely short calls might contain too little information to predict suicide risk.

Step 3. Perform Complete Analysis: ordinal regression model cannot deal with missing values. As a result, we have to remove all the observations with at least one missing value.

Step 4. Remove Suicide Risk Cannot be Evaluated Group (CANTE): The research objective of the final dataset is to use the known suicide risk groups (NSR/LSR/MSR/HSR,/ISR) to construct the best model for the prediction of suicide risk among the unknown suicide risk group (CANTE). Hence, we have to remove the CANTE group from the dataset.

After completing the data filtering criteria, 5,966 inputs remain.

### 2.3 Variable Definition

The variable definition is recorded in Table 2. In each variable, if the sample size in a specific category is too small or cannot not directly help the intepreation of the analysis, we will combine the categories. For example, Wenz (1977) mentions that callers have higher suicide risk in the winter. Since the predictor Call Month originally consists of 12 levels (January to December), and these levels cannot help to identify the seasonal effect on suicide risk, we transform the variable into 2 levels (Spring/Summer and Fall/Winter). Another example is the predictor Call Time. Valtonen et al. (2006) discuss the time pattern of attempted suicide. As a result, we transfer the variable from 24 hours levels to 2 levels (Daytime/Night). Doing so can help observe the effects of call time on suicide risk.

Table 2. Variable Definition

| Variables | Detail |
|---|---|
| **Outcome Suicide Risk** | 1 = No Suicide Risk (NSR), 2 = Low Suicide Risk(LSR), 3 =High Suicide Risk(HSR) |
| **Demographic Characteristics** | |
| Gender | 0 = Male, 1 = Female |
| Age | 0 = <= 29 y, 1 = 30-39 y, 2 = 40-49 y, 3 = > 50y |
| Marriage Status | 0 = Married/Married/Cohabitation, 1 = Single, 2 = Separated/Divorced/Widowed |
| Residential Types | 0 = Countryside, 1 = City |
| Living company | 0 = Live with Family/Friends/Others, 1 = No Company |
| **Socioeconomic Characteristics** | |
| Education | 0 = College/Graduate, 1 = Under Senior High School |
| Job | 0 = Employd, 1 = Unemployed |
| Income | 0 = Job/Saving/Pension, 1 = Subsidy/Family Support |
| **Clinical Characteristics** | |
| Mental Health Status | 0 = Normal, 1 = Mental Illness/Suspect Mental Illness |
| Substance Abuse | 0 = No, 1 = Alcohol/Medicine/Drug Addiction |
| Exposure to Violence | 0 = No Exposure, 1 = Abuser/Victim/Hurting Each Other |
| **Suicidal Factor** | |
| Primary Concern | 0 = Emational Issue, 1 = Interpersonal Issue, 2 = Mental Health Issue, 3 = Financial Issue, 4 = Major Loss/ Life-Adjustment, 5 = Substace Abuse, 6 = Other |
| Access Care for Primary Concern | 0 = Not Seeking/Refuse for Medical Support, 1 = Not Regularly/ Already Seek for Medical Support |
| Previous Suicide Attempt | 0 = No, 1 = Yes |
| **Call Information** | |
| Call Frequency | 0 = First-Time Caller, 1 = Repeated Caller |
| Call Duration | 0 = Less than 10 minutes, 1 = More than 10 minutes |
| Call Month | 0 = Srping/Summer, 1 = Fall/Winter |
| Call Day | 0 = Weekend, 1 = Weekday |
| Call Time | 0 = Daytime, 1 = Night |

## 3 Method

Agresti (2013) suggested that treating ordered categorial variable as ordinal rather than nominal data could increase power, achieve model parsimony, and is easier to intepreat the model. Hence, in our study, Ordinal Regression Model should be used because the outcome suicide risk has explict ordering nature (no suicide risk, low suicide risk, and high suicide risk).

Ordinal data is inherently quantitative, thus the models and measures of the association share many resemblances to continuous variables. One way to deal with ordinal data is to assume the existence of an underlying unobserved continuous variable associated with each response (Winship and Mare, 1984). Such a variable, also called latent variable, is often assumed to be drawn from a continuous distribution with a mean value that is modeled as a linear function of the respondent's covariate vector. (Albert, 1999).

### 3.1 Frequentist Approach: Ordinal Regression Model

We may extend the concepts of the latent variable for the outcome variable $Y$ with $C$ ordered categories. The $C$ ordered categories are considered as $c$ non-overlapping and exhaustive intervals on the continuous scale. In this study, $C$ is set up to 3. Denote $y = (y_1, y_2, \dots, y_n)$ is the observed response of the outcome variable and $n$ is the sample size of the dataset.

#### 3.1.1 Latent Variable

Next, assume that $Z$ is the continuous latent variable with unknown cutoff points $\gamma_0, \gamma_1, \dots, \gamma_c$ where $-\infty \equiv \gamma_0 < \gamma_1 \leq \gamma_2 \leq \cdots \gamma_{c-1} \leq \gamma_c \equiv \infty$. Using this mapping concept of the continuous interval can avoid the imposition of an arbitrary discrete scoring system for the categories. The distribution of the latent variable $Z$ can be expressed by $Z = X^t\beta + \epsilon$ where $X$ is the covariates. The distribution of the random error $\epsilon$ depends on the link function of the ordinal regression model.

Let $F$ be the cummuative distribution of the random error $\epsilon$. Then, we can obtain

$$\theta_c = \mathrm{P}(Y \leq j \mid x) = \mathrm{P}(Z \leq \gamma_c \mid x) = \mathrm{P}(\epsilon \leq \gamma_c - x^t\beta) = F(\gamma_c - x^t\beta) \qquad (1)$$

#### 3.1.2 Link Function

Logistic link means that the random error $\epsilon$ is drawed from a standard Logistic distribution. If we define the model is

$$log\left(\frac{\theta_c}{1-\theta_c}\right) = \gamma_c - x^t\beta \implies \frac{\theta_{1c}/(1-\theta_{1c})}{\theta_{2c}/(1-\theta_{2c})} = exp[-(\boldsymbol{x}_1 - \boldsymbol{x}_2)^t\beta] \qquad (2)$$

The choice of logistic link is most frquently used for ordinal data because of its ease of intepreation. This model is also called Proportional Odds Model which proposed by McCullagh (1980).

Another link function often used to model the cumulative probability is the standard normal distribution. With such a line, Eq (1) can be expressesed by

$$\Phi^{-1}(\theta_c) = \gamma_c - x^t\beta \qquad (3)$$

This model is referred to as the Ordinal Probit Regression Model.

### 3.2 Bayesian Appraoch: Ordinal Regression Model

Now, we can define $p_i = (p_{i1}, p_{i2}, \dots, p_{ic})$ where each element of $p_{ic}$ denotes the probability that subject $i$ is classified into category c. Then, the likelihood is

$$L(\beta,\gamma) = \prod_{i=1}^{n} p_{iy_i} = \prod_{i=1}^{n} F\left(\gamma_{y_i} - x^t\beta\right) - F(\gamma_{y_i-1} - x^t\beta) \tag{4}$$

The Eq.(4) can be represseed in terms of the latent variable $Z$ as

$$p(\beta,\gamma,Z|y) = \prod_{i=1}^{n} F(Z_i - x^t\beta) I(\gamma_{y_i-1} \leq Z_i < \gamma_{y_i}) \tag{5}$$

Where $I(\cdot)$ indicates the indicator function.

#### 3.2.1 Markov Chain Monte Carlo (MCMC): Gibbs Sampling

For ordianl deta, probit link is a great choice because the prior assumption of Normal distribution in MCMC will significantly reduce the computation time. To find the posterior distribution of $(\beta,\gamma)$, one of the efficient method is to perform Gibbs Sampling based on MCMC strategy. (Albert, 1993)

If we use probit link function, from the joint distribution in Eq (5), the conditional posterior distribution of $Z_i|y,\beta,\gamma$ is a truncated Normal distribution with mean $x_i^t\beta$ and variance 1. The density function can be obtained

$$p(Z_i|y,\beta,\gamma) = \frac{\Phi(Z_i - x_i^t\beta)}{\int_{\gamma_{y_{i-1}}}^{\gamma_{y_i}} p(\beta,\gamma,Z_i|y)dz_i} I\left(\gamma_{y_i-1} \leq Z_i < \gamma_{y_i}\right) = \frac{\phi(Z_i - x_i^t\beta)}{\Phi\left(\gamma_{y_i} - x_i^t\beta\right) - \Phi\left(\gamma_{y_{i-1}} - x_i^t\beta\right)} I\left(\gamma_{y_i-1} \leq Z_i < \gamma_{y_i}\right) \tag{6}$$

The conditional posterior distribution of $\beta|y,Z,\gamma$ is a multivariate Normal distribution with mean $(X^TX)^{-1}(X^TZ)$ and covariance matrix $(X^TX)^{-1}$.

