Original Article
http://dx.doi.org/10.3349/ymj.2014.55.1.254
pISSN: 0513-5796, eISSN: 1976-2437
Yonsei Med J 55(1):254-263, 2014
YMJ

Psychological and Social Factors Affecting Internet Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends

Tae Min Song,[1] Juyoung Song,[2] Ji-Young An,[3] Laura L. Hayman,[4] and Jong-Min Woo[5]

[1]Research Center on Psychosocial Health, Korea Institute for Health and Social Affairs, Seoul;
[2]International Center for Criminal Justice, Korean Institute of Criminology, Seoul;
[3]u-Healthcare Design & Healthcare Service Design Development Program, Design Institute, Inje University, Seoul, Korea.
[4]College of Nursing & Health Sciences, University of Massachusetts-Boston, Boston, MA, USA.
[5]Department of Psychiatry, Seoul Paik Hospital, Inje University, Seoul, Korea.

Received: June 21, 2013
Revised: October 14, 2013
Accepted: October 16, 2013
Corresponding author: Dr. Ji-Young An,
u-Healthcare Design & Healthcare Service
Design Development Program,
Design Institute, Inje University,
Room # 710, 31 Supyo-ro, Jung-gu,
Seoul 100-032, Korea.
Tel: 82-2-2270-0904, Fax: 82-2-2270-0517
E-mail: ajy0130@inje.ac.kr

This study was presented at the Spring Symposium of the Korean Society of Medical Informatics, "Big Data & Cloud for Hospitals" which was held in Seoul, Korea (June 13-14, 2013).

· The authors have no financial conflicts of interest.

Purpose: The average mortality rate for death by suicide among OECD countries is 12.8 per 100000, and 33.5 for Korea. The present study analyzed big data extracted from Google to identify factors related to searches on suicide in Korea. **Materials and Methods:** Google search trends for the search words of suicide, stress, exercise, and drinking were obtained for 2004-2010. Analyzing data by month, the relationship between the actual number of suicides and search words per year was examined using multi-level models. **Results:** Both suicide rates and Google searches on suicide in Korea increased since 2007. An unconditional slope model indicated stress and suicide-related searches were positively related. A conditional model showed that factors associated with suicide by year directly affected suicide-related searches. The interaction between stress-related searches and the actual number of suicides was significant. **Conclusion:** A positive relationship between stress- and suicide-related searches further confirmed that stress affects suicide. Taken together and viewed in context of the big data analysis, our results point to the need for a tailored prevention program. Real-time big data can be of use in indicating increases in suicidality when search words such as stress and suicide generate greater numbers of hits on portals and social network sites.

Key Words: Internet, suicide, prevention and control, psychological stress, statistical models

INTRODUCTION

In 2010, the suicide rate in Korea was 33.5 per 100000, far above the average of 12.8 per 100000 reported in Organization for Economic Co-operation and Development (OECD) countries.[1] With an increase of 101.8% between 2000 and 2010, Korea has experienced the highest increase in suicide rate among OECD countries. Additionally, suicide rates among teenagers and the elderly were reported to be extraordinarily high compared to other OECD countries.[2] The reported causes of suicide in Korea include economic issues, family problems, depression, and anxiety about the future. According to one survey, 15.6% of adults from the general population seriously considered suicide at least once in their lives.[3] In 2007, data obtained from the National Statistical Office reported that 56.5% of teenagers were distressed by heavy study loads and that 95% of employees were suffering from stress.[4]

The Korean Statistical Information Service defines suicide as "intentional self-harm." As well, "accidental poisoning" or "poisoning with undetermined intent" can also be deemed to fall under the category of suicide.[5] Rather than being a one-time event, suicide is actually a series of steps or a process that may include "suicidal ideation→suicide planning→suicidal attempts→suicidal behavior",[6] thus intervention to prevent this series of suicidal processes is needed.[5,7]

To date, research on suicide has focused on psychological, biological, medical, and socio-environmental factors in order to determine risk factors and causes. As noted above, suicide is generally the result of a compounding of social and psychological stressors associated with depression or alcoholism.[8] Such disorders have been shown to progress negatively over time, resulting in suicide.[9,10] Hence, this study was conducted within a theoretical framework based on the stress-vulnerability model in an attempt to explicate the causes of suicide in Korean adults.

Being exposed to extreme amounts of stress, either acute or chronic, can threaten not only an individual's capabilities or resources, but also the person's well-being; such threats are signs of suicidal ideation and suicidal behaviors.[9-16] Reportedly, suicidal behaviors are a result of an inability to handle what is requested or expected in an individual's life, which stems from the stress that arises from the perception of such a dire reality.[10,17] Even though individuals may experience stress of the same intensity, its effects with regard to suicidal behaviors vary between individuals. Many studies have examined potential factors that may mediate the relationship between stress and suicidal behaviors: cognitive factors,[10,14] such as self-worth, family support, and social support, as well as lifestyle factors,[16,18,19] such as smoking, drinking, physical exercise, and nutrition, have been reported in the literature.

