

An Exploratory Study of Health Inequality Discourse Using Korean Newspaper Articles: A Topic Modeling Approach

Jin-Hwan Kim^{1,2}

¹Department of Health Policy and Management, Seoul National University College of Medicine, Seoul, Korea; ²Graduate School of Public Health, Seoul National University, Seoul, Korea

Objectives: This study aimed to explore the health inequality discourse in the Korean press by analyzing newspaper articles using a relatively new content analysis technique.

Methods: This study used the search term “health inequality” to collect articles containing that term that were published between 2000 and 2018. The collected articles went through pre-processing and topic modeling, and the contents and temporal trends of the extracted topics were analyzed.

Results: A total of 1038 articles were identified, and 5 topics were extracted. As the number of studies on health inequality has increased over the past 2 decades, so too has the number of news articles regarding health inequality. The extracted topics were public health policies, social inequalities in health, inequality as a social problem, healthcare policies, and regional health gaps. The total number of occurrences of each topic increased every year, and the trend observed for each theme was influenced by events related to its contents, such as elections. Finally, the frequency of appearance of each topic differed depending on the type of news source.

Conclusions: The results of this study can be used as preliminary data for future attempts to address health inequality in Korea. To make addressing health inequality part of the public agenda, the media’s perspective and discourse regarding health inequality should be monitored to facilitate further strategic action.

Key words: Health inequality, Newspaper, Discourse, Topic modeling, Korea

INTRODUCTION

The number of studies on health inequality has been steadily increasing worldwide [1]. Similarly, the amount of research on

health inequality has increased over time in Korea [2], where many researchers and activists have struggled to address health inequalities [3]. As a result of these efforts, since the publication of the Health Plan 2010, the Korean government has designated the improvement on health equity as one of the overarching goals of its national health promotion plan [4]. Although the discussion of health inequality has seemingly become somewhat mainstream, the realities are quite different from the portrayal of this subject.

Based on comparative analysis of the policy process, Cobb and Elder conceptualized the agenda-setting process into 4 stages, termed the social problem, social issue, public agenda, and formal agenda (government agenda). Under this frame-

Received: August 3, 2019 Accepted: October 14, 2019

Corresponding author: Jin-Hwan Kim, MD
Department of Health Policy and Management, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea

E-mail: jhkim00011@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

work, a social problem is an issue considered to be in a state far from desirable, a social issue refers to a social problem that is subject to controversy due to diverging opinions in the nature of the problem or how to resolve it, the public agenda consists of items that deserve the attention of the general public and are justified targets of government intervention, and the formal agenda refers to issues for which the government has decided to take action. They termed a portion of the formal agenda the pseudo-agenda, which refers to items that the government has placed on the agenda largely for symbolic reasons, meaning that there is little chance of pseudo-agenda items being translated into action [5]. In Korea, health inequality is highly likely to be part of the pseudo-agenda in that the social determinants of health are not reflected in actual policies and in that improvement on health equity is only a symbolic goal, without concrete action strategies detailed in the published Health Plans [6]. The efforts of the Korean Society for Equity in Health (KSEH) to influence the status of health inequality as an agenda item in the 2018 local elections can be viewed as a move to turn health inequality into a social problem or a social issue with the rationale that health inequality has not yet been properly put on the agenda. The media are a natural outlet for such efforts to leverage upcoming political opportunities, and the relationship between the news media and elections was also demonstrated in a study conducted at Chapel Hill in the USA, as well as follow-up studies [7,8].

The result of such efforts to problematize health inequalities in Korean society can be explained by Whitehead's action spectrum. Whitehead [9] proposed the existence of an action spectrum that progresses through the steps of measurement, recognition, awareness-raising, will to take action, isolated initiatives, more structured developments, and comprehensive coordinated policy. In Korea, health inequality as an issue seemed to have passed the measurement and recognition stages, but it has not reached the point of awareness-raising [10]. Although there are many theories on the role of media in the formation of public opinion, it is crystal clear that the media plays an important role at each stage of the agenda-setting process. The media would also play an important role if action on health inequality did not pass the stage of awareness-raising and instead fell into indifference or denial.

