Prediction by data mining, of suicide attempts in Korean adolescents: a national study

Sung Man Bae1
Seung A Lee2
Seung-Hwan Lee2,3
1Department of Counseling Psychology, The Cyber University of Korea, Seoul, South Korea; 2Clinical Emotion and Cognition Research Laboratory, Goyang, South Korea; 3Department of Psychiatry, Ilsan Paik Hospital, Inje University College of Medicine, Goyang, South Korea

Objective: This study aimed to develop a prediction model for suicide attempts in Korean adolescents.

Methods: We conducted a decision tree analysis of 2,754 middle and high school students nationwide. We fixed suicide attempt as the dependent variable and eleven sociodemographic, intrapersonal, and extrapersonal variables as independent variables.

Results: The rate of suicide attempts of the total sample was 9.5%, and severity of depression was the strongest variable to predict suicide attempt. The rates of suicide attempts in the depression and potential depression groups were 5.4 and 2.8 times higher than that of the non-depression group. In the depression group, the most powerful factor to predict a suicide attempt was delinquency, and the rate of suicide attempts in those in the depression group with higher delinquency was twice higher than in those in the depression group with lower delinquency. Of special note, the rate of suicide attempts in the depressed females with higher delinquency was the highest. Interestingly, in the potential depression group, the most impactful factor to predict a suicide attempt was intimacy with family, and the rate of suicide attempts of those in the potential depression group with lower intimacy with family was 2.4 times higher than that of those in the potential depression group with higher intimacy with family. And, among the potential depression group, middle school students with lower intimacy with family had a 2.5-times higher rate of suicide attempts than high school students with lower intimacy with family. Finally, in the non-depression group, stress level was the most powerful factor to predict a suicide attempt. Among the non-depression group, students who reported high levels of stress showed an 8.3-times higher rate of suicide attempts than students who reported average levels of stress.

Discussion: Based on the results, we especially need to pay attention to depressed females with higher delinquency and those with potential depression with lower intimacy with family to prevent suicide attempts in teenagers.

Keywords: predictor, severity of depression, suicide, delinquency, intimacy with family

Introduction
The rate of completed suicide in South Korea in 2011 was 33.3 in 100,000, which is significantly higher compared to that in other developed countries, including the United States (12.5), the United Kingdom (6.7), Germany (10.8), Australia (10.1), and Japan (20.9).1 Suicide attempts in adolescents, in particular, are increasing rapidly and suicide is becoming one of the serious social issues in South Korea. Actually, the rate of completed suicide of Korean adolescents was 4.4% in 2012.2

In previous studies, depression was considered to be the main predictor of suicide attempts, and it is known that suicide attempts for all ages are affected by depression.1,3,4 Adolescents, who are more impulsive and emotionally unstable compared to adults, are likely to attempt suicide unexpectedly, either when they are very depressive or...
shortly after recovery from a depressive state.\textsuperscript{5} Daily stress is another main predictor of suicide attempts, and it could affect suicide attempts directly or through interaction with depression.\textsuperscript{4} The stress level from daily life, such as from academic competition or conflict with parents or the peer group, was found to significantly predict suicide attempts in adolescents.\textsuperscript{7}

Some studies have suggested delinquency as an important predictor of suicide attempts.\textsuperscript{8,9} Juvenile delinquency occurs frequently along with depression,\textsuperscript{10} and an interaction between them would have some causal effects on completed suicide. Also, juvenile delinquency could be predicted by relationships with peers, parents, and teachers, as well as by stress and adjustments at school.\textsuperscript{11} According to the problem behavior theory of Jessor,\textsuperscript{12} perceived environment variables such as parent–child conflict and peer relationship problems affect problem behaviors (eg, delinquency, conduct problems), and problem behaviors can impact other emotional and behavioral problems (eg, depression, internet addiction, suicide attempt).\textsuperscript{9}