Factors related to lifestyle and problematic drinking behaviors are especially reported to be associated with higher odds of abusing other addictive substances and provoking suicidal impulses in situations where it is difficult to maintain self-control.[18-20] Among alcoholics, suicidal attempt rates are 10 times higher than those of their non-alcoholic counterparts.[20] On the contrary, among lifestyle-related behaviors associated with better health, exercise is reported to be a protective factor for preventing diseases or facilitating recovery from diseases, as well as for maintaining and promoting optimal mental health.

News from newspapers, broadcasts, and the Internet focusing on suicide may popularize suicides.[21] Celebrity suicides, in particular, which usually garner detailed news coverage, can cause the Werther Effect.[21-25] It has been reported that this can increase the risk of suicide up to five times or even 14.3 times.[22,24] In Korea, the number of suicides in March 2005 was 1309, which was after a celebrity had just committed suicide, almost twice that of the previous month (736 cases).[26] According to a study,[23] after a celebrity suicide case in 2008, a much higher number of individuals who had attempted suicide using the same method as the celebrity used were admitted to the emergency room compared to the previous period.

Also, among cases of divorce or bereavement, the suicide rate is higher than that of their counterparts.[27] As a result of analyzing the association of suicide and socio-economic variables such as divorce rate, birth rate, female labor force participation rate, migration, income, education level, etc., a high level of urbanization was shown to be associate with lower rates of suicide.[28,29]

As mentioned above, establishment of systematic strategies to prevent suicides, as well as conducting studies that include macroscopic variables (time sequential, spatial) that consider social influences, is needed. It is difficult; however, to respond effectively in the early stages of suicidal behavior, particularly where social and psychological factors have an extensive impact. Thus, it may prove difficult to establish a plan for predicting and managing suicide because of the complexity of suicidal impulses. In such situations, analysis of big data may be effective in managing suicide at the national level.[30,31]

Big data can be understood as extremely large data sets, which cannot be collected, stored, managed, and analyzed via conventional approaches by the use of, for instance, database management systems. In public sectors, big data has been used in disease prevention, prediction, treatment, and patient management through sharing genetic and biological resources.[32,33]

Over the recent years, the amount of data transmitted via smart devices and social network services has exponentially increased. Currently, data are recognized as a economic assets:[32,34] By analyzing big data, the Obama administration and other government bodies, as well as even multi-national information technology companies all over the world, are able to generate meaningful and useful information.[34]

For example, the National Institutes of Health in the U.S. reduced their expenses by $50 million per year by means of a Pillbox service that employs two-way interaction between manufacturers and users to provide information on various medications at the users' request. Google provides a real-time flu forecast service by analyzing the spread of the flu worldwide through analysis of search words by users on the web. Such applications of big data and the development of analysis methods may allow for more accurate predictions of various aspects of our society. For this reason, big data is recognized as a cost-effective resource.[33]

The existing methods for identifying suicide related factors, such as questionnaire surveys and clinical studies, have the advantage of enabling investigation of related variables based on data obtained from an individual; however, these also involve limitations. The magnitude of the association between suicide and these investigated variables remains to be clarified. Suicide is a complex and social phenomenon, and its diffusion effect is macroscopic under both time sequential and spatial aspects, as suggested by the Werther Effect, which highlights the importance of applying big data to the analysis of suicide related risk factors. Nevertheless, no studies have analyzed big data related to suicide. The purposes of this study were to utilize big data obtained from Google statistics and to analyze the determinants of suicide-related searches by means of a multi-dimensional analysis.

MATERIALS AND METHODS

Research design and keyword search

The research design of this study was based on reviewed literature demonstrating that stress is closely related to depression and suicide, and that lifestyle factors related to health have mediating effects on the relationship between stress and suicide. This study, therefore, was designed to analyze the determinants of the number of suicide-related searches in Korea using big data. The institutional review board of Inje University Seoul Paik Hospital approved this research.

This study employed Google search trends, a method that utilizes analyses of search-word inputs by users worldwide to provide standardized statistics for the number of searches on a specific search word carried out in a specific region at a specific time extracted according to a certain conditions. The Google search words used for this study, both English and Korean, included "stress" and "스트레스"; "drinking" and "음주"; "exercise" and "운동"; and lastly, "suicide" and "자살".

By means of a multi-dimensional analysis, an attempt was made to discover whether suicide rate by year, unemployment rate, after reports of a celebrity suicide, and the number of searches by month for stress, drinking, and exercise were compared with the number of searches on suicide (Fig. 1). In conducting the study, the following research questions were assessed:

1) Is there a difference between the amount of stress of the Korean people and the number of suicide-related searches by year?

2) Do monthly factors (stress, drinking, and exercise according to number of related searches) affect suicide-related search numbers?

3) Do monthly factors and yearly factors of Korean people affect suicide-related search numbers?

4) Is there an interaction between monthly factors and yearly factors on the number of suicide-related searches in Korea?

5) Do suicide-related search numbers affect the stress-related search numbers in Korea?

