

소셜 빅 데이터를 활용한 자살검색 요인 다변량 분석

송태민*, 송주영**, 안지영***, 진달래*

* 한국보건사회연구원

** 웨스트조지아대학교

*** 인제대학교

Multivariate Analysis of Factors for Search on Suicide Using Social Big Data

Tae Min Song*, Juyoung Song**, An Ji-Young***, Jin Dallae*

* Korea Institute for Health and Social Affairs

** University of West Georgia

*** Inje University

<Abstract>

Objectives: The study is aimed at examining the individual reasons and regional/environmental factors of online search on suicide using social big data to predict practical behaviors related to suicide and to develop an online suicide prevention system on the governmental level. **Methods:** The study was conducted using suicide-related social big data collected from online news sites, blogs, cafés, social network services and message boards between January 1 and December 31, 2011 (321,506 buzzes from users assumed as adults and 67,742 buzzes from those assumed as teenagers). Technical analysis and development of the suicide search prediction model were done using SPSS 20.0, and the structural model, and multi-group analysis was made using AMOS 20.0. Also, HLM 7.0 was applied for the multilevel model analysis of the determinants of search on suicide by teenagers. **Results:** A summary of the results of multivariate analysis is as follows. First, search on suicide by adults appeared to increase on days when there were higher number of suicide incidents, higher number of search on drinking, higher divorce rate, lower birth rate and higher average humidity. Second, search on suicide by teenagers rose on days when there were higher number of teenage suicide incidents, higher number of search on stress or drinking and less fine dust particles. Third, the comparison of the results of the structural equation model analysis of search on suicide by adults and teenagers showed that teenagers were more likely to proceed from search on stress to search on sports, drinking and suicide, while adults significantly tended to move from search on drinking to search on suicide. Fourth, the result of the multilevel model analysis of determinants of search on suicide by teenagers showed that monthly teenagers suicide rate and average humidity had positive effect on the amount of search on suicide. **Conclusions:** The study shows that both adults and teenagers are influenced by various reasons to experience stress and search on suicide on the Internet. Therefore, we need to develop diverse school-level programs that can help relieve teenagers of stress and workplace-level programs to get rid of the work-related stress of adults.

Key words: Social big data, Suicide search prediction model, SEM, HLM

I. 서론

우리나라는 최근 스마트폰의 보급이 확산됨에 따라 모바일 인터넷과 SNS 이용은 급속히 증가하고 있다. 2011년

7월 현재 우리나라의 만 3세 이상 인구의 인터넷 이용율은 78.0%이며 이중 만 6세 이상 인터넷 이용자의 66.5%가 1년 이내 SNS를 이용하고 있으며(Korea Communications Commission·Korea Internet & Security Agency, 2011), 만 6

Corresponding author : Juyoung Song

1925 Monroe dr. 1157 Atlanta GA 30324, University of West Georgia Assistant Professor of Criminology

* 이 논문은 ‘송태민(2013). 한국보건사회연구원(2013). 소셜 빅 데이터 분석을 통한 자살검색 예측모형 개발. 보건복지포럼(2013. 8)’의 내용을 수정, 보완한 것임을 밝히는 바이다.

Tel: 678-839-6337, Fax: 678-839-6506, E-mail: juyoung81@gmail.com

▪ 투고일: 2013.07.30

▪ 수정일: 2013.09.11

▪ 게재확정일: 2013.09.15

세~ 19세의 인터넷 이용자 중 78.9%가 친구관계를 강화하거나 일상의 스트레스를 해소하기 위해 SNS를 이용(Lee & Sung, 2012)하고 있다. 소셜미디어¹⁾는 ‘개개인의 주관적인 생각 또는 경험을 바탕으로 한 정보를 공유 및 재가공하는 등’ 참여, 소통, 공유’ 기반의 뉴미디어를 의미²⁾으로 대표적인 소셜미디어로는 페이스북과 트위터 등 SNS(Social Network Service)가 있다(National Information Society Agency, 2011). SNS는 개인, 집단, 사회의 관계를 네트워크로 파악하는 사회관계망 서비스를 의미하며³⁾ 실시간성과 가속성이라는 특징을 지녔기 때문에 어떠한 매체보다 이슈에 대한 확산속도가 빨라, 개개인의 단순한 생활 내용뿐만 아니라 정치, 경제, 사회·문화 등 사회 전반의 문제에 대한 이슈가 SNS를 통해 확산된다(National Information Society Agency, 2011). 특히, SNS는 청소년들이 일상생활 속에서 갖는 우울한 감정이나 스트레스, 고민을 들을 수 있고 행태를 이해할 수 있는 공간으로 SNS 상에서 나타나는 자살에 대한 감정표현이나 심리적 위기 행태들을 분석하게 되면 위험징후와 유의미한 패턴을 감지하여 자살을 예방하는데 긍정적 효과를 발휘할 수 있다(National Information Society Agency, 2012). 이러한 SNS를 통하여 전송되는 데이터의 양이 기하급수적으로 증가하여 데이터가 경제적 자산으로서의 가치를 인정받기 시작하면서 빅 데이터의 활용과 분석이 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하며 다양한 부문에서 빅 데이터의 적극적인 활용을 시도하고 있다⁴⁾. 빅 데이터는 기존의 데이터베이스 관리시스템(database management system)으로는 수집·저장·관리·분석이 어려운 정형 또는 비정형의 대용량 데이터로 의미 있는 새로운 가치를 추출해낼 수 있는 데이터라는 개념으로 이해할 수 있다(McKinsey Global Institute, 2011; National Information Society Agency, 2012). 우리나라의 보건복지 분야에서는 이미 수많은 빅 데이터가 정부 및 공공기관에서 관리되고 있으나 정보 접근의 어려움으로 활용이 미흡한 반면 민간기관의 검색포털이나 SNS에

서 관리되고 있는 빅 데이터의 분석과 활용은 활발히 이루어지고 있다.

한편, 우리나라는 급격한 사회·경제적 변화속에 자살에 의한 사망률(자살률)은 2010년 인구 10만 명당 33.5명으로 OECD(Organisation for Economic Co-operation and Development) 회원국 평균자살률 12.8명에 비해 높게 나타나고 있으며(OECD Health Data, 2012), 특히, 2000~2010년 기간 중 우리나라 자살률은 OECD 국가 중 가장 높은 증가율인 101.8%로 나타나고 있으며, 특히, 청소년과 노인의 자살사망률 증가폭은 다른 OECD 회원국에 비해 매우 큰 것으로 보고되었다(Chang, 2012). 우리나라의 자살의 원인은 계층별로 다양하게 나타나고 있다. 청소년은 학교폭력과 생활 스트레스로 인해 자살을 선택하는 비율이 높으며, 대학생은 2007년 이후 학자금 대출로 인한 신용불량자의 증대와 함께 생활고 및 취업난 등의 경제문제와 학업문제 등으로 자살률이 급속히 증가하고 있다⁴⁾. 2012년 청소년 통계조사에서 청소년의 8.8%가 자살을 생각해 본 적이 있는 것으로 나타났다(Statistics Korea, 2012). 베이비붐(baby boom) 세대는 1998년 IMF 경제위기와 2008년 글로벌 금융위기를 겪으면서 제대로 된 노후 준비도 없이 삶의 터전이 무너지고 조기은퇴로 인한 어려움으로 자살률은 증가하고 있으며, 노인은 신체적 질병이나 경제상태의 악화, 사회와 가족으로부터의 고립 등으로 우울이 증가하면서 자살을 선택하는 비율이 증가하고 있다(Song, 2012). 2010년 통계청 조사(Statistics Korea, 2010)에서 베이비붐 세대의 7.1%가 자살을 생각해 본 적이 있는 것으로 나타났다. 이와 같이 우리나라는 모든 계층에서 자살이 증가하고 있는 것으로 나타남에 따라 온라인/오프라인 빅 데이터의 분석을 통한 자살의 원인과 위험징후를 파악하여 자살을 예방할 수 있는 정부차원의 대책이 시급한 실정이다.

그동안 자살에 대한 연구는 자살원인을 밝히기 위하여 정신과적 요인, 생물학 및 의학적 요인, 그리고 사회·환경적 요인 등에 초점이 맞추어져 왔다(Bae & Woo, 2011). 자

1) 소셜미디어 유형은 블로그(blog), 소셜네트워크서비스(SNS), 콘텐츠커뮤니티(contents communities), 위키(wiki), 마이크로블로그(micro blog) 등(National IT Industry Promotion Agency, 2006)이 있으나 본 연구에서는 SNS로 통일함.