The conditional posterior distribution of $\gamma_j|y,Z,\beta$ is a uniform distribtution with boundary $[max\{Z_i: y_i = j\}, min\{Z_i: y_i = j+1\}]$. The boundary would cause the convergence process of Gibbs Sampling slow down.

### 3.2.2 Hybrid Monte Carlo Sampling (HMC)

To accerlerating MCMC convergence for Bayesian ordianl model, Cowles (1996) proposed a Hybrid Monte Carlo (HMC) method for Bayesian Probit Regression Model. Based on the idea of Metropolis–Hastings algorithm, the proposal function $g$ follow a Normal distribtuion with mean $\gamma_j$ and variance $\sigma_{MH}^2$ which is truncated on the interval $(g_{j-1}, \gamma_{j+1})$, for $j = 2, \dots, c-1$. The density can be expressed by

$$g_j \sim \frac{1}{\sigma_{MH}} \frac{\Phi\left(\frac{g_j - r_j}{\sigma_{MH}}\right)}{\Phi\left(\frac{\gamma_{j+1}-\gamma_j}{\sigma_{MH}}\right) - \Phi\left(\frac{g_{j-1}-\gamma_j}{\sigma_{MH}}\right)} \tag{7}$$

Here, we only need to find the condition posterior distribution for $Z, \gamma | y, \beta$. Using Bayes Theorem, we can obtain $p(Z, \gamma | y, \beta) = p(Z | y, \beta, \gamma) p(\gamma | y, \beta)$. The distribution for $\gamma | y, \beta$ can be derived from Eq (4).

$$p(\gamma | y, \beta) \propto \prod_{i=1}^{n} \left[\Phi\left(\gamma_{y_i} - x_i^t \beta\right) - \Phi\left(\gamma_{y_i - 1} - x_i^t \beta\right)\right] \tag{8}$$

From Eq (6) and Eq (8), it is easier to draw the samples from $p(Z, \gamma | y, \beta)$.

Now, the Metropolis–Hastings algorithm can thus be summarized as follows:

1. Use Maximum Likelihood Estimator to obtain the Initial value, $\beta^{(0)}$ and $\gamma^{(0)}$. Set up the iteration time $k = 1$ and $\sigma_{MH}^2 = 0.25/c^2$
2. Sample $g_j$ from $N(\gamma_j^{(k-1)}, \sigma_{MH}^2)$ truncated to the interval $(g_{j-1}, \gamma_{j+1}^{(k-1)})$
3. Using the Independence Chain setting, we calculate the acceptance ratio

$$\mathrm{R} = \prod_{i=1}^{n} \frac{\Phi\left(g_{y_i} - x_i^t \beta^{(k-1)}\right) - \Phi\left(g_{y_i - 1} - x_i^t \beta^{(k-1)}\right)}{\Phi\left(\gamma_{y_i}^{(k-1)} - x_i^t \beta^{(k-1)}\right) - \Phi\left(\gamma_{y_i-1}^{(k-1)} - x_i^t \beta^{(k-1)}\right)} \times \prod_{j=2}^{c-1} \frac{\Phi((\gamma_{j+1}^{(k-1)} - \gamma_j^{(k-1)})/\sigma_{MH}) - \Phi((g_{j-1} - \gamma_j^{(k-1)})/\sigma_{MH})}{\Phi((g_{j+1} - g_j)/\sigma_{MH}) - \Phi((\gamma_{j-1}^{(k-1)} - g_j)/\sigma_{MH})}$$

4. Set $\gamma^{(k)} = g$ with probability R. Otherwise, take $\gamma^{(k)} = \gamma^{(k-1)}$
5. Sample $Z_i^{(k)}$ from a Normal distribution with mean $x_i^t \beta^{(k-1)}$ and variance 1, truncated to the interval $(\gamma_{y_i-1}^{(k)}, \gamma_{y_i}^{(k)})$
6. Sample $\beta^{(k)}$ from a multivariate normal distribution with mean $(X^T X)^{-1}(X^T Z^{(k)})$ and covariance matrix $(X^T X)^{-1}$
7. Increase $k$ to $k+1$ and then repeat steps (1)-(3) until a sufficient number of sampled values have been obtained.

# 4 Result

In this study, the objective is to explore the most important predictors. Hence, we do not transform the categorical covariates into dummy variables. To solve the interpretation issue, in most of the predictors, we assign the reference group 0 to reflect that the callers should have lower suicide risk. This code rules will depend on the clinical decisions and the literature review. For example, the response “No Exposure” is coded as the reference group 0 in the predictor Substance Abuse. The response “Live with Family/Friends/Others” is coded as the reference group 0 in the predictor Living Company. Then, after obtaining the parameter estimation, the significant primary predictors would be decided by the absolute value of the parameter estimation.

We use R software to peform the analysis. In Frequentist Approach, Ordianl Regression Model is analyzed by using the polr function in package MASS. In Bayesian Approach, MCMC is perfomed by the BUGS code and package rjags after finding the full conditional posterior distribution. The computation time is around 8 hours for either probit link or losgitic link. However, HMC is much more efficient because it does not need to estimate from the one-dimensional conditional posterior of cutoff points, which is a complex density. Using the brm function in package brms, HMC only took 7 minutes and 25 minutes to product the result for Probit and Logistic link, respectively. The saving of the computational expense is the major advantage of this specific method (i.e. HMC).

In MCMC and HMC method, we only construct one Markov Chain. We set up 3000 iterations for adaption. The number of total iterations is 30000 with thin number 3 (Actually, we will produce 30000/3 = 10000 final sample in the chain). The R code is shown in Appendix. The comparsion of Ordinal Regression Model using Frequentist Approach and Bayesian Approach are summarized in the following sections.

## 4.1 Exploratory Data Analysis

First, we produce the detailed contingency tables to explore the relationship between suicide risk and each predictor. (See Table 3) Most of the predictors have significant effects on suicide risk (the p-Values of the chi-square independence test are all close to 0.000). Here, we identify the essential characteristics of high suicide risk from the information in the contingency tables. If the predictor effect increases as suicide risk increases for groups with larger sample sizes, we will consider that the predictor does affect the suicide risk. To summarize, the subjective characteristics of people who have suicide risk are included in the following index.

1. Male, Separated/Divorced/Widowed, Living in Countryside
2. Low Education Level, Unemployed
3. Mental Illness, Substance Addiction, Perpetrator of Abuse, Previous Suicide Attempt
4. Call Duration more than 10 minutes, Call in Night Hours (07PM-06AM)

