

# 텍스트마이닝을 통한 노인장기요양보험 키워드 네트워크 분석

서종근<sup>1</sup>, 남정모<sup>2</sup>, 김태현<sup>3</sup>, 박소희<sup>2</sup>

<sup>1</sup>국민건강보험공단 수원동부지사 주임, <sup>2</sup>연세대학교 보건정보통계학과 교수, <sup>3</sup>연세대학교 의료경영학과 교수

## Keyword Network Analysis on Long Term Care Insurance Using Text Mining

Jonggeun Seo<sup>1</sup>, Chung Mo Nam<sup>2</sup>, Tae Hyun Kim<sup>3</sup>, So-Hee Park<sup>2</sup>

<sup>1</sup>Assistant Manager, Eastern Suwon branch of National Health Insurance Service, Suwon; <sup>2</sup>Professor, Department of Health Information, Yonsei University, Seoul; <sup>3</sup>Professor, Department of Hospital Administration, Yonsei University, Seoul, Korea

**Objectives:** This study conducted research using big data in order to overcome the limitations of existing qualitative research or analysis research. By analyzing keywords, the flow and role of long-term care insurance in society were analyzed. **Methods:** Issues were searched through text mining, one of the big data techniques, and the flow of agendas by period was examined by 3 time points (institutional settlement period, 1st basic plan, 2nd basic plan). Using R and NetMiner, Daum News (news.daum.net) and Naver News (news.naver.com) were web-scraped to collect 20,965 news articles, 4,994 articles were filtered for keyword extraction and analysis. Result: Looking at the characteristics of each data type, in all data types, long-term care institutions (including nursing homes) and care providers appear as the top keywords, and the keyword subgroup characteristics are ① grade/service, ② institution management, and ③ the employee group includes the keyword subgroup. **Conclusions:** This study is based on the subject of long-term care insurance for the elderly and applies big data analysis techniques, and can be used as a decision-making tool in establishing policies and systems.

**Key words:** Text mining, Big data, Keyword network, Long term care, News article

## 서 론

경제협력개발기구(Organisation for Economic Co-operation and Development, OECD) 가입국 중 한국은 최단기간 고령사회로 진입하였고, 2026년 초고령사회로 접어들 전망이다. 더 나아가, 2067년 국내 인구는 1,300만 명 감소('67년 3천 9백만 명)하지만, 전체 인구 중 46.5% ('19년 기준 14.9%)는 노인이 될 것으로 추정하고 있다[1]. 이에 급속한 고령화로 인한 노인 부양부담에 대한 논의는 2000년부터 활발히 시작되었고, 2008년 노인장기요양보험 출범으로 국가가 노인 부양의무를 이행하게 되었다. 노인장기요양보험은 노인 및 부양가족의 심리적·경제적·신체적 부담을 경감시키고[2,3], 가족의 사회활동 및 여가활동 시간을 증대[4,5]하는 효과가 관찰되었다. 노인장기요양보험의 사회적 욕구는 높

아져, 2008년 대비 2016년 등급자가 3.5배 높아지고[6] 제도 지출액이 2008년 대비 2017년 10배 상승하였다[7].

노인장기요양보험은 매년 양적 성장이 이루었고, 다양한 근거를 활용해 제도의 사회적 욕구 탐색 및 정책 수립을 하였다. 먼저, 다양한 선행연구가 수행되며 제도의 사회적 역할을 탐색하였고 정책 개발에 근거로 활용하였다. 이와 더불어, 언론 등 미디어 매체로 아젠다(Agenda)가 형성하였고, 정책 수립에 중요한 영향을 미쳤다. 언론보도는 다양한 노인장기요양보험의 문제를 다뤘고, 이를 기반으로 아젠다가 구성되어 노인장기요양 정책으로 제안되었다[8]. 정책이 및 연구자의 이론·실천 분석이 사회적 욕구 탐색 및 정책 개발에 큰 역할을 하였지만, 언론보도 또한 정책 수립에 결정적 역할을 하였다.

언론보도는 제도 수립 및 아젠다 형성에 중요한 역할을 했음에도,

**Corresponding author:** So-Hee Park

50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea  
Tel: +82-2-2228-1532, E-mail: SOHEEPARK@yuhs.ac

Received: May 1, 2021 Revised: June 12, 2021 Accepted: June 30, 2021

No potential conflict of interest relevant to this article was reported.

**How to cite this article:**

Seo J, Nam CM, Kim TH, Park SH. Keyword network analysis on long term care insurance using text mining. J Health Info Stat 2021;46(3):257-266. Doi: <https://doi.org/10.21032/jhis.2021.46.3.257>

© It is identical to the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permit unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2021 Journal of Health Informatics and Statistics

언론보도 등과 같은 텍스트 데이터를 활용한 연구는 없었다. 기존 선행연구는 전문가, 이해관계자, 이용자, 대중 등을 대상으로 한 조사연구가 이뤄졌고, 더불어 전문 문헌을 이용한 분석 연구가 이뤄졌다. 반면, 노인장기요양보험 제도는 상당량의 뉴스가 실시간으로 축적되고 빅데이터(big data)로서 필수요소 3V (규모 Volume, 다양성 Variety, 속도 Velocity) [9]를 모두 갖추었기에 충분한 연구 가치가 있지만, 언론보도를 연구 대상으로 보려는 시도가 없었다. 제도 출범 10년 이상이 지났고 막대한 양의 언론보도가 디지털 데이터로 축적되었기에, 이를 빅데이터 관점에 분석하는 것이 필요하다.

빅데이터 분석은 노인장기요양 제도의 이슈를 살펴보는 데 탁월하고, 이미 다양한 분야에서 분석의 효과성이 입증되었다. 빅데이터 분석은 비즈니스 영역에서 선택이 아닌 필수요소로 자리 잡았고, 기업 이슈 탐색 및 의사결정 근거로 활용되고 있다. 빅데이터 기반 운영은 업무처리 효율성, 경쟁 우위, 조직적 민첩성을 가져오고[10], 기업 간 이슈를 정확히 판단해 문제해결에 탁월하다[11]. 빅데이터는 빠른 데이터 축적 속도, 가공의 어려움, 방대한 데이터양 등으로 다양한 분석 전문성이 요구되지만, 빅데이터 분석방법 중 텍스트 마이닝(text mining) 분석이 정교하게 발전하면서 복잡한 비정형 데이터를 과학적 방법으로 가공할 수 있게 되었다. 지금까지 작성된 노인장기요양보험 관련 기사 수만 건을 특정 개인이 객관적으로 핵심 이슈를 정리하기에 한계가 있다. 텍스트 마이닝 분석은 모든 언론보도를 독해하지 않아도, 분석 기법만으로 전체 텍스트의 주요 이슈를 선별하고 이슈 중요도를 수치화할 수 있다.