2) <http://ko.wikipedia.org/wiki/>. 2013. 7. 14.

3) 공공부문에서 유전자와 생명연구자원 공유를 통한 질병 예방 및 예측, 치료, 그리고 환자 관리 등에 활용되고 있다. 또한, 다국적 IT(information technology) 기업들과 웹(web) 검색 포털(portal) 사이트들은 서버에 저장된 빅 데이터를 분석함으로써 다양한 가치 정보를 생산해 내고 있다(Policy Exchange, 2012).

4) <http://ask.nate.com/qna/view.html?n=8975979>, <http://wbkn.tv/?bid=news&m=bbs&uid=2436>

살의 원인을 설명하기 위한 모델로는 사회학적 모델, 심리학적 모델, 유전·생물학적 모델, 스트레스취약성 모델 등 여러 가지가 있다. 그 중 스트레스취약성 모델은 자살은 단편적으로 이루어지는 것이 아니라 사회·심리적 자원과 스트레스 요인이 상호작용하여 우울증이나 알코올 중독 등의 질환이 발생하고(Alloy, Hartlage, & Abramson, 1988) 이들 질환이 시간을 두고 점차 진행되면서 자살에 이르는 일련의 과정을 거치게 된다는 것이다(Izadinia, Amiri, Jahromi, & Hamidi, 2010; Park, Lim, & Yoon, 2010).

스트레스는 개인의 능력 혹은 자원을 초과하여 안녕을 위협할 수 있으며(Lazarus & Folkman, 1984), 자살행동은 개인의 삶에 대한 요구 및 기대치에 적절하게 대처하지 못한 결과로 개인이 지각하는 현실에서의 스트레스에 인한 것으로 볼 수 있다(Dixon, Heppner, & Anderson, 1991). 동일한 강도의 스트레스를 경험하더라도 개인에 따라 자살행동에 미치는 그 영향 정도는 각기 다르기에 스트레스와 자살행동의 중재요인을 규명하기 위한 연구 또한 많이 진행되어 왔다. 선행 연구에서는 자아존중감과 같은 인지적 소질이나 가족지지와 사회적지지(Wilbum & Smith, 2005), 그리고 흡연, 음주, 운동, 영양 등의 생활습관을 중재요인으로 보고하고 있다. 생활습관 관련된 요인 가운데 특히 문제음주 행위는 다른 중독성 물질을 남용할 가능성이 높을 뿐만 아니라 자기통제가 힘든 상황에서 자살충동을 쉽게 유발하는 것으로 보고되었으며(Chung, 2011; Tapert et al., 2003; Yoon, 2011), 이와는 반대로 건강을 유지·증진하기 위한 생활습관 행위 가운데 운동 요인은 질병 예방 뿐 아니라 질병으로부터의 빠른 회복, 그리고 현재 건강상태를 유지·증진시킴으로써 궁극적으로 삶의 질을 향상시켜 스트레스를 감소시킬 수 있다고 보고되고 있다(Bae & Woo, 2011; Chapman & Beaulieu, 1983). 배우자의 이혼이나 사별한 경우 정상적인 가정보다 자살률이 높고(Smith, 1988; Stack, 1981) 경제적 어려움을 겪는 노인과 실업노인들 일수록 새로운 생활의 적응능력이 부족한 상황에서 자살의 요인이 된다고 하고 있다(McIntosh, 1995). 이혼율, 출생률, 여성의 경제활동

참가율, 인구이동, 소득, 교육수준 등의 사회적 변수를 사용하여 분석을 시도한 결과 도시화 수준이 자살과 강한 연관관계가 있으며(Kowalski et al., 1987) 높은 도시화의 수준이 자살을 낮춘다는 연구결과(Stack, 2000)가 있다. 오존과 일산화탄소와 같은 대기오염인자가 증가할수록(Biermann et al., 2009) 자살을 증가시키며 미세먼지가 증가할수록(Kim et al., 2010) 기온이 높을수록(Durkheim, 2008; Yang, Albert, Shin-Jen, & Norden, 2011) 자살이 증가한다는 연구결과가 있다.

이와 같이 그동안 자살의 연구는 통계자료와 조사데이터의 분석을 통하여 국가 간 자살률 비교나 자살요인 등에 초점을 맞춘 연구가 진행되어 왔으나 이러한 연구는 자살에 대한 개별적 변인을 보는 데에는 장점이 있으나, 개별 대상자로부터 파악한 변수들이 지역변수나 사회·환경적인 변수와 얼마나 어떻게 관련되는 지는 불분명하며, 실시간으로 원인을 분석하는데는 한계가 있다. 따라서 본 연구는 소셜 빅 데이터를 활용하여 자살검색의 개인별 요인과 지역·환경적 요인을 검증함으로써 자살 관련된 실질적인 행동을 예측하여 정부차원의 온라인 자살예방 대응체계를 마련하고자 한다.

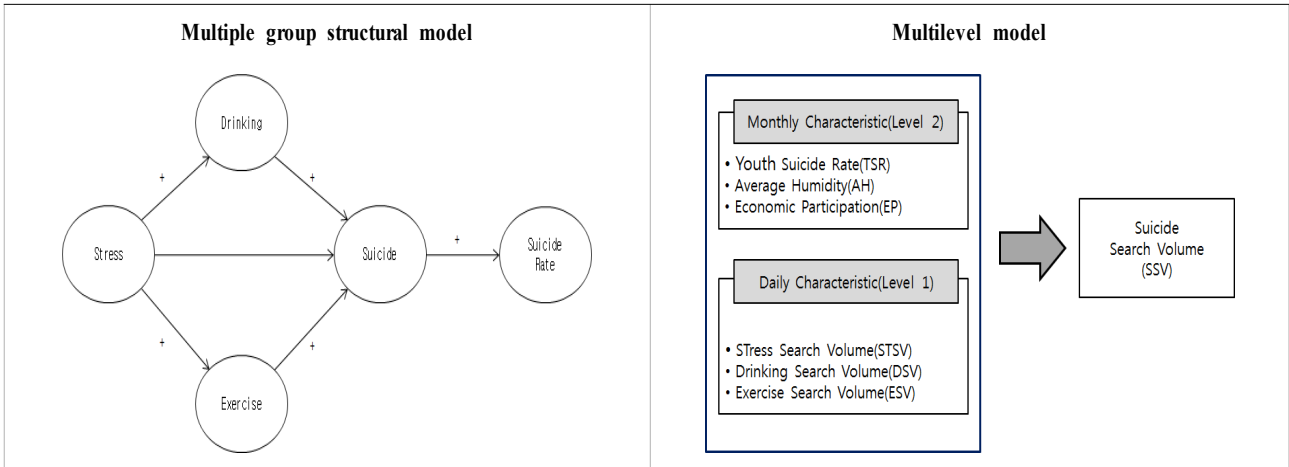
II. 연구 방법

1. 연구모형 및 연구가설

본 연구는 소셜 빅 데이터를 활용하여 우리나라의 자살 검색 요인과 자살검색 예측모형 개발을 위한 다변량 분석이 목적으로 성인과 청소년 집단의 자살요인의 비교 분석은 다중집단 구조모형⁵⁾으로 구성하였고[Figure 1]6), 월별 청소년 자살률, 기후, 경제활동과 일별 스트레스, 음주, 운동 검색량이 청소년 자살 검색의 결정요인에 미치는 영향을 분석하기 위하여 다층모형(Multi-level Model)으로 구성하였다[Figure 1]. 본 연구의 다층모형의 월별 요인은 이혼율, 출생률, 경제활동참가율 등이 자살과 연관이 있다는

5) 본 연구의 구조모형은 그동안의 연구에서 제안된 스트레스가 건강생활실천요인(음주, 운동)을 매개하여 자살에 영향을 미칠 것이라는 스트레스 취약모형을 적용하였으며, 다중집단 분석은 성인의 자살 검색량과 청소년의 자살 검색량을 비교 분석하였다.
6) 구조모형에 사용된 Stress, Drinking, Exercise, Suicide 변수는 소셜 빅 데이터에서 수집된 일일 검색량이며, Suicide Rate는 통계청 사망원인 통계에서 산출된 일일 자살자 수를 나타낸다. 그리고, 다중회귀분석에 사용된 Divorce와 Fertility는 월별 이혼자 수와 출생아 수를 나타내며 Humidity와 Dust는 일일 평균습도와 미세먼지량을 나타낸다.