This study specifically focuses on the term "health inequality." There are 3 major expressions used to refer to health inequality: health disparity, health inequality, and health inequity. The first 2 terms refer simply to differences in health, while the third

signifies an unfair or unjust difference in health [11,12]. Despite the ethical implications of health inequity, health inequality is the term that is associated with the most ideological tension in Korea. Similar to attempts in the USA to replace the terms "health inequality" and "health inequity" with "health disparity" during the Clinton administration [13], major Korean policy stakeholders have been reported to consciously avoid using the expression "health inequality." The following quote from one policy stakeholder exemplifies this phenomenon [14]:

"As I said before, *equity and equality are completely different*. If we talked about *equality*, I wouldn't be here. There's no reason for me to sit here" (emphasis added).

Considering the ease of data collection and the politicized nature of the term "health inequality," this study aimed to identify how health inequality is reflected in newspaper articles, through the following specific research questions: (1) Has there been any temporal change in the number and content of newspaper articles on health inequalities? (2) What topics can be derived from newspaper articles when applying the topic modeling technique? and (3) Are there any differences in topics according to the type of news source?

METHODS

Text Mining and Topic Modeling

Text mining is an analytical technique that examines the frequency and distribution of words by processing large quantities of text data. This technique involves 5 stages: extraction, pre-processing, corpus formation, text analysis, and visualization [15]. A topic modeling technique was selected to identify topics by clustering mutually-associated words, since the purpose of this study was to explore patterns in word choice. The latent Dirichlet allocation (LDA) model was used for this purpose. The model assumes that each document consists of words, and each word is associated with a particular topic. A single document can be viewed as a distribution of words that describe multiple topics. The researchers can designate the number of topics covered in a set of documents either in advance or empirically. The probabilities of words appearing in their associated topics follow the Dirichlet distribution [16].

Data Collection and Analysis

Articles from 51 press outlets published between 2000 and 2018 were collected using "health inequality" as a search term in *BigKinds*, which is a big data analysis platform provided by

the Korea Press Foundation. The articles were classified as being from either a major or a local newspaper. News sources were divided into major and local newspapers instead of adopting the usual classification of conservative, centrist, and progressive for 2 reasons. The first was that the journals typically classified as conservative rarely published articles using the phrase “health inequality” (only 4 total articles were found from 3 major conservative press outlets). The political orientation of a journal, at least with regard to health inequality, was therefore evident through the presence or absence of articles rather than through the content of articles, which contrasts with other issues that reveal political discrepancies in their content. The second is that major and local newspapers have different reporting patterns, especially when dealing with regional issues [17]. As the analysis shows, the region has become an important area of focus in relation to health inequality. In addition, articles from a pre-compiled list of medicine, nursing, and pharmacology journals (hereafter referred to as specialized news sources) were also collected using the same search term (Table 1). After isolating the Web addresses containing the original texts, duplicates were removed, and the original texts were extracted. The total number of articles pre-processed was 1038, which consisted of 424, 336, and 278 articles from major, local, and specialized news sources, respectively. Of these, only 5 articles were from 2001, but this number had increased to 198 by 2018 (Figure 1).

Stop words, which contain essentially no semantic value, were removed, and words with the same semantic value but that can be used differently were standardized for pre-processing. Then, nouns alone were extracted using a Korean-language natural language processing package for R software (R Foundation for Statistical Computing; R Core Team, Vienna, Austria). Among the extracted words, misclassified words that represented units or sequences were additionally processed as stop words. Finally, words that appeared in most documents and those that appeared fewer than 5 times were excluded from the data, as they may not provide useful information in the extraction of an associated topic. After pre-processing, a term-document matrix and a document-term matrix were created using the R packages *tm*, *stringr*, *snowball*, *topicmodels*, and *lda*, and the perplexity was calculated to help determine the appropriate number of topics. To extract the potential topics based on the pre-determined number, topic modeling was performed with the *lda* package, which uses the LDA algorithm through a collapsed Gibbs sampling method. Then, the proba-