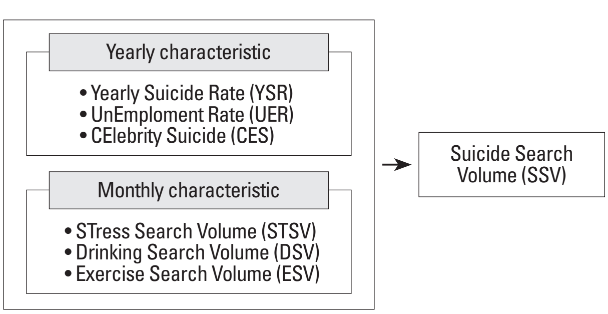

Fig. 1. Research model.

Data analysis

This study analyzed the relationship between the number of searches for the four keywords mentioned above at a primary level (month) and Korea's suicide rate, unemployment rate, and after reports of a celebrity suicide or not at a secondary level (year) from January 1, 2004 to December 31, 2010, building a multi-level model (Fig. 2). In the figure, the peak refers to the point of suicide incident when a famous politician committed suicide in Oct. 2010.

A multi-level model is commonly called a hierarchical linear model (HLM), which is used when predictions are to be made of dependent variables not only with primary-level variables but also with higher ranked variables. A multi-level model can reflect all varieties and characteristics of both lower and higher levels of data collection.[35]

To estimate parameters in this study, restricted maximum likelihood was used, which considers decreases in the degree of freedom in a fixed effect during the process of estimating the variance in a random effect.[35] For the final estimation of the fixed effect, a robust standard error was applied, which does not assume that the distribution of dependent variables is a normal distribution. Stress, exercise, and drinking-related searches at the primary level were entered as group means and suicide rates at the secondary level as the grand mean. Then, the variables were entered into the model. SPSS software, version 20.0 (SPSS Inc., Chicago, IL, USA), was used for the descriptive statistics analysis, and HLM software, version 7.0 (SSI Inc., Chicago, IL, USA), was used for the multi-level model analysis. The basic form of HLM is as follows:

<Level 1 model>

$Y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+r_{ij}$

Y_{ij}: Suicide search numbers per month i of year j

β_{0j}: Intercept of year j

β_{1j}: Regression coefficient of variable X for year j

X_{ij}: Independent variables (stress, drinking, and exercise-related search numbers) in month i of year j

r_{ij}: Residual at level 1 (month) that is not explained by level 1 prediction variables owing to random effects in month i of year j

<Level 2 model>

$\beta_{0j}=\Upsilon_{00}+\Upsilon_{01}W_j+u_{0j}$

$\beta_{1j}=\Upsilon_{10}+\Upsilon_{11}W_j+u_{1j}$

Υ_{00}, Υ_{10}: Level 2, in other words, year model's intercept

Υ_{01}, Υ_{11}: Regression coefficient at level 2

W_j: Prediction variables at level 2

u_{0j}, u_{1j}: Residual by year that did not explain the characteristics of level 2 (year) due to random effects at level 2.

For a multi-dimensional analysis to examine the determinants of suicide-related searches, the following four models were employed in this study:

Model 1 (Basic model): $SSV_{ij}=\Upsilon_{00}+u_{0j}+r_{ij}$

Model 2 (Unconditional slope model):

$SSV_{ij}=\Upsilon_{00}$
$+\Upsilon_{10}*STSV_{ij}$
$+\Upsilon_{20}*DSV_{ij}$
$+\Upsilon_{30}*ESV_{ij}$
$+u_{0i}+u_{1j}*STSV_{ij}+u_{2j}*DSV_{ij}+u_{3j}*ESV_{ij}+r_{ij}$

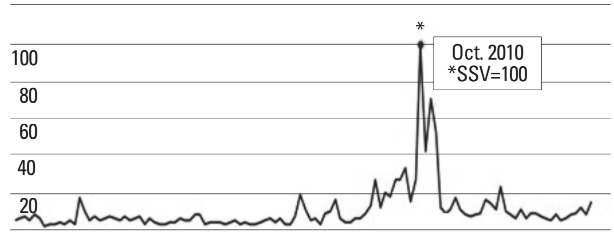

Fig. 2. Suicide-related search results from Google Trends (http://www.google.co.kr/trends/) in Korea. The axis of X refers to month by year. The axis of Y refers to suicide search volume. *Peak refers to the point of suicide incident when a famous politician committed suicide.

Model 3 (Condition model):

$SSV_{ij}=\Upsilon_{00}+\Upsilon_{01}*YSR_j$
$+\Upsilon_{10}*STSV_{ij}$
$+\Upsilon_{20}*DSV_{ij}$
$+\Upsilon_{30}*ESV_{ij}$
$+u_{0j}+u_{1j}*STSV_{ij}+u_{2j}*UER+u_{3j}*CES+r_{ij}$

Model 4 (Interaction model):

$SSV_{ij}=\Upsilon_{00}+\Upsilon_{01}*YSR_j$
$+\Upsilon_{10}*STSV_{ij}+\Upsilon_{11}*YSR_j*STSV_{ij}$
$+\Upsilon_{20}*DSV_{ij}$
$+\Upsilon_{30}*ESV_{ij}$
$+u_{0j}+u_{1j}*STSV_{ij}+r_{ij}$

RESULTS

Descriptive statistics for major study variables

A descriptive analysis was conducted to test the normality of variables (Table 1). Skewness and kurtosis appeared to meet the normality assumptions.[36] The suicide rate in Korea showed an increasing trend, and the volume of Google searches related to suicide showed a trend similar to the actual suicide rate of Korea. In particular, the volume of suicide-related searches increased in 2005, 2008, and 2010 after reports of celebrity suicides, indicating a risk for copycat suicides (Fig. 3). Table 2 shows the suicide rates and monthly suicide-related search volume (SSV) of Korea and other OECD countries. For Korea, both "suicide" and "자살" were used as search words.