연구(Kowalski et al., 1987; Stack, 1981)에 근거하였다.



[Figure 1] Research model

본 연구모형에 따른 구체적인 연구가설은 다음과 같다.
첫째, 성인과 청소년의 자살검색의 예측요인은 무엇인가?

둘째, 성인과 청소년의 자살검색 구조모형은 차이가 있는가?

셋째, 성인과 청소년의 자살검색 구조모형에서 건강생활실천요인(음주, 운동)의 매개효과는 있는가?

넷째, 청소년의 자살검색에 영향을 주는 일별 요인과 월별 요인은 무엇인가?

2. 연구대상 및 측정도구

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅 데이터⁷⁾를 대상으로 하였다. 2011. 1. 1~2011. 12. 31(365일)까지 해당 채널에서 자살관련 토픽(성인 추정 Buzz: 321,506건, 청소년 추정 Buzz: 67,742건)을 수집

하였으며, 수집된 토픽 중 스트레스(성인 Buzz: 14,504건, 청소년 Buzz: 9,232건), 음주(성인 Buzz: 2,241, 청소년 Buzz: 2,818건), 운동(성인 Buzz: 6,256, 청소년 Buzz: 6,629건) 토픽을 추출하여 분석하였다⁸⁾. 그리고 자살토픽의 수집은 요일별, 주말, 휴일을 고려하지 않고 매 시간단위로 수집하였다⁹⁾. 본 연구의 측정도구 중 종속변수로는 소셜 빅 데이터에서 수집된 자살 검색량을 사용하였으며 독립변수로는 소셜 빅 데이터에서 수집된 스트레스·음주·운동 검색량과 통계청 사망원인 통계자료의 2011년 일별 전체 자살자 수와 청소년(19세 미만) 자살자 수, 기상청의 일별 평균습도와 미세먼지량을 사용하였다. 그리고 자살검색 예측모형에는 월별 이혼자 수와 출산아 수 통계자료를 사용하였고, 다층분석의 일별 독립변수로는 스트레스·운동·음주 검색량 사용하였으며, 월별 독립변수로는 통계청의 2011년 월별 청소년 자살자 수, 월별 경제활동 참가율과 기상청의 2011년 월별 평균습도를 사용하였다.

7) 본 분석에서는 214개의 온라인 뉴스사이트, 4개의 블로그(네이버, 네이버, 다음, 티스토리), 2개의 카페(네이버, 다음), 2개의 SNS(트위터, 미투데이), 6개의 게시판(네이버 지식인, 네이버 지식, 다음 지식, 네이버톡, 네이버판, 뽀뿌) 등 총 229개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서를 소셜 빅 데이터로 정의함.

8) 본 연구를 위한 소셜 빅 데이터의 수집 및 토픽 분류는 '(주) SK텔레콤 스마트 인사이트'에서 수행하였으며 자료의 수집은 Crawler를 사용하였고, 토픽의 분류는 Text Mining 기법을 사용하였다.

9) 요일별 버즈 분석결과 청소년의 자살검색은 일요일과 월요일에 증가하여 금요일과 토요일에는 감소하는 추세를 보이는 것으로 나타났다.

3. 통계분석

본 연구의 구조모형 적합도 비교에는 증분적합지수 (Incremental Fit Index)인 NFI(Normed Fit Index), CFI (Comparative Fit Index)와 절대적합지수(Absolute Fit Index)인 RMSEA(Root Mean Squared Error of Approximation)를 사용하였다. 일반적으로 CFI를 비롯한 증분적합지수들은 0.9보다 크면 모형 적합도가 양호하다고 해석한다(Hair, Black, Babin, Anderson, & Tatham, 2006). RMSEA는 대표 본이나 다수의 관측변수들로 인해 발생하는 χ^2 통계량의 문제점을 보완하기 위해 개발된 적합지수이다. 일반적으로 RMSEA가 0.05이하이면 적합도가 매우 좋고, 0.05와 0.08사이의 값을 나타내면 양호하다고 해석하며, 0.10이상이면 적합도가 좋지 않다고 해석한다(Bollen & Long, 1993). 모형추정은 최대우도법(Maximum Likelihood [ML])을 사용하였으며, 결측치 문제의 해결은 관찰된 모든 변인의 모든 정보를 활용하여 결측자료의 특성을 고려하는 완전정보최대우도법(Full Information Maximum Likelihood [FIML])을 사용하였다. 본 연구의 간접효과에 대한 유의성 검증은 모든 자료가 정규성 분포를 따른다는 가정하에 유의성을 검증하는 Sobel Test(Preacher & Hayes, 2004)를 실시하였다. 그리고 본 연구의 다층모형의 모수 추정 방식은 무선효과(random effect)의 분산을 추정하는 과정

에서 고정효과(fixed effect)의 자유도 감소를 고려하는 한 정최대우도추정법(REstricted Maximum Likelihood [REML])을 사용하였다(Raudenbush & Bryk, 2002). 고정효과와 최종추정은 종속변수의 분포를 정상분포로 가정하지 않는 표준오차(Robust Standard Error)를 적용하였다. 본 연구의 기술분석과 자살검색 예측모형 개발은 SPSS 20.0을 사용하였고, 구조모형과 다중집단 분석은 AMOS 20.0을 사용하였다. 그리고 청소년 자살검색 결정요인의 다층모형 분석은 HLM 7.0을 사용하였다.

III. 연구 결과

1. 주요 변인들의 기술통계

본 연구의 주요 변인들의 다변량 정규성을 확인한 결과 소셜 빅 데이터에서 수집된 모든 변인(자살, 스트레스, 음주, 운동)의 첨도가 10이상인 것으로 나타나, 상용로그(lg10)로 치환하여 사용하였다. 그리고 모든 독립변수(전체 자살자수, 청소년 자살자 수, 평균습도, 미세먼지량, 경제활동참가율, 월별 이혼율, 월별 출산아 수)도 상용로그(lg10)으로 치환하여 사용하였다<Table 1>.

<Table 1> Descriptive statistics of variables

Year (Suicide Rate)	Suicide			Stress			Drinking			Exercise		
	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}
Adult	909.7±981.9	11.55	2.90	39.7±63.4	38.01	5.47	6.1±9.3	74.86	6.92	17.1±37.2	59.26	6.98
Youth	185.6±445.8	38.05	5.96	25.3±22.1	46.31	5.69	7.7±5.7	57.71	5.44	18.2±14.4	44.75	5.60
	Total Suicide Rate			Youth Suicide Rate			Humidity			Dust		
	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}	Mean±S.D.	K ^{a)}	S ^{b)}
	43.58±10.64	.31	.32	1.02±1.03	1.87	1.18	67.4±13.37	-.92	-.11	47.83±28.6	33.82	4.31

Note: ^{a)}Kurtosis, ^{b)}Skewness

2. 자살 검색량 예측모형

본 연구에서는 성인 자살 검색량의 예측을 위하여 전체

자살률, 음주검색량, 이혼율, 출산율, 평균습도를 사용¹⁰⁾ 하였으며 청소년 자살 검색량을 예측하기 위하여 청소년

10) 본고의 자살검색 예측 모형에 사용된 독립변수 중, 자살률은 일일 자살자 수를 의미하며, 이혼율과 출산율은 각각 월별 이혼자

자살률, 스트레스 검색량, 음주 검색량, 미세먼지량을 사용하였다¹¹⁾. 성인자살 검색의 예측모형은 <Table 2>와 같다. 모형 1은 2011년 일별 전체 자살률이 성인 자살검색에 미치는 영향을 예측하는 것으로 성인 자살자 수가 많을수록(자살률이 높을수록) 자살 검색을 증가하는 것으로 나타났다. 모형 2는 일별 전체 자살률과 음주 검색이 자살검색에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 음주 검색이 많을수록 성인의 자살 검색은 증가하는 것으로 나타났다. 모형 3은 일별 전체 자살률, 음주 검색, 이혼율이 자살검색이 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 음주 검색이 많을수록, 이혼율이 높을수록 자살 검색은 증가하는 것으로 나타났다. 모형 4는 일별 전체 자살률, 음주 검색, 이혼율, 출산율이 자살검색

에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 음주 검색이 많을수록, 이혼율이 높을수록, 출산율이 낮을수록 성인의 자살 검색은 증가하는 것으로 나타났다. 모형 5는 일별 전체 자살률, 음주 검색, 이혼율, 출산율, 평균습도가 자살검색에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 음주 검색이 많을수록, 이혼율이 높을수록, 출산율이 낮을수록, 평균습도가 높을수록 성인의 자살 검색은 증가하는 것으로 나타났다. <Table 2>에서 자살 관련 Buzz가 급속히 전파되는 1주간의 시차(time lag)를 적용하여¹²⁾ 자살률(Independent Variable: Suicide rate(1 week time lag))에 미치는 영향을 살펴본 결과, 예측도가 조금 향상된 것으로 나타났다¹³⁾. 성인의 자살검색을 예측하는 모형의 그래프는 [Figure 2]과 같다.