Optimism is considered to be an important variable of mental health. Optimistic persons experience positive emotions more frequently, and their level of worrying, anxiousness, or depression was found to be lower compared to those of pessimistic people.\textsuperscript{13} Hirsch et al reported that optimism decreased suicidal ideation even after covariates such as age, sex, and depression symptoms are controlled.\textsuperscript{14} Also, Seo reported that optimism, family environment, school life, or community activities affect suicidal ideation by mediating depression and hopelessness in high school students.\textsuperscript{15}

In adolescents, relationship with parents is an important predictor of suicide ideation.\textsuperscript{16,17} Close intimacy with parents helps adolescents to maintain emotional stability and develop self-esteem. Furthermore, it was revealed that good intrafamilial relationships prevented suicide attempts in depressed adolescents.\textsuperscript{18,19} Self-esteem could be another major predictor of suicide. Adolescents with a high level of self-esteem showed a lower frequency of suicide ideation and attempts.\textsuperscript{20}

Joiner insisted that suicide risk is increased by thwarted belongingness and perceived burdensomeness from relationships with others.\textsuperscript{21} It could be assumed that community support and interaction with the community could become predictors of suicide ideation and attempts. A person can experience self-worth and feel less loneliness and alienation when community support and interaction are abundant. These would act as buffers for negative thoughts when people are in stressful situations and eventually reduce the risk of completed suicide.\textsuperscript{22,23}

As happiness and satisfaction of adolescents are closely related to school life, an adolescent’s suicide attempt is highly linked to academic achievement and adjustment at school. Conflict with peers is one of the primary factors for stress, and peer conflicts could have significant effects on suicide attempts in adolescents.\textsuperscript{24,25} Good relationships with teachers also have buffering effects on stress. Especially for Korean adolescents, academic achievement is another major stressful factor which could affect suicide attempts in adolescents.\textsuperscript{26}

Previous studies have selected suicide ideation as the dependent variable. However, considering that completed suicide follows after suicide ideation and suicide attempt,\textsuperscript{18} selecting suicide attempt as the dependent variable would allow us to make more accurate predictions of completed suicide and develop a theoretical framework for early intervention. Moreover, most studies did not consider sociodemographic and psychological variables together relevant to suicide. Particularly, sex,\textsuperscript{27} socioeconomic status,\textsuperscript{28} and parents’ academic backgrounds have been revealed as important variables that could predict suicide attempts or completed suicide.

We assumed that interaction between sociodemographic and psychological variables could be important to predicting suicide attempts. The current study tried to develop a model to predict suicide attempts in Korean adolescent using decision tree analysis, which is a well-known useful statistical method to analyze all interaction effects of multiple variables. The purpose of this study was the development of a predictive model of suicide attempts. We considered all interactions of sociodemographic variables and eight intrapersonal and extrapersonal variables (depression, stress, delinquency, self-esteem, intimacy with family, optimism, school adaptation, and community support and interaction) for middle and high school students.

**Methods**

**Participants and procedure**

The current study utilized data from the survey of mental health in 2011 from the National Youth Policy Institute of Korea. Sixteen metropolitan cities and provinces (Seoul, Busan, Daegu, Incheon, Kwangju, Daejeon, Ulsan, Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, Jeonnam, Gyeongbuk, Gyeongnam, Jeju) and school level (middle or high) were selected as stratifying variables. Population size was determined by proportional allocation.
Data were collected from a class in one grade level of each school located in each city. The total sample consisted of 2,754 middle and high school students (1,110 males, 1,644 females). Sociodemographic characteristics of the participants are shown in Table 1.

The current study attempted to develop a prediction model for suicide attempts by including eleven sociodemographic variables and eight psychological variables. In order to investigate which variables affect suicide attempts in different levels of depression, students were divided into three groups based on depression severity, which is the primary predictor variable of suicide attempt or completed suicide. This study was approved by the Institutional Review Board of Inje University Ilsan Paik Hospital (Goyang, South Korea). The study was exempted from the requirement of informed consent by the board.