본 연구는 노인장기요양보험 제도의 이슈를 탐색하기 위해, 언론보도를 활용해 텍스트 마이닝 분석을 수행하고자 한다. 본 연구는 지난 12년간(2008년 7월-2020년 6월) 축적된 노인장기요양 관련 언론보도를 수집하여 연구에 활용하였다. 노인장기요양보험 이슈의 종단적 특성을 살펴보기 위해, 3 시점별(제도정착기, 1차 기본계획, 2차 기본계획) 주요 이슈를 탐색하였다. 추출된 키워드의 특성 및 관계를 키워드 네트워크 분석으로 살펴볼 예정이다. 본 연구는 텍스트 마이닝 분석으로 노인장기요양 제도의 현안을 파악하고, 사회적 요구를 체계적으로 도출하는 데 그 목적이 있다. 연구 분석을 기반으로 노인장기요양 제도의 동향을 파악하고, 추후 제도 수립의 기초자료로써 활용을 기대할 수 있다.

## 연구 방법

### 연구설계

본 연구는 노인장기요양보험 관련 언론보도의 키워드를 추출해, 핵심 키워드 빈도를 살펴보았다. 더 나아가 추출 키워드 간 관계를 살펴기 위해, 키워드 네트워크 분석을 수행하였고, 연결중심성, 연결강도로

키워드 특성을 탐색하였다. 키워드의 전체 특성을 살펴보고, 더 나아가 종단적 특성을 살펴기 위해 시점별 변화를 살펴보았다. 시점 분류는 보건복지부의 노인장기요양보험 기본계획 적용연도를 기준으로 하였다. 기본계획이 발표되기 전을 제도정착기(2008년 7월-2012년)로 보았고, 이후 보건복지부가 발표한 1차 기본계획(적용연도 2013-2017년), 2차 기본계획(적용연도 2018년-2020년 6월)으로 분류해 시점별 분석을 수행하였다.

### 분석 자료

본 연구는 노인장기요양보험 언론보도 분석을 위해, 제도 출범일인 2008년 7월 1일 00:00부터 2020년 6월 30일 23:59까지의 12년간 온라인 기사를 수집하였다. 뉴스 플랫폼 사이트인 다음 뉴스(news.daum.net)와 네이버 뉴스(news.naver.com)에서 ‘노인장기요양보험’으로 검색어를 설정해, 웹 크롤링(web crawling) 하였다. 웹 크롤링 수행을 위해, 통계 R 4.0.2 rvest와 dplyr 패키지를 사용해 텍스트 데이터를 수집하였고, R 웹 크롤링 코드는 Seo [16]에서 볼 수 있다.

웹 크롤링으로 언론보도 24,942건(다음 뉴스 15,307건, 네이버 뉴스 20,965건)을 수집하였고, 다음 뉴스와 네이버 뉴스에 중복 게재된 뉴스 5,665건을 삭제하였다. 뉴스 기사 19,277건을 분석해 부적합 여부(중복 기사, 불필요한 기사)를 탐색하였다. 먼저, 특수문자, 문단 등을 제거했을 때 중복기사 1,681건이 추가로 발견되어 제외하였고, 노인장기요양보험과 무관한 기사를 다음 3가지 기준에 따라 제외하였다. ① 민간 실버보험 등의 상업·광고성 기사이거나, ② 노인장기요양보험 검색어가 포함되었지만, 유관 기관 동정(홍보, 인사)이거나, ③ 노인장기요양보험 검색어만 포함되었을 뿐, 노인장기요양보험과 무관한 기사일 경우 분석에서 제외하였다.

노인장기요양보험과 연계된 민간 실버보험 홍보 기사가 상당수 있었고, 장기요양기관의 광고·홍보 기사, 국민건강보험공단 동정(인사채용, 사내 대외활동 등)에 관련 기사를 제외하였다. 또한 노인 일자리 정책 기사 등에서 노인장기요양보험 키워드만 포함될 뿐, 관련된 내용이 없는 기사는 분석에 제외하였다. 이처럼 주제어만 포함될 뿐 내용상 관련 없는 기사 12,603건을 제외하였고, 필터링 작업으로 기사 4,993건이 분석 대상으로 선정되었다. 선정된 언론보도 4,993건에 대한 일별 보도 건수를 시각적으로 살펴보기 위해, R 4.0.2의 ggplot2를 활용해 calendar heatmap을 작성하였다.

### 분석절차

#### 사전(Dictionary) 구축

핵심 키워드 추출을 위해, 언론보도 4,993건에 대한 용어 사전을 구축하였다. 본 연구는 불용어 사전, 고유어 사전, 유사어 사전을 작성해,

불필요한 키워드를 제거하고 유의미한 키워드를 추출하였다. 불용어 사전을 구성해, 연구에 불필요한 키워드를 분석에서 제외하였다. “[특종]”, “[영상]” 등의 불용어가 상당수 포함되어 있었고, 기자 성명, 언론사명 등을 불용어로 지정하였다. 고유명사를 키워드로 추출하기 위해 고유어 사전을 구축하였다. 다음으로, 고유어 사전 제작을 위해, ①국립국어원 우리샘사전[12], ②다음 백과사전[13], ③장기요양급여비용 청구 관련 고시[14], ④노인장기요양 용어사전[15], ⑤연구자 제작 고유어 사전을 통합해 고유어 사전으로 제작하였다. 또한, 유사어 사전을 작성하였는데, 동의어임에도 표기법 차이로 다른 키워드 인식하는 것을 방지하였다. 실제로 “국민건강보험공단” 단어를 축약해 “건보공단”으로 표기되는 경우가 잦았는데, 유사어 사전을 구축해 “건보공단”과 같은 동의어를 “국민건강보험공단”이라는 하나의 단어로 인식할 수 있게 하였다. 본 연구에서 구축된 용어 사전은 Seo [16]에서 찾아볼 수 있다.