<Table 2> Prediction model of suicide search volume in adults

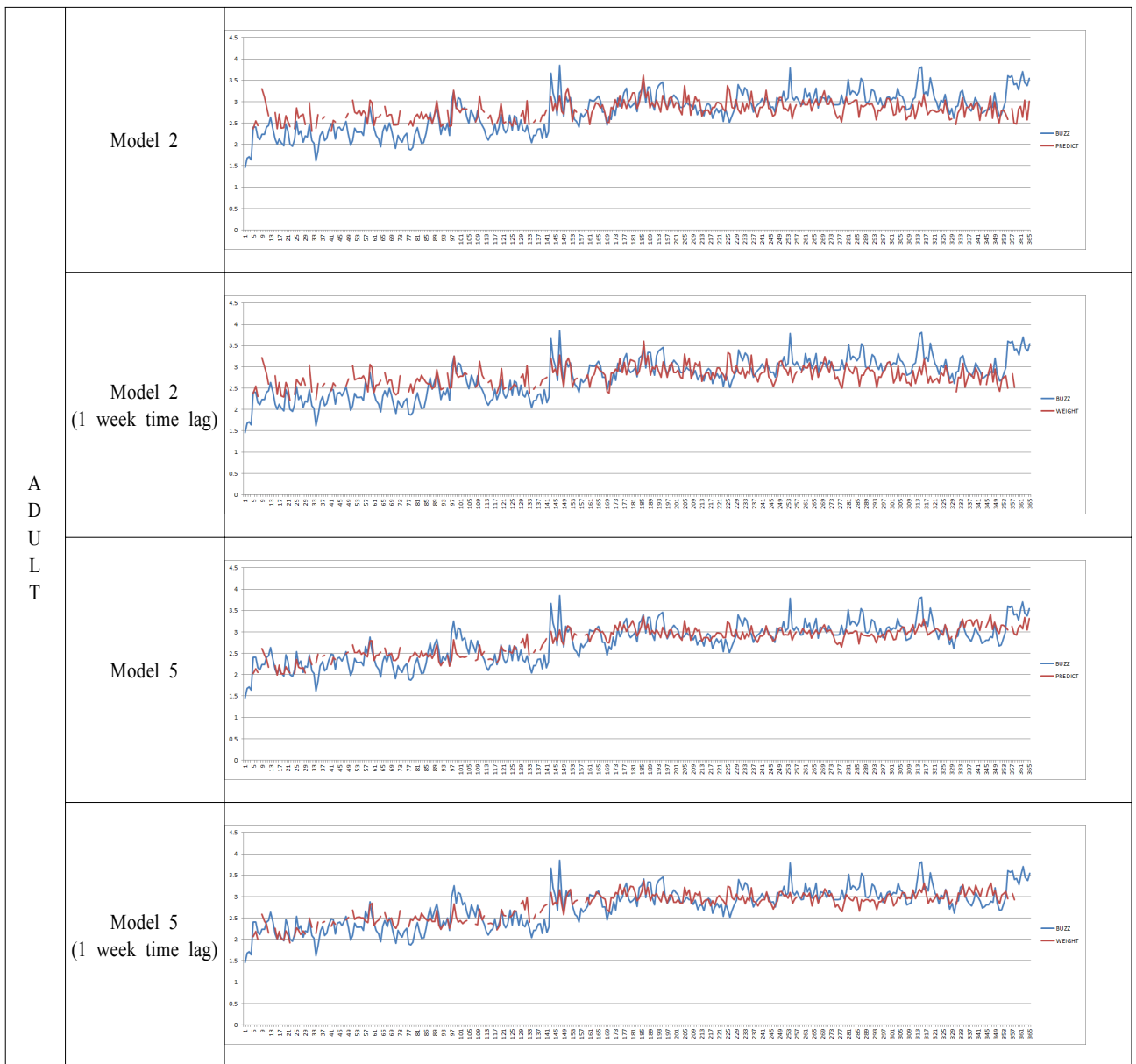
Independent Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	β	t	β	t	β	t	β	t	β	t
(Constant)	1.409	4.21***	1.736	5.91***	-25.351	-9.91***	4.421	1.01	2.574	.59
Suicide rate	.827	4.03***	.448	2.47**	.503	3.20***	.329	2.26**	.249	1.70*
Drinking			.526	10.34***	.407	8.98***	.373	8.96***	.347	8.18***
Divorce					6.802	10.64***	4.436	6.78***	4.346	6.70***
Fertility							-4.368	-8.06***	-4.063	-7.41***
Humidity									.522	2.73***
Adjusted R ²	.040		.270		.457		.546		.555	
F	16.212***		61.901***		93.121***		99.787***		82.916***	
(Constant)	1.326	3.98***	1.423	4.80***	-24.590	-9.68***	1.716	.393	-.699	-.16
Suicide rate (1 week time lag)	.869	4.26***	.626	3.50***	.744	4.75***	.559	3.78***	.475	3.21***
Drinking			.531	10.81***	.416	9.45***	.382	9.28***	.351	8.39***
Divorce					6.508	10.30***	4.513	6.94***	4.358	6.78***
Fertility							-3.929	-7.16***	-3.481	-6.23***
Humidity									.619	3.19***
Adjusted R2	.046		.297		.470		.543		.555	
F	18.153***		68.897***		96.334***		96.479***		81.442***	

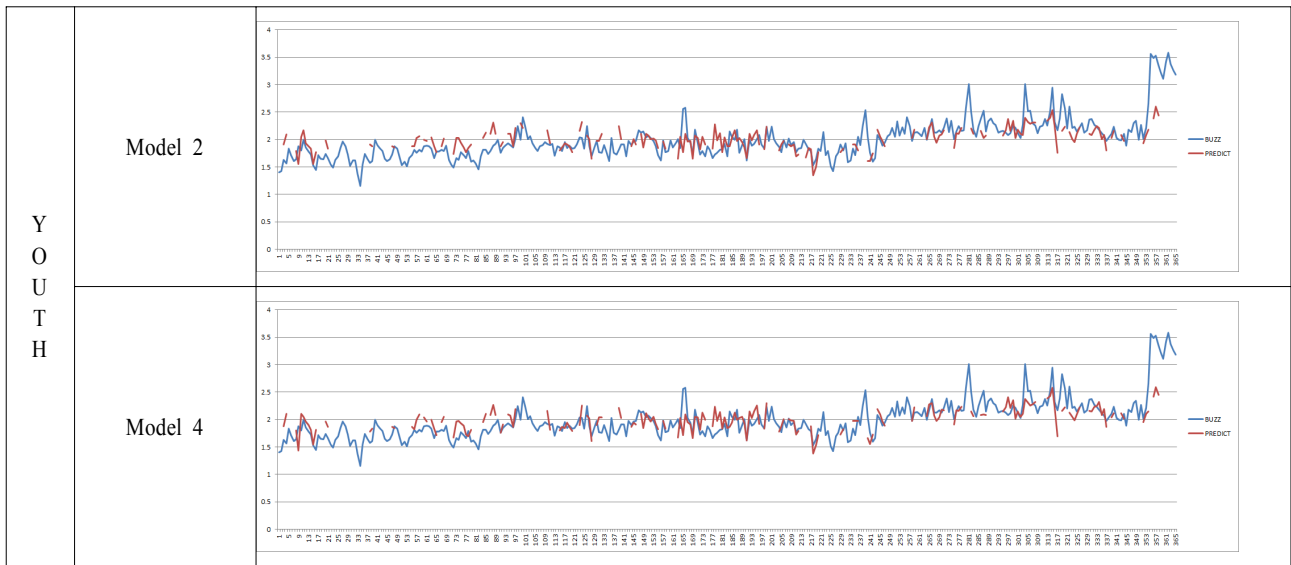
수와 출산아 수를 일별 단위로 공동으로 적용하였다.