**Measures**

**Suicide attempt**
Suicide attempt was measured by a single-item scale (“Did you attempt suicide in the last 1 year?”). Students were required to make a two-alternative forced choice (“yes” or “no”).

**Sociodemographic variables**
In the current study, the nine sociodemographic variables predicting suicide attempt were selected as follows: sex, age, school level (middle or high school), school record (excellent, good, average, poor, bad), academic backgrounds of father and mother, employment statuses of father and mother, location of school (metropolitan, micropolitan, rural areas), socioeconomic status (high, middle, low), and family structure (two parents, single parent, divorced parent, custodial grandparents, etc).

**Depression**
The Beck Depression Inventory (BDI) was used to measure depression levels in the adolescents. There are 21 items with a four-point scale, totaling 63 points. Scores greater than 18 are taken to indicate depression. Internal consistency of the current data was 0.900.

**Stress**
A daily stress questionnaire developed by the National Youth Policy Institute of South Korea was used. Daily stress was measured based on 12 categories (relationships with parents, siblings, friends, the opposite sex, colleagues, and teachers; appearance; physical health; mental health; financial problems; and academic stress) with a four-point scale (not at all influenced, not very influenced, somewhat influenced, very influenced). Internal consistency of the current data was 0.910.

**Delinquency**
A delinquency questionnaire developed by the National Youth Policy Institute of South Korea was used. Delinquent behavior within the last 1 year was measured by ten items (smoking, drinking alcohol, theft, use of violence, extortion, absence without leave, running away from home, watching pornography, gambling, and damaging public property) with a four-point scale (none, one to three times, three to four times, over five times). Internal consistency of the current data was 0.812.

**Self-esteem**
A self-esteem scale developed by the National Youth Policy Institute of South Korea was used to measure self-esteem. The level of self-esteem was evaluated with five items with a four-point scale. A higher score denotes a higher level of self-esteem. Internal consistency of the current data was 0.902.

**Optimism**
An optimism questionnaire developed by the National Youth Policy Institute of South Korea was used. Three items with a four-point Likert scale were used to measure whether students believed themselves to be able to control their own lives and whether they viewed themselves and global affairs

<table>
<thead>
<tr>
<th>Table 1 Sociodemographic characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>School level</td>
</tr>
<tr>
<td>Middle school</td>
</tr>
<tr>
<td>High school</td>
</tr>
<tr>
<td>Location of school</td>
</tr>
<tr>
<td>Metropolitan</td>
</tr>
<tr>
<td>Micropolitan</td>
</tr>
<tr>
<td>Rural areas</td>
</tr>
<tr>
<td>Socioeconomic status</td>
</tr>
<tr>
<td>High</td>
</tr>
<tr>
<td>Middle</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Non-response</td>
</tr>
</tbody>
</table>
positively and optimistically. Internal consistency of the current data was 0.794.

Intimacy with family
A questionnaire developed by the National Youth Policy Institute of South Korea was used for evaluation of intimacy with family. Intimacy with family was measured by five items with a four-point scale. A higher score denotes closer intimacy with family. Internal consistency for the current data was 0.924.

Community support and interaction
A questionnaire developed by the National Youth Policy Institute of South Korea was used for evaluation of community support and interaction. Three items for community support and another three items to measure interaction with community, totaling six items, were used. Community support indicates a person (except family, friends, or teachers) or organization which one can trust or seek help from when in need. Interaction with community implies involvement in volunteering or activities with community organizations. Internal consistency for community support was 0.828, for interaction with community was 0.765, and for total data was 0.811.

Adjustment to school
A questionnaire developed by the National Youth Policy Institute of South Korea was used for evaluation of adjustment to school. Twelve items with a four-point scale measured the level of each adolescent’s adjustment to school. There are four indexes, covering attitude toward class, relationships with teachers and with peers, and school life. Each of the four indexes is comprised of three items. Internal consistency for school class is 0.814, for relationship with teacher is 0.875, for relationship with peers is 0.843, for school life is 0.712 and for total data is 0.861.