Table 1. List of the journals included in the analysis

Name (English) ¹	Type	Name (English) ¹	Type
The Kyunghyang Shinmun	Major	Kangwondominilbo	Local
Kukminilbo	Major	Kangwonilbo	Local
Naeil Shinmun	Major	Kyeonggiilbo	Local
The Dong-A Ilbo	Major	Gyeongnamdominilbo	Local
Munhwailbo	Major	Knnews	Local
Seoul Shinmun	Major	Kyeongsangilbo	Local
Segyeilbo	Major	Kyeonginilbo	Local
The Chosunilbo	Major	Kjdaily	Local
Korea Joongang Daily	Major	Kwangjuilbo	Local
Hankyoreh	Major	Kookjenews	Local
The Korea Times	Major	Daeguilbo	Local
Maeil Business News Korea	Major	Daejeonilbo	Local
Moneytoday	Major	Maeil shinmun	Local
Sedaily	Major	Moodeungilbo	Local
The Asia Business Daily	Major	Busanilbo	Local
Aju Business Daily	Major	Yeongnamilbo	Local
The financial news	Major	Ulsanmaeil	Local
Dailymedi	Specialized	Jnilbo	Local
doc3news	Specialized	Dominilbojb	Local
Rapportian	Specialized	Jeonbukilbo	Local
MEDICAL Observer	Specialized	Jeminilbo	Local
The medical news	Specialized	Joongdoilbo	Local
MedicalTimes	Specialized		
Medipana	Specialized		
Doctor's news	Specialized		

¹For journals with no official English name, the company name and the official social networking service account name were recorded. If this was not possible, a name as close as possible to the Korean meaning was recorded.

bility of each topic appearing in articles each year was checked to examine how the extracted topics changed in frequency over time. Additionally, a network of the extracted words was drawn, and each word was marked with its associated topic. For the data sampling with *BigKinds*, Python (Python Software Foundation, Beaverton, OR, USA) was used with the Beautiful-Soup module, and R version 3.6.0 (<https://cran.r-project.org/bin/windows/base/old/3.6.0/>) was used for text data analysis.

Ethics Statement

As the study used publicly open data retrieved from *BigKinds*, research ethics review was not required.

RESULTS

The top 50 extracted words (nouns) and their frequencies are given in Figure 2. Excluding the search terms “health” and