- 11) 본 연구의 자살예측 모형의 개발을 위하여 사용된 예측변인들은 모델링 과정을 거쳐 유의미한 변인들만 모형에 적용하여, 성인과 청소년의 예측변수는 다르다는 것을 밝힌다.
- 12) URL을 통해 확산되는 자살관련 Buzz는 약 3주 정도의 생명주기를 가지며 발행 후, 첫 주에 급속히 전파되는 경향을 보이고 있다(National Information Society Agency, 2012).
- 13) Table 2와 모형 5의 경우, 예측도는 89.59%에서 89.64%로 조금 증가한 것으로 나타났다.

Independent Variable	Model 1		Model 2		Model 3		Model 4	
	β	t	β	t	β	t	β	t
(Constant)	1.957	68.86***	.648	6.48***	.624	6.22***	.939	5.75***
Suicide rate	.290	2.53**	.187	2.17**	.173	2.01**	.192	2.24**
Stress			1.003	13.40***	.913	10.94***	.935	11.25***
Drinking					.178	2.34**	.176	2.34**
Dust							-.214	-2.43**
Adjusted R ²	.023		.448		.456		.467	
F	6.421***		95.430***		65.464***		51.629***	

Note: *** p<0.01, ** p<0.05, * p<0.1





[Figure 2] Prediction model of suicide search volume in adults

청소년 자살 검색의 예측모형은 <Table 3>과 같다. 모형 1은 2011년 일별 청소년(19세이하) 자살률이 청소년 자살 검색에 미치는 영향을 예측하는 것으로 청소년 자살자 수가 많을수록(자살률이 높을수록) 자살 검색을 증가하는 것으로 나타났다. 모형 2는 일별 청소년 자살률과 스트레스 검색이 자살검색에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 스트레스 검색이 많을수록 청소년의 자살 검색은 증가하는 것으로 나타났다. 모형 3은 일별 청소년 자살률, 스트레스 검색, 음주 검색이 청소년 자살 검색에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 스트레스 검색이 많을수록, 음주 검색이 많을수록 청소년 자살 검색은 증가하는 것으로 나타났다. 모형 4는 일별 청소년 자살률, 스트레스 검색, 음주 검색, 미세먼지량이 청소년 자살 검색에 미치는 영향을 예측하는 것으로 자살자 수가 많을수록, 스트레스 검색이 많을수록, 음주 검색이 많을수록, 미세먼지량이 적을수록 청

소년 자살 검색은 증가하는 것으로 나타났다. 청소년의 자살검색을 예측하는 모형의 그래프는 [Figure 2]와 같다.

3. 다중집단 구조모형 분석

자살요인의 다중집단 구조모형 분석은 연구모형의 적합성을 검증한 후, 집단 간 등가제약 과정을 거쳐 경로계수 간 유의미한 차이를 검증하였다. 다중집단 구조모형 분석을 위한 연구모형의 적합도는 $\chi^2(df, p)=59.411(8, p<.000)$, NFI=0.954(양호), TLI=0.847(보통), CFI=.959(양호), RMSEA=0.094(보통)으로 대부분의 적합도에서 적합한 것으로 나타났다. 성인과 청소년 두 집단 모두 스트레스 검색에서 운동·음주·자살 검색으로 가는 경로와 자살 검색에서 자살률로 가는 경로가 청소년의 경우 음주 검색에서 자살 검색으로 가는 경로를 제외하고 모든 경로에서 정적(+)으로 유의한 영향을 미치는 것으로 나타났다 <Table 3>.

<Table 3> Parameters of adult and youth

Path	Adult			Youth		
	B(β)	C.R.	P	B(β)	C.R.	P
Stress→Drinking	.335(.385)	7.610	**	.572(.513)	11.347	**
Stress→Suicide	.432(.457)	10.986	**	.808(.525)	8.749	**
Stress→Exercise	.599(.555)	12.468	**	.721(.761)	22.379	**
Drinking→Suicide	.247(.228)	6.257	**	.051(.037)	.868	.385
Exercise→Suicide	.271(.310)	7.830	**	.348(.214)	3.827	**
Suicide→Suicide Rate	.052(.204)	3.980	**	.093(.172)	2.671	**

Note: ** p<0.01, * p<0.05

구조모형 내 자살요인 변수 간의 인과관계에 있어 두 집단(성인과 청소년) 사이에 유의미한 차이가 존재할 수 있어 모형 내 존재하는 모든 경로계수에 대해 각각 동일성 제약을 가한 모형을 기저모형과 비교하기 위해 집단 간 구조모형 분석을 실시하였다. 두 집단 사이의 경로에 동일성 제약을 가한 모형은 Stress→Drinking, Stress→Suicide, Stress→Exercise, Drinking→Suicide 경로의 모형에서 집단 간 유의미한 차이를 보였다. 따라서, 스트레스 검색에서 운동·음주·자살 검색 검색으로 가는 경로에 청소년이 더 강하게 영향을 받고 있으며, 음주 검색에서 자살 검색으로 가는 경로는 성인만 유의한 영향을 미치는 것으로 나타났다. 그리고 운동 검색에서 자살 검색으로 가는 경로와 자살 검색이 자살률에 미치는 영향은 두 집단 모두 비슷하게 영향을 주는 것으로 나타났다<Table 4>.

스트레스 검색과 자살 검색의 경로에 음주 검색과 운동 검색의 매개효과를 살펴보기 위해 효과분해를 실시한 결과 성인과 청소년 모두 부분 매개효과(partial mediation)가 있는 것으로 나타났다. 따라서 우리나라 성인과 청소년 모두 스트레스를 경험할 경우 건강생활실천요인(음주, 운동)을 많이 찾게 되고 이러한 건강생활실천요인이 자살 검색에 영향을 미치고 궁극적으로 자살률에 영향을 주는 것으로 나타났다.

4. 청소년 자살 검색 요인 다층모형 분석

청소년의 자살검색의 결정요인에 대한 다층분석을 위해 3개의 분석 모형 함수를 검증하였다.

<Table 4> Comparisons of the differences in adults and youths

Path	χ^2	df	$\Delta\chi^2$	C.R. ¹⁾
Default model	59.411	8		
Stress→Drinking	71.338	9	11.927	3.540**
Stress→Suicide	74.449	9	15.038	3.746**
Stress→Exercise	63.622	9	4.211	2.113**
Drinking→Suicide	66.600	9	7.189	-2.781**
Exercise→Suicide	59.997	9	.586	.788
Suicide→Suicide Rate	60.545	9	1.134	1.112
All constrains	110.579	14	51.168	

Note: ¹⁾ Critical ratios for differences
 ** p<0.05, * p<0.1

<Model 1> 기초모형(Unconditional Model)

$$LN\text{전체 } Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

<Model 2> 무조건적 기울기 모형(Unconditional Slope Model)

$$LN\text{전체 } Y_{ij} = \gamma_{00} + \gamma_{10} * LSTR_Y_{ij} + \gamma_{20} * LDRL_Y_{ij} + \gamma_{30} * LEXE_Y_{ij} + u_{0i} + u_{1j} * LSTR_Y_{ij} + u_{2j} * LDRL_Y_{ij} + u_{3j} * LEXE_Y_{ij} + r_{ij}$$

<Model 3> 조건적 기울기 모형(Conditional Slope Model)

$$LN\text{전체 } Y_{ij} = \gamma_{00} + \gamma_{01} * TSR_j + \gamma_{02} * AH_j + \gamma_{03} * EP_j + \gamma_{10} * LSTR_Y_{ij} + \gamma_{20} * LDRL_Y_{ij} + \gamma_{30} * LEXE_Y_{ij} + u_{0j} + u_{1j} * LSTR_Y_{ij} + r_{ij}$$

기초모형(Model 1: Unconditional Model)은 설명변수(독립변수)를 투입하지 않은 상태에서 청소년의 일별 자살 검색량에 대한 월별 분산을 분석함으로써 이후 모형에서 투입된 다른 독립변수들의 설명력을 살펴볼 수 있게 한다. 즉, 기초모형은 다층분석을 통해 일별 자살 검색량이 월별로 차이를 보이고 있는지 검증하고자 하는 것이다. Model 1에서 고정효과를 살펴보면, 일별 자살 검색량의 전체 평균은 184.18건으로 95% 신뢰구간(p=.026)에서¹⁴⁾ 기초 통계치 자료에서 산출된 평균(185.59)와 동일함을 알 수 있다. 이는 한국 청소년의 월 평균 자살검색량이 184.18건이 될 확률을 나타내는 것으로 통계적으로 유의하였다¹⁵⁾(β=1.99, p<.001). 무선평가를 살펴보면, 일 수준의 자살 검색량의 분산은 .06이며, 월 수준의 분산은 .08으로 통계적으로 유의한 것으로 나타났다(χ²=423.87, p<.001). 따라서 자살 검색량은 월별로 상당한 변량이 존재하고 있다고 할 수 있다.

