

# 텍스트 마이닝과 토픽 모델링을 이용한 성조숙증 관련 인터넷 건강상담 분석

박수잔<sup>\*†</sup>, 서영균<sup>\*\*</sup>, 김재성<sup>\*\*\*</sup>, 권진원<sup>\*\*\*\*</sup>

<sup>\*</sup>경북대학교 약학연구소 연구초빙교수, <sup>\*\*</sup>경북대학교 IT 대학 컴퓨터학부 교수  
<sup>\*\*\*</sup>경북대학교 IT 대학 컴퓨터학부 석사과정, <sup>\*\*\*\*</sup>경북대학교 약학대학 교수

## Analysis for internet health counseling about precocious puberty using text mining and topic modeling

Susan Park<sup>\*†</sup>, Young-Kyoon Suh<sup>\*\*</sup>, Jaesung Kim<sup>\*\*\*</sup>, Jin-Won Kwon<sup>\*\*\*\*</sup>

<sup>\*</sup>Research professor, Research Institute of Pharmaceutical Sciences, Kyungpook National University  
<sup>\*\*</sup>Professor, The school of computer Science and Engineering, Kyungpook National University  
<sup>\*\*\*</sup>Master's student, The school of computer Science and Engineering, Kyungpook National University  
<sup>\*\*\*\*</sup>Professor, College of Pharmacy, Kyungpook National University

**Objectives:** This study aimed to investigate the characteristics of the demand for precocious puberty-related health information using the text data of online health counseling through the techniques of text mining and topic modeling. **Methods:** The input data of this study were question and answer documents that were searched for Naver Jisik-iN using the keywords 'precocious puberty'. The documents were automatically collected using a web crawler from January 2017 to June 2019. We performed text pre-processing, after which all nouns were extracted using a morphological analyzer. LDA(Latent Dirichlet Allocation) topic modeling was performed to find hidden themes in the text. **Results:** A majority of questions were submitted during the school vacation. Approximately half of all questions were answered by doctors. We used LDA topic modeling to yield three themes regarding precocious puberty-related questions: 1) asking for expert opinion on diagnosis or hospital selection, 2) questions regarding diagnostic tests and treatment, and 3) questions considering obesity and/or diet, as well as timing of puberty onset. **Conclusion:** This study identified the major precocious puberty-related concerns. In light of the increasing incidence of precocious puberty, individuals with this condition, as well as parents, should be provided with appropriate education to promote awareness and management.

**Key words:** precocious puberty, text mining, topic modeling, health education

### I. 서론

성조숙증은 2차 성징이 기대 연령보다 빨리 이루어지는 것으로 정의되며, 임상적으로 여아는 8세 미만, 남아는 9세 미만에서 2차 성징이 나타나는 것을 말한다(Berberoglu,

2009; Kim, 2018). 아동의 성조숙증은 신체적 문제 뿐 아니라 또래와 다른 체형으로 인한 스트레스, 이로 인한 교우관계 문제 및 성범죄의 노출 가능성 등 여러 심리사회적 문제를 초래할 수 있다(Johansson & Ritzén, 2005; Tremblay & Frigon, 2005). 전세계적으로 최근 출생 코호트가 과거에

Corresponding author: Susan Park

Research Institute of Pharmaceutical Sciences, Kyungpook National University, 80, Daehak-ro, Buk-gu, Daegu, 41566, Republic of Korea

주소: (41566) 대구광역시 북구 대학로 80번지 경북대학교 약학연구소

Tel: +82-53-950-8580, Fax: +82-53-950-8557, E-mail: [lovelysu99@knu.ac.kr](mailto:lovelysu99@knu.ac.kr)

※ 이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2019R1I1A1A01046991)

• Received: February 29, 2020

• Revised: July 26, 2020

• Accepted: August 2, 2020

비해 성적 성숙이 빨라지는 경향은 공통적으로 관찰되어 왔으나(Kim, 2018), 국내 성조숙증 발생률의 증가세는 굉장히 가파르다. 건강보험심사평가원의 보험청구자료를 분석한 김 (2019)의 연구에 따르면 국내 성조숙증 발생률은 2010년에서 2014년까지 여아는 10만명당 89.4명에서 415.3명으로, 남아는 10만명당 1.6명에서 14.7명으로 증가하였다고 보고하였다(Kim et al., 2019). 연구대상과 시기에 따라 편차가 있지만 해외 성조숙증 발생률이 인구 10만명당 10명(González, 1982; Krstevska-Konstantinova et al., 2001) 수준임을 고려하면 국내 성조숙증 환자 발생률은 매우 높은 수준임 알 수 있다.

대부분의 성조숙증은 특발성으로 직접적인 원인을 알지 못하는 경우가 많다. 그러나 일부는 중추신경계나 내분비계의 종양이 원인이 된 경우도 있어 전문의의 정확한 진단과 치료는 중요하다(Lee et al., 2011). 성조숙증의 진단과정에는 X-ray, 뇌 MRI, 복부 초음파, GnRH(gonadotropin-releasing hormone) 자극검사 등 다양한 검사들이 수반되고 (Lee et al., 2011; Rhie & Lee, 2015), 치료를 위해서는 통상 2년 이상의 성호르몬 억제제(GnRHa)의 주사 투여가 필요하다(Lee et al., 2011). 높은 질환 발생률, 복잡한 진단 검사와 장기간의 침습적 치료과정은 환아와 가족들의 정보 요구도를 높일 것으로 예상되지만 이에 대한 연구는 매우 제한적이다. 일부에서 이루어진 국내 선행연구들은 주로 환아와 아동을 돌보는 어머니의 경험에 국한되어 있고 연구대상자도 일개 병원 방문자와 같은 편의추출에 의존하고 있다 (Chun & Jung, 2019; Lee, Ju, & Kwak, 2015).

통신기술의 발전과 여러 디지털 기기의 보급으로 방대한 양의 텍스트 데이터가 온라인 상에 축적되고 있다. 이에 따라 최근에는 보건의로 분야에서도 온라인 텍스트 정보를 전염병 감시(surveillance) 혹은 환자의 건강상담/치료 경험 분석에 활용하고자 하는 시도들이 이루어지고 있다 (Adusumalli et al., 2015; Park & Hong, 2018; Sadah, Shahbazi, Wiley, & Hristidis, 2016). 국내에서 주로 활용된 온라인 자료원으로는 트위터, 인터넷 블로그, 카페 (Song, Song, Seo, Jin, & Kim, 2017; Song, Song, Seo, & Jin, 2016; Woo et al., 2016)와 네이버 지식인(Kim, Park, & Ji, 2016; Kim & Lim, 2013; Oh, Shin, & Hwang, 2018) 등이 있으며, 이 중 네이버 지식인은 연간 2천만건 이상의 질문과 6천건 이상의 답변이 등록되는 국내 대표적인

지식교류 서비스이다(Naver). 건강관련 질의응답도 활발히 이루어져 2002년 서비스 시작 이후 최근까지 건강관련 누적 질문 수는 2억건을 상회한다.