### 키워드 추출

본 연구는 키워드 추출 분석 도구로 NetMiner를 사용하였다. 키워드 특성을 살펴보고, 연결망을 시각화하기 위해, 측정 가능한 수준에 맞춰 핵심 키워드를 선별하였다. 연구자마다 핵심 키워드 선별기준에는 차이가 있고, Gwak [17]은 전체 키워드 중 7%를 핵심 키워드로 보기도 하였다. 텍스트의 특성 및 속성에 따라 핵심 키워드의 선별기준에 차이가 있기에, 연구 목적에 적합한 기준을 설정해 핵심 키워드를 선별하는 작업이 필요하다[18].

본 연구는 핵심 키워드의 선별기준을 다음 4가지로 설정하였다. ①고유명사, ②키워드 출현 빈도 상위 10% 이상, ③키워드별 출현 문헌(언론보도) 건수 상위 10% 이상, ④키워드 간 연결강도 상위 50% 이상 인지 살펴보았고, 해당 기준에 충족한 키워드를 네트워크 분석에 활용하였다. 먼저, “일반”, “사람”과 같은 일반명사를 분석에 제외하였고, 특성 및 의미를 명확히 유추할 수 있는 고유명사를 핵심 키워드로 선별하였다. 실제로, 고유명사인 “장기요양기관”은 그 의미를 정확히 추론할 수 있지만, 일반명사 “기관”은 단일 키워드만으로 복지기관인지 의료기관인지를 추론할 수 없다. 또한 키워드 전체 등장 횟수 및 키워드별 출현 문헌 건수가 상위 10%에 해당되는 키워드를 NetMiner로 추출해 분석에 활용하였다.

### 빈도분석

키워드 네트워크 분석 전, 키워드 빈도를 살펴보기 위해 워드 클라우드(word cloud)를 작성하였다. 워드 클라우드는 키워드 빈도에 따라 글자 크기가 커지도록 시각화하였고, R 4.0.2의 wordcloud2 패키지를 활용하였다.

### 키워드 네트워크 분석

본 연구에서는 키워드 간 연결 관계를 살펴보기 위해 NetMiner를 활용해 네트워크 분석하였다. 빈도분석은 키워드 출현 정도를 탐색할 수 있을 뿐, 키워드가 어떤 맥락과 의미로 사용되었는지 파악할 수 없다. 반면, 키워드 네트워크 분석은 키워드 간 연결망을 형성해 키워드의 의미를 탐색하고, 특정 이슈가 어떤 맥락으로 사용되었는지 파악할 수 있다. 이를 위해 분석 대상인 언론보도 4,993건에서 핵심 키워드를 앞서 설명한 4가지 기준에 맞춰 추출해 키워드 연결망을 구축하였다. 관계망 측정을 위해 언론보도-키워드에 대한 2-mode 매트릭스(matrix)를 구성하여 네트워크 분석을 하였다. 네트워크 분석은 각 키워드가 다른 키워드와 맺고 있는 관계에 따라 중요성이 평가되고, 중심성(Centrality)으로 각 키워드의 영향력을 평가한다. 네트워크 특성에 따라 각 키워드는 다양한 중심성을 갖게 되는데, 본 연구에서는 연결중심성(Degree Centrality)을 사용해 각 키워드 중요성을 평가하였다. 연결중심성은 특정 키워드가 다른 키워드와 관계를 맺고 있는지 수치화하는 것으로, 연결 빈도를 표준화해 산출한다. 연결중심성은 관계를 가장 직관적으로 볼 수 있기에, 해당 지수를 연구에 활용하였다. 실제로, “치매”라는 키워드가 ‘치매→노인’, ‘치매→인정자’, ‘치매↔노인’과 같은 관계를 맺고 있다면, 치매 키워드의 연결강도는 3(노인 2회 연결, 인정자 1회 연결)이 되고 이를 0부터 1까지 값으로 표준화시킨 수치가 연결중심성이다[17]. 중심성 지수를 통해, 각 키워드의 특성을 파악하고 전체 네트워크에서 어떤 역할을 하는지 살펴보았다. 더불어 네트워크의 모듈성(Modularity)을 활용해 키워드의 하위그룹을 분류하였다. 키워드 간 연결강도가 각기 다르고, 특정 키워드 집단이 높은 연결강도로 응집해 연결될 수 있다. 이때, 특정 키워드 집단의 응집력을 가지고 하위그룹으로 나눌 수 있는데, 본 연구에서 커뮤니티(Community) 분석으로 하위집단 특성을 살펴보았다.