Data analysis
The current study used the Answer Tree 3.0 program to conduct a decision tree analysis of data mining. Data mining is efficient for analyzing big data because every possible interaction between a great number of variables, including continuous and dichotomous variables, can be analyzed. In contrast to typical statistical hypothesis testing, data mining does not establish a specific hypothesis. In data mining, explanation of the results from the big data analysis is essential.

For splitting criteria in the decision tree analysis, we employed chi-squared automatic interaction detection (CHAID). Although CHAID can predict both continuous and dichotomous target variables, it is more apt for prediction of a dichotomous target variable. As the target variable in the current study was dichotomous (did attempt suicide or did not attempt suicide), multway split was conducted by chi-square test. In multway split, more than two child nodes are separated from a parent node. The stopping rule for maximum tree depth is a value of 5, and the minimum numbers of cases for parent and child nodes are 50 and 25, respectively. Gain charts and risk charts are applied to test the fitness of the model. Gain chart is the ratio of the target category (suicide attempt for the current study) in a particular node. Risk chart is the probability that a model misclassifies an observation.

We divided the total sample into training data (70% of total sample) and testing data (30% of total sample). Firstly, to develop the best prediction model for suicide attempts, we utilized training data and identified the possibility of generalization of the prediction model through testing data. If there is no difference in risk estimates of the model between training data and testing data, generalization of training data can be assumed.

Results
The outcomes from the decision tree analysis of the training sample were as follows (Figure 1). Firstly, approximately 9.64% (n=186) out of a total of 1,910 students (training sample) reported that they had attempted suicide. Those who had never attempted suicide was 90.36% (n=1,744). Secondly, depression was the primary variable predicting suicide attempts in adolescents (chi-square: 101.3982, P<0.0001). The rate of suicide attempts was 20.34% in the depression group (n=531), 7.32% in the potential depression group (n=710), and 3.77% in the non-depression group (n=689).

The rate of suicide attempts in the depression group was different according to delinquency level (chi-square: 13.6599, P<0.0028). Of depressed students with a high score for delinquency (delinquency >19, n=78), 35.90% showed a suicide attempt, while 17.66% of depressed students with a low score for delinquency (delinquency \(\leq19, n=453\)) showed a suicide attempt. The rate of suicide attempts of those in the depression group with a high score for delinquency was different according to sex (chi-square: 4.6897, P<0.0303). The rate of suicide attempts of females with a high score for delinquency in the depression group (n=32) was 50.00%, while the rate of suicide attempts of males with a high score for delinquency was 20.34%.
Prediction of suicide attempts in Korean adolescents

Figure 1  Decision tree for prediction of suicide attempt.

Notes: "Yes" means "suicide attempt", while "No" is "no suicide attempt". Scores for stress, family intimacy, and delinquency are scores of inventory used in this study to measure these variables. Numbering 1–20 is each node of the decision tree analysis.

Abbreviation: SES, socioeconomic status.
Table 2 Gain index of predicting suicide attempts

<table>
<thead>
<tr>
<th>Node</th>
<th>Gain index</th>
<th>Accumulated gain index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Node: n</td>
<td>Node: %</td>
</tr>
<tr>
<td>20</td>
<td>32</td>
<td>1.7</td>
</tr>
<tr>
<td>19</td>
<td>46</td>
<td>2.4</td>
</tr>
<tr>
<td>18</td>
<td>306</td>
<td>15.9</td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>3.6</td>
</tr>
<tr>
<td>12</td>
<td>68</td>
<td>3.5</td>
</tr>
<tr>
<td>17</td>
<td>147</td>
<td>7.6</td>
</tr>
<tr>
<td>16</td>
<td>87</td>
<td>4.5</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>554</td>
<td>28.7</td>
</tr>
<tr>
<td>6</td>
<td>58</td>
<td>3.0</td>
</tr>
<tr>
<td>11</td>
<td>210</td>
<td>10.9</td>
</tr>
<tr>
<td>13</td>
<td>293</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Notes: Node: n = each node number. Node: n = sample size of each node. Node: % = rate of target category of total sample. Gain: n = sample size of target category of each node. Gain: % = rate of target category of each node of total target category. Resp: % = rate of sample size of target category of sample size of each node. Index: % = resp (%) versus rate of target category of sample size.