Multi-level model analysis

The results of a multi-dimensional analysis for the determinants of suicide searches are shown in Table 3. By analyzing yearly level variance regarding monthly SSV when independent variables were not entered in the examination of research question 1, Model 1 was employed to test if there is a difference in SSV by year through a multi-level analy-

Table 1. Descriptive Statistics for Study Variables (unit: %, search volume)

Yr (Suicide rate)	Suicide			Stress			Rinking			Exercise		
	Mean±SD	K	S	Mean±SD	K	S	Mean±SD	K	S	Mean±SD	K	S
2004 (29.5)	45.2±8.3	-1.86	-0.22	116.8±32.9	0.0	0.50	74.6±23.0	2.04	-1.11	135.6±24.2	-0.92	-0.55
2005 (29.9)	58.9±18.5	3.38	1.37	139.2±15.8	0.39	1.28	78.8±20.5	0.76	1.11	143.8±27.5	-1.44	-0.41
2006 (26.2)	42.2±7.2	-1.03	-0.55	124.5±24.9	-0.32	0.58	74.3±16.9	-1.40	0.25	128.4±17.2	-0.64	0.03
2007 (28.7)	47.3±7.9	-0.62	-0.14	99.5±14.4	-0.35	-0.53	64.0±9.4	-0.40	-0.01	96.8±14.4	-0.18	-0.21
2008 (29.9)	55.8±22.3	6.93	2.52	99.9±15.2	-1.76	-0.19	74.3±8.4	-0.98	0.02	104.2±16.8	-0.41	-0.68
2009 (33.8)	58.8±19.1	4.54	1.96	109.0±25.0	-0.43	-0.55	85.8±18.4	0.43	0.87	115.3±21.0	-1.12	-0.38
2010 (33.5)	78.3±20.1	2.34	1.16	106.8±19.9	-1.37	0.22	92.5±21.7	-0.57	0.09	116.3±22.4	-0.86	0.60

K, Kurtosis; S, Skewness.

Table 2. Suicide Rate and Suicide-Related Search Volume among OECD Countries (unit: %, search volume)

Country	2005 yr		2006 yr		2007 yr		2008 yr		2009 yr		2010 yr	
	Suicide*	SSV†	Suicide*	SSV†	Suicide*	SSV†	Suicide*	SSV†	Suicide*	SSV†	Suicide*	SSV†
United States	11.2	74.2	11.3	64.2	11.7	59.7	12.0	61.0	-	57.5	-	57.4
United Kingdom	6.7	87.0	6.7	77.0	6.3	63.8	6.9	67.0	6.8	56.8	6.7	53.4
Australia	10.3	79.1	10.4	69.6	10.8	60.2	10.8	54.5	10.5	49.9	10.6	51.0
South Korea	29.9	58.9	26.2	42.2	28.7	47.3	29.0	55.8	33.8	58.8	33.5	78.3

OECD, Organization for Economic Co-operation and Development.
*Suicide rates per 100,000 (OECD Health Data, 2012).
†SSV refers to Monthly Suicide Search Volume, which means the number of searches on suicide compared to the total number of searches performed on Google (likelihood for search in a specific region at a specific time).

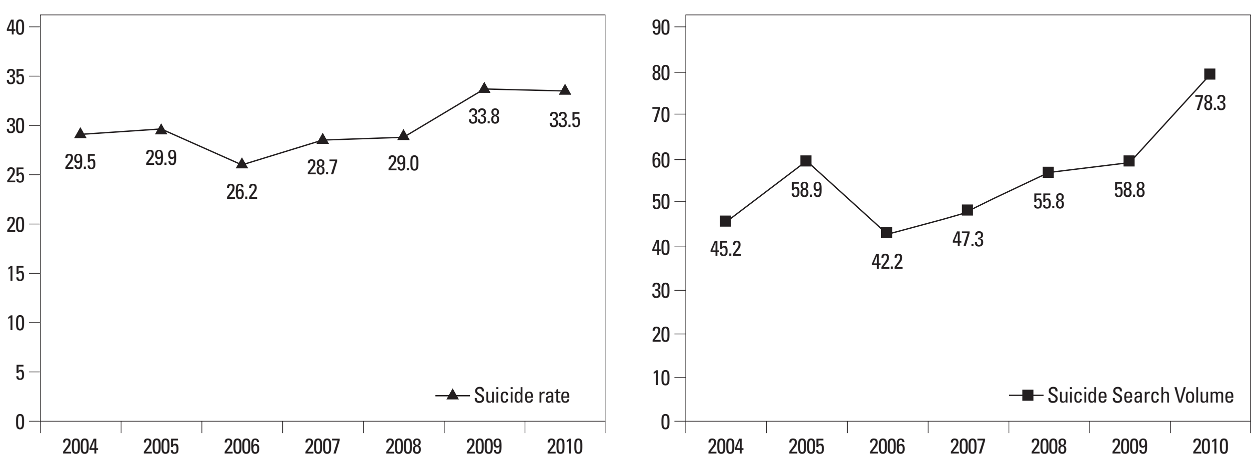

Fig. 3. Suicide rate and suicide-related search volume in Korea.