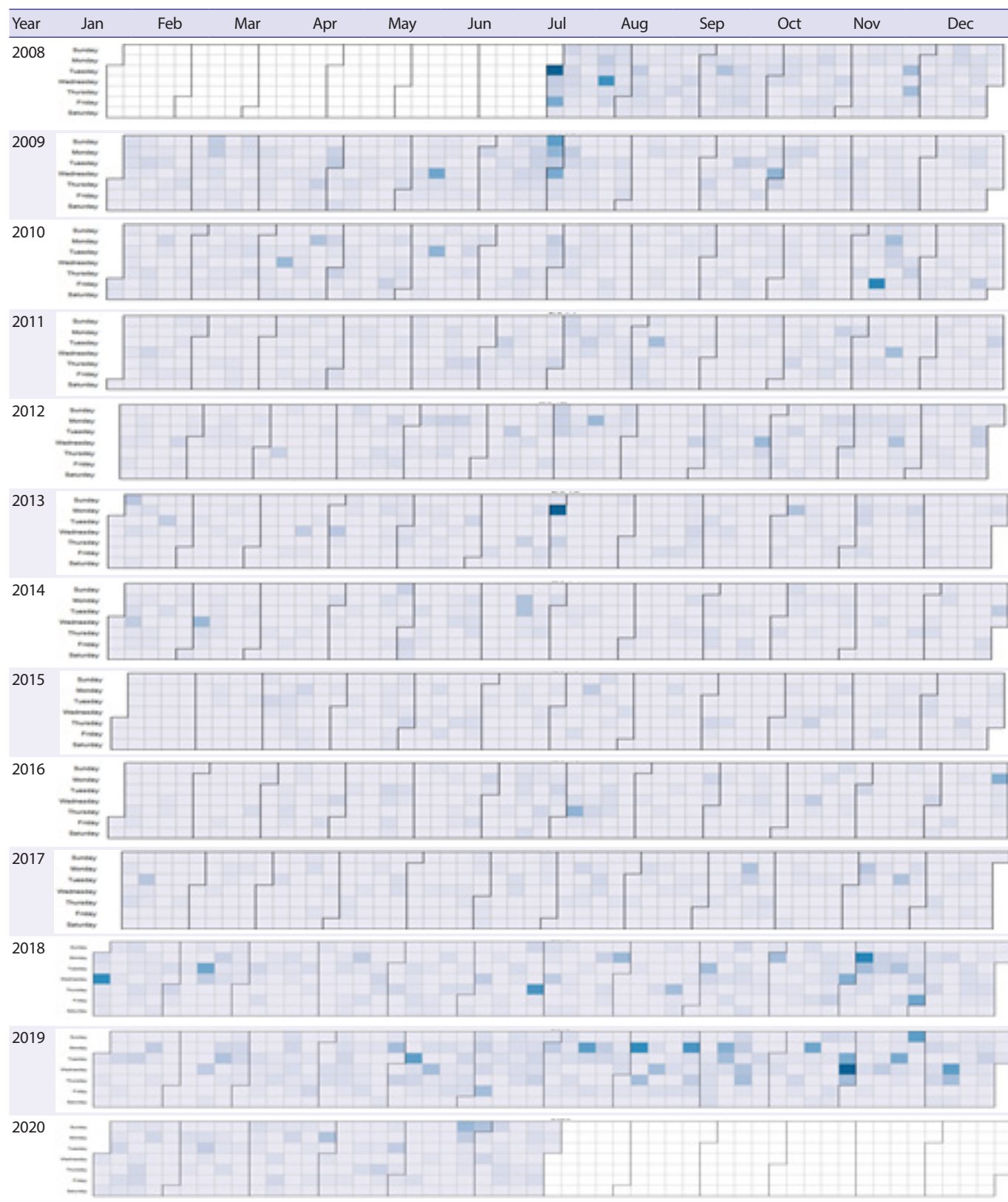
## 연구 결과

### 일별 언론보도 빈도

추출된 언론보도의 일별 건수를 살펴보기 위해, Figure 1과 같이 Heatmap Calendar를 작성하였다. Heatmap Calendar로 12년간 기사 건수를 달력으로 시각화하였고, 기사 건수에 따라 채도를 높였다. Figure 1을 살펴보면, 노인장기요양보험 일별 기사 건수는 정책, 이슈, 행사 등에 영향을 받았다. 언론보도가 높게 나타난 특정 시기를 간략하게 살펴보면 다음과 같다.

언론보도 건수가 높은 일자를 살펴보면, 장기요양보험 제도가 출범한 2008년 7월 첫 주에 기사 건수가 높게 나타났고, 2009년 6월 말 제도 출범 1주년 평가 기사가 여러 건 작성되었다. 2010년 11월 13일은 요양





**Figure 1.** The number of articles per day by heatmap calendar.

일별 언론보도 패턴을 보면, 제도정착기 초반인 2008-2009년 언론

다음으로, 정책 추진 시점별로(제도정착기, 1차 기본계획, 2차 기본



계획) 빈도의 특성을 살펴보았다. 제도정착기의 언론보도 1,852건에서 키워드 9,268개를 추출하였고, 핵심 키워드 138개를 선별해 빈도분석하였다. 제도정착기는 요양보호사( $n=3,618$ ) 빈도가 가장 높았고, 요양원( $n=3,091$ ), 장기요양기관( $n=2,104$ ), 국민건강보험공단( $n=1,957$ ), 장기요양등급( $n=1,022$ ) 등의 키워드 빈도가 높았다. 요양보호사 관련 언론보도가 활발히 다뤄졌고, 열악한 처우에 대한 키워드(성폭행 빈도=141, 저임금 빈도=73 등)와 같이 등장하며 키워드 빈도가 높았다. 제도정착기는 다른 시점보다 장기요양등급, 보건복지부, 부당청구 키워드 빈도가 상대적으로 높게 나타났다.

다음으로, 1차 기본계획 시점에서 작성된 언론보도 2,149개를 대상으로 키워드 9,645개를 추출하였다. 고유명사, 출현빈도 49개 이상, 키워드별 동시출현 기사 건수 34건 이상인 키워드를 추출하였고, 핵심 키워드 162개가 빈도분석에 활용되었다. 요양원( $n=4,317$ ), 장기요양기관( $n=2,737$ ) 키워드의 출현빈도가 가장 높았다. 1차 기본계획은 전체 언론보도의 패턴과 유사한 특성을 보였다. 1차 기본계획에는 치매특별등급( $n=988$ ) 키워드가 높은 빈도로 새롭게 등장했다. 또한 치매 관련 키워드(치매환자,  $n=13,919$ ; 치매노인,  $n=872$ ; 경증치매,  $n=457$  등)가 두드러지게 나타났다.

마지막으로, 2차 기본계획의 언론보도 993건을 대상으로 키워드 8,756개가 추출되었다. 고유명사, 키워드 출현빈도 41개 이상, 키워드 동시출현 기사 건수 26건 이상을 기준으로 선별하였고, 키워드 126개가 빈도분석에 활용되었다. 다른 시점의 언론보도와 유사하게 요양원( $n=2,408$ ), 요양보호사( $n=2,074$ ), 국민건강보험공단( $n=1,384$ ), 장기요양기관( $n=1,319$ ), 인정자( $n=1,057$ ), 보건복지부( $n=756$ ) 등의 키워드가 높은 빈도로 나타났다. 2차 기본계획 키워드 특징으로, 케어( $n=410$ ), 지역사회( $n=554$ ), 커뮤니티 케어( $n=408$ ), 통합돌봄( $n=283$ ), 돌봄서비스( $n=384$ )가 워드 클라우드에서 새롭게 등장했다. 2차 기본계획에서 노인 지역사회 거주 및 관리에 관련된 주제로 새로운 키워드가 등장하였다. 이는 2차 기본계획에 접어들며 공급자가 아닌 소비자 중심 개념이 나타났다 볼 수 있다.