Among students with a middle level of stress in the non-depression group, the rate of suicide attempts of students who reported that their school record was excellent (n=28) was 7.14%. On the other hand, the rate of suicide attempts for those whose school record was below excellent (good, average, poor, bad, n=293) was 0%.

The gain chart in Table 2 provides legends for each node of target category (suicide attempt). The gain chart consists of node-by-node statistics (gain index) and cumulative statistics (accumulated gain index). The highest gain score for the current study was the 20th node, which was the depressed females with a high level of delinquency. The number of cases in the 20th node was 32 out of the total 1,930 samples, which is 1.7% of the total subjects. Among cases in the 20th node (n=32), the number of cases which corresponded to the target category (gain: n) was 16 students, which is 50% of the total number of cases in the 20th node (response: %). Response (%) in the 20th node was five times greater than the rate of the target category (9.64%) of total samples (1,930). Finally, the index (%) of the 20th node was 518.8%. Although there are no clear criteria for index (%), index (%) over 200% at a particular node is generally considered to be a high gain score.

The risk estimate of the final model for training data is shown in Table 3. Risk estimates are the prediction error rates of the final model. The prediction error rate in the current study was 9.64%, and the accuracy of classification was over 90%.

Table 3 Risk estimate for prediction of suicide attempt

<table>
<thead>
<tr>
<th>Risk estimates</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0963731</td>
<td>0.00671728</td>
</tr>
</tbody>
</table>
In the current study, validity testing for training data was conducted to ensure generalization of the prediction model for suicide attempts. The validity testing of training data through testing data (n=824) showed that the risk estimate of the testing sample was 0.092233 and it was not significantly different from the risk estimate (0.0963731) of the training data. Therefore, it seems the prediction model for suicide attempts in the current study has a high possibility of generalization.

**Discussion**

The current study attempted to develop the best-fit model for the prediction of suicide attempts in Korean adolescents by analyzing every possible interaction between eleven sociodemographic and eight psychological variables. The results showed that: 1) the primary predictive variable for suicide attempts in adolescents was severity of depression; 2) juvenile delinquency was the primary variable for prediction of suicide attempts in the depression group; 3) intimacy with family was the primary variable in the potential depression group; and, finally, 4) stress was the primary predictive variable for suicide attempts in the non-depression group.

Decision tree analysis showed that severity of depression was the primary variable for prediction of suicide attempts in Korean adolescents. The rate of suicide attempts was 2.8 times higher in the depression group compared to the potential depression group. Compared to the non-depression group, the depression group had a 5.4-times higher rate of suicide attempts. These results are consistent with previous findings that have revealed depression as the strongest predictive variable for suicide ideation, plans, and attempts in all ages.31,32

Juvenile delinquency was the primary factor in the depression group. Those in the depression group with a high level of delinquency had a two-times higher risk for suicide attempts than those in the depression group with a low level of delinquency. Whereas predictors like depression or stress have been well studied, the influence of delinquency on suicide attempts has been found in relatively few studies.33,34 Moreover, coexistence rate of conduct disorder, which includes delinquency and depression is high.35 Our results suggest that risk for suicide attempts would rapidly increase if depressed adolescents showed delinquent behaviors in addition.

Furthermore, in the depression group, girls with a high level of delinquency had a 50% rate of suicide attempts. Interestingly, the rate of suicide attempts was 1.9 times higher in depressed females with a high level of delinquency compared to depressed males with a high level of delinquency, although there were more males with delinquent behavior. Combined with previous studies, the current results show that sex as well as delinquency are important predictive factors for suicide attempts.36 Because female delinquency is less likely than male delinquency, female delinquency could be deemed more serious by friends or others.34 Also, delinquency in female students could make it more difficult for them to adjust at school and lead to them experiencing depression or attempting suicide.