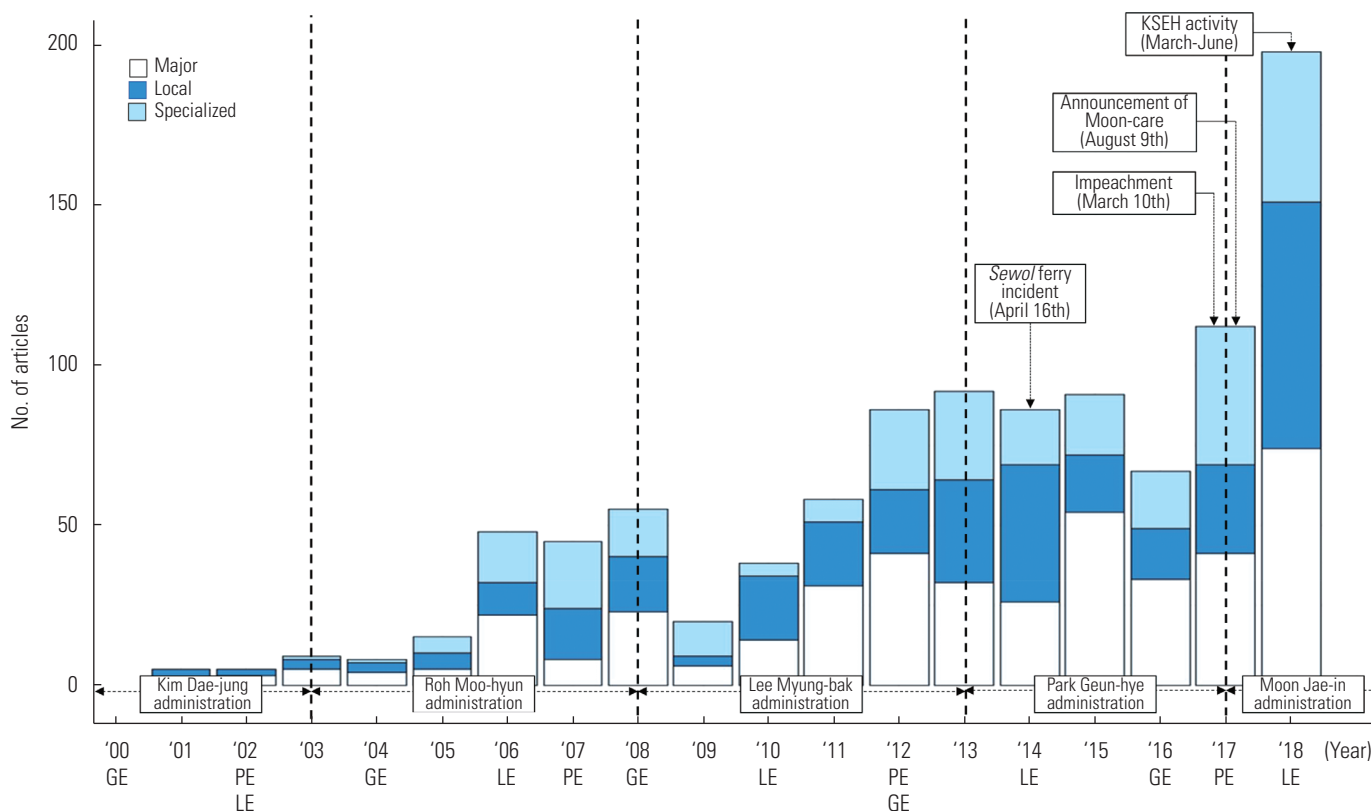


Figure 1. Trends of news articles on health inequalities. KSEH, Korean Society for Equity in Health; PE, presidential election; GE, general election; LE, local election.

“inequality,” this list consists of words related to government-centered healthcare policies and regional health gaps. Given the calculated perplexity, 5 topics were extracted. The list of the top 20 words for each topic and the probability of each topic appearing across all documents are shown in Supplemental Material 1. The themes, which were assigned based on the words contained in each topic, were found to be public health policies, social inequalities in health, inequality as a social problem, healthcare policies, and regional health gaps. The first and fourth topics reflect current health policies, the second and fifth refer to ways health inequalities have been described, and the third indicates whether inequality is regarded as a major social problem. Figure 3 shows the relationships between the extracted words, with the rough area occupied by each topic corresponding to its frequency.

The first topic relates to efforts to build public hospitals to address health inequalities. The issue of the Daejeon city hospital is captured in this topic, as that hospital has been discussed for a long time. The second topic shows current trends in health inequality studies that have addressed differences or gaps in health outcomes such as mortality, disease prevalence,

and smoking rates depending on gender and socioeconomic position (including income and educational status). The third theme displayed the highest probability and shows that health inequality is not addressed only in isolation, but rather is also considered within its global context. This topic shows that health inequality is linked to social institutions, such as politics, economy, labor, welfare, and education. The appearance of the name of Korea’s incumbent president also suggests that this is a political problem. The terms associated with the fourth theme indicate that the Korean health system is centered on the national health insurance system, and specific policy changes in coverage are closely linked to elections. The last theme reflects the efforts of the KSEH to make health inequalities part of the agenda for the 2018 local elections. Because the KSEH distributed information regarding the health profile of each municipality (*si, gun, gu*)—including calculations of the life expectancy of each area and visualization of gaps based on gender and socioeconomic position—around the time of the 2018 local election, words related to their work appeared in topic 5 [3].