sis. As a result of the fixed effect analysis, the probability that the number of Google searches in Korea per month would reach an average of 55.20 times was statistically significant (β=55.20, p<0.001). As a result of the random effect analysis, both the monthly level variance (δ^2=256.61) and yearly level variance (δ^2=127.98) appeared to be statistically significant (χ^2=41.91, p<0.001).

The calculation of the variance ratio of yearly SSV through an intra-class correlation coefficient (ICC), which shows similarity among the lower levels belonging to the same level, yielded the following results:

Variance ratio that is explained by a difference in level 2 (yearly)

=[Level 2 (year) variance]/[Level 1 (month) variance+Level 2 (year) variance]

=127.98/(256.61+127.98)

=0.33

This showed that yearly level variance accounted for about 33.2% of the total variance explained regarding monthly SSV; consequently, the monthly level variance was shown to make up about 66.8% of the total variance explained.

Table 3. Multi-Level Model Analysis of Suicide-Related Searches

Parameter	Model							
	Model 1 Unconditional model		Model 2 Unconditional Slope model		Model 3 Conditional model		Model 4 Interaction model	
Fixed effect	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Level 1								
Intercept (Υ_{00})	55.20	4.28 (p<0.001)	55.20	4.28 (p<0.001)	55.93	36.16 (p=0.220)		
STSV			0.38	0.17 (p=0.071)	0.24	0.15 (p=0.161)		
DSV			-0.08	0.10 (p=0.437)	-0.07	0.10 (p=0.490)		
ESV			-0.22	0.11 (p=0.104)	-0.13	0.11 (p=0.214)		
Level 2								
YSR					3.35	0.87 (p=0.031)		
UER					-1.79	10.24 (p=0.872)		
CES					13.01	4.00 (p=0.048)		
Interaction								
STSV×YSR							0.07	0.02 (p=0.032)
Random effect	δ^2	χ^2	δ^2	χ^2	δ^2	χ^2	δ^2	χ^2
Level 2, u_0	127.98	41.91 (p<0.001)	135.34	50.32 (p<0.001)	9.11	4.83 (p=0.183)	45.85	17.10 (p=0.005)
Level 1, r	256.61		213.73		226.29		227.33	
STSV			0.20	12.41 (p=0.053)	0.07	16.40 (p=0.012)	0.04	9.91 (p=0.077)
DSV			0.02	5.29 (p>0.500)				
ESV			0.08	2.52 (p>0.500)				
ICC	0.332		0.387		-		0.168	
Deviance	710.25		707.12		685.79		707.33	

STSV, stress-related search volume; DSV, drinking-related search volume; ESV, exercise-related search volume; YSR, yearly suicide rate; UER, unemployment rate; CES, celebrity suicide; ICC, interclass correlation coefficient.

This study therefore rejected the null hypothesis of χ^2, which states that suicide-related searches would vary the same amount across years as the monthly averages for suicide-related searches varied for a single year. In the model, deviance was revealed to be 710.25.

Generally, if ICC is greater than 0.05, one can suppose that there is an intergroup variation, and even if ICC is less than 0.05, a multi-level analysis can be conducted when there is an experiential research result regarding intergroup variation.[37] This result supports the idea that analyzing a multi-level model is valid if all the variables are entered at monthly and yearly levels; nevertheless, although SSV is affected by monthly factors, the influence of yearly factors cannot be ignored.

Model 2 was performed to address research question 2. The effects of monthly factors on SSV were estimated through fixed effects, and the random effects were analyzed to see if these individual factors showed differences by year. As a result of fixed effects, drinking-related search volume (DSV) and exercise-related search volume (ESV) appeared not to have any effects on SSV, but stress-related search volume (STSV) appeared to have some effect (β=0.38, p=0.071). As a result of random effects, STSV was statistically significant (χ^2=12.41, p=0.053), and there were differences by year (χ^2=50.32, p<0.001). Therefore, the necessity of entering yearly variables was supported. In other words, SSV increased as STSV increased monthly, and this effect showed that there were differences by year. The ICC for yearly SSV was calculated as 0.39 and the deviance was 707.12.

Model 3 was designed to test research question 3, and analyzed SSV by including both yearly and monthly factors in the model. This model also included DSV and ESV, which were processed as fixed unknowns in the analysis because they were not significantly associated with STSV. As a result of fixed effects regarding the suicide search results, none of the monthly factors was statistically significant. Yearly suicide rate (YSR) appeared to affect SSV to a statistically significant degree (β=3.35, p=0.031). In other words, increases in YSR indicate increases in SSV. The number of suicides by a celebrity for a year showed a statistically significant effect on SSV (β=13.01, p=0.048). However, unemployment rate by year did not. As a result of random effects, although STSV was statistically significant (χ^2=16.40, p=0.012), it did not show a difference by year (χ^2=4.83, p=0.183). In the model, deviance was 707.12.