Table 3. Exploratory Data Analysis

| | | | No Suicide Risk (NSR) | Low Suicide Risk(LSR) | High Suicide Risk(HSR) | P-Value |
|---|---|---|---|---|---|---|
| | | | N=2866 | N=2895 | N=205 | |
| Gender | 0 | Male | 903 (31.5%) | 1174 (40.6%) | 127 (62.0%) | <0.001 |
| | 1 | Female | 1963 (68.5%) | 1721 (59.4%) | 78 (38.0%) | |
| Age | 0 | < = 29 y | 635 (22.2%) | 584 (20.2%) | 33 (16.1%) | <0.001 |
| | 1 | 30-39 y | 917 (32.0%) | 834 (28.8%) | 52 (25.4%) | |
| | 2 | 40-49 y | 855 (29.8%) | 841 (29.1%) | 49 (23.9%) | |
| | 3 | >= 50 y | 459 (16.0%) | 636 (22.0%) | 71 (34.6%) | |
| Marriage Status | 0 | Married/Cohabitation | 624 (21.8%) | 628 (21.7%) | 34 (16.6%) | <0.001 |
| | 1 | Single | 1471 (51.3%) | 1274 (44.0%) | 69 (33.7%) | |
| | 2 | Separated/Divorced/Windowed | 771 (26.9%) | 993 (34.3%) | 102 (49.8%) | |
| Residential Types | 0 | Countryside | 1457 (50.8%) | 1348 (46.6%) | 123 (60.0%) | <0.001 |
| | 1 | City | 1409 (49.2%) | 1547 (53.4%) | 82 (40.0%) | |
| Living company | 0 | Live with Family/Friends/Others | 2178 (76.0%) | 2115 (73.1%) | 146 (71.2%) | 0.022 |
| | 1 | No Company | 688 (24.0%) | 780 (26.9%) | 59 (28.8%) | |
| Education | 0 | College/Graduate | 1282 (44.7%) | 1032 (35.6%) | 33 (16.1%) | <0.001 |
| | 1 | Under Senior High School | 1584 (55.3%) | 1863 (64.4%) | 172 (83.9%) | |
| Job | 0 | Employed | 1536 (53.6%) | 1250 (43.2%) | 56 (27.3%) | <0.001 |
| | 1 | Unemployed | 1330 (46.4%) | 1645 (56.8%) | 149 (72.7%) | |
| Income | 0 | Job/Saving/Pension | 1031 (36.0%) | 1086 (37.5%) | 66 (32.2%) | 0.198 |
| | 1 | Subdidy/Family Support | 1835 (64.0%) | 1809 (62.5%) | 139 (67.8%) | |
| Mental Health Status | 0 | Normal | 294 (10.3%) | 116 (4.01%) | 2 (0.98%) | <0.001 |
| | 1 | Mental Illness/Suspect Mental Illness | 2572 (89.7%) | 2779 (96.0%) | 203 (99.0%) | |
| Substance Abuse | 0 | No | 2345 (81.8%) | 1892 (65.4%) | 76 (37.1%) | <0.001 |
| | 1 | Alcohol/Medicine/Drug Addiction | 521 (18.2%) | 1003 (34.6%) | 129 (62.9%) | |
| Exposure to Violence | 0 | No Exposure | 2255 (78.7%) | 1971 (68.1%) | 108 (52.7%) | <0.001 |
| | 1 | Abuser/Victim/Hurting Each Other | 611 (21.3%) | 924 (31.9%) | 97 (47.3%) | |
| Primary Concern | 0 | Emational Issue | 753 (26.3%) | 735 (25.4%) | 51 (24.9%) | . |
| | 1 | Interpersonal Issue | 475 (16.6%) | 235 (8.12%) | 15 (7.32%) | |
| | 2 | Mental Health Issue | 1206 (42.1%) | 1440 (49.7%) | 98 (47.8%) | |
| | 3 | Financial Issue, | 88 (3.07%) | 157 (5.42%) | 22 (10.7%) | |
| | 4 | Major Loss/ Life-Adjustment | 279 (9.73%) | 210 (7.25%) | 10 (4.88%) | |
| | 5 | Substace Abuse | 31 (1.08%) | 63 (2.18%) | 6 (2.93%) | |
| | 6 | Other | 34 (1.19%) | 55 (1.90%) | 3 (1.46%) | |
| Access Care for Primary Concern | 0 | Not Seeking/Refuse for Medical Support | 499 (17.4%) | 382 (13.2%) | 28 (13.7%) | <0.001 |
| | 1 | Not Regularly/ Already Seek for Medical Support | 2367 (82.6%) | 2513 (86.8%) | 177 (86.3%) | |
| Previous Suicide Attempt | 0 | No | 1406 (49.1%) | 679 (23.5%) | 19 (9.27%) | <0.001 |
| | 1 | Yes | 1460 (50.9%) | 2216 (76.5%) | 186 (90.7%) | |
| Call Frequency | 0 | First-Time Callers | 129 (4.50%) | 95 (3.28%) | 5 (2.44%) | 0.031 |
| | 1 | Repeated Callers | 2737 (95.5%) | 2800 (96.7%) | 200 (97.6%) | |
| Call Duration | 0 | Less than 10 minutes | 1354 (47.2%) | 845 (29.2%) | 71 (34.6%) | <0.001 |
| | 1 | More than 10 minutes | 1512 (52.8%) | 2050 (70.8%) | 134 (65.4%) | |
| Call Month | 0 | Spring/Summer | 1546 (53.9%) | 1586 (54.8%) | 117 (57.1%) | 0.608 |
| | 1 | Fall/Winter | 1320 (46.1%) | 1309 (45.2%) | 88 (42.9%) | |
| Call Day | 0 | Weekdend | 794 (27.7%) | 775 (26.8%) | 54 (26.3%) | 0.700 |
| | 1 | Weekday | 2072 (72.3%) | 2120 (73.2%) | 151 (73.7%) | |
| Call Time | 0 | Daytime | 1367 (47.7%) | 1294 (44.7%) | 91 (44.4%) | 0.065 |
| | 1 | Night | 1499 (52.3%) | 1601 (55.3%) | 114 (55.6%) | |

### 4.2 Parameter Estimation

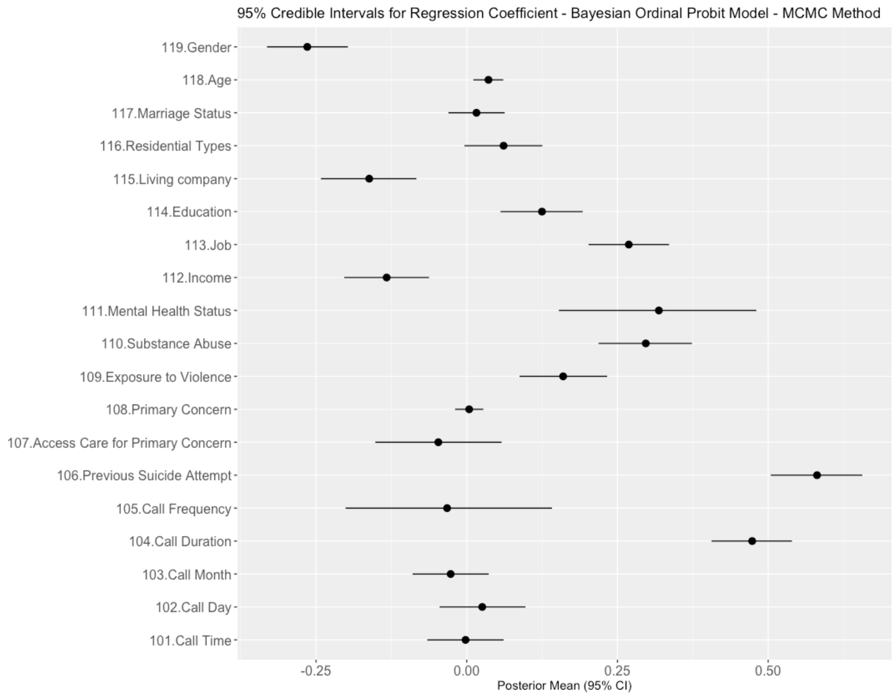
Table 4 summarizes the comparsion of Ordinal Regression Model using Frequentist Approach and Bayesian Approach. It is found that the estimated values of the predictors are quite similar between these two approaches while using same link function. The estimates of MCMC method are also close to HMC.