동일한 수준에 속한 하위수준간의 유사성을 보여주는

집단 내 상관계수(Interclass Correlation Coefficient [ICC])를 통해 월별 자살 검색량의 분산비율을 계산해 보면 0.571(.08/(.08+.06)¹⁶⁾로 이는 곧 일별 자살 검색량에 대한 총 분산 중 월별 수준의 분산이 차지하는 비율이 57.1%임을 나타내는 것이며, 따라서 일별 수준의 분산이 차지하는 비율은 약 42.9%임을 알 수 있다. 이 결과는 모든 월의 자살 검색량이 같다는 X²의 귀무가설(null hypothesis)을 기각함으로써 일별 자살 검색량의 평균이 월별로 통계적으로 유의한 차이가 있음을 보여주는 것이다(p<.001). 이는 자살 검색량이 일별 요인에 의해서도 영향을 받지만, 월별 요인에도 영향을 받는다는 것을 나타냄으로써 일별, 월별 검색량의 차이를 분석하기 위해서는 일별 변수 및 월별 변수를 모두 투입하여 다층모형 분석을 실시하는 것이 타당함을 입증해 주는 결과라 할 수 있다.

Model 2의 검증은 일별 요인들이 일별 자살 검색량에 미치는 영향에 있어 월별로도 차이가 발생하는가를 알아 보고자 하는 것이다. 따라서 일별 스트레스·음주·운동 검색량이 자살 검색량에 미치는 영향을 고정효과를 통해 파악하고 이들 개별 요인이 월별로 따라 차이를 보이는가에 대해 무선평가를 통해 살펴보았다<Table 5>. 일별(Level 1) 자살 검색량에 대한 고정효과를 분석한 결과, 음주 검색량은 자살 검색량에 영향을 주지 않는 것으로 나타났고 스트레스 검색량과 운동 검색량은 통계적으로 유의미하여 자살 검색량에 영향을 주는 것으로(스트레스: β=.40, p<.05, 운동: β=.25, p<.05) 나타났다. 확인된 일별 특성변수와 자살 검색량의 관련성에 있어 각 변수가 월별 차이가 나는지에 대해 무선평가 검증을 실시한 결과, 스트레스 검색량의 적합도(χ²=28.97, p<.01), 음주 검색량의 적합도(χ²=36.72, p<.001), 운동 검색량의 적합도(χ²=19.13, p<.05)가 통계적으로 유의미한 것으로 확인되어 일별 수준의 변수들이 자살 검색량에 미치는 영향에 있어 월별 간 차이가 있음(χ²=90.99, p<.001)을 확인하였다. 따라서 이는 분석에 있어 월별 요인의 투입이 필요함을 입증하는 결과라 할 수 있겠다. 따라서 일별 특성에서 스트레스 검색량이 높아질수록 자살 검색량이 높아지는 것으로 이러

14) 일일 자살 검색량의 전체 평균은 실제 검색량의 무조건모형 분석결과 회귀계수는 184.18(표준오차: 71.58, t-ratio: 2.57, p=.026)로 유의하게 나타났다.

15) Table 5의 상용로그로 치환한 자살검색량의 회귀계수와 유의수준을 참조함.

16) 종속변수가 연속형인 경우 ICC=(τ₀₀(2수준분산))/(τ₀₀(2수준분산)+σ²(1수준분산))으로 산출한다.

한 효과는 월별에 따라서 차이가 있음을 보여주고 있다. 그리고 무선희과 검증에서 유의미한 일별 특성 변수(스트레스·음주·운동 검색량)은 조건적 모형 검증에서 미지수로 투입하여 분석할 필요가 있는 것으로 나타났다. 연도별 자살 검색량의 ICC는 .903로 산출되었다.

Model 3은 청소년의 일별 요인(스트레스·음주·운동 검색량)과 월별 청소년 자살률, 평균습도, 경제활동 참가율이 자살 검색량에 미치는 영향을 분석한 것이다. 즉, 앞서 Model 2에서 월별 변수를 투입할 수 있는 일별 요인 변수인 스트레스·음주·운동 검색량을 동시에 투입하는 연구 모형을 검증한 것이다. 일별 요인과 월별 요인을 동시에 고려하였을 때 자살 검색량에 영향을 미치는 요인의 영향력을 검증하기 위해 자살 검색량에 대한 고정효과를 분석한 결과, 수준 1인 일별요인 변수는 무조건적 기울기

모형(Model 2)의 검증과 비슷한 결과를 보였다. 수준 2인 월별 청소년 자살률은 자살 검색량에 정(+)적 영향($\beta = 2.16, p < .05$)을 미치는 것으로 나타났다. 이는 월별 청소년 자살률이 증가하면 자살 검색량도 증가한다는 것으로 의미한다. 그리고 평균습도는 자살 검색량에 정(+)적 영향($\beta = 1.83, p < .01$)을 미치는 것으로 나타났다. 이는 월별 평균습도가 증가하면 자살 검색량은 증가한다는 것으로 의미한다. 월별 경제활동 참가율은 자살 검색량에 영향을 미치지 않는 것으로 나타났다. 무선희과 검증 결과, 일별 스트레스 검색량($\chi^2 = 28.94, p < .01$)과 운동 검색량($\chi^2 = 19.44, p < .05$)은 월별 차이가 있는 것으로 나타났다. 조건적 모형에서 연도별 자살 검색량의 ICC는 .857으로 산출되었다.

<Table 5> Multi-level model analysis in suicide research factors

Parameter	Model	Model 1 Unconditional model		Model 2 Unconditional Slope model		Model 3 Conditional model	
	Fixed effect	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Level 1	Intercept (γ_{00})	1.99	.08***	1.12	.16***	-3.77	6.81
	STSV			.40	.11*	.42	.11**
	DSV			-.00	.07	-.01	.07
	ESV			.25	.09*	.24	.09*
Level 2	TSR					2.16	.92*
	AH					1.83	.47**
	EP					.68	4.19
Random effect		σ^2	χ^2	σ^2	χ^2	σ^2	χ^2
	level 2, u_0	.08	423.87***	.28	90.99***	.18	70.50***
	level 1, r	.06		.03		.03	
	STSV			.11	28.97**	.10	28.94**
	DSV			.06	36.72***	.06	36.54***
	ESV			.07	19.73*	.08	19.44*
	ICC		.571		.903		.857
	Deviance		65.21		-184.89		-204.58

Note: *** p<.001, ** p<.01, * p<.05, # p<.1

IV. 논의

본 연구는 소셜 빅 데이터를 활용하여 우리나라의 자살검색 요인과 자살검색 예측모형을 개발하기 위해 다변량 분석을 실시하였다. 자살검색 예측모형의 개발은 다중 회귀분석을 적용하였으며, 성인과 청소년 집단의 자살검색이 실제 자살률에 미치는 영향은 다중집단 구조모형으로 검증하였다. 그리고 자살검색에 미치는 일별 요인과 월별 요인은 다층모형으로 검증하였다.