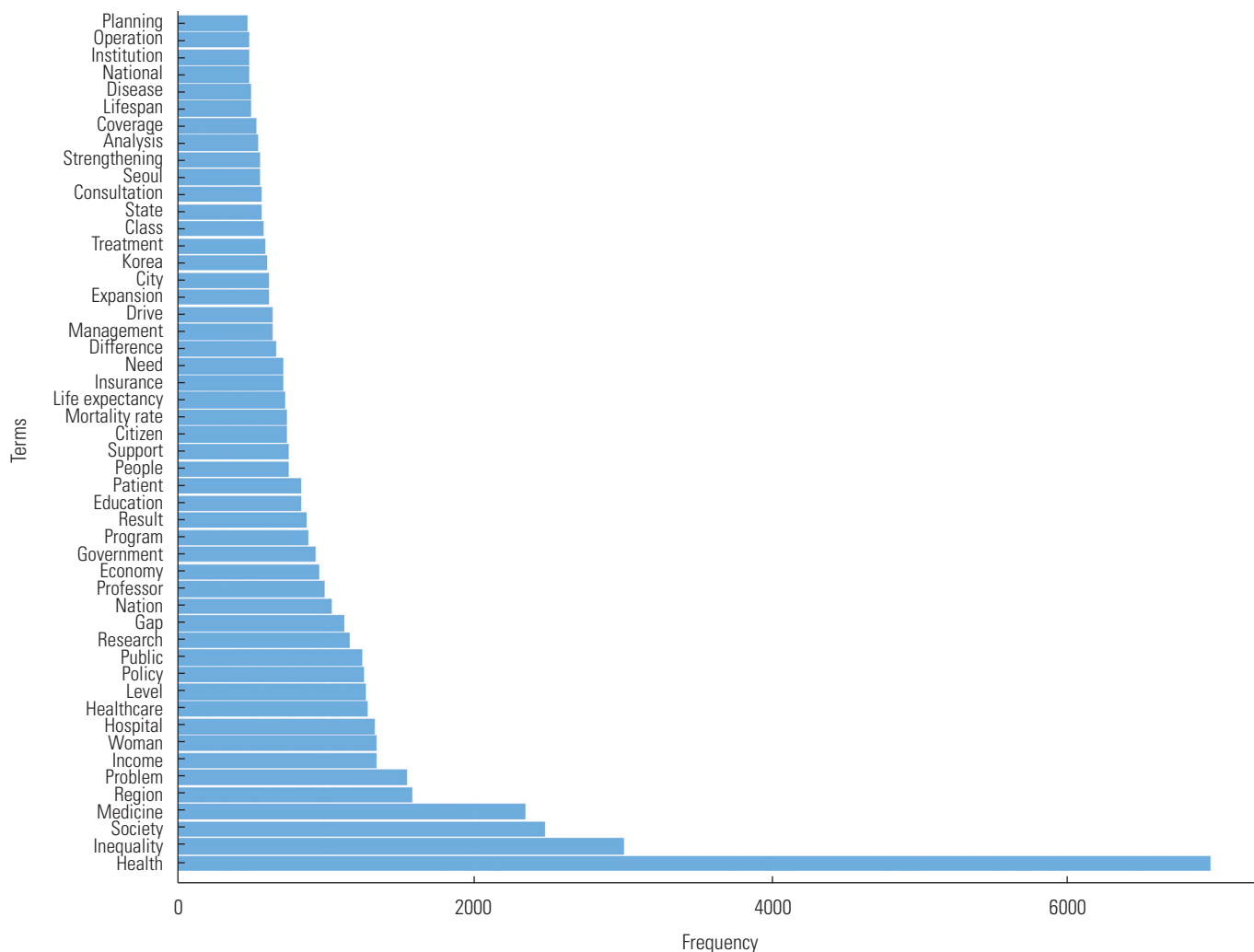


Figure 2. The top 50 most frequent keywords in newspaper articles.

Temporal Trends

Figure 1 shows how the number of articles concerning health inequality has changed over time, and Figure 4 shows how the trends for the topics differ. The number of articles increased significantly over time overall, but dropped sharply in 2009 and 2016. This number increased around the time of local elections (2006, 2010, and 2018). The exception was the 2014 local election, which may be explained by the Sewol ferry incident in April 2014 and the shutdown of the Jinju public hospital in July 2013. Still, the frequency of mentions of health inequality in the local press increased from 2013 to 2014.

Figure 4A indicates the probability of each topic appearing in articles on a yearly basis. Since the number of articles increased significantly each year, the total volume of each topic was calculated by multiplying the probability of the topic appearing by the number of articles each year (Figure 4B). Based

on the probability of appearance, topic 3 decreased over the course of the study, topic 1 peaked in 2010, topic 2 in 2015, topic 4 in 2007 and 2016, and topic 5 in 2018. Based on total volume, all 5 topics increased over the course of the study, but the rising trends of topic 3 and topic 1 were the most noticeable, whereas topic 2 declined in volume over the past few years. Topic 5 sharply increased recently, and topic 4 increased in accordance with presidential election cycles (2007, 2012, and 2017) and declined in 2018. Finally, the activity of the KSEH was effective enough to be captured as a single independent topic (topic 5).