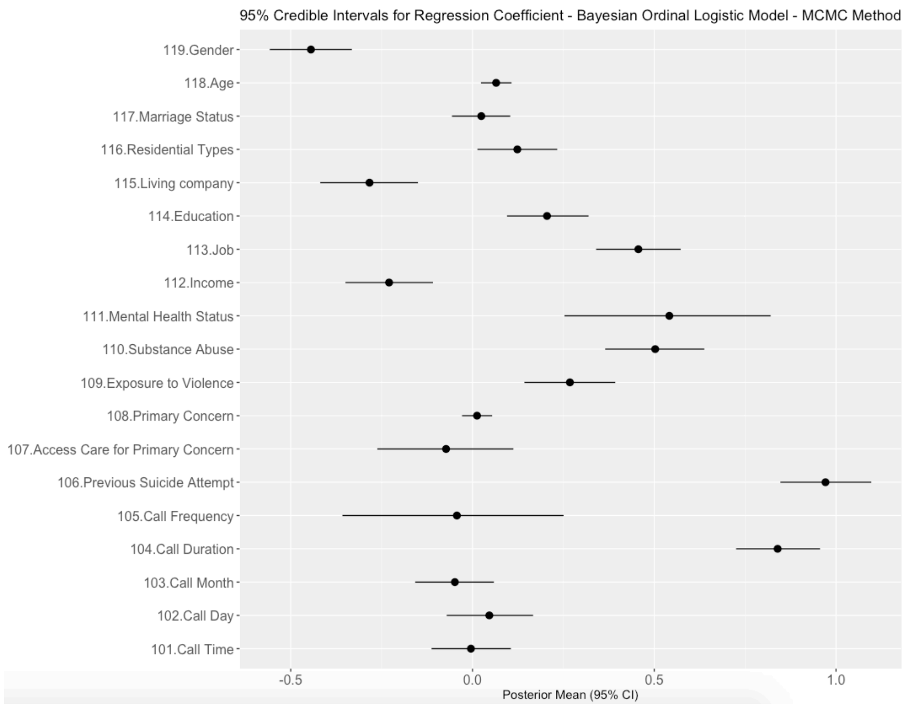
Table 4. Estimate of the risk factors using Frequentist and Bayesian Ordinal Regression Model

| | Probit Link | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Classical Approach | | | | | Bayesian Approach - MCMC | | | | Bayesian Approach - HMC | | | |
| Predictors | Estimate | SD | LCI | UCI | $p$ | Mean | SD | LCI | UCI | Mean | SD | LCI | UCI |
| **Demographic Characteristics** | | | | | | | | | | | | | |
| Gender | -0.2642 | 0.0342 | -0.3313 | -0.1972 | 0.0000 | -0.2646 | 0.0340 | -0.3315 | -0.1971 | -0.2646 | 0.0344 | -0.3319 | -0.1981 |
| Age | 0.0360 | 0.0126 | 0.0114 | 0.0606 | 0.0041 | 0.0360 | 0.0127 | 0.0109 | 0.0606 | 0.0361 | 0.0125 | 0.0115 | 0.0602 |
| Marriage Status | 0.0154 | 0.0240 | -0.0315 | 0.0623 | 0.5204 | 0.0161 | 0.0240 | -0.0304 | 0.0630 | 0.0152 | 0.0237 | -0.0308 | 0.0611 |
| Residential Types | 0.0619 | 0.0327 | -0.0023 | 0.1260 | 0.0589 | 0.0611 | 0.0328 | -0.0039 | 0.1254 | 0.0618 | 0.0331 | -0.0050 | 0.1263 |
| Living company | -0.1620 | 0.0402 | -0.2408 | -0.0833 | 0.0001 | -0.1618 | 0.0403 | -0.2420 | -0.0835 | -0.1618 | 0.0407 | -0.2419 | -0.0812 |
| **Socioeconomic Characteristics** | | 0.0344 | | | | | | | | | | | |
| Education | 0.1247 | 0.0344 | 0.0574 | 0.1921 | 0.0003 | 0.1248 | 0.0349 | 0.0560 | 0.1924 | 0.1248 | 0.0342 | 0.0575 | 0.1919 |
| Job | 0.2688 | 0.0347 | 0.2008 | 0.3370 | 0.0000 | 0.2688 | 0.0344 | 0.2020 | 0.3357 | 0.2693 | 0.0352 | 0.1996 | 0.3388 |
| Income | -0.1331 | 0.0361 | -0.2038 | -0.0624 | 0.0002 | -0.1329 | 0.0358 | -0.2033 | -0.0626 | -0.1329 | 0.0361 | -0.2024 | -0.0608 |
| **Clinical Characteristics** | | | | | | | | | | | | | |
| Mental Health Status | 0.3157 | 0.0835 | 0.1524 | 0.4799 | 0.0002 | 0.3187 | 0.0842 | 0.1527 | 0.4804 | 0.3165 | 0.0836 | 0.1539 | 0.4804 |
| Substance Abuse | 0.2974 | 0.0399 | 0.2192 | 0.3757 | 0.0000 | 0.2970 | 0.0400 | 0.2185 | 0.3736 | 0.2981 | 0.0396 | 0.2206 | 0.3768 |
| Exposure to Violence | 0.1602 | 0.0371 | 0.0874 | 0.2329 | 0.0000 | 0.1599 | 0.0370 | 0.0876 | 0.2329 | 0.1605 | 0.0372 | 0.0863 | 0.2327 |
| **Suicidal Factor** | | | | | | | | | | | | | |
| Primary Concern | 0.0041 | 0.0123 | -0.0200 | 0.0282 | 0.7374 | 0.0040 | 0.0121 | -0.0194 | 0.0276 | 0.0041 | 0.0122 | -0.0200 | 0.0280 |
| Access Care for Primary Concern | -0.0454 | 0.0538 | -0.1507 | 0.0600 | 0.3979 | -0.0472 | 0.0536 | -0.1517 | 0.0579 | -0.0457 | 0.0534 | -0.1491 | 0.0591 |
| Previous Suicide Attempt | 0.5799 | 0.0384 | 0.5047 | 0.6553 | 0.0000 | 0.5811 | 0.0389 | 0.5043 | 0.6562 | 0.5802 | 0.0384 | 0.5040 | 0.6559 |
| **Call Information** | | | | | | | | | | | | | |
| Call Frequency | -0.0334 | 0.0901 | -0.2095 | 0.1437 | 0.7106 | -0.0326 | 0.0885 | -0.2012 | 0.1414 | -0.0322 | 0.0899 | -0.2086 | 0.1440 |
| Call Duration | 0.4733 | 0.0345 | 0.4057 | 0.5410 | 0.0000 | 0.4735 | 0.0341 | 0.4061 | 0.5395 | 0.4737 | 0.0344 | 0.4072 | 0.5417 |
| Call Month | -0.0276 | 0.0321 | -0.0906 | 0.0352 | 0.3891 | -0.0267 | 0.0326 | -0.0899 | 0.0365 | -0.0275 | 0.0320 | -0.0895 | 0.0367 |
| Call Day | 0.0250 | 0.0360 | -0.0454 | 0.0955 | 0.4865 | 0.0257 | 0.0362 | -0.0451 | 0.0974 | 0.0252 | 0.0362 | -0.0461 | 0.0959 |
| Call Time | -0.0021 | 0.0327 | -0.0662 | 0.0620 | 0.9483 | -0.0020 | 0.0323 | -0.0656 | 0.0611 | -0.0021 | 0.0328 | -0.0667 | 0.0619 |
| **Latent Model Cutoff Points** | | | | | | | | | | | | | |
| NSR/LSR | 1.0101 | 0.1148 | | | | 1.0142 | 0.1154 | 0.7809 | 1.2291 | 1.0127 | 0.1163 | 0.7860 | 1.2442 |
| LSR/HSR | 3.1096 | 0.1210 | | | | 3.1151 | 0.1216 | 2.8690 | 3.3454 | 3.1152 | 0.1215 | 2.8809 | 3.3574 |