### 키워드 네트워크 분석

통합 시점과 시점별 키워드 네트워크 분석을 Figure 3과 같이 수행하였다. 빈도분석은 키워드 출현 통계량을 살펴봤다면, 키워드 네트워크 분석은 각 키워드 간 관계를 살펴볼 수 있다. 네트워크 분석에서 각 키워드를 노드(Node)라 칭하고, 연결관계를 링크(Link)로 부른다. Figure 3을 살펴보면, 키워드 빈도에 따라 키워드 노드 크기를 설정하였고, 키워드 간 연결강도가 강할수록 링크의 연결선이 두꺼워 지도록 설정해 시각화하였다. 키워드별 하위그룹을 분류하기 위해, 하위그룹마다 채색과 도형 모양을 다르게 설정하였다. 키워드 네트워크 분석에

필요한 핵심 키워드 선별을 위해 연결강도 상위 50%인 링크만 분석에 포함하였고, 키워드 연결이 단절된 노드는 분석에서 제외하였다.

먼저, 전체 키워드를 대상으로 키워드 네트워크 분석을 실시하였다. 전체 시점의 키워드 네트워크 분석은 빈도분석 키워드 203개 중 연결강도 40 이상인 연결망만 분석에 포함하였고, 연결망이 유지된 키워드 37개로 네트워크 분석하였다. 연결중심성(centrality, C)은 빈도분석과 유사한 형태를 보였고, 요양원( $C=0.4$ ), 요양보호사( $C=0.3$ ), 장기요양기관( $C=0.3$ ), 국민건강보험공단( $C=0.3$ ), 인정자( $C=0.3$ )가 높았다. 키워드 간 연결강도(link weight, W)는 ‘부당청구-장기요양기관’이 가장 높았고( $W=465$ ), 다음으로 ‘요양원-요양병원’의 연결강도가 높았다( $W=361$ ). ‘부당청구-장기요양급여’의 연결강도( $W=179$ )가 높았는데, ‘부당청구-장기요양급여-장기요양기관’과 ‘부당청구-장기요양급여-국민건강보험공단’의 삼각관계가 강하게 나타났다. 요양보호사 키워드는 종사자(간호사, 사회복지사) 및 요양원, 재가급여와 높은 연결강도를 보였다. 요양보호사 자격 발급에 대한 문제로 인해, ‘요양보호사-교육기관’의 연결강도( $W=344$ )가 높았다. 또한 매년 언론보도에서 요양보호사의 임금에 대한 문제를 다뤘기에 ‘요양보호사-인건비’ 연결강도( $W=58$ ) 또한 높았다. “인정자” 키워드는 “장기요양등급”과 연결강도( $W=210$ )가 가장 높았고, 다음으로 국민건강보험공단( $W=103$ )이었다. 치매환자를 인정자로 확대하며 ‘치매특별등급-장기요양등급’의 연결강도( $W=167$ )가 높게 나타났다. 키워드 출현빈도는 낮지만, 치매환자와 치매노인 간 관계 특이성이 나타났다. 두 단어 모두 요양원과 강한 연결강도를 갖지만, 치매노인만 주간보호센터( $W=45$ ) 링크가 형성됐고, 더 나아가 단기보호시설과 간접관계를 맺었다. 즉, 치매노인은 재가서비스에 대한 연결망이 형성되었지만, 치매환자는 그렇지 못했다. 노인과 환자의 개념 인식에 따라, 단어 확장성이 다르게 나타났다. 노인과 환자의 개념에 따라 재가서비스 이용에 대한 관계가 다르게 나타나는 것을 볼 수 있다. 또한 “시범사업” 키워드는 재가급여만 연결( $W=44$ )되었는데, 재가급여 확대를 위해 다양한 시범사업의 시행을 추론할 수 있다. “현지조사” 키워드는 출현빈도는 낮았지만, 장기요양기관, 공단, 부당청구와 연결관계를 가졌다.

시점별 키워드 네트워크의 특성은 Figure 3과 같다. 시점별 빈도분석에서 추출된 키워드 중 연결강도 상위 50%인 것만 선별해 키워드 네트워크 분석 수행하였다. Figure 3에서 시점별 빈도와 연결중심성 키워드 양상이 대체로 비슷하였다. 상위 5개 키워드(요양보호사, 장기요양기관, 요양원, 인정자, 국민건강보험)은 동일하게 높은 연결중심성을 보였다. 시점별 언론보도의 상위 연결강도를 살펴보면, 모든 시점에서 ‘부당청구↔장기요양기관’ 키워드 간 연결강도가 매우 높게 등장하였다(제도정착기  $W=172$ , 1차 계획  $W=213$ , 2차 계획  $W=80$ ). 또한 매 시점마다 ‘부당청구-장기요양급여-장기요양기관’의 삼자 연결관계가 나타났다.