본 연구의 구체적인 결과와 해석은 다음과 같다. 첫째, 성인의 자살검색 예측모형에서 음주검색이 많을수록 자살 검색이 높은 것은 스트레스로 인해 우울이나 알코올 중독이 발생하고 자살에 이른다라는 기존의 연구(Izadinia et al., 2010; Park et al., 2010)를 지지하는 것으로 나타났다. 이혼율이 높을수록 자살 검색이 증가하는 것은 배우자의 이혼이나 사별한 경우 자살률이 높다는 기존의 연구(Smith, 1988; Stack, 1981)를 지지하는 것으로 나타났다. 출산율이 낮을수록 자살 검색이 증가하는 것은 핵가족화와 도시화로 인해 가족구성원의 지지가 부족한 것에 원인을 들 수 있다. 평균습도가 높을수록 자살 검색이 증가하는 것은 습도가 높으면 실내 공기 중의 농도가 증가하여 궁극적으로 대기오염이 증가하기 때문에 대기오염이 높으면 자살률이 높다는 기존의 연구(Biermann et al., 2009)를 지지하는 것으로 나타났다. 둘째, 청소년의 자살검색 예측모형에서 스트레스 검색이 많으면 자살검색이 증가하는 것은 스트레스가 자살의 원인이 된다는 기존의 연구(Dixon et al., 1991; Lazarus & Folkman, 1984)를 지지하는 것으로 나타났다. 음주 검색이 많을수록 자살 검색이 증가하는 것은 문제음주는 다른 중독성 물질을 남용할 수 있고 자기통제가 어려워 쉽게 자살충동을 유발할 수 있다는 기존의 연구(Chung, 2011; Tapert et al., 2003; Yoon, 2011)를 지지하는 것으로 나타났다. 미세먼지가 적을수록 자살률이 증가하는 것은 미세먼지가 증가할수록 자살이 증가한다는 기존의 연구(Kim et al., 2010)를 지지하지는 않았으나, 우리나라 청소년의 경우 성인과 달리 강도 높은 학업 등으로 인해 대부분 실내에서 활동함에 따른 결과로 보고 있다. 셋째, 구조모형 분석결과 스트레스 검색에서 자살 검색으로 가는 경로에 스트레스 취약요인이 매개하여 기존의 스트레스 취약요인의 연구

(Chapman & Beaulieu, 1983; Izadinia et al., 2010)를 지지하는 것으로 나타났다. 특히 스트레스 검색에서 음주·운동·자살 검색으로 가는 경로가 청소년이 더 강하게 영향을 미친다는 것은 청소년은 학업문제, 학교폭력, 취업, 경제문제 등의 다양한 문제로 인해 성인보다 생활 스트레스가 더욱 심각하다는 것을 나타내고 있다. 넷째, 다층모형 분석에서 청소년의 일별 자살검색은 자살검색 예측모형과 같이 일별 요인(스트레스·운동 검색량)과 월별 요인(청소년 자살률, 평균습도)에 영향을 받는 것으로 나타났다. 경제활동 참가율은 청소년 자살검색에 영향이 없는 것은 인간관계가 빈번하고 사회생활이 집중될 때 자살이 많이 발생한다는 기존의 연구(Durkheim, 2008)나 노인의 경제활동이 줄어들 때 자살이 증가한다는 기존의 연구(Kim, 2004)의 결과와는 반대로 향후 다른 경제변수와의 분석을 통한 추가 연구가 있어야 할 것으로 본다. 본 연구는 소셜 빅 데이터에서 수집된 스트레스·음주·운동 검색량과 오프라인 조사 데이터(자살률, 이혼율, 출산율, 기후 등)를 적용하여 성인과 청소년의 자살 검색에 대한 예측모형을 개발하였고, 스트레스 검색은 스트레스 취약요인(음주, 운동)을 매개하여 자살 검색에 영향을 미치고 궁극적으로 실제 자살률에 영향을 준다는 것을 구조모형을 통하여 검증하였으며, 청소년의 일일 자살검색은 일일 스트레스·음주·운동 검색량 뿐만 아니라 월별 요인인 청소년 자살률과 평균습도에도 영향을 받는다는 것을 다층모형을 통하여 검증하였다는 점에서 분석 방법론적이나 정책적으로 의미를 가진다고 할 수 있다. 그리고 빠르고 효과적으로 실제적인 내용을 파악하여 사회 통계가 지닌 한계를 보완할 수 있는 새로운 조사방법으로서의 빅 데이터의 가치를 확인하였다는 점에서(National Information Society Agency, 2012) 조사 방법론적 의의를 가진다고 할 수 있다. 또한 소셜 빅 데이터와 조사 데이터와의 연계를 통한 자살 검색의 이론적 모형을 제시하였다는 점에서 학술적 의의가 있으며, 향후 데이터마이닝 분석과 네트워크 분석을 통한 자살위험 예측모형의 정교화를 위한 연구가 꾸준히 추진되어야 할 것으로 본다. 그럼에도 불구하고 본 연구에서는 다음과 같은 제한점이 있다. 첫째, 일 단위로 생산되는 조사통계의 부족으로 인해 월 단위 변수를 일 단위로 적용(본 연구에서의 출산율, 이혼율, 경제활동참가율 등)함에 따른 생태학적 오류가

(Ecological Fallacy)가 발생할 수 있다. 둘째, 본 연구에서 자살 검색이 자살률에 미치는 영향은 구조모형으로 밝혔으나, 2012년 자살률에 대한 데이터의 미 공표로 인해 자살 검색예측 모형에 대한 예측을 제시하지 못하였으며 연예인 자살이나 동반자살, 물가지수, 지역 특성 등의 다양한 요인을 반영하지 못하였다는 한계가 있다. 이는 2010년부터 연차별 소셜 빅 데이터의 분석과 통계청 사망원인 통계 등 조사 통계와의 연계를 통하여 다양한 요인이 반영된 예측분석이 가능할 것으로 본다. 셋째, 본 연구의 구조모형에서 제시된 소셜 빅 데이터 변수들의 인과관계를 횡단조사나 종단조사에서 연구된 이론적 모형을 적용하는데 대한 이론적 근거를 제시하지 못하였다. 이는 본 연구에서 사용된 변수가 일일 검색량에 기인한 것으로, 향후 개인별 버즈에 대한 심층연구를 수행한다면 연구에 사용된 소셜 빅 데이터의 변수 간 인과관계를 밝힐 수 있을 것으로 본다. 넷째, 본 연구의 다중회귀 모형에서는 자살률과 다른 환경변수를 예측모형으로 사용하여 소셜 빅 데이터의 자살 검색에 영향을 미칠 것이라는 가정을 하여 구조모형과의 연관관계를 충분히 설명하지 못하였다. 이는 본 연구에서의 다중회귀분석은 소셜 빅 데이터의 자살 검색에 미치는 예측요인을 살펴보기 위해 본 연구의 연구모형과 연구가설과는 별도로 분석된 것으로, 향후 소셜 빅 데이터 자살 검색이 자살률에 미치는 영향에 대한 연구가 수행되어야 할 것으로 본다.

V. 결론

본 연구는 소셜 빅 데이터의 담론에서 자살과 관련된 빅 데이터와 오프라인 통계데이터의 연계를 통하여 우리나라 자살 검색의 예측요인을 다변량 분석으로 살펴보았다. 이를 위해 소셜 빅 데이터에서 2011. 1. 1~2011. 12. 31 언급된 자살관련 토픽(스트레스, 음주, 운동, 자살)을 추출하여 분석하였다. 다변량 분석결과를 요약하면 다음과 같다.

첫째, 일일 자살자수가 많을수록, 음주 검색이 많을수록, 이혼율이 높을수록, 출산율이 낮을수록, 평균습도가 높을수록 성인의 자살 검색은 증가하는 것으로 나타났다. 그리고, 자살관련 담론이 급속히 확산되는 1주간의 시차

(time lag)를 실제 자살률에 적용한 예측도는 조금 향상(89.59%→89.64%)된 것으로 나타났다.

둘째, 일일 청소년의 자살자 수가 많을수록, 스트레스 검색이 많을수록, 음주 검색이 많을수록, 미세먼지량이 적을수록 청소년의 자살 검색은 증가하는 것으로 나타났다.

셋째, 성인과 청소년의 자살검색의 구조모형 분석결과 두 집단 모두 스트레스 검색에서 운동·음주·자살 검색으로 가는 경로와 자살 검색에서 자살률로 가는 대부분의 경로가 정적(+)으로 유의한 영향을 미치고 있으며, 스트레스 검색에서 운동·음주·자살 검색으로 가는 경로에 청소년이 더 강하게 영향을 받고 있으며, 음주 검색에서 자살검색으로 가는 경로는 성인만 유의한 영향을 미치는 것으로 나타났다.

넷째, 스트레스 취약요인(음주, 운동) 매개효과를 분석한 결과 성인과 청소년 모두 스트레스 취약요인의 부분 매개효과(partial mediation)가 있는 것으로 나타났다.

다섯째, 청소년 자살 검색의 결정요인에 대한 다중모형 분석결과 일별 요인(스트레스·운동 검색량)은 유의미한 차이가 있는 것으로 나타났다. 월별 청소년 자살률과 평균습도는 자살 검색량에 정적인 영향을 미치는 것으로 나타났다.