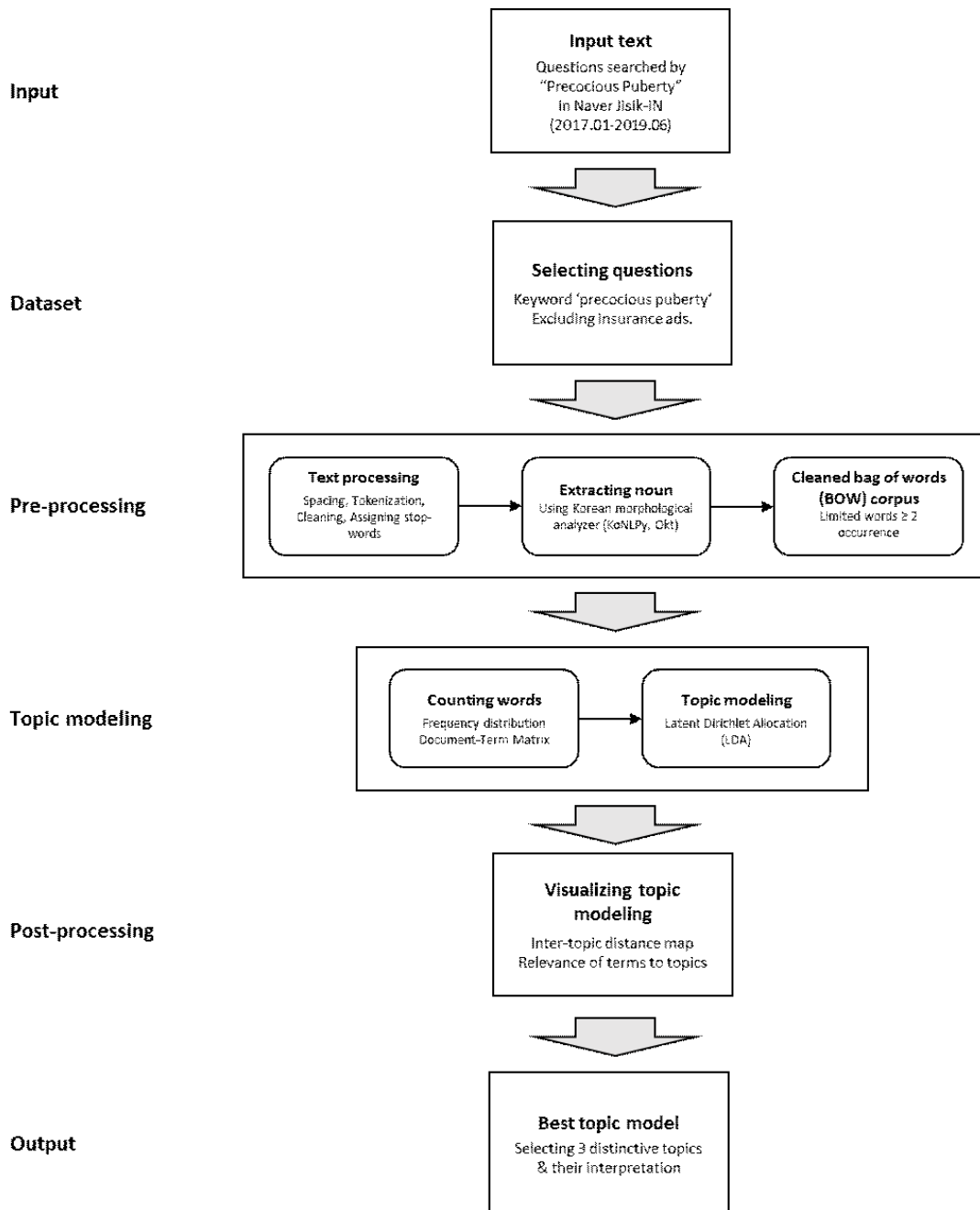
이와 같은 방대한 양의 온라인 텍스트 데이터에서 의미있는 주제와 내용을 도출하기 위해 활용할 수 있는 방법으로는 텍스트 마이닝(text mining)기법을 들 수 있다. 텍스트 마이닝은 비정형 데이터를 수집하고 분석하기 위한 기법으로, 키워드 분류, 나열, 요약의 일련의 과정을 통해 새로운 패턴을 찾아내는 과정이다(Lee & Lee, 2015). 이 중 LDA (Latent Dirichlet Allocation) 토픽 모델링(topic modeling) (Blei, Ng, & Jordan, 2003)은 베이지안 추론(Bayesian inference)을 기반으로 텍스트 데이터의 특정 주제를 추출하는 알고리즘으로, 최근의 학술연구에서 활용빈도가 높은 대표적인 텍스트 마이닝 분석방법 중 하나이다(Kim, Cho, & Kang, 2016).

따라서 본 연구는 네이버 지식인에 등록된 성조숙증 관련 질의 응답을 텍스트 마이닝과 토픽 모델링을 통해 분석함으로써 성조숙증 환자와 가족들이 필요로 하는 건강정보의 특성을 살펴보고, 이들의 주요 주제와 핵심어를 도출하고자 하였다.

## II. 연구방법

### 1. 연구방법 개요

[Figure 1]은 연구방법의 흐름을 단계별로 제시하였다. 1 단계는 입력(input) 자료 생성과정으로, 이 연구는 네이버 지식인에 '성조숙증'의 키워드로 검색된 질의 문서들을 웹 크롤러(Web crawler)를 이용하여 자동 수집하였다. 2단계는 실제 텍스트 분석에 활용할 데이터셋(Dataset)를 생성하는 과정으로 전체 질의 문서 중 텍스트 내에 '성조숙증' 키워드를 포함한 문서만 선정하였다. 3단계는 전처리 과정으로 자동 띄어쓰기, 정제, 불용어 처리 등의 텍스트 전처리 후 명사만을 추출하여 토픽모델링에 사용할 Bag of words (BoW)를 생성하였다. 4단계로 LDA 토픽 모델링 수행한 후, 5단계에서 해당 결과들을 시각화하여 토픽 간 독립성과 토픽 키워드의 연관성을 확인하는 과정을 거쳤다. 본 연구는 경북대학교 생명윤리위원회의 심의를 받았다(승인번호: KNU 2018-0141).



[Figure 1] Work flow of text mining and topic modeling

## 2. 데이터 수집 및 구축과정

### 1) 웹 크롤링(Web crawling)

본 연구는 2017년 1월부터 2019년 6월까지 네이버 지식인에 '성조숙증'을 키워드로 검색된 질문과 답변을 대상으로 하였다. 관찰기간 동안 네이버에 등록된 질문과 답변은

웹 크롤링(web crawling) 기법을 이용하여 수집하였고, 파이썬(Python)의 requests 와 BeautifulSoup 모듈을 사용하였다. 크롤링된 질문의 수는 2017년 2950건, 2018년 2027건, 2019년 822건으로 대상기간 동안 등록된 질문의 총 수는 5799건 이었다. 그러나 연구진의 검토결과 네이버 지식인에서 '성조숙증'으로 검색된 질문 중 상당수 실제 성조숙

증에 대한 질문이 아니었다. 따라서 이 증 질문의 제목과 내용 중 '성조숙증'이란 단어가 한번도 언급되지 않은 3565건의 질문을 1차 배제하였고, 연구자가 질문의 내용을 확인하여 성조숙증 환자와 가족의 교육 요구와 무관한 질문(ex. 단순 보험상품 문의/광고, 성조숙증 영어 번역 요청 등) 81건을 추가 제외하였다. 네이버 지식인은 질문자의 개인 정보를 습득할 수 없으므로, 동일한 사람이 여러 개의 중복 질문을 올렸을 가능성이 있다. 따라서 중복질문을 배제하기 위하여 TF-IDF(Term Frequency-Inverse Document Frequency) (Manning, Raghavan, & Schütze, 2008) 벡터(vector)의 코사인 유사도(Cosine Similarity) (Manning et al., 2008)를 이용하였고, 유사도 90% 이상의 문서(718건)는 동일 질문으로 간주하여 분석 자료에서 제외하였다. 이를 통해 최종적으로 선정된 성조숙증 관련 질문은 1435건이었다.