Model 4 was designed to test the interaction of STSV as a monthly factor and YSR as a yearly factor. The result was statistically significant (β=0.7, p=0.032). This result indicated that there is a difference in the relationship between stress-related search volume and suicide-related search volume according to YSR. In other words, if there are a large number of stress-related searches, the number of suicide-related searches increases as well. This result also demonstrated that YSR has the effect of increasing the number of suicide-related searches. The ICC of yearly SSV was 0.168, and its deviance was 707.33.

Thus far, statistical tests were conducted on factors affecting SSV via a multi-level analysis based on the stress-vulnerability model. The fact that stress-related searches affected suicide-related searches was supported in this study, but there remained a need to further define this result and provide evidence for a good circular model: suicide-related searches affect stress-related searches and stress-related searches affect suicide-related searches. Therefore, another analysis was conducted to check for differences by year in relation to the effect of monthly suicide-related searches on stress-related searches (Table 4).

In the fixed effect of Model 1, the total average number of stress-related searches was 113.67, and significant variations existed by year, given the significant variances at monthly and yearly levels (t=22.62, p<0.001). ICC was 25.3%, and there were differences in STSV by year, as there was a significant variance at the year level (χ^2=30.42, p<0.001).

As a result of the fixed effect analysis regarding STSV in Model 2, STSV was significantly different by year (t=22.62, p<0.001), but SSV did not appear to affect STSV. This supports the validity of the one-way model: that is, stress-related searches increase suicide-related searches, but not vice versa. Yet in the results from the random effect analysis, SSV appeared to affect stress-related searches (χ^2=15.29, p=0.018). There was a difference in STSV by year (χ^2=33.69, p<0.001). The ICC of yearly suicide search numbers was calculated to be 0.28, and the deviance was 761.69.

DISCUSSION

The ultimate goal of this study was to analyze the determinants of searches concerning suicide utilizing big data. To this end, the stress-vulnerability model was employed as a theoretical basis for this study. As shown in this study, the suicide rate of Korea and the volume of suicide-related searches on Google showed similar trends, and further demonstrated that copycat suicides may result from broadcasting celebrity suicides. This finding was consistent with previous studies that have reported the serious impact of suicide-related press reports.[25,27]

Recent suicide rates and suicide-related search numbers in the major OECD countries have either remained stable or decreased, whereas Korea's suicide rate has risen to almost three times the average suicide rate of other OECD members. Based on the results of this study, Koreans conduct suicide-related searches on Google 55.2 times a month on average. Moreover, suicide-related searches and stress-related searches showed a year-by-year difference.

In the analysis to test the effects of stress, drinking, and exercise-related searches on suicide-related searches, stress-related searches appeared to have the greatest influence on suicide-related searches. In other words, more stress-related searches were associated with more suicide-related searches. This result is consistent with previous studies, reporting that stress directly affects suicide.[12-14,17-20] Therefore, it is

Table 4. Multi-Level Model Analysis of Stress-Related Searches

Parameter	Model					
	Model 1 Unconditional model			Model 2 Unconditional Slope model		
Fixed effect	Coef.	S.E.	t-ratio	Coef.	S.E.	t-ratio
Level 1						
Intercept, Υ_{00}	113.67	5.02	22.62 (p<0.001)	113.67	5.02	22.62 (p<0.001)
SSV				0.07	0.24	0.286 (p=0.785)
Random effect	SD	δ^2	χ^2	SD	δ^2	χ^2
Level 2, u_0	12.86	165.49	30.42 (p<0.001)	13.02	169.45	33.69 (p<0.001)
Level 1, r	22.09	488.02		20.99	440.64	
SSV				0.52	0.27	15.29 (p=0.018)
ICC		0.253			0.278	
Deviance		761.68			761.69	

SSV, suicide-related search volume; ICC, interclass correlation coefficient.

possible that, when there is a buzz or when many keyword searches related to stress or suicide are on search portals or social network services, suicide impulses could be curtailed by analyzing various types of data, including the age and search patterns of an individual user, in order to provide timely intervention.

Additionally, although monthly factors did not affect the number of suicide-related searches, the actual suicide rate and yearly factors strongly affected it. This result may be attributable to the fact that monthly factors were affected by controlling yearly suicide rate. It also can be interpreted that yearly suicide-related factors directly affected suicide-related searches. The interactions between stress-related searches and suicide rates were significant: a large number of stress-related searches in a year reporting high suicide rates were shown to affect suicide-related searches.