본 연구에서는 소셜 빅 데이터와 오프라인 빅 데이터의 연계를 통해 성인과 청소년의 자살 검색에 대한 요인을 검증하고 예측모형을 개발하였다. 본 연구를 근거로 우리나라의 자살예방과 관련하여 다음과 같은 정책적 함의를 도출할 수 있다.

첫째, 본 연구의 결과는 성인과 청소년 모두 다양한 원인들에 의해 스트레스를 경험하면서 자살을 검색을 하게 된다. 따라서 청소년의 스트레스를 해소할 수 있는 학교 차원의 다양한 프로그램의 마련과 함께 성인의 경제활동으로 인한 스트레스를 해소 시킬 수 있는 직장차원의 프로그램이 개발되어야 할 것이다.

둘째, 성인과 청소년은 온라인상에서 자살과 관련한 담론을 주고 받고 있으며 이러한 언급이 실제적인 자살과 관련된 심리적·행동적 특성으로 노출될 수 있기 때문에 자살 예측모형에 따른 위험징후가 예측되면 실시간으로 개입할 수 있는 애플리케이션(가칭: 생명존중 온라인 게이트키퍼)이 개발되어야 할 것이다. ‘생명존중 온라인 게이트키퍼’는 자살에 대한 위험징후가 예측되면 소

설 담론의 분석에서 추측된 위험요인을 줄일 수 있는 맞춤형 프로그램을 실시간으로 제공할 수 있도록 개발되어야 할 것이다.

셋째, 지역별 소셜 빅 데이터와 지역 요인의 연계를 통한 지역별 자살 예측모형을 개발하여 실시간으로 자살을 사전에 방지할 수 있는 시스템(가칭: 지역 생명존중 예보 시스템)을 구축하여야 할 것이다. ‘지역 생명존중 예보 시스템’은 지역별 생명존중 관련기관에 지역별 자살예보를 1주 또는 월 단위로 제공하여 지역별 자살행동의 원인에 대한 적극적 대응을 위한 대국민 홍보활동을 지속적으로 실시함으로써 지역 주민의 생명존중 인식을 강화할 수 있을 것이다.

넷째, SNS에서 주고 받는 자살관련 소셜 담론은 개인이 일상생활에서 갖는 우울한 감정이나 고민이 기록되는 온라인 심리적 부검보고서라고 할 수 있다. 핀란드가 오프라인 심리적 부검보고서를 바탕으로 국가 차원의 자살 예방 대책을 마련하여 자살률을 줄였다면, 우리나라는 세계 최고수준의 IT 기술과 소셜 빅 데이터의 분석으로 국가 차원의 자살예방 대책이 마련되어야 할 것이다.

참고문헌

- Alloy, L. B., Hartlage, S., & Abramson, L. Y. (1988). Testing the cognitive diathesis-stress theories of depression: Issues of research design, conceptualization, and assesment. In L. B. Alloy (Ed.), *Cognitive precesses in depression* (pp. 31-73). New York, NY: The Guilford Press.
- Bae, S. B., & Woo, J. M. (2011). Suicide prevention strategies from medical perspective. *Journal of the Korean Medical Association*, 54(4), 386-391.
- Biermann, T., Nikolaos, S., Stefan, B., Norbert, T., Johannes, K., & Udo, R. (2009). The hypothesis of an impact of ozone on the occurrence of completed and attempted suicides. *Medical Hypotheses*, 72, 338-341.
- Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage
- Chang, Y. S. (2012). Actual conditions of suicide and policy challenges in Korea 2012. *Issue & Focus*, 165, 46.
- Chapman, N. J., & Beaullet, M. (1983). Environmental predictor of well-being for at risk older adults in mid-sized city. *Journal of Gerontology*, 38(2), 237-244.
- Chung, S. (2011). 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. *Journal of Korean Alcohol Science*, 12(1), 15-28.
- Dixon, W. A., Heppner, P. P., & Anderson, W. P. (1991). Problem-solving appraisal, stress, hopelessness, and suicide ideation in a college population. *Journal of Counseling Psychology*, 38(1), 51-56.
- Durkheim, E. (2008) *Suicide: A study in sociology*, Paju, Korea: Chungabook.
- Hair, J. F. Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (Eds.). (2006). *Multivariate Data Analysis*(6th ed.). Upper Saddle River, NJ: Prentice-Hall International.
- Izadinia, N., Amiri, M., Jahromi, R. G., & Hamidi, S. (2010). A study of relationship between suicidal ideas, depression, anxiety, resiliency, daily stresses and mental health among Teheran university students. *Procedia-Social and Behavioral Sciences*, 5, 1615-1619.
- Kim, C., Jung S. H., Kang, D. R., Kim, H. C., Moon, K. T., Hur N. W., . . . Suh, I. (2010). Ambient particulate matter as a risk factor for suicide, *The American Journal of Psychiatry*, 167(9), 1100-1107.
- Kim, S. Y. (2004). Trends in elderly suicide rates and social factors in Korea. *Social Welfare Policy*, 19, 181-205
- Korea Communications Commission·Korea Internet & Security Agency (2012). *2011 Survey on the Internet Usage*. Seoul, Korea: Author.
- Kowalski, G. S., Charles, E. F., & Paul, D. S. (1987). Urbanism and suicide: A study of American Counties. *Social Forces*, 66(1), 85-101.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal and Coping*. New York, NY; Springer Publishing Company.
- Lee, C. H., Sung, Y. S. (2012). *A Study on the SNS Use of Young People*. Seoul, Korea: National Youth Policy Institute.
- McIntosh, J. L. (1995). Suicide prevention in the elderly. *Suicide and Life Threatening Behavior*, 25(1), 180-192.
- McKinsey Global Institute (2011). *Big data: The next frontier for innovation, competition, and productivity*. San Francisco, CA: McKinsey & Company.
- National Information Society Agency (2011). *Public sympathy policy measure based on social media analysis for the advancement of the national administration*. Seoul, Korea: Author.
- National Information Society Agency (2012). *Implications for Suicide Prevention Policy of Youth Described in the Social Analysis*. Seoul, Korea: Author.
- National Information Society Agency (2012). *Evolved into the Big Data World (Case Study of Global for Big Data)*. Seoul, Korea: Author.
- National IT Industry Promotion Agency (2006). What is Social

- Media?. Seoul, Korea: Author.
- Park, J. Y., Lim, Y. O., & Yoon, H. S. (2010). Suicidal impulse caused by stress in Korea: Focusing on mediation effects of existent spirituality, family support, and depression. *Korean Journal of Social Welfare Studies*, 41(4), 81-105.
- Policy Exchange (2012). The big data opportunity: Making government faster, smarter and more personal. London, England: Policy Exchange.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717-731.
- Raudenbush, S. W., & Bryk, A. S. (eds.) (2002). *Hierarchical Linear Models: Application and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Smith, J. C. (1988). Marital status and the risk of suicide. *American journal of public Health*, 78, 78-80.
- Song, T. M. (2012). *Study on the baby boomer and echo generation suicide characteristics*. Seoul, Korea: Korea Institute for Health and Social Affairs.
- Stack, S. (1981). Divorce and suicide. *Journal of family Issues*, 2, 77-90.
- Stack, S. (2000). Suicide: A 15-year review of the sociological literature part II: Modernization and social integration perspectives. *Suicide & Life - Threatening Behavior*, 30(2), 163-176.
- Statistics Korea. (2010). *Characteristics of baby boomer generation using social survey 2010*. Seoul, Korea: Author.
- Statistics Korea. (2012). *Youth Statistics*. Seoul, Korea: Author.
- Tapert, S., Colby, S., Barnett, N., Spirito, A., Rohsenow, D., Myers, M., & Monti, P. (2003). Depressed mood, gender, and problem drinking in youth. *Journal of Child & Adolescent Substance Abuse*, 12(4), 55-68.
- Yang, A. C., Tsai, S. J., & Huang, N. E. (2011). Decomposing the association of completed suicide with air pollution, weather, and unemployment data at different time scales. *Journal of Affective Disorders*, 129, 275-281.
- Yoon, M. S. (2011). Suicidal ideation among alcoholics moderating effect of alcohol use. *Mental Health & Social Work*, 38, 113-140.
- Wilbum, V. R., & Smith, D. E. (2005). Stress, self-esteem, and suicidal ideation in late adolescents. *Adolescence*, 40(157), 33-46.