## 2) 텍스트 전처리(Pre-processing)

선정된 질문들은 텍스트 전처리 과정을 통해 이후 사용될 분석기법에 적합한 형태의 데이터로 변환되어야 한다. 네이버 지식인의 질문들은 일반인들이 자발적으로 올린 사적인 상담 글이기 때문에 공식적 문서들과는 달리 은어, 비속어와 같은 비표준어 사용이 빈번하고, 띄어쓰기가 온전하지 않은 글도 상당히 많다. 크롤링에서 얻어진 텍스트 데이터는 일차적으로 Python의 kospacing 모듈을 이용하여 자동 띄어쓰기 처리 수행하였다. 이 후 특수기호, 영어, 숫자, 한글자 단어는 제외하였고, '성조숙증', '네이버', '내공' 등의 상담내용과 관련없는 네이버 이용자의 특정 용어, 지역명과 특정 병원 이름은 불용어(stop-words) 처리하였다. 형태소 분석(Manning et al., 2008)은 형태소를 비롯한 다양한 언어적 속성의 구조를 파악하는 것으로, 본 연구에서는 3개 형태소 분석기(Khail, ETRI, Okt)의 품사 태깅 결과를 검토한 후 Konlpy 패키지내의 Okt 를 연구에 최종 사용하였다. Okt를 사용한 어간추출(Steming)과 품사태깅 작업을 통해 기타 품사를 제외한 명사만을 이후 분석에 사용하였다.

## 3. 토픽 모델링의 적용

### 1) LDA 토픽 모델링

추출된 명사는 출현 빈도(frequency) 기반의 단어 집합인

Bag of Words(BoW)로 변환 후 LDA 토픽 모델링을 위해서 다른 문서들의 BoW를 결합한 문서 단어 행렬(Document-Term Matrix, DTM)을 생성하였다. LDA 분석을 위해 문서 간 중복성이 없는 단어는 분석에서 제외한 후 2회 이상의 빈출 단어만을 분석에 사용하였다.

LDA 토픽 모델링(Blei et al., 2003)은 성조숙증 환자와 가족의 상담문서들의 숨겨진 주제를 확률을 기반으로 객관적으로 찾기 위해 사용되었다. LDA의 알고리즘은 각 문서는 잠재적 토픽들이 확률적으로 혼합되어 있다고 간주하며, 각 토픽은 단어들의 분포로써 특징 지어진다. 이후 베이즈안 추론(Baysian inference)을 기반으로 실제 텍스트 데이터 분포와 사전확률을 토대로 반복적인 샘플링 과정을 통해 토픽과 단어들의 확률을 연산한다. LDA 분석은 R 프로그램의 LDA 패키지를 이용하여 수행하였다(Team, 2013).

### 2) 사후 처리(Post-processing)

최적의 토픽 모델을 선정하기 위하여 최소 2개에서 최대 20개까지 다양한 토픽 수를 지정 하여 LDA 토픽 모델링을 수행 후 결과물을 비교검토 하였다. 이 작업에는 LDA 토픽 모델의 시각화 모듈인 LDavis 활용하여(Sievert & Shirley, 2014), 다차원의 주제간 거리 지도(Inter-topic distance map via multidimensional scale)와 연관성 추정치(relevance estimate)를 기준으로 상위 랭크된 키워드를 기준으로 주제와 단어의 관련성을 확인하였다.

## 4. 토픽 별 최종 키워드와 상담사례 제시

LDavis의 결과를 토대로 토픽간의 독립성이 가장 높으며, 토픽의 키워드간 관련성이 높은 모델을 최적의 모델로 선정하여, 최종적으로 3개의 주제별 20개의 키워드를 도출하였다.

본 연구는 단일 단어(uni-gram)를 기본 단위로 토픽 모델링을 수행하여, 문장 혹은 문서 전체의 맥락을 파악하는 데는 한계가 있다. 따라서 각 문서 별로 가장 높은 확률의 토픽을 기록한 후, 각 토픽 별로 할당된 문서의 원본 질문의 전체 질문 내용을 확인하여 토픽 별 질문의 전체 맥락과 특성을 파악하고자 하였다.

### III. 연구 결과

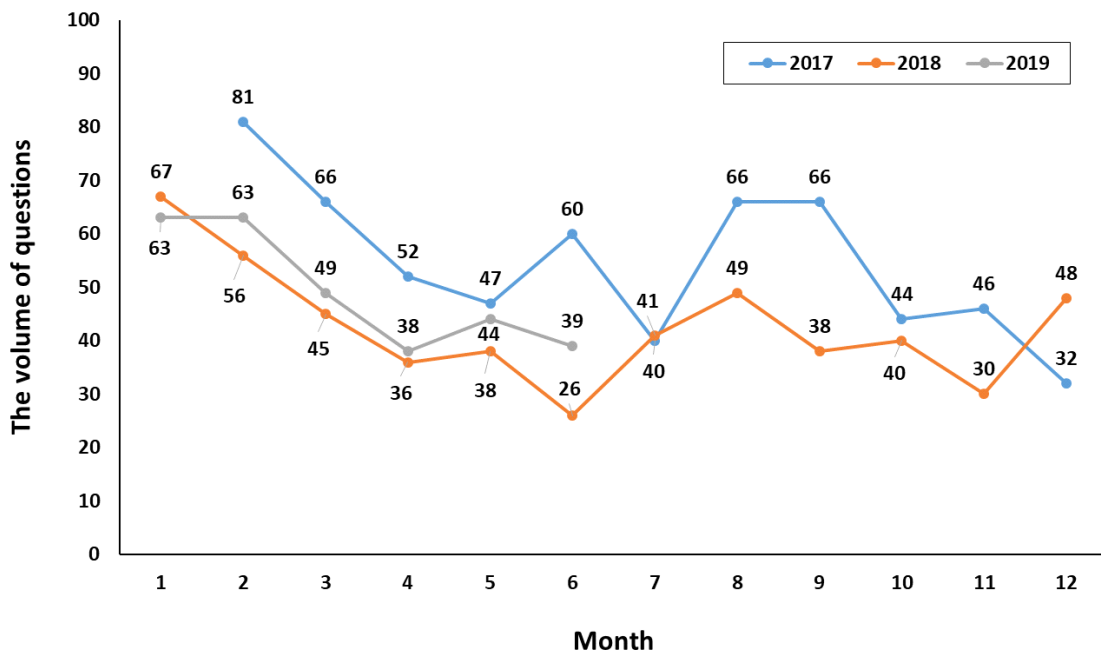
#### 1. 네이버 지식인에 등록된 성조숙증 관련 질문의 양과 답변자 특성, 시기별 분포

2017년 1월에서 2019년 6월까지 네이버 지식인에 등록된 성조숙증 관련 질문은 2017년 625건, 2018년 514건, 2019년 296건으로 총 1435건이 확인되었다(Table 1). 이들 중 1개 이상의 답변이 등록된 질문은 총 1418건으로 전

체 질문의 99%가 답변이 등록되어 있었다. 응답자의 분포를 살펴보면 일반인이 55.9%, 네이버 등록 전문가는 44.1%로 전문가 답변이 절반가량을 차지하였다. 전문가 응답자의 특성을 살펴보면 전체 관찰 기간 응답자의 분포는 의사가 많았으나, 한의사의 응답 비중이 최근 3년 사이 빠르게 증가하고 있는 것을 확인 할 수 있었다. [Figure 2]는 성조숙증 관련 질문의 연도별·월별 분포를 나타낸다. 네이버 지식인의 질문 등록은 학생들의 방학시기인 1-2월과 8월에 가장 높았다.

〈Table 1〉 The number of questions and their responders about precocious puberty in NAVER Jisik-iN

Questions	Qusetion with response	Characteristics of responders			
		Doctor of oriental medicine	Doctor of medicine	Non-professional	
		n (%)	n (%)	n (%)	
2017	625	615	9 (1.5)	242 (39.3)	364 (59.2)
2018	514	512	45 (8.8)	175 (34.2)	292 (57.0)
2019	296	291	78 (26.8)	77 (26.5)	136 (46.7)
Total	1,435	1,418	132 (9.3)	494 (34.8)	792 (55.9)



[Figure 2] Annual and monthly trend of question volume about precocious puberty

## 2. 성조숙증 관련 질문의 키워드와 주요 주제 탐색

성조숙증 관련 질문에서 추출된 명사 중 최다 빈출 단어는 ‘호르몬’, ‘몸무게’, ‘초등학교’, ‘뼈 나이’, ‘성장판’으로 초등학생의 성조숙증 진단 검사와 관련된 질문이 많았다(Supplementary table 1). 질문에서 언급된 단어들을 보다 자세히 확인하기 위해 5회 이상의 모든 빈출 단어들은 워드 클라우드(Word cloud)로 제시하였다(Supplementary figure 1).