In contrast to the depression group, the potential depression group showed that intimacy with family was the primary variable in prediction of suicide attempts. Those in the potential depression group with a low level of family intimacy showed a 2.5-times higher rate of suicide attempts compared to those in the potential depression group with a high level of family intimacy. These results are consistent with previous studies that also found family-related variables are important in the prediction of adolescent suicide.37,38 The intimacy with family variable was included in the predictive model of our study instead of self-esteem, optimism, community support and interaction, and school adjustment. This implies that family-related variables should receive more focus in the prediction of suicide attempts in adolescents. Intimacy with family and psychological support from family can be buffers for stressful situations such as an entrance examination for college.32 Therefore, it is strongly suggested that clinicians should pay more attention to depression and family function in preventing suicide attempts and completed suicide.

In the potential depression group, middle school students with low family intimacy had a 2.5-times greater risk for suicide attempts compared to high school students with low family intimacy. In middle school students, stresses from school work, peer relationships, and school adjustment are radically increased, while their self-identity and cognitive abilities (eg, integrated thinking, cognitive flexibility) are less developed. In this situation, intimacy with family can be an important factor to buffer these stresses arising from psychological and environmental issues. It is important for parents to share emotional intimacy with them in order to prevent completed suicide in adolescence.

In the non-depression group, those with high levels of stress had a 3.8-times greater risk for suicide attempts compared to those with middle levels of stress,39,40 while those with low levels of stress had an almost equal propensity to attempt suicide as those with high levels of stress. Research by Feskanich et al in 94,110 female nurses, similar to this study, showed that the rate of suicide attempts of those with minimal and moderate levels of stress was higher than that of people who reported light levels of stress.41 Why is there a U-shaped curve (the
rate of suicide attempt of person with mild or middle level of stress) around stress and suicide attempts for this without depression? Adolescents who reported mild levels of stress in this study may deny stress, and a coping mechanism of denial may be another risk factor of suicide attempts.42

In addition, those who reported middle levels of stress in the non-depression group showed a different rate of suicide attempts according to their school record. Students with the “highest” (excellent) school record had a seven-times greater risk for suicide attempts compared to those with “under highest” (good, average, poor, and bad). Students who have been very good at schoolwork are relatively more likely to experience academic stress than other students. Some of these students become frustrated and can perform extreme behaviors even if their grade falls a little.

In this study, self-esteem and school adjustment variables were excluded in the present model, despite those variables having been considered to be important predictive factors in previous studies.18,20 This implies that these excluded variables may play a less significant role in interaction with other variables for prediction of suicide attempts.

The clinical values of this study are as follows. Firstly, suicide attempt was used as the dependent variable, which is the most relevant variable with completed suicide. As a result, more accurate prediction of completed suicide could be made. Bridge et al showed that a person who has attempted suicide is at an over 60-times greater risk for committing completed suicide compared to those who have never attempted suicide.43 Secondly, decision tree analysis made it possible to analyze every possible interaction between sociodemographic and psychological variables to improve the precision of the prediction model. Thirdly, considering depression severity in predicting suicide attempts seems be a new direction. Results of the current study indicate that a different approach according to severity of depression is required in order to prevent suicide attempts in adolescents.

**Conclusion**

Our model suggests that adolescents with different severities of depression show distinctive pathways to suicide attempts. Severity of depression should be considered above all, to predict suicide attempts. Delinquency and sex were important variables to predict suicide attempts in depressed adolescents. Intimacy with family was a significant variable to predict suicide attempts in the potential depression group. Interestingly, school performance record was a significant variable to predict suicide attempts in the non-depression group of Korean adolescents.

**Acknowledgment**

This study used the data from the survey of mental health in 2011 from the National Youth Policy Institute. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and technology (No. 2012R1A1A2043992).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**