There are few previous studies on suicide risk factors utilizing big data. From this study, it was revealed that suicide searches and suicide rates were interrelated. Therefore, it may be possible to set up a systematic plan for preventing suicide at the governmental level. For instance, in Finland, governmental departments made an comprehensive effort to prevent suicide by implementing a national project (1986-1996) based on 1397 psychological autopsy reports related to suicide (including medical and social security data, as well as family, friends, and doctor interview materials, etc.). Following its implementation, the suicide rate in Finland was found to be decreasing as follows: 30.3 suicides per 100000 people in 1990, 20.4 suicides in 2004, and 17.3 suicides in 2010. Considering that the decrease in Finland's suicide rates was based on a review of offline reports at the government level, Korea may also now be in a position to utilize big data to set up a nationwide suicide prevention plan.[38]

For the health and welfare sector in Korea, governmental and public institutions are already managing many different types of big data; however, their utilization and application are still in the initial steps. To present efficient strategies for preventing suicide and providing tailored service by the use of big data, we suggest the following: first, collaborative and systematic approaches should be developed between governmental departments and private institutes to manage health and welfare-related big data in an integrated way. Currently, in Korea, the National Health Insurance Corporation, the Korean Food & Drug Administration, and other national research institutes manage big data. Such data are mainly collected and stored through web search portals run by private institutes or social network services. In the society, therefore, classifying data-driven information and securing personal data should be also considered to prepare another societal challenge.

Second, there is an urgent need to further develop a way of analyzing and processing big data in health and welfare sectors. To promote greater development, education programs are needed to train professionals in data management, so called "data scientists," in an effort strengthen their ability to find hidden information from large-scale, unstructured big data. Also, development of clouding computer services for which to store, classify, and analyze non-relational and atypical data should be at the core of this whole process. Additionally, preferential attention should be given to the development of relevant technology that can enable "collection→storage→analysis→deduction" of big data in the sectors of health and welfare.

Third, a policy should be set in place for data security and privacy protection regarding personal information while utilizing big data. However, at this point, relevant laws and systems are next to non-existent and are not even discussed in Korea. Therefore, for the prevention of privacy related crimes such as a cyber-civil-rights violation, strict controls on data processing, information access, and anonymity assurance must accompany the use of big data. To this effect, a policy should be established to ensure the balance of utilizing and protection of big data in health and welfare sectors.

For the prevention of suicide, the policy implications of this study are as follows: first, age-friendly intervention programs for suicide prevention, such as school-wide programs for youth, workplace programs for workers, and aging-friendly programs for the elderly, warrant development. Second, applications such as a "Respect for Life Online Gatekeeper" should be developed to reduce suicide-associated risks as revealed through the analysis of big data and to provide tailored programs in real time if the warning signs of suicide are detected. Third, suicide prediction models by region such as "Respect for Life Prediction System" should be developed to prevent suicide in real time, by analyzing suicidal behaviors by region.

Finally, the limitations of this study in the following: first, rather than analyzing individual characteristics, this study analyzed entire groups as a whole; therefore, applying the results to individuals may result in ecological errors. Second, this study used data obtained from Google search trends statistics, so its representativeness may be questioned. Therefore, it is recommended that big data also be collected through various channels to analyze factors related to searches on suicide, such as other web search portals, blogs, Internet cafes, various types of social network services, web boards, etc. Third, suicide BUZZ via the Internet shows a tendency to spread rapidly for the first week after its onset and has about the life cycle of three weeks.[39] Employing a time lag of one to three weeks for the analysis is therefore necessary; however, this study did not apply such a time lag. Future research should consider this issue. Fourth, this study analyzed search results obtained from unspecified individuals; therefore, it would be impossible to predict suicidal behavior therefrom. Future research should consider this issue by analyzing social big data including emotions and psychological behaviors related to suicide. Fifth, in this study, unemployment rate by year was not statistically significant in the model. Nevertheless, this result warrants further study.

In conclusion, this study utilized big data to analyze suicide related factors and was ultimately intended to suggest the use of big data, which exists in various forms in different areas of society, to establish suicide prevention plans at the national level. Training talented professionals and preparing related infrastructure should be systematically implemented so that big data can be employed for various purposes in health and welfare sectors in the future.

ACKNOWLEDGEMENTS

This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-413-G00006).