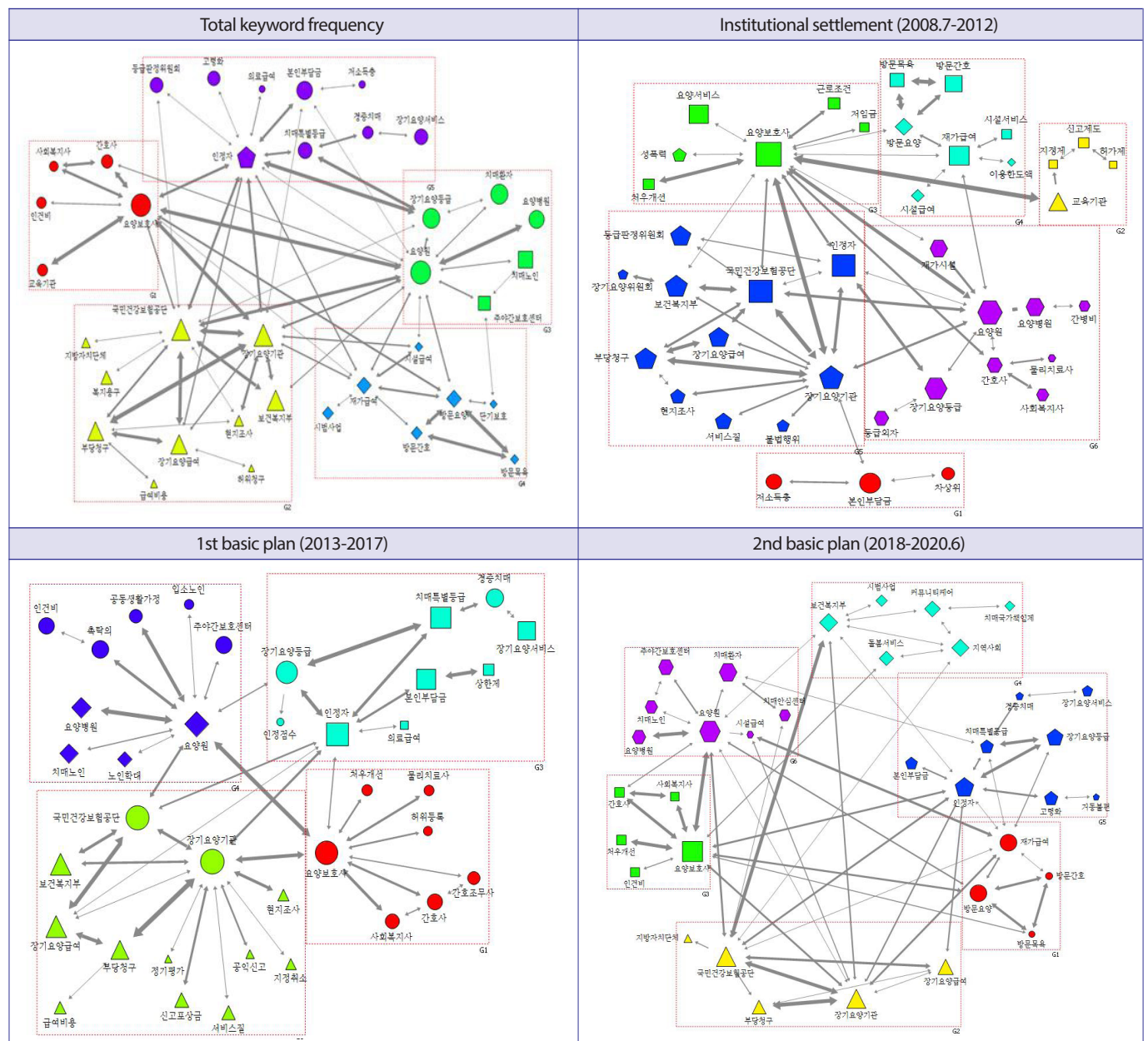


Figure 3. Keyword network by period.

다. 장기요양기관이 장기요양 급여를 통해 부담청구를 했고, 부담청구는 현지조사를 통해 적발되었다는 것을 키워드 네트워크 구조에서 나타냈다. 제도정착기에는 ‘교육기관↔요양보호사’(W=267)의 연결강도가 가장 높았는데, 요양보호사 교육기관의 허가제/지정제 이슈로 두 키워드가 같이 다루어졌다. 1차 기본계획에 접어들면서, ‘요양병원↔요양원’(W=208), ‘요양원↔요양보호사’(W=142)의 연결강도가 상대적으로 높게 나타났고, 2차 기본계획까지 높은 연결강도가 유지되었다. 특히, 요양원 종사 요양보호사와 관련된 언론보도가 1차 기본계획에서 활발

히 다루어졌다. 1차 기본계획에서 ‘장기요양등급↔치매특별등급’의 연결강도(W=106)가 높게 나타났는데, 장기요양등급이 치매특별등급(5등급)까지 확대되면서 언론에서 해당 이슈가 활발히 다뤘다. 2차 기본계획에서 ‘요양보호사↔치우개선’이 상위 연결 강도(W=72)로 나타났고, 제도정착기부터 2차 기본계획까지 두 키워드 연결강도는 꾸준히 높았다. 하지만, 2차 기본계획에 접어들면서 두 키워드의 연결강도가 더 높아졌는데, 요양보호사에게 지급되었던 치우개선비 폐지 이슈가 가시화되면서 이를 반대하는 다양한 언론보도가 2차 기본계획에 등

장하였다.

각 시점별로 주목할만한 세부 특성을 Figure 3에서 살펴보면 다음과 같다. 제도정착기에서 요양보호사와 부정적인 키워드(처우개선, 근로조건, 저임금, 성폭력)가 연결망을 형성하는 특성이 나타났다. 이후, 1차 기본계획 키워드 네트워크에 촉탁의와 공동생활가정 개념이 새롭게 등장하였다. 촉탁의와 연결된 키워드로 인건비 부분이 나타난 것을 볼 수 있다. 특히, 1차 기본계획에서 5등급 체계로 개편되어 ‘장기요양 등급-인정자-치매특별등급’의 새로운 연결망이 형성되었다. 2차 기본계획으로 넘어가면서 장기요양기관의 파생 이슈(신고포상금, 정기평

가, 공익신고, 지정취소, 서비스 질 등)가 다양하게 등장하였다. 그중 2차 기본계획 특징으로 지역사회 관련 키워드가 새롭게 등장하였다. 국민건강보험공단과 지방자치단체와의 관계가 나타났는데, 이는 치매 국가책임제 시행으로 인해 공단과 치매안심센터를 주로 다뤘다. 또한 “커뮤니티 케어” 및 “지역사회” 키워드가 새롭게 등장하였고, 해당 키워드는 보건복지부 키워드와 관계가 형성되었다. 있다. 또한 2차 기본계획에서 인지원등급(6등급) 신설로 인해 인정자-치매특별등급-경증치매의 관계망이 형성되었다.

시점별 키워드 네트워크 하위그룹을 정리하면, Table 1과 같이 4가지

**Table 1.** Subgroup of keywords by period

	Institutional settlement (modularity=0.49)	1st basic plan (modularity=0.57)	2nd basic plan (modularity=0.41)
1	<p><b>Institutional management</b></p> <p>[Illegal activities] LTC institution, Approval member, NHIS, Ministry of Health and Welfare, Unfair claims, LTC committee, Illegal activities, LTC benefits, Service quality, Field investigation, LTC association</p> <p>[Care worker training] Training center, Notification system, Permission system, Designation system</p>	<p>[Illegal activities] LTC institution, NHIS, LTC benefits, Ministry of Health and Welfare, Unfair claims, Monetary incentive, Service quality, Usage cost, Field investigation, Cancellation of designation, Public interest report, Regular evaluation</p>	<p>[Illegal activities] NHIS, LTC institution, LTC benefits, Unfair claims, Local government</p>
2	<p><b>Institution/Service characteristics</b></p> <p>[Institution/Facility comparison] Nursing home, LTC rating, Nursing hospital, In-home institution, Nurse, Dropout, Social worker, Nursing expenses, Physical therapist</p> <p>[Type of service] In-home care, Visit care, Visit nursing, Visit bathing, Facility benefit, facility service, Usage limit</p> <p>[Out of pocket payment] Out of pocket payment, Low income, The secondary poor</p>	<p>[Residential facilities] Nursing home, Nursing hospital, Wage, Dementia elderly, Commissioned doctor, Facilities elderly, Day care center, Elder abuse, Senior congregate house</p>	<p>[Home Service] Visit care, In-home care, Visit nursing, Visit bathing</p>
3	<p><b>Dementia</b></p>	<p>[Dementia special rating] Approval member, Upper limit, Approval score, Dementia special rating, Medical benefits, LTC rating, Copay, LTC service, Mild dementia</p>	<p>[Dementia National Responsibility] Ministry of Health and Welfare, Community, Community care, Care service, Pilot project, Dementia National Responsibility</p> <p>[Dementia special rating] Approval member, LTC rating, Aging, Dementia special rating, Copay, LTC service, Mild dementia, Impaired mobility</p> <p>[Dementia related services] Nursing home, Dementia patient, Day care center, Nursing Hospital, Dementia elderly, Dementia center, Facility benefit</p>
4	<p><b>Treatment of workers</b></p> <p>Care worker, Care service, sexual abuse, Working condition, Low income, Treatment improvement</p>	<p>Care worker, nurse, Social worker, Nursing assistant, Treatment improvement, Physical therapist, False registration</p>	<p>Care worker, Nurse, Wage, Social worker, Treatment improvement</p>