| | Logistic Link | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Classical Approach | | | | | Bayesian Approach - MCMC | | | | Bayesian Approach - HMC | | | |
| Predictors | Estimate | SD | LCI | UCI | $p$ | Mean | SD | LCI | UCI | Mean | SD | LCI | UCI |
| **Demographic Characteristics** | | | | | | | | | | | | | |
| Gender | -0.4437 | 0.0589 | -0.5593 | -0.3284 | 0.0000 | -0.4444 | 0.0580 | -0.5579 | -0.3317 | -0.4451 | 0.0585 | -0.5623 | -0.3329 |
| Age | 0.0647 | 0.0215 | 0.0226 | 0.1068 | 0.0026 | 0.0650 | 0.0216 | 0.0233 | 0.1074 | 0.0650 | 0.0218 | 0.0224 | 0.1082 |
| Marriage Status | 0.0246 | 0.0411 | -0.0559 | 0.1051 | 0.5493 | 0.0242 | 0.0414 | -0.0567 | 0.1040 | 0.0246 | 0.0411 | -0.0554 | 0.1059 |
| Residential Types | 0.1216 | 0.0559 | 0.0120 | 0.2312 | 0.0297 | 0.1233 | 0.0558 | 0.0134 | 0.2332 | 0.1214 | 0.0562 | 0.0102 | 0.2310 |
| Living company | -0.2811 | 0.0693 | -0.4172 | -0.1455 | 0.0000 | -0.2833 | 0.0692 | -0.4190 | -0.1501 | -0.2826 | 0.0691 | -0.4169 | -0.1480 |
| **Socioeconomic Characteristics** | | | | | | | | | | | | | |
| Education | 0.2029 | 0.0582 | 0.0889 | 0.3169 | 0.0005 | 0.2050 | 0.0577 | 0.0949 | 0.3194 | 0.2025 | 0.0583 | 0.0878 | 0.3161 |
| Job | 0.4548 | 0.0592 | 0.3390 | 0.5710 | 0.0000 | 0.4563 | 0.0596 | 0.3399 | 0.5728 | 0.4558 | 0.0595 | 0.3402 | 0.5705 |
| Income | -0.2288 | 0.0620 | -0.3504 | -0.1075 | 0.0002 | -0.2295 | 0.0618 | -0.3496 | -0.1087 | -0.2305 | 0.0616 | -0.3536 | -0.1101 |
| **Clinical Characteristics** | | | | | | | | | | | | | |
| Mental Health Status | 0.5405 | 0.1431 | 0.2619 | 0.8233 | 0.0002 | 0.5415 | 0.1459 | 0.2530 | 0.8204 | 0.5436 | 0.1454 | 0.2588 | 0.8321 |
| Substance Abuse | 0.5008 | 0.0691 | 0.3655 | 0.6365 | 0.0000 | 0.5026 | 0.0694 | 0.3649 | 0.6377 | 0.5020 | 0.0694 | 0.3627 | 0.6375 |
| Exposure to Violence | 0.2674 | 0.0640 | 0.1419 | 0.3929 | 0.0000 | 0.2680 | 0.0642 | 0.1426 | 0.3928 | 0.2670 | 0.0641 | 0.1426 | 0.3936 |
| **Suicidal Factor** | | | | | | | | | | | | | |
| Primary Concern | 0.0122 | 0.0210 | -0.0290 | 0.0535 | 0.5614 | 0.0125 | 0.0209 | -0.0289 | 0.0541 | 0.0123 | 0.0211 | -0.0289 | 0.0539 |
| Access Care for Primary Concern | -0.0717 | 0.0930 | -0.2542 | 0.1106 | 0.4407 | -0.0727 | 0.0962 | -0.2619 | 0.1126 | -0.0715 | 0.0935 | -0.2527 | 0.1124 |
| Previous Suicide Attempt | 0.9681 | 0.0646 | 0.8418 | 1.0951 | 0.0000 | 0.9710 | 0.0638 | 0.8470 | 1.0971 | 0.9723 | 0.0641 | 0.8479 | 1.0996 |
| **Call Information** | | | | | | | | | | | | | |
| Call Frequency | -0.0524 | 0.1525 | -0.3504 | 0.2481 | 0.7314 | -0.0429 | 0.1586 | -0.3580 | 0.2506 | -0.0537 | 0.1547 | -0.3489 | 0.2520 |
| Call Duration | 0.8370 | 0.0591 | 0.7214 | 0.9532 | 0.0000 | 0.8394 | 0.0594 | 0.7250 | 0.9563 | 0.8388 | 0.0601 | 0.7201 | 0.9567 |
| Call Month | -0.0490 | 0.0550 | -0.1567 | 0.0587 | 0.3726 | -0.0485 | 0.0553 | -0.1577 | 0.0589 | -0.0489 | 0.0549 | -0.1566 | 0.0588 |
| Call Day | 0.0469 | 0.0615 | -0.0736 | 0.1675 | 0.4458 | 0.0463 | 0.0609 | -0.0711 | 0.1668 | 0.0475 | 0.0628 | -0.0763 | 0.1690 |
| Call Time | -0.0056 | 0.0558 | -0.1151 | 0.1038 | 0.9200 | -0.0043 | 0.0556 | -0.1128 | 0.1057 | -0.0051 | 0.0564 | -0.1162 | 0.1032 |
| **Latent Model Cutoff Points** | | | | | | | | | | | | | |
| NSR/LSR | 1.7487 | 0.1960 | | | | 1.7646 | 0.1987 | 1.3757 | 2.1518 | 1.7538 | 0.1985 | 1.3702 | 2.1535 |
| LSR/HSR | 5.5130 | 0.2122 | | | | 5.5349 | 0.2147 | 5.1246 | 5.9546 | 5.5271 | 0.2150 | 5.1124 | 5.9572 |

There are 11 out of 19 predictors that exhibit significant association under the significance level of 0.05. Previous Suicide Attepmt and Call Duration (larger than 10 minutes) are the two crucial risk factors of suicide risk. Male, Unemployed, Having Mental Health Issue or Substance Abuse Behavior are also strong predictors. The inferential results basically match the trend of Exploratory Data Analysis. See Figure 2 below.




For the cutoff points of the latent variable model, Frequentist Approach and Bayesian Approach yield the similar results. Using probit link finction, the cutoff points for NSR/LSR and LSR/HSR are 1.01, and 3.11, respectively. In contrast, when using logistic link finction, the cutoff points for NSR/LSR and LSR/HSR are 1.75~1.76, and 5.51~5.53.

### 4.3 Decision Tree Model Results

To verify the significant primary predictors, we construct Decision Tree Model. This model is performed by using the rpart function in package rpart. The result is revealed in Figured 3. The primary six predictors are Precious Suicide Attempt, Call Duration, Having Substance Abuse Behavior, Gender, Access Care, Marriage Status, and Job. In conclude, the Decision Tree model produces similar results with Ordinal Regression Model.

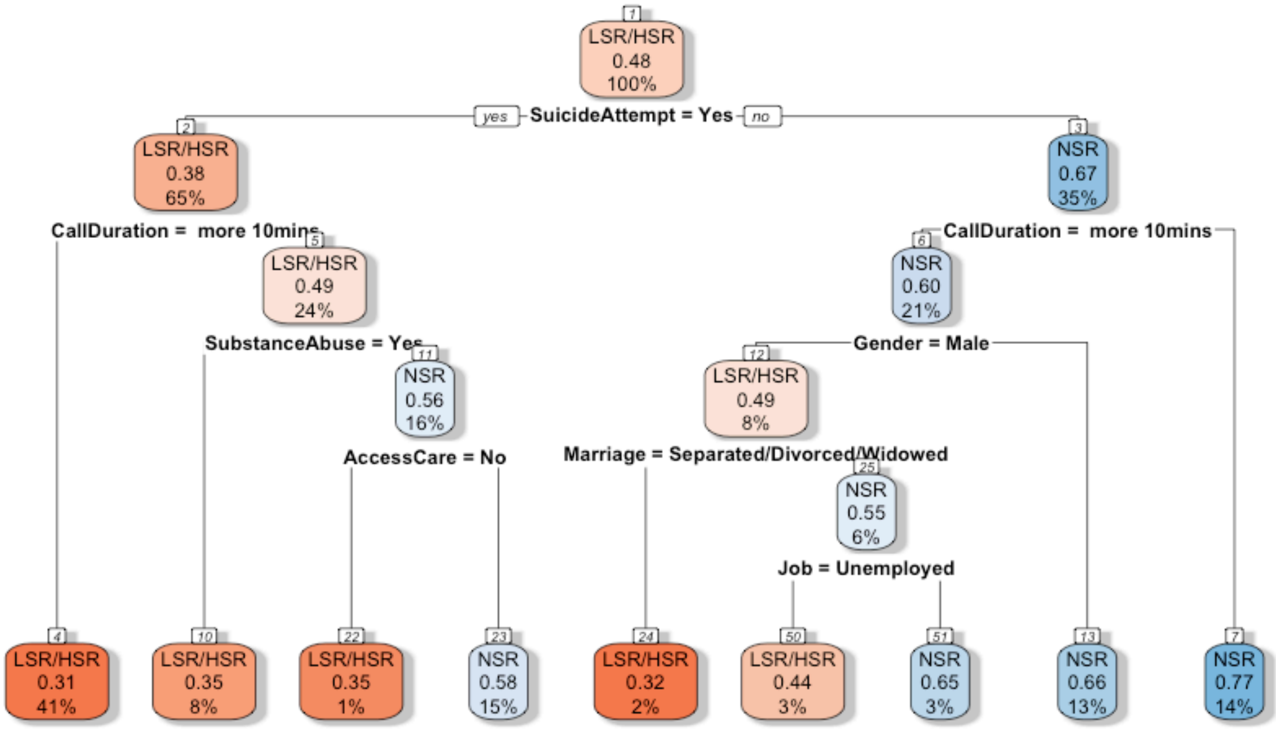



Figure 3. Decision Tree Model (Total Sample Size = 5,966)

### 4.4 Diagnostic for MCMC method

We perform model diagnostic of MCMC method for Bayesian Regression Model. Figure 4-5 reveals the Trace plot and Autocorrelation plot, respectively. The sample results do achieve convergence and looks like independent sample.