REFERENCES

1. OECD Health Data 2012. [accessed on 2013 April 7]. Available at: http://www.oecd-ilibrary.org/social-issues-migration-health/suicides_20758480-table10.
2. Chang YS. Actual conditions of suicide and policy challenges in Korea 2012. Issue Focus 2012;165:46.
3. Ministry of Health and Welfare. Epidemiological Investigation of Psychiatric Disorder Reality; 2011.
4. Stress Republic of Korea.The Korea Economic Daily. [accessed on 2013 January 11]. Available at: http://www.hankyung.com/news/app/newsview.php?aid=2007050274801&intype=1.
5. Bae SB, Woo JM. Suicide prevention strategies from medical perspective. J Korean Med Assoc 2011;54:386-91.
6. Harwood D, Jacoby R. Suicidal Behaviour Among the Elderly. In: Hawton K, van Heeringen K, editors. The International Handbook of Suicide and Attempted Suicide. West Sussex, UK: John Wiley & Sons, Ltd.; 2008. p.275-91.
7. United Nations. Prevention of Suicide: Guidelines for the Formulation and Implementation of National Strategies. United Nations; 1996.
8. Alloy LB, Hartlage S, Abramson LY. Testing the cognitive diathesis-stress theories of depression: Issues of research design, conceptualization, and assessment. In: Alloy LB, editor. Cognitive Processes in Depression. New York, NY: The Guilford Press; 1988. p.31-73.
9. Izadinia N, Amiri M, Jahromi RG, Hamidi S. A study of relationship between suicidal ideas, depression, anxiety, resiliency, daily stresses and mental health among Teheran university students. Procedia Soc Behav Sci 2010;5:1515-9.
10. Park JY, Lim YO, Yoon HS. Suicidal impulse caused by stress in Korea: focusing on mediational effects of existent spirituality, family support, and depression. Korean J Soc Welf Stud 2010; 41:81-105.
11. Jeong DH, Jeong MH. Path Analysis in a study on youth' stress, delinquency and suicide ideals. Korean J Fam Welf 2011;16:171-87.
12. Lazarus RS, Folkman S. Stress, Appraisal and Coping. New York: Springer Publishing Company; 1984.
13. Park BK. An Analysis of the developmental path of Adolescent' depression and suicidal ideation. Ment Health Soc Work 2009:31; 207-35.
14. Wilburn VR, Smith DE. Stress, self-esteem, and suicidal ideation in late adolescents. Adolescence 2005;40:33-45.
15. Woo CY, Park AC, Jeong HH. The structural relationship among human relations, stress, depression and suicidal ideation of adolescents by genders and level of school. J Educ Psychol 2010;24:19-38.
16. Yoon MS, Lee HS. The relationship between depression, job preparing stress and suicidal ideation among college students: moderating effect of problem drinking. Korean J Youth Stud 2012;19: 109-37.
17. Dixon WA, Heppner PP, Anderson WP. Problem-solving appraisal, stress, hopelessness, and suicide ideation in a college population. J Couns Psychol 1991;38:51-6.
18. Chung S. Influence of early onset of drinking and problem drinking on suicide ideation and attempt among Korean adolescents: Analysis of 2009 Korean Youth Risk Behavior Survey. J Korean Alcohol Sci 2011;12:15-27.
19. Tapert SF, Colby SM, Barnett NP, Spirito A, Rohsenow DJ, Myers MG, et al. Depressed mood, gender, and problem drinking in youth. J Child Adolesc Subst Abuse 2003;12:55-68.
20. Yoon MS. Suicidal ideation among alcoholics moderating effect of alcohol use. Ment Health Soc Work 2011;38: 113-40.
21. Kim IS. Third-person effect of celebrity suicide news: exploring the relationships among attitude toward celebrity suicide news, media use, the third-person effect and media regulation. J Commun Sci 2009;9:5-36.
22. Kim BS, Yun DH. Prevention of suicide infection on part of advance guard in media. Newsp Broadcast 2011;7:22-6.
23. Song YJ. The comparative study of change in characteristics of suicidal attempts before and after the celebrity suicide. Unpublished master's thesis, Graduate School of Yonsei University, Seoul, Korea; 2011.
24. Stack S. Suicide in the media: a quantitative review of studies based on non-fictional stories. Suicide Life Threat Behav 2005;35: 121-33.
25. Tousignant M, Mishara BL, Caillaud A, Fortin V, St-Laurent D. The impact of media coverage of the suicide of a well-known

Quebec reporter: the case of Gaëtan Girouard. Soc Sci Med 2005;60:1919-26.
26. Korea National Statistical Office. 2005 Cause-of-Death Statistics Results; 2006.
27. Stack S. Divorce and suicide: a time series analysis, 1933-1970. J Fam Issues 1981;2:77-90.
28. Kowalski GS, Faupel CE, Starr PD. Urbanism and suicide: a study of American counties. Soc Forces 1987;66:85-101.
29. Stack S. Suicide: a 15-year review of the sociological literature. Part II: modernization and social integration perspectives. Suicide Life Threat Behav 2000;30:163-76.
30. Koh SJ, Jeong YH. Health risk prediction using big health data. Health Welf Policy Forum 2012;193:43-52.
31. Song TM. Multivariate analysis on factors affecting suicide utilizing big data: Applying Google search trend. Issue Focus 2012; 186:1-8.
32. McKinsey Global Institute. Big data: The Next Frontier for Innovation, Competition, and Productivity; 2011.
33. National Information Society Agency. Evolved into the Big Data World (Case Study of Global for Big Data); 2012.
34. Policy Exchange. The Big Data Opportunity: Making Government Faster, Smarter and More Personal; 2012.
35. Raudenbush SW, Bryk AS. Hirearchical Linear Models: Application and Data Analysis Methods. 2nd ed. Thousand Oaks, CA: Sage; 2002.
36. Kline RB. Principles and Practice of Structural Equation Modeling. 3rd ed. New York, NY: Guilford Press; 2010.
37. Heck R, Thomas S. An Introduction to Multi-level Modeling Techniques. 2nd ed. New York, NY: Routledge; 2009.
38. Three issues of Korea. The Dong-A Daily. [accessed on 2013 January 11]. Available at: http://news.donga.com/3/all/20121029/50456912/1.
39. National Information Society Agency. Implications for Suicide Prevention Policy of Youth described in the Social Analysis; 2012.