Proportions by Journal Type

Supplemental Material 2 shows the proportions of topics by type of news source (i.e., the probability of each topic appearing in each journal type). Topics 1 and 5 appeared most fre-

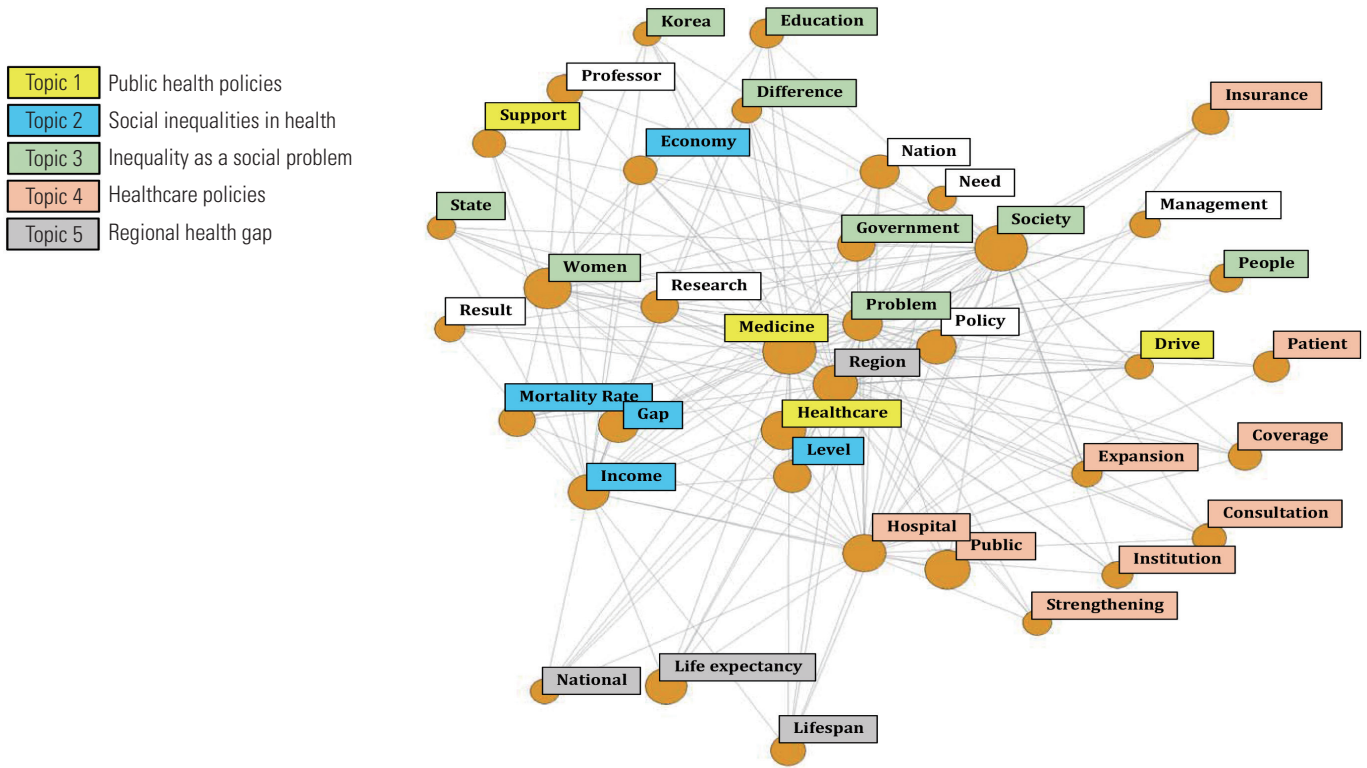


Figure 3. The relationship between the extracted words. The isolated vertices were removed after leaving relatively strong edge (>2.5% of the maximum value). The theme and its color of each topic are shown in the top-left corner and the words not included in a specific topic were left white.