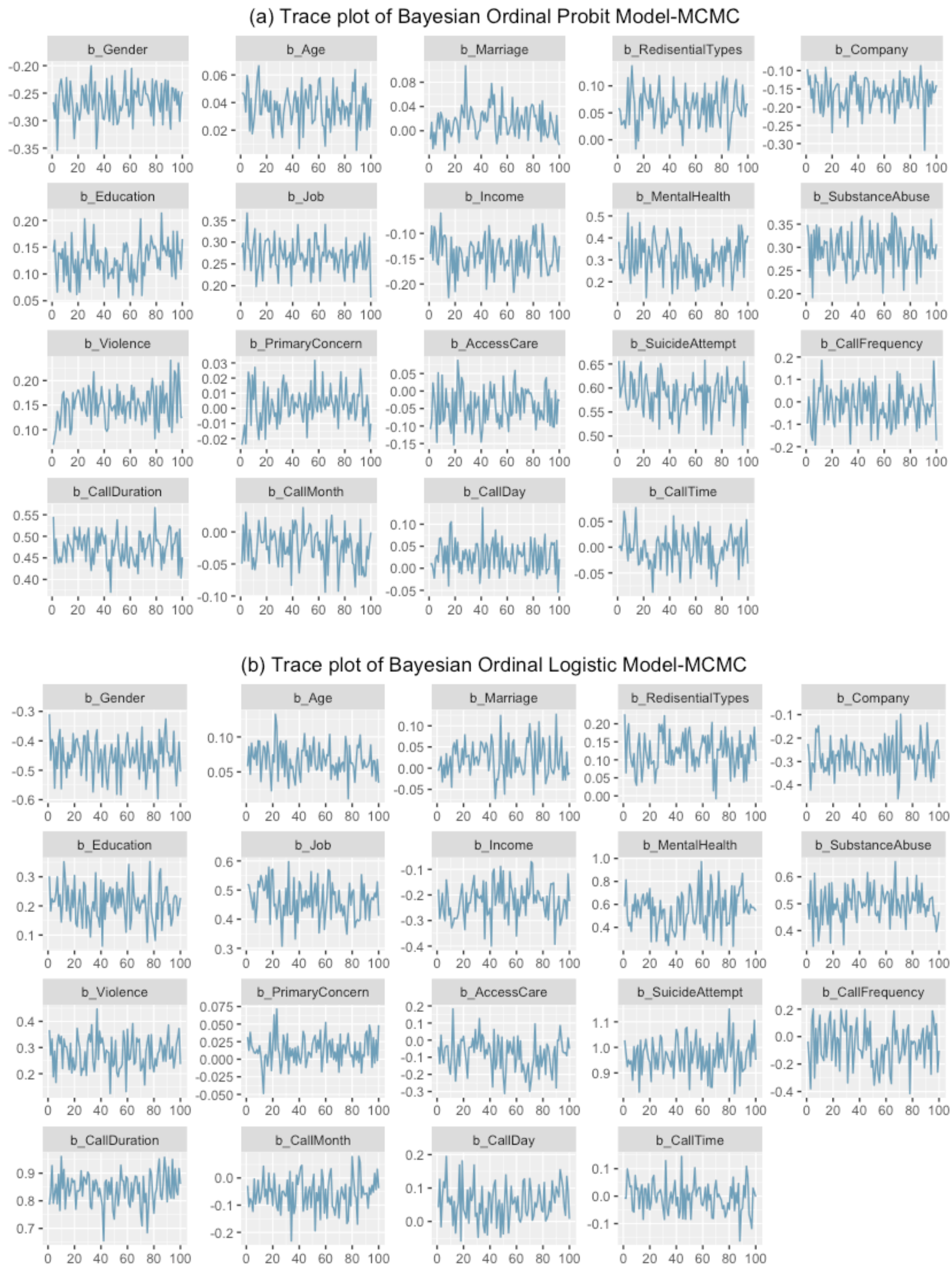



Figure 4. Trace Plot - Bayesain Probit/Logistic Model using MCMC method

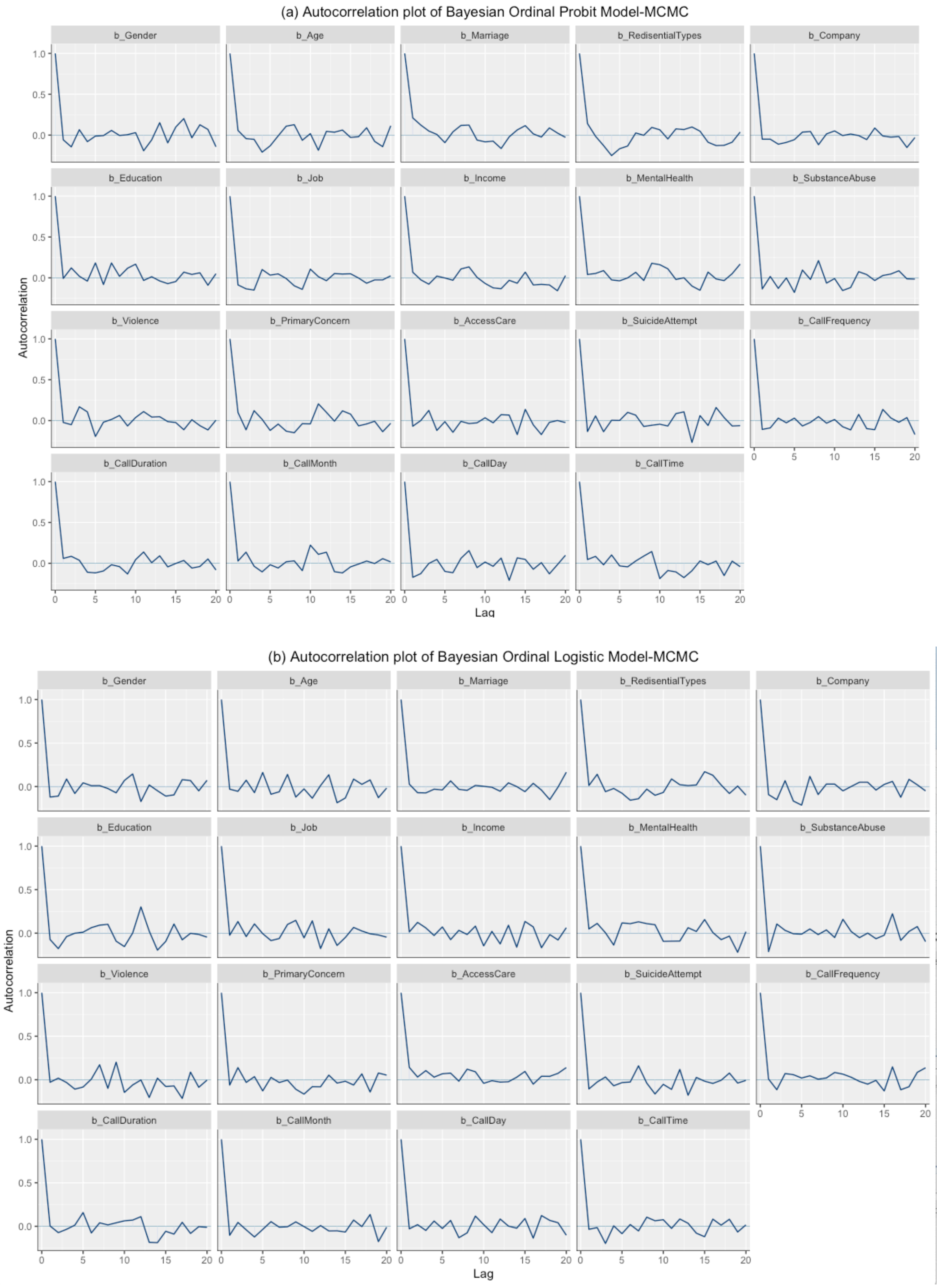



Figure 5. Auto Correlation Plot - Bayesain Probit/Logistic Model using MCMC method

## 5 Conclusion

In this project, we construct Ordinal Regression Model by applying Frequentist/Bayesian Approaches. Both methods yields the similar parameter estimation results while assigning them the same link function, namely, the estimation results are not affected by the choice of the link function nor the selection of the model approach.