특성(기관 관리, 기관/서비스 특성, 치매, 종사자 처우)으로 키워드 그룹이 분류되었다. 커뮤니티 분석으로 키워드 간 응집성을 분석하였고, 이를 근거로 하위그룹을 설정했고 그 특성은 다음과 같다.

첫 번째, “기관 관리” 특성은 2가지 주제(기관 불법행위, 요양보호사 교육)로 키워드 그룹이 형성되었다. 하위그룹인 “기관 불법행위”는 시점별로 언론보도 양상이 달라졌다. 제도정착기에는 기관의 부당행위와 관련된 키워드가 그룹을 형성했지만, 1차 기본계획에서는 불법행위 포착 방법에 관한 키워드(신고포상금, 지정취소, 공익신고 등)가 등장하였다. 2차 기본계획에 접어들며 지방자치단체 키워드가 등장하였는데, 지방정부와 협력해 불법행위를 적발하는 보도가 나타났다. 다른 하위그룹인 “요양보호사 교육”은 제도정착기에만 나타났는데, 교육기관의 지정제/신고제에 대한 이슈로 인해 이 시기에 하위그룹이 생겨났다.

두 번째, “기관/서비스” 특성은 각 시점에 따라 주요 이슈가 변화하였다. 제도정착기에는 기관 및 서비스 이용 안내에 초점이 맞춰져 있다. 기관/시설 비교, 서비스 종류, 본인부담금을 안내 및 설명하는 키워드들로 하위그룹이 형성되었다. 기관 특징 비교(요양원, 요양병원, 재가시설)하는 하위그룹이 있는 반면, 서비스 종류(재가서비스, 입소서비스)를 안내하는 하위그룹 또한 나타났다. 더불어, 서비스별 본인부담금을 안내하는 하위그룹이 나타났다. 1차 기본계획에 접어들면서, 입소기관 특성 및 실태를 다루는 키워드가 나타났다. 2차 기본계획에는 이전 시점과 상반되게 재가서비스 관련 키워드로 하위그룹이 형성되었다.

세 번째로, “치매” 특성은 제도정착기에 하위그룹이 형성되지 않았지만, 1차 기본계획에 접어들면서 나타났다. 1차 기본계획에서 치매노인 대상으로 5등급 신설되어, 치매특별등급 관련 언론보도가 나타났다. 2차 기본계획에 접어들면서, 더욱 세분화되어 3개 주제(치매국가책임제, 치매특별등급, 치매 이용 서비스)로 하위그룹이 형성되고, 치매국가책임제 발표와 높은 관련성이 있었다.

네 번째로, “종사자 처우” 특성은 매 시점마다 나타났다. 제도정착기는 주로 요양보호사의 열악한 근로환경 관련 키워드로 구성되었다면, 1차 기본계획에서는 기관 종사자 전체로 확대되어 키워드가 구성되었다. 특히, 이와 더불어 기관 수익성을 위해 종사자를 허위등록하는 행태도 키워드(허위등록)로 나타났다. 2차 기본계획에서는 요양보호사 처우 개선비 폐지 이슈가 나타나면서, 인건비 키워드가 새롭게 등장하였다.

## 고찰 및 결론

본 연구는 노인장기요양보험의 특성을 살펴보기 위해, 시점별로 언론보도의 키워드 및 네트워크 특성을 분석하였다. 본 연구 결과를 바

탕으로 노인장기요양보험 정책적 논의 및 제언은 다음과 같다.

첫 번째, 노인장기요양기관 및 서비스의 투명성 재고를 언론보도에서 지속적으로 언급하였다. ‘장기요양기관↔부당청구’ 간 연결강도는 가장 높게 나타났고, 매 시점마다 최상위 연결강도를 보였다. 또한, “장기요양기관” 키워드는 2개 키워드(부당청구, 장기요양급여)와 높은 삼자 관계가 나타났다. “장기요양기관” 키워드는 신고 포상금, 내부종사자, 불법행위, 노인학대, 현지조사, 지정취소 등의 부정적 키워드들과 높은 관계를 보였다. 기관 부당행위에 대한 언론보도가 활발히 이뤄진 반면, 기관이 법적으로 어떤 결과가 나타났는지에 대한 후속 기사 및 키워드(처벌, 기관폐쇄, 환수, 승소 등)는 찾아볼 수 없었다. 언론보도 이슈에서 드러난 것처럼, 기관 관리의 투명성 재고가 사회적 요구로 나타나는 것을 살펴볼 수 있다.

두 번째, 시점별 키워드 중 요양보호사 처우개선 문제가 지속적으로 대두되었다. ‘요양보호사↔처우개선’ 연결강도가 높았고, 처우 관련 키워드(성폭력, 인건비, 저임금, 근로조건 등)와 같은 파생 키워드로 나타났다. 제도 초반부터 현재까지 요양보호사 처우개선 문제는 언론보도로 꾸준히 소비되어 왔다. 언론보도뿐만 아니라, 기존 선행연구에서도 요양보호사를 감정노동자로 바라보고, 처우개선을 위한 제도개선 및 보호조치가 이뤄질 것을 권고하고 있다[19,20]. 요양보호사의 근로환경이 열악하고, 처우개선의 필요성은 현재까지 부각되고 있다.