〈Table 2〉는 LDA 토픽 모델링을 통해 도출한 성조숙증 관련 질문들의 3개 주요 주제와 주제별 최다 빈출 단어 10개를 나열한 것이다. 주요 단어들을 토대로 Topic 1을 살펴보면, 초등학생, 여자 어린이의 성조숙증 관련하여 전문가 의견을 구하는 질문이 많은 것으로 추정된다. 또한 Topic 1의 키워드 중 한의원이 3번째로 많이 나타난 것을 고려할 때, 성조숙증의 한방치료에 대한 관심이 상당히 높은 것으로 보인다. Topic 2는 호르몬, 뼈 나이, 엑스레이, 억제제, 초음파 등 주로 성조숙증 검사와 치료에 대한 질문이 주를 이루는 것으로 보인다. Topic 3은 몸무게, 줄넘기, 다이어트의 키워드를 토대로 할 때 비만과 다이어트 관련 질문이 성조숙증 환자와 보호자들의 주된 관심영역 중 하나임을 확인 할 수 있었고, 중학교, 여드름, 겨드랑이, 브래지어 등의 키워드를 고려할 때 성조숙증 치료 후 2차 성징 발현 시

기에 대한 질문도 상당수 존재하는 것으로 생각된다. 각 토픽간 다차원의 주제간 거리지도와 연관성 추정치(relevance estimate)를 기초로 한 주요 단어리스트는 부록에 제시하였다(Supplementary figure 2-3).

〈Table 3〉은 각 토픽에 해당하는 성조숙증 관련 질문의 실제 사례를 제시한 결과이다. Topic 1의 주요 사례들을 살펴보면, 성조숙증을 의심하는 어린이가 본인의 신체발달 상황 혹은 성적 성숙과정이 성조숙증 질환에 해당하는지 의견을 구하는 질문과 성조숙증이 의심되는 어린이의 성인 가족이 아이의 증상과 신체발달 상황을 서술하며 좋은 치료기관을 알려달라는 질문들이 많은 것을 확인 할 수 있었다. Topic 2의 질문은 성조숙증 진단검사에 대한 질문, 진단 후 치료를 받아야 하는지 상담하는 질문과 성조숙증 치료과정에서 주의해야 할 점이나 부작용 등을 묻는 질문들이 주로 분포하고 있었다. Topic 3는 비만한 아동의 다이어트 관련 질문과 더불어 비만으로 인해 성조숙증이 생길 수 있는지 걱정하는 상담내용과 성조숙증 치료 후 2차 성징이 언제쯤 나타날지 혹은 현재의 신체 상태가 2차 성징에 해당하는지 묻는 질문으로 구성되어 있음을 확인할 수 있었다. 각 토픽 간의 독립성을 표현하는 다차원 주제간 거리지도와 주요 연관 단어 30개의 리스트들은 부록에 제시하였다(Supplementary figure 2-3).

〈Table 2〉 The 3 distinctive topics of questions about precocious puberty

Topics	Topic name	Topic words
Topic 1	Asking for expert opinion on diagnosis or hospital selection	‘초등학교’, ‘성장판’, ‘한의원’, ‘사춘기’, ‘여자아이’, ‘인터넷’, ‘어린이’, ‘부작용’, ‘일주일’, ‘전문가’ Elementary school, Growth plate, Oriental clinic, Puberty, Girl, Internet, Child, Side effect, One week, Expert
Topic 2	Questions about diagnostic tests and GnRHα	‘호르몬’, ‘뼈나이’, ‘선생님’, ‘엑스레이’, ‘대학병원’, ‘부작용’, ‘억제제’, ‘소아과’, ‘비타민’, ‘초음파’ Hormone, Bone age, Sir/Ma'am, X-ray, University hospital, Side effect, GnRHα, Pediatrics, Vitamins, Sonogram
Topic 3	Questions about obesity/diet & puberty timing	‘몸무게’, ‘중학교’, ‘스트레스’, ‘여드름’, ‘겨드랑이’, ‘여학생’, ‘부모님’, ‘줄넘기’, ‘다이어트’, ‘브래지어’ Body weight, Middle school, Stress, Pimple, Armpit, Female student, Parents, Jump rope, Diet, Brassiere

**Note.** Keywords are presented in both Korean and English to show the original meaning and structure.

<Table 3> Samples of original questions about precocious puberty according to the 3 distinctive topics

Topics	Topic name	Questions <sup>a</sup>
Topic 1	Asking for expert opinion on diagnosis or hospital selection	<p>“제가 초등학교 4학년인데 가슴에 몽우리가 엄청 딱딱하고 3~4cm 정도로 잡힌것 같아요 성조숙증 인가요 ??부모님께 말씀드리긴 싫어요 ㅠㅠ ”</p> <p>“I am in the 4th grade in elementary school, by the way, my breast lump is very hard and about 3 ~ 4cm in size. Am I with precocious puberty? I don't want to tell my parents.”</p>
		<p>“XX<sup>b</sup> 성조숙증 문의 좀 드릴게요. 또래들보다 키는 작은데 몸무게도 많이 나가고 몸은 완전 어른이 예요. 조숙증 치료를 한의원에서도 한다고 들어서 치료를 좀 받아볼까 하구요. 효과 좋은 XX성조숙증 한의원 있으면 좀 알려주세요”</p> <p>“I would like to ask about the precocious puberty in XX. Although I am shorter than my peers, I am much weight and my body is almost like an adult. I heard that treatment of precocious puberty is also provided in the oriental clinic, so I want to try. Please tell me which oriental clinic is good at treatment of precocious puberty in XX.”</p>
		<p>“안녕하세요.제 조카가 가슴에 멍울이 생겨서 병원에 갔는데 성 조숙증일 가능성이 크므로 대학병원에 가서 검사를 해보라고 하는데 한달이나 기다려야 한다고 합니다.작년에도 멍울이 있어서 병원에서 약을 먹고 없어졌습니다.한달동안 이대로 방치를 해도 되는건지 너무 걱정이 됩니다.한의원에 방문을 하더라도 반드시 검사를 받은 후에 가야 하는지 아니면 내원을 해서 진료를 통해 진단이 되는지 궁금합니다.만약 성조숙증이라면 치료 기간을 어떻게 되는지 치료는 되는것인지 너무 궁금합니다.한달간 기다리자니 혹시 치료 시기를 놓치지 않는을까 걱정이 됩니다.제가 키와 체중을 정확히 몰라서 \0\으로 하였습니다.키와 몸무게는 보통 입니다.”</p> <p>“Hello. My niece went to the hospital because of her breast lump, and doctor suggested an examination because she is likely to be with precocious puberty, but we have to wait a month. Last year, she had also taken a medicine because of her breast lump, and it disappeared. I'm so worried if she will be ok without any treatment for a month. I am wondering if she must visit a oriental clinic after taking the examination? Or is she diagnosed through doctor's meeting at the clinic? I'm worried that she will lose treatment timing while waiting for a month. I don't know her height and weight exactly, so I wrote it as \0\'. Height and weight are normal.”</p>
Topic 2	Questions about diagnostic tests and GnRHa	<p>“성조숙증 주사 맞아야 할까요?며칠전 엑스레이와 피검사를 대학병원에서 하고왔습니다.현재 아이나는 만9세이며 여아 뼈검사결과 만11세 2년정도빠르게나왔고 호르몬검사결과수치가 2.8정도나왔습니다 호르몬수치는 5가넘어야 성조숙증이라고 하더라고요. 병원에선 주사를 맞는 걸 권하는 상태고 요아이키는 145 몸무게가 44키로 비만이어서 성장판이 빨리닫힐거라고 합니다 부모입장으로 주사가 만약부작용이 있을지 걱정이고 혹시나 주사의 남용에 휘둘리는 건 아닌지 ..병원의 권유로 주사를 맞아야는지 아니면 식이조절과 운동으로 지내야하는지 궁금합니다 참고로 가슴멍울이 생겼습니다”</p> <p>“Should my child get an injection for the treatment of precocious puberty? A few days ago, my child had an X-ray and a blood test at a university hospital. My child is 9 years old, and the bone age was 11 years old, which is about 2 years fast, and the hormone test result was about 2.8. I heard that hormone level above 5 is diagnosed as the precocious puberty. The hospital recommended the injection, and the doctor said that her growth plate would close quickly because of obesity (her height 145cm and weight 44kg). I am wondering as a parent if the injection has side effects and if it is abused. I am wondering if she should get the injection according to doctor's recommendation or if she needs diet and exercise. In addition, she has a breast lump.”</p>
		<p>“저는 지금 성조숙증 치료를 받고 있는 초등학교 5학년 이에요. 6월달에 XX 교수님<sup>b</sup>이 치료 끊는 거 추천 한다고 하거든요.. 교수님 말씀 으로는 1년 반 동안 뼈나이를 같게 해서 2차 성징을 늦추는 치료라고 하세요..치료를 끊으면 언제 쯤 초경을 시작하게 될까요? 갑자기 일어나서 보면 너무 겁이 날것 같아요ㅠㅠ 정확하진 않아도 대충 알수 있었으면 좋겠습니다..”</p> <p>“I am in the 5th grade in elementary school, being treated because of precocious puberty. Prof. XX recommended stopping the treatment in June. The professor said that it's a treatment that slows puberty by making bone age the same of my age for a year and a half. When do I start menarche after stopping the treatment? I think it's too scary when it starts suddenly. I want to know even though it's not accurate.”</p>

Topics	Topic name	Questions <sup>a</sup>
Topic 2	Questions about diagnostic tests and GnRH <sub>a</sub>	<p>“딸아이가 10살인데 성조숙증진단받아서 첫호르몬주사맞았는데음식을 조심하고 먹이지말라는게 많은데 어떻게주의해야되는지알려주세요”</p> <p>“My daughter is 10 years old, and had her first hormone injection due to precocious puberty. There's a lot of foods that I should not serve. Please let me know how I should be careful.”</p>
		<p>“제가 성조숙증인데요 나이 13살에 키153 몸무게 60kg입니다 몸무게 보시다시피 비만이고요 지금 병원에서호르몬 막는?? 제가 생리를 해서 생리를 막는 주사를 맞고 있는데.. 키 크는 주사를 맞아야 할까요? 아니면 지금 맞는 주사를 계속 맞으면 되나요?? 그리고 성조숙증은 다이어트 가능 한가요? ㅠㅠ 나올 수는 있겠죠??”</p> <p>“I am diagnosed with precocious puberty, 13 years old, height 153 cm, and weight 60kg. As you know, I am obese and I am getting injections to prevent hormone or menstruation because I am menstruating. Should I get an injection to grow my height or keep getting the same shot? Can I diet even if I have precocious puberty? Can it be cured?”</p>
Topic 3	Questions about obesity/diet & puberty timing	<p>“초경에 관한 질문 합니다.저는 2학년때 가슴 몽우리가 나와 성조숙증 치료를 받았습니다. 6학년때 까지 받았구요, 치료를 마치고 나서 가슴이 나오기 시작했습니다. 그래서 가슴이 나온지는 1년 가까이 됐어요. 저는 14살에 키는 158~9정도에 몸무게는 45kg정도 나갑니다. 속옷에 냉은 나온지 좀 오래됐습니다. 불투명하거나 노랗게 냉이 나와요. 6학년 때 속옷에 갈색 냉이 나온적이 있었습니다. 그런데 그 이후로는 불투명하고 노란 냉만 나왔습니다. 음모도 조금 났습니다. 그리고 자주 배가 아플 때가 있습니다.질문· 위 증상이 초경 직전 증상인가요?/초경은 언제쯤 할까요?· 만약 학교 수업 시간때 첫 생리가 터지면 어떻게 해야하나요?/친구와 같이 화장실을 가는게 좋은가요?· 생리를 시작하면 샌 피 또는 생리대 처리는 어떻게 해야하나요?”</p> <p>“I am asking about menarche. I had a treatment for precocious puberty due to breast lump when I was in the second grade. I was treated until the sixth grade, and my breast was growing after I finished the treatment. So it's been almost a year since my breast grew. I am 14 years old, height 158~9cm, and weight 45kg. It has been almost a year since vaginal discharge was found in my underwear. It is opaque or yellow. When I was in the sixth grade, brown vaginal discharge appeared in my underwear. Some hairs grew in the pubic area. I often have an abdominal pain. Question. Are these symptoms just before menarche? When do I start menarche? What should I do if my first period occurs during class? Should I go to the bathroom with a friend? What should I do with the blood or sanitary napkin when I start my period?”</p> <p>“내공100안녕하세요 중학생 여자 입니다 제가 초등학교 때 성조숙증이라는 거가 있어서 성조숙증 주사를 맞았었습니다 아마 중학교 올라오고나선 안 맞았었던 거 같아요 근데 전엔 주사 맞아서 가슴이 안 크는구나 시간 지나면 크겠지 했는데 졸업이 다가와도 전혀 안 커요 그냥 남자 수준입니다 거의 너무 스트레스 받아요 그냥 이제 제 성장이 끝난건가 싫어요 저희 부모님도 그냥 작진 않고 보통 사이즈 인데 전 이거 맞아서 안 크는건가요 그냥 안 크는건가요 정말 스트레스 받아요 .. 제발 알려주세요ㅠㅠㅠㅠ”</p> <p>“Naegong 100. Hello. I am a middle school student. I was diagnosed with precocious puberty and got injections when I was in elementary school. Perhaps I stopped taking the injections after entering middle school. I used to think that my breast did not grow because of the injections and it will grow someday, but It's not growing at all, even when it's almost time to graduate. I am almost like a man. I am very stressful. I just think my growth is over. My parents aren't small, but I am not growing because of the injections? I am really stressful. Please let me know.”</p>

**Note.** <sup>a</sup> Sample questions are presented in both Korean and English to show the original meaning and structure. Original text of question was presented without pre-processing the text (ex. spacing, cleaning) to confirm the actual content. Sample questions could be duplicated in each topic because some questions had more than one theme.

<sup>b</sup> Typical names were masked as 'XX'.



## IV. 논의

본 연구는 성조숙증 환자와 가족들이 네이버 지식인에 올린 질문들의 텍스트 분석을 통해 이들의 온라인 상담 특성을 파악하고, 상담의 주요 주제별로 키워드와 실제 질문사례들을 살펴보았다. 연구결과 연간 500건 이상의 성조숙증 관련 상담 질문이 네이버 지식인에 게시되고 있었고 대다수의 질문에 답변이 등록되어 온라인 건강상담이 활발히 이루어지고 있는 것을 확인할 수 있었다. 질문 응답자의 절반가량은 전문가 집단으로 조사되었고, 한의사의 건강상담 사례는 일반 의사보다 적었으나 최근 3년간 빠르게 증가하는 경향을 보였다. 빈출 단어 분석에서도 한의원이 8위에 랭크되었다. 질문의 주요 주제는 초등학교 여자 어린이의 성조숙증 관련 전문가 의견 요청(치료기관, 부작용 포함), 성조숙증 검사결과와 치료관련 문의, 비만/다이어트와 2차 성징이 나타나는 시기에 대한 것이었다.

네이버 지식인의 건강 상담 사례들을 살펴보면, 주요 빈출 단어 10위안에 ‘한의원’이 랭크 되어 한방 치료에 대한 관심이 높음을 확인할 수 있었다. 성조숙증의 침습적 치료 과정에서 느끼는 부담감과 장기 호르몬 치료의 부작용 발생 우려가 대안적 치료방법 중 하나인 한방치료를 대한 관심을 높인 것으로 추정된다. 그러나 답변자 중 한의사의 비율이 최근 급격히 증가하고 있음을 고려할 때, 한방치료를 대한 관심은 환자와 보호자의 요구뿐 아니라 한의사들의 적극적 자문활동도 영향을 미치고 있는 것으로 생각된다. 실제로 네이버 지식인의 한방 상담현황을 분석한 선행 연구에서도 한의학 분야의 전문가 수는 일반 의사에 비해 소수였으나, 2012년 기준 한의사 1인당 답변 수가 평균 약 90건에 이르는 등 적극적으로 활동을 하고 있다고 보고된 바 있다(Kim & Lim, 2013). 온라인 상담으로 인한 한방 진료의 증가가능성에 대해서도 향후 연구를 통해 추가적인 논의가 필요할 수 있다.

성조숙증 관련 질문들의 주요 주제는 아동의 현재 신체 발달 상황이 성조숙증에 해당하는지에 대해 전문가 의견을 묻는 것이었다. Topic 1의 빈출 단어에서 확인한 바와 같이 질문의 대상연령은 초등학생, 성별은 여자아이였다. 해당 토픽의 등록 질문의 전체 텍스트를 확인해보면, 보호자가 올린 질문뿐 아니라 아동 본인이 자신의 신체적 변화가 성조숙증에 해당하는지 묻는 질문도 상당수 존재하였다. 학령

기 아동은 신체적 차이점을 인식하게 되어 주변 친구들과의 신체적 차이에 대해 민감하게 반응하게 되나(Chun & Jung, 2019), 신체변화를 수용하고 이해할 수 있는 능력은 부족하여 신체적 변화에 대한 두려움과 혼란스러운 감정을 경험하게 된다(Mendle, Turkheimer, & Emery, 2007). 초등학교의 학교 성교육은 연간 15시간 이상을 의무 실시하도록 규정하고 있으나, 실제 교육 현장에서 성교육 시간을 편성하여 운영하는 비율은 매우 낮았다(Jin & Park, 2011; Kim, Kang, & Lee, 2018). 따라서 이 결과는 성 인지도가 낮은 초등학교 저학년 아동들의 미충족된 교육 요구가 온라인 매체를 통해 표출되고 있는 것으로 해석할 수 있다. 따라서 본 연구결과를 토대로 할 때 성조숙증에 대한 이해를 높이기 위한 초등학교 학생과 학부모를 대상으로 한 성교육이 필요할 것이다.

Topic 2는 성조숙증 검사·치료 과정에서 보호자들의 올린 질문이 주를 이루고 있었다. 등록된 질문의 전체 내용을 살펴보면, 환자의 보호자들이 검사결과와 수치들을 제시하며, 검사 결과의 의미와 진단의 정확성을 재확인 하는 내용이 많았다. 이 결과는 치료기관에서 성조숙증 환자의 보호자들이 충분한 정보를 제공받지 못하고 있음을 시사한다. 실제로 성조숙증 아동의 어머니를 대상으로 한 연구에서도 이들의 성조숙증 관련 교육 경험율은 낮았고, 교육 요구도는 매우 높은 것으로 조사되었다(Lee et al., 2015). 더불어 성조숙증 치료 종결 이후의 생활과 신체적 성장에 대해서도 많은 질문이 있었던 점을 고려할 때, 치료 완료 후 환자와 보호자들이 관련한 정보를 획득할 수 있는 적절한 정보전달 통로가 부족함을 보여준다. 따라서 의료기관에서는 환자와 보호자를 대상으로 적절한 건강정보를 제공할 수 있는 프로그램 마련이 필요하며, 치료 종결 후의 추후 관리 방안에 대한 모색도 필요하다.

그러나 본 연구는 다음의 제한점을 가진다. 첫째, 일개 인터넷 포털 사이트의 질의내용을 연구대상으로 하여 인터넷 포털 사이트의 선호 경향에 따라 연구대상자의 선택편향이 발생했을 수 있다. 또한 온라인 매체에 대한 접근성이 떨어지는 취약계층은 연구대상에서 배제되었을 가능성이 크다. 둘째, 본 연구는 uni-gram 단위의 명사만을 분석대상으로 한정하여, 문장의 전체적인 맥락을 정확히 파악하기 어려운 한계를 가진다. 이 연구에서는 이러한 한계점을 극복하기 위하여, 토픽별로 할당된 문서의 전체 질의 내용을 재확인

하여 질문의 전체적인 내용과 주제 도출간의 일치도를 높이고자 하였다. 그러나 단일 단어들의 조합으로 주요 주제를 추론하는 과정은 연구자의 주관성이 온전히 배제하기 어렵다. 추후 연구에서는 두개(bi-) 혹은 세개(tri-) 단어 단위의 분석을 통해 문장의 맥락을 반영한 토픽 모델링을 수행할 필요가 있다.

## V. 결론

이 연구는 성조숙증 관련 인터넷 상담내용을 분석하여, 온라인 건강상담의 현황과 이들의 주요한 주제를 파악할 수 있었다. 온라인 상담의 절반 가량이 관련 전문가에 의해 답변이 이루어지고 있는 점은 보건교육의 매체로써 온라인 건강상담이 유용한 방안이 될 수 있음을 시사한다. 그러나 동시에 절반 가량의 성조숙증 관련 질문이 비전문가에 의한 답변임을 고려할 때 부정확한 건강정보의 확산으로 인한 폐해도 우려될 수 있다. 또한 상담 내용의 주요 주제가 진단 검사 및 치료에 대한 질문임을 고려할 때 성조숙증 환아와 보호자에게 의료진의 적절한 건강정보 제공이 이루어지고, 의료기관에서 체계적인 교육 프로그램을 마련할 필요성이 있음을 확인할 수 있었다. 최근 높아지는 성조숙증 발생률을 고려하여 초등학교 저학년 학생과 학부모가 정상적인 신체적 성숙과정과 성조숙증을 구분할 수 있도록 보다 적극적인 성교육이 이루어져야 할 것이다. 향후 소셜 네트워크를 비롯한 다양한 온라인 매체를 활용한 후속 연구들이 이루어진다면, 성조숙증 환아와 가족을 위한 실제적 교육 프로그램 마련의 기초자료로 활용할 수 있을 것이다.

## References

- Adusumalli, S., Lee, H., Hoi, Q., Koo, S.-L., Tan, I. B., & Ng, P. C. (2015). Assessment of web-based consumer reviews as a resource for drug performance. *Journal of medical Internet research*, 17(8), e211. doi: 10.2196/jmir.4396.
- Berberoğlu, M. (2009). Precocious puberty and normal variant puberty: Definition, etiology, diagnosis and current management. *Journal of clinical research in pediatric endocrinology*, 1(4), 164-174. doi: 10.4274/jcrpe.v1i4.3.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Chun, S. M., & Jung, H. Y. (2019). Experience of precocious puberty in primary school girls with hormone therapeutics. *The Journal of Korean Academic Society of Nursing Education*, 25(4), 459-470. doi: 10.5977/jkasne.2019.25.4.459.
- González, E. R. (1982). For puberty that comes too soon, new treatment highly effective. *The Journal of the American Medical Association*, 248(10), 1149-1155. doi: 10.1001/jama.1982.03330100003001.
- Jin, H. S., & Park, H. R. (2011). Sexual knowledge and perception and current status of sex education among parents of first and second grade elementary schoolers. *Child Health Nursing Research*, 17(3), 198-205. doi: 10.4094/jkachn.2011.17.3.198.
- Johansson, T., & Ritzén, E. M. (2005). Very long-term follow-up of girls with early and late menarche. In H. A. Delemarre-van de Waal (Ed.), *Abnormalities in puberty* (Vol. 8, pp. 126-136). Basel: Karger Publishers. doi: 10.1159/000084098.
- Kim, H., Park, J. H., & Ji, E. (2016). Pattern of pharmacist consultation among the health counseling services via Internet portal sites: A pilot study. *Korean Journal of Clinical Pharmacy*, 26(4), 324-329.
- Kim, S.-J., Kang, S.-R., & Lee, J. M. (2018). Development of a sexual abuse prevention education program for elementary school students using a hybrid application. *Child Health Nursing Research*, 24(1), 109-118. doi: 10.4094/chnr.2018.24.1.109.
- Kim, S. G., Cho, H. J., & Kang, J. Y. (2016). The status of using text mining in academic research and analysis methods. *The Journal of Information Technology and Architecture*, 13(2), 317-329. UCI: G704-SER000010357. 2016.13.2.010.
- Kim, S. Y. (2018). Etiology and treatment of central precocious puberty. *Journal of the Korean Medical Association*, 61(10), 591-598. doi: 10.5124/jkma.2018.61.10.591.
- Kim, Y.-K., & Lim, B. (2013). Internet health counseling for Korean medicine in the NAVER Jisik-iN. *Journal of Society of Preventive Korean Medicine*, 17(1), 51-63. UCI: G704-002177.2013.17.1.006.
- Kim, Y. J., Kwon, A., Jung, M. K., Kim, K. E., Suh, J., Chae, H. W., . . . Kim, H. S. (2019). Incidence and prevalence of central precocious puberty in Korea: an epidemiologic study based on a national database. *The Journal of Pediatrics*, 208, 221-228. doi: 10.1016/j.jpeds.2018.12.022.

- Krstevska-Konstantinova, M., Charlier, C., Craen, M., Du Caju, M., Heinrichs, C., de Beaufort, C., . . . Bourguignon, J. P. (2001). Sexual precocity after immigration from developing countries to Belgium: Evidence of previous exposure to organochlorine pesticides. *Human Reproduction*, 16(5), 1020-1026. doi: 10.1093/humrep/16.5.1020.
- Lee, J. H., & Lee, H. K. (2015). A study on unstructured text mining algorithm through R programming based on data dictionary. *Journal of the Korea Industrial Information Systems Research*, 20(2), 113-124. doi: 10.9723/jksis.2015.20.2.113.
- Lee, J. Y., Ju, H. O., & Kwak, M. J. (2015). Disease-related characteristics of children with precocious puberty and their mothers educational needs in a general hospital of Korea. *Journal of the Korean Society of Maternal and Child Health*, 19(2), 279-288. doi: 10.21896/jksmch.2015.19.2.279.
- Lee, K. Y., Kim, C. J., Kim, H. S., Shin, C. H., Yang, S., Lim, J. S., . . . Hwang, J. S. (2011). *Clinical guidelines for precocious puberty*. The Korean Society of Pediatric Endocrinology.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge university press.
- Mendle, J., Turkheimer, E., & Emery, R. E. (2007). Detrimental psychological outcomes associated with early pubertal timing in adolescent girls. *Developmental review*, 27(2), 151-171. doi: 10.1016/j.dr.2006.11.001.
- Naver. Naver Jisik-iN. Retrieved from <https://kin.naver.com/>
- Oh, S. H., Shin, N. R., & Hwang, S. J. (2018). Analysis of online questions about dental hygienist: case report in Naver Intelligent Network. *Oral Biology Research*, 42(1), 62-66. UCI: I410-ECN-0102-2018-500-004081092.
- Park, S. H., & Hong, S. H. (2018). Identification of primary medication concerns regarding thyroid hormone replacement therapy from online patient medication reviews: Text mining of social network data. *Journal of Medical Internet Research*, 20(10), e11085. doi: 10.2196/11085.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rhie, Y. J., & Lee, K. H. (2015). Overview and treatment of precocious puberty. *Journal of the Korean Medical Association*, 58(12), 1138-1144.
- Sadah, S. A., Shahbazi, M., Wiley, M. T., & Hristidis, V. (2016). Demographic-based content analysis of web-based health-related social media. *Journal of Medical Internet Research*, 18(6), e148. doi: 10.2196/jmir.5327.
- Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Jason Chuang, Spence Green, Marti Hearst, Jeffrey Heer, Philipp Koehn (Eds.), *The Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63-70). Maryland, BA: Association for Computational Linguistics.
- Song, J., Song, T. M., Seo, D.-C., Jin, D.-L., & Kim, J. S. (2017). Social big data analysis of information spread and perceived infection risk during the 2015 Middle East respiratory syndrome outbreak in South Korea. *Cyberpsychology, Behavior, and Social Networking*, 22(1), 22-29. doi: 10.1089/cyber.2016.0126.
- Song, J., Song, T. M., Seo, D.-C., & Jin, J. H. (2016). Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. *Journal of Adolescent Health*, 59(6), 668-673. doi: 10.1016/j.jadohealth.2016.07.025.
- Tremblay, L., & Frigon, J.-Y. (2005). Precocious puberty in adolescent girls: a biomarker of later psychosocial adjustment problems. *Child Psychiatry and Human Development*, 36(1), 73-94.
- Woo, H., Cho, Y., Shim, E., Lee, J.-K., Lee, C.-G., & Kim, S. H. (2016). Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *Journal of Medical Internet Research*, 18(7), e177. doi: 10.2196/jmir.4955.

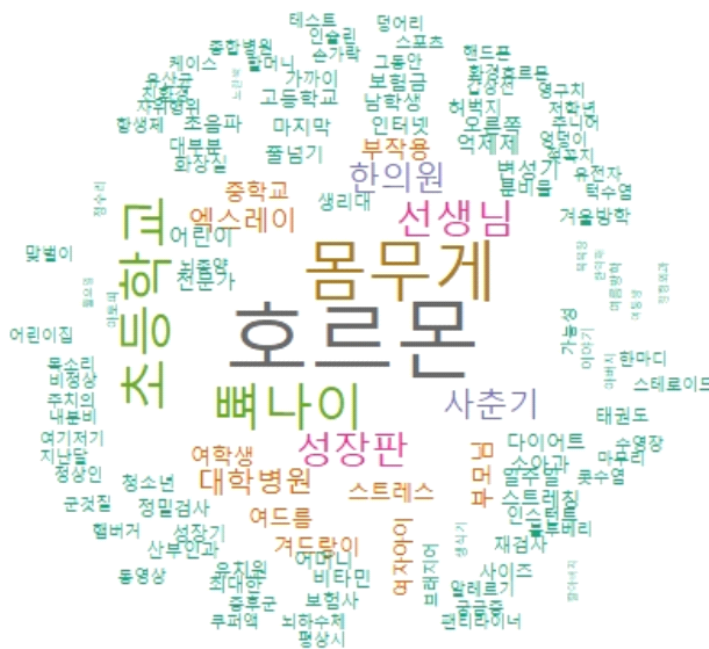
■ Susan Park	<a href="http://orcid.org/0000-0002-7082-8554">http://orcid.org/0000-0002-7082-8554</a>
■ Young-Kyoon Suh	<a href="http://orcid.org/0000-0003-3124-2566">http://orcid.org/0000-0003-3124-2566</a>
■ Jaesung Kim	<a href="http://orcid.org/0000-0003-3663-6177">http://orcid.org/0000-0003-3663-6177</a>
■ Jin-Won Kwon	<a href="http://orcid.org/0000-0003-3467-7805">http://orcid.org/0000-0003-3467-7805</a>

Supplementary

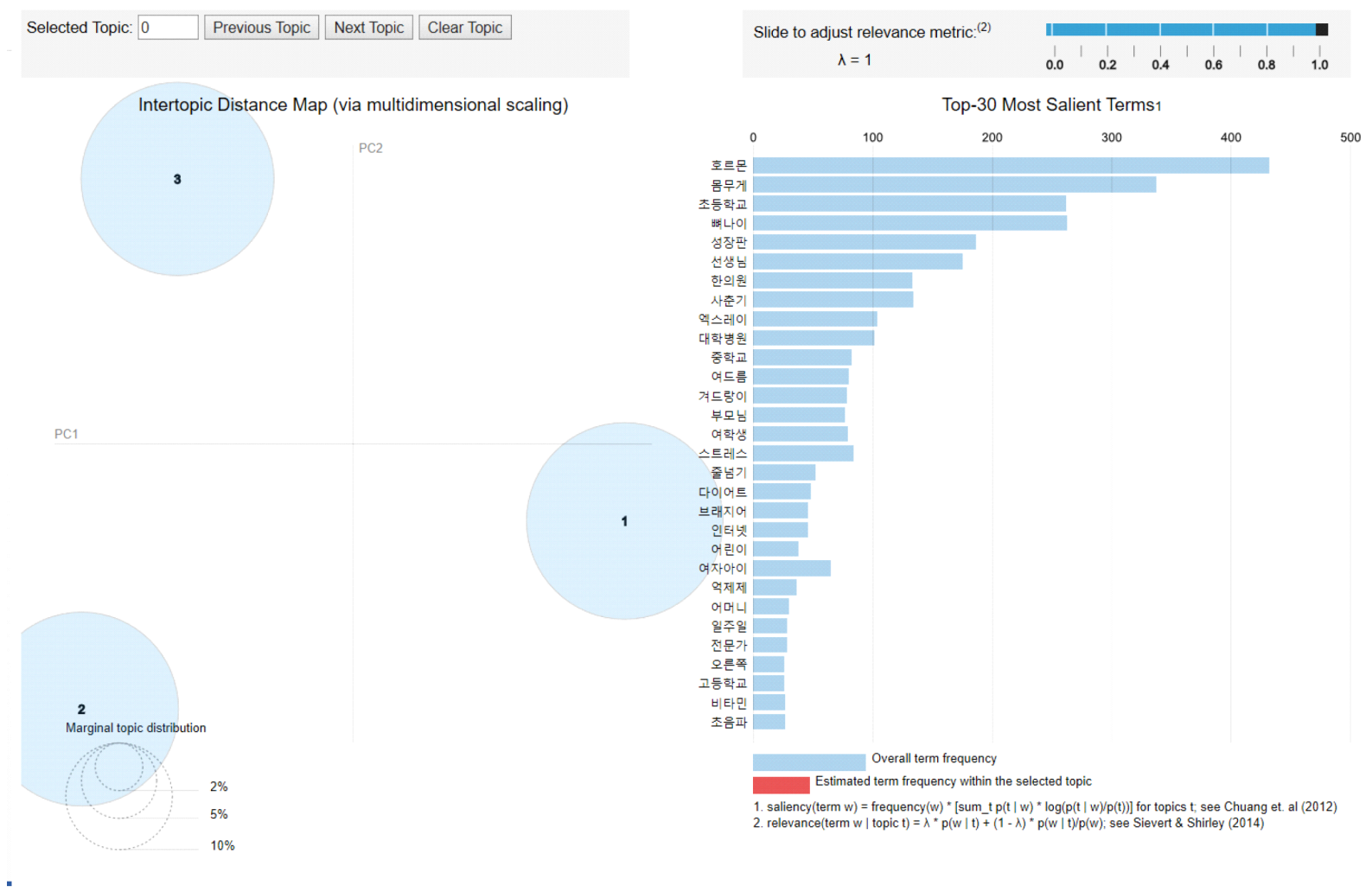
〈Supplementary table 1〉 Top 20 keywords of questions about precocious puberty

Rank	Keywords	Frequency
1	‘호르몬’ Hormone	437
2	‘몸무게’ Body weight	341
3	‘초등학교’ Elementary school	271
4	‘뼈나이’ Bone age	266
5	‘성장판’ Growth plate	193
6	‘선생님’ Sir/Ma’am	177
7	‘사춘기’ Puberty	139
8	‘한의원’ Oriental hospital	138
9	‘엑스레이’ X-ray	105
10	‘대학병원’ University hospital	102
11	‘스트레스’ Stress	85
12	‘중학교’ Middle school	83
13	‘여드름’ Pimple	81
14	‘여학생’ Female student	80
15	‘겨드랑이’ Armpit	79
16	‘부모님’ Parents	78
17	‘부작용’ Side effect	72
18	‘여자아이’ Girl	67
19	‘줄넘기’ Jump rope	53
20	‘다이어트’ Weight loss	49

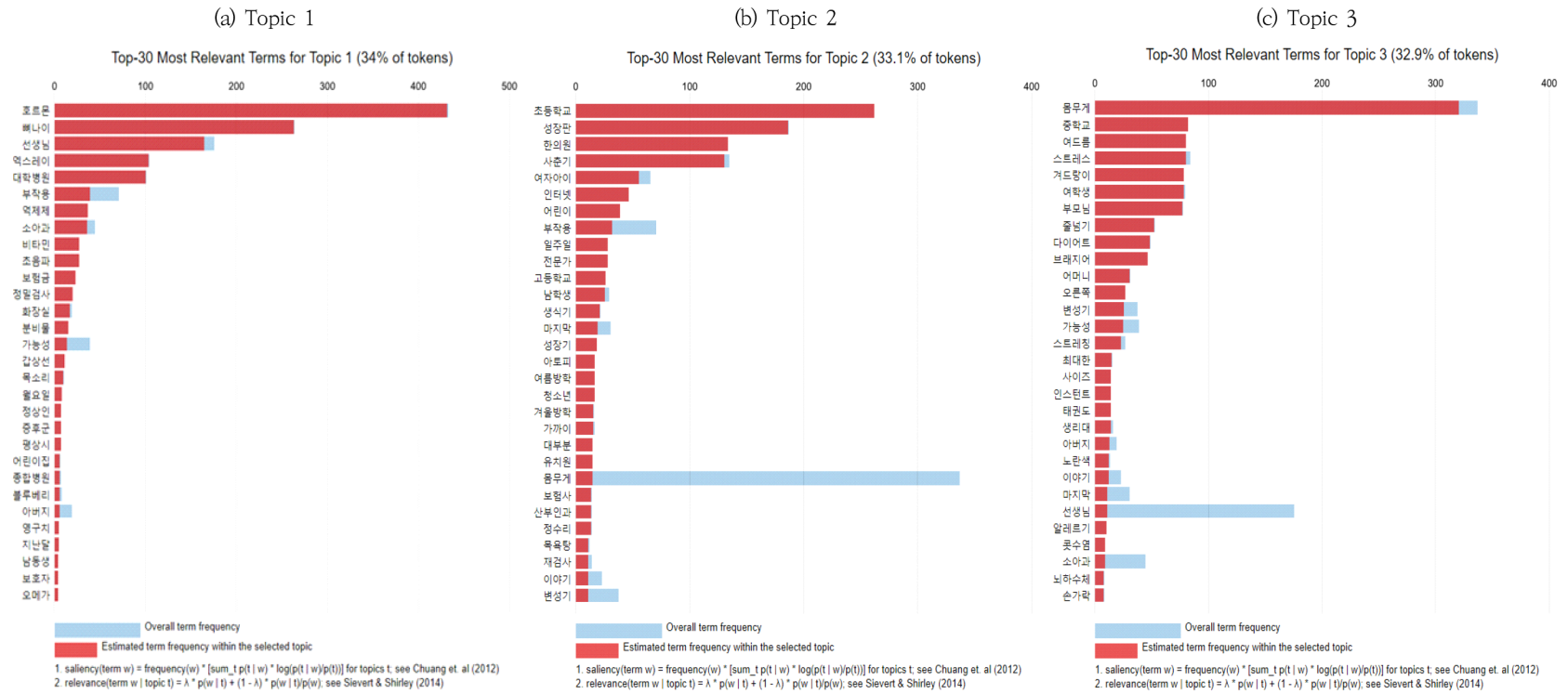
Note. Keywords are presented in both Korean and English to show the original meaning and structure.



[Supplementary figure 1] Word cloud of nouns used more than 5 times in questions about precocious puberty



[Supplementary figure 2] The inter-topic distance map and most important words using LDAvis module in R.



[Supplementary figure 3] Top 30 relevant words by each topic using LDAvis module in R.