A total of 19 predictors are tested, 11 of which are significant. Among these 11 predictors, the primary six predictors are Precious Suicide Attempt (posterior mean in HMC/Probit link, PM, is 0.5802), Call Duration (PM = 0.4737), Having Mental Health Issue (PM = 0.3165), Having Substance Abuse Behavior (PM = 0.2981), Job (PM = 0.2693), and Gender (PM = - 0.2646). In addition, compared with the Decision Tree model, most of the primary predictors of the Ordinal Regression Model are still present except the one "Having Mental Health Issue. The Tree model verifies the significant primary predictors of the Ordinal Regression Model. (See Table 5)

Table 5. Comparison of Ordinal Regression and Decision Tree Model

| Priority of Significant Predictors | Ordinal Regression | Decision Tree |
|---|---|---|
| Precious Suicide Attempt | 1 | 1 |
| Call Duration | 2 | 2 |
| Having Mental Health Issue | 3 | - |
| Having Substance Abuse Behavior | 4 | 3 |
| Job | 5 | 5 |
| Gender | 6 | 5 |
| Access Care | - | 4 |
| Marriage Status | - | 5 |

To sum up, we suggest that TNSPH volunteers should employ these crucial predictors to improve their consulting skills while interacting with the callers. The priority of the significant predictors can serve to identify the callers who possibly have suicide risk. This project demonstrates that it is feasible to utilize Ordinal Regression Model by Bayesian approach to effectively identify and reach individuals in need of assistance, saving countless lives in the future.

# 6 Reference


Agresti, A. (2013). Categorical Data Analysis. Wiley Johnsons.

Albert, J. (1999). Ordinal Data Modeling. Springer.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 88 (422): 669-679.

Christopher Lukas, H. M. S. (2007). Silent Grief. Jessica Kingsley Publishers.

Cowles M. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. Statistics and Computing, 6: 101-111.

McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society, 42(2):109-142.

Shaw, F. F.-T. and Chiang, W.-H. (2019). An evaluation of suicide prevention hotline results in Taiwan: Caller profiles and the effect on emotional distress and suicide risk. Journal of Affective Disorders, 244:16–20.

Valtonen, H., Suominen, K., Partonen, T., Ostamo, A., and L¨onnqvist, J. (2006). Time patterns of attempted suicide. Journal of Affective Disorders, 90(2-3):201–207.

Wenz, F. V. (1977). Seasonal suicide attempts and forms of loneliness. Psychological Reports, 40(3):807–810.

Winship, C. and Mare, R. D. (1984). Regression models with ordinal variables. American Sociological Review, 49(4):512.

Yen, P. (2019). Using Machine Learning Classification Methods to Predict Suicide Risk of National Taiwan Suicide Prevention Hotlines Callers. Course Report.

# Appendix

## A.1 MCMC – Bayesian Ordinal Probit Model

```
############################################ Bayesian Ordinal Probit Model -  MCMC
library(R2jags, rjags)
# 1. Model
# Construct bug file (model.bug)
modelstring <- "
model{
  for (i in 1:n) {
    suci_want[i] ~ dcat(p[i,])
    p[i,1] <- phi(c[1] - mu[i])
    p[i,2] <- phi(c[2] - mu[i]) - phi(c[1] - mu[i])
    p[i,3] <- 1 - phi(c[2] - mu[i])
    mu[i] <-b[1]*gender[i] + b[2]*age[i] + b[3]*marriage[i] + b[4]*location[i] + b[5]*living_stat[i] + b[6]*education[i] + b[7]*job[i] + b[8]*revenue[i] + b[9]*mind_stat[i] + b[10]*obj_abuse
  }

  c[1] <- g[1]
  c[2] <- g[1] + g[2]

  g[1] ~ dnorm(0, 1.0E-6)
  g[2] ~ dgamma(0.001, 0.001)

  for (j in 1:m) {
    b[j] ~ dnorm(0, 1.0E-6)
  }
}
"
writeLines(modelstring, con = "model.bug")
rm(modelstring)

# 2. Data
data = list(n = nrow(SOS_f_KSR_C), m = ncol(X), suci_want=SOS_f_KSR_C$suci_want)
for (var.name in colnames(X)) {eval(parse(text = paste0("data$", var.name, " <- X$", var.name)))} ;rm(var.name) # Put all variables into data list

# 3. Initial Values (1 chain) (for Gibbs Sampling)
inits = list(g = c(0, 1), b = rep(0, 19))

# JAGS
jags = jags.model(file = "model.bug", data = data, inits = inits, n.adapt = 1000, n.chains = 1)
# burn-in number = 1000
update(jags, n.iter = 2000, progress.bar = "text")
# thinning = 1
out = coda.samples(jags, variable.names = c("b", "c"), n.iter = 10000*3, thin = 3)
summary(out)
```

## A.2 MCMC – Bayesian Ordinal Logistic Model

```
SOS_f_KSR_C = read.csv("/Users/peishanyen/Dropbox/UIC BSTT 566 Baysian Methods/Project/R code & Data/Data/SOS_f_KSR_C.csv", header=T, sep=",")
SOS_f_KSR_C = SOS_f_KSR_C[,-1]
############################################ Bayesian Logistic Model -  MCMC
library(R2jags, rjags)
# 1. Model
# Construct bug file (model.bug)
modelstring <- "
model{
  for (i in 1:n) {
    suci_want[i] ~ dcat(p[i,])

    p[i,1] = 1/( 1 + exp( -c[1] + mu[i] ) )
    p[i,2] = 1/( 1 + exp( -c[2] + mu[i] ) ) - 1/( 1 + exp( -c[1] + mu[i] ) )
    p[i,3] = 1- 1/( 1 + exp( -c[2] + mu[i] ) )


    mu[i] <-b[1]*gender[i] + b[2]*age[i] + b[3]*marriage[i] + b[4]*location[i] + b[5]*living_stat[i] + b[6]*education[i] + b[7]*job[i] + b[8]*rev
  }

  c[1] <- g[1]
  c[2] <- g[1] + g[2]

  g[1] ~ dnorm(0, 1.0E-6)
  g[2] ~ dgamma(0.001, 0.001)

  for (j in 1:m) {
    b[j] ~ dnorm(0, 1.0E-6)
  }
}
"
writeLines(modelstring, con = "model.bug")
rm(modelstring)

# 2. Data
data = list(n = nrow(SOS_f_KSR_C), m = ncol(X), suci_want=SOS_f_KSR_C$suci_want)
for (var.name in colnames(X)) {eval(parse(text = paste0("data$", var.name, " <- X$", var.name)))} ;rm(var.name) # Put all variables into data lis

# 3. Initial Values (1 chain) (for Gibbs Sampling)
inits = list(g = c(0, 1), b = rep(0, 19))

# JAGS
jags = jags.model(file = "model.bug", data = data, inits = inits, n.adapt = 1000, n.chains = 1)
# burn-in number = 1000
update(jags, n.iter = 2000, progress.bar = "text")
# thinning = 1
out = coda.samples(jags, variable.names = c("b", "c"), n.iter = 10000*3, thin = 3)
summary(out)
```

## A.3 MCMC – Bayesian Ordinal Probit/Logistic Model – Trace Plot

```
# Trace PLot
library(bayesplot)
Interval <- 100 # Sampling Sample Gap
Probit_Data <- read.table("/Users/peishanyen/Dropbox/UIC BSTT 566 Baysian Methods/Project/R code & Data/Output Ta
Logistic_Data <- read.table("/Users/peishanyen/Dropbox/UIC BSTT 566 Baysian Methods/Project/R code & Data/Output

mcmc_trace(Probit_Data[c(seq(1, 10000, Interval)), -c(1, 21, 22)], size = 0.5) +
  ggtitle("(a) Trace plot of Bayesian Ordinal Probit Model-MCMC") +
  theme(plot.title = element_text(hjust = 0.5)) # line size
mcmc_trace(Logistic_Data[c(seq(1, 10000, Interval)), -c(1, 21, 22)], size = 0.5) +
  ggtitle("(b) Trace plot of Bayesian Ordinal Logistic Model-MCMC") +
  theme(plot.title = element_text(hjust = 0.5)) # line size

################################################
# AutoCorrelation PLot
mcmc_acf(Probit_Data[c(seq(1, 10000, Interval)), -c(1, 21, 22)])+
  ggtitle("(a) Autocorrelation plot of Bayesian Ordinal Probit Model-MCMC") +
  theme(plot.title = element_text(hjust = 0.5)) # line size

mcmc_acf(Logistic_Data[c(seq(1, 10000, Interval)), -c(1, 21, 22)])+
  ggtitle("(b) Autocorrelation plot of Bayesian Ordinal Logistic Model-MCMC") +
  theme(plot.title = element_text(hjust = 0.5)) # line size
```

## A.4 HMC – Bayesian Ordinal Probit/Logistic Model

```
library(brms)
set.seed(20200501)
# HMC - Bayesian Ordianl Probit Model
HybridMC_fit_Probit = brm(suci_want ~ ., data = SOS_f_KSR_C, family = cumulative("probit"), thin = 3, chains = 1, warmup =3000, iter=33000)
summary(HybridMC_fit_Probit)
View(summary(HybridMC_fit_Probit)[["fixed"]])
HMC_Probit = summary(HybridMC_fit_Probit)[["fixed"]]
write.csv(HMC_Probit, file = "/Users/peishanyen/Dropbox/UIC BSTT 566 Baysian Methods/Project/R code & Data/HMC_Probit.csv")

# HMC - Bayesian Ordianl Logistic Model
HybridMC_fit_Logistic = brm(suci_want ~ ., data = SOS_f_KSR_C, family = cumulative("logit"), thin = 3, chains = 1, warmup =3000, iter=33000)
summary(HybridMC_fit_Logistic)
View(summary(HybridMC_fit_Logistic)[["fixed"]])
HMC_Logistic = summary(HybridMC_fit_Logistic)[["fixed"]]
write.csv(HMC_Logistic, file = "/Users/peishanyen/Dropbox/UIC BSTT 566 Baysian Methods/Project/R code & Data/HMC_Logistic.csv")
```

## A.5 Frequentist Approach – Ordinal Probit/Logistic Model

```
library(MASS)
SOS_f_KSR_C$suci_want = as.factor(SOS_f_KSR_C$suci_want)
# Probit model:
M1 = polr(suci_want ~ ., data = SOS_f_KSR_C, method = "probit")
ParameterM1 = coef(summary(M1))
pvalue1 = pnorm(abs(ParameterM1[, "t value"]), lower.tail = FALSE) * 2
CI_M1 = confint(M1)  # confint.default(M1)
CI_M1 = matrix(c(as.numeric(CI_M1)[1:19], c("NA","NA"), as.numeric(CI_M1)[20:38], c("NA","NA")),nrow=21,byrow=F)
M1_Output =data.frame("Parameter"=ParameterM1[,1], "SD"=ParameterM1[,2], "Tvalue"=ParameterM1[,3],"Pvalue" = pvalue1 ,"LCL"=CI_M1[,1],"UCL"=CI_M1[,2])

# Logit link:
M2 = polr(suci_want ~ ., data = SOS_f_KSR_C, method = "logistic")
ParameterM2 = coef(summary(M2))
pvalue2 = pnorm(abs(ParameterM2[, "t value"]), lower.tail = FALSE) * 2
CI_M2 = confint(M2)  # confint.default(M1)
CI_M2 = matrix(c(as.numeric(CI_M2)[1:19], c("NA","NA"), as.numeric(CI_M2)[20:38], c("NA","NA")),nrow=21,byrow=F)
M2_Output =data.frame("Parameter"=ParameterM2[,1], "SD"=ParameterM2[,2], "Tvalue"=ParameterM2[,3],"Pvalue" = pvalue2 ,"LCL"=CI_M2[,1],"UCL"=CI_M2[,2])
```

## A.6 Decision Tree Model

```
library(rpart)
library(rpart.plot)

SOS_f_KSR_C$Gender[SOS_f_KSR_C$Gender == 1] = "Female"
SOS_f_KSR_C$Gender[SOS_f_KSR_C$Gender == 0] = "Male"

SOS_f_KSR_C$SuicideAttempt[SOS_f_KSR_C$SuicideAttempt == 1] = "Yes"
SOS_f_KSR_C$SuicideAttempt[SOS_f_KSR_C$SuicideAttempt == 0] = "No"

SOS_f_KSR_C$CallDuration[SOS_f_KSR_C$CallDuration == 1] = " more 10mins"
SOS_f_KSR_C$CallDuration[SOS_f_KSR_C$CallDuration == 0] = "less than 10mins"

SOS_f_KSR_C$SubstanceAbuse[SOS_f_KSR_C$SubstanceAbuse == 1] = "Yes"
SOS_f_KSR_C$SubstanceAbuse[SOS_f_KSR_C$SubstanceAbuse == 0] = "No"

SOS_f_KSR_C$Age[SOS_f_KSR_C$Age == 0] = "<=29 Y"
SOS_f_KSR_C$Age[SOS_f_KSR_C$Age == 1] = "30-39Y"
SOS_f_KSR_C$Age[SOS_f_KSR_C$Age == 2] = "40-49Y"
SOS_f_KSR_C$Age[SOS_f_KSR_C$Age == 4] = "50-59Y"

SOS_f_KSR_C$Education[SOS_f_KSR_C$Education == 0] = "College or College above"
SOS_f_KSR_C$Education[SOS_f_KSR_C$Education == 1] = "Below College"

SOS_f_KSR_C$RedisentialType[SOS_f_KSR_C$RedisentialType == 0] = "HAve Company"
SOS_f_KSR_C$RedisentialType[SOS_f_KSR_C$RedisentialType == 1] = "No Company"

SOS_f_KSR_C$Marriage[SOS_f_KSR_C$Marriage == 0] = "Married/Cohabitation"
SOS_f_KSR_C$Marriage[SOS_f_KSR_C$Marriage == 1] = "Single"
SOS_f_KSR_C$Marriage[SOS_f_KSR_C$Marriage == 2] = "Separated/Divorced/Widowed"

SOS_f_KSR_C$SuicideRisk[SOS_f_KSR_C$SuicideRisk == 1] = 0
SOS_f_KSR_C$SuicideRisk[SOS_f_KSR_C$SuicideRisk == 2] = 1
SOS_f_KSR_C$SuicideRisk[SOS_f_KSR_C$SuicideRisk == 3] = 1

SOS_f_KSR_C$SuicideRisk[SOS_f_KSR_C$SuicideRisk == 0] = "NSR"
SOS_f_KSR_C$SuicideRisk[SOS_f_KSR_C$SuicideRisk == 1] = "LSR/HSR"

SOS_f_KSR_C$Job[SOS_f_KSR_C$Job == 0] = "Employed"
SOS_f_KSR_C$Job[SOS_f_KSR_C$Job == 1] = "Unemployed"

SOS_f_KSR_C$AccessCare[SOS_f_KSR_C$AccessCare== 0] = "No"
SOS_f_KSR_C$AccessCare[SOS_f_KSR_C$AccessCare == 1] = "Yes"

tree2 = rpart(SuicideRisk~., SOS_f_KSR_C, cp=0.005)
rpart.plot(tree2, box.palette="RdBu", shadow.col="gray", nn=TRUE)
```