세 번째, 2차 기본계획에 접어들면서 지역사회 관리 및 연대 이슈가 제기되었고, 보험자의 지역사회 관리강화 필요성을 언론보도에서 언급하였다. 2017년 전(1차 기본계획)까지 지역사회 이슈는 주목받지 않았지만, 2018년(2차 기본계획)에 지역사회 협력 관리와 관련된 언론보도가 활발히 이루어졌다. 2차 기본계획에 접어들며, 지역사회 이슈를 중심으로 하위그룹 3개(치매국가책임제, 재가서비스, 지역사회 불법행위 적발)가 출현하였다. 하위그룹 중 “재가서비스” 특성은 2차 기본계획에 접어들며 급부상하였다. 1차 기본계획 전까지, “입소시설” 키워드가 주목을 받았으나 2차 기본계획에 접어들며 재가서비스로 이슈가 전환되었다. 2차 기본계획에서는 노인장기요양보험의 치매 관리를 위해, 지역사회 자원(보건소, 치매안심센터)과 연계가 필요함을 언론보도 하였다. 또한, 장기요양기관 불법행위 관리에 있어 공단-지방자치단체의 협력을 강조하는 하위그룹이 형성되었다. 1차 기본계획까지 노인장기요양제도의 특성만 언론보도가 되었으나, 2차 기본계획에 접어들며 지역사회 협력 및 관리를 촉구하는 키워드가 나타났다.

본 연구는 노인장기요양보험이라는 주제로 빅데이터 분석 기법을 적용한 최초의 연구로서, 노인장기요양보험 정책과 현안을 진단하고 이슈의 변화를 살펴보았다. 노인장기요양보험에 대한 첫 텍스트 마이닝 연구이기에, 심층적 탐색하기 위한 다양한 텍스트 마이닝 후속 연구가 필요하다. 실례로, 키워드 네트워크 분석으로 “치매국가책임제”

키워드의 중요성을 추론할 수 있지만, 언론보도에서 해당 키워드가 어떤 감정(긍정·중립·부정)으로 유통되었는지 알 수 없었다. 후속연구는 감정분석 등을 통해 키워드 사용의 긍정/부정을 파악해, 심층적인 여론 동향을 살펴보는 것이 필요하다. 또한 본 연구는 네이버·다음 뉴스 포털을 통해 텍스트 데이터를 수집하였는데, 두 포털 외에도 다양한 온라인 매체를 활용한 데이터 분석이 필요하다. 아울러, 후속연구에서는 문헌고찰을 통해 본 연구보다 확장된 용어사전을 구축해 분석되어야 한다. 본 연구는 노인장기요양보험에 대한 언론보도 키워드 분석으로, 정책 수립 근거로 활용되길 기대한다.

## ORCID

Jonggeun Seo <https://orcid.org/0000-0001-8455-4497>  
 Chung Mo Nam <https://orcid.org/0000-0003-0985-0928>  
 Tae Hyun Kim <https://orcid.org/0000-0003-1053-8958>  
 So-Hee Park <https://orcid.org/0000-0001-8513-5163>

## REFERENCES

1. Statistics Korea. 2015, World and Korean population status and forecast. Press release (2019.9.2) (Korean).
2. Korea Institute for Health and Social Affairs. Plan of evaluating and redesign of long-term care insurance system. Press release (2016.10.31) (Korean).
3. Gyeonggi Welfare Foundation. Achievements and challenges of long-term care insurance in 10 year. Suwon: Gyeonggi Welfare Foundation, 2017 (Korean).
4. Lee JA. A study on the impact of in-home service on the families of beneficiaries of long-term care insurance: focusing on their experience of labour and leisure activities. Korean J Gerontol Soc Welf 2015;70(1): 81-112 (Korean).
5. Lee HJ. Evaluating the effects of long-term care insurance on labor supply among beneficiary households. Korean Soc Security Assoc 2015; 31(2):185-208 (Korean).
6. Ministry of Health and Welfare. 2018-2022 The 2nd basic plan for long-term care insurance. Press release (2018.8.3) (Korean).
7. National Assembly Budget Office. Financial outlook for long-term care insurance in 2018-2027. Seoul: National Assembly Budget Office, 2018 (Korean).
8. Lee JS, Cho EY. An analysis of policy making process of long term care insurance. J Soc Sci 2012;23(1):3-22 (Korean).
9. Beyer MA, Laney D. The importance of 'big data': a definition. Stamford, CT: Gartner, 2012.
10. Lee GH. A study on firm's strategic performance leveraged by big data analytics (BDA): focused on organizational structural dimensions [dissertation]. Seoul National University; Korea, 2017.
11. Han BJ, Yang DW. A empirical study on effects of dynamic capabilities and entrepreneurial orientation of SMEs on big data utilization intention. J Dig Converg 2018;16(11):237-253 (Korean). DOI: 10.14400/JDC.2018.16.11.237
12. K-ICT Big Data Center. NIADic. Available at <https://kbig.kr/portal/kbig/knowledge/files.page> [accessed on October 01, 2020].
13. DAUM. DAUM Encyclopedia Information collected by Web-Crawling Method in R packages. Available at <https://100.daum.net> [Web-Crawling on October 1, 2020].
14. Ministry of Health and Welfare. Announcement on standards for providing long-term care benefits and methods of calculating benefits. Press release (2019.10.27) (Korean).
15. National Health Insurance Service. NHIS longterm care glossary. Available at <https://www.longtermcare.or.kr/npbs/e/b/601/openIrmFtb?menuId=npe0000000990&zoomSize=> [accessed on September 30, 2020].
16. Seo JG. Keyword network and LDA analysis on long term care insurance using big data [dissertation]. Yonsei University; Korea, 2021.
17. Gwak GY. Social network analysis. Seoul: Cheong Ram; 2014 (Korean).
18. Julia S, David R. Text mining with R. 1st ed. Boston, MA: O'Reilly Media, Inc.; 2019.
19. Kim KE, Park YJ, Koo EJ. A study on experiences of homecare worker as a emotional labor: focusing on providing improved treatment for homecare workers. Korea Care Manage Res 2014;12(1):103-119 (Korean).
20. Song MY, Kim HK, Ann JS. A case study on elderly carer-oriented working conditions: focused on the region and types difference of management of elderly care institution. J Welf Counsel Educ 2015; 4(1):57-80 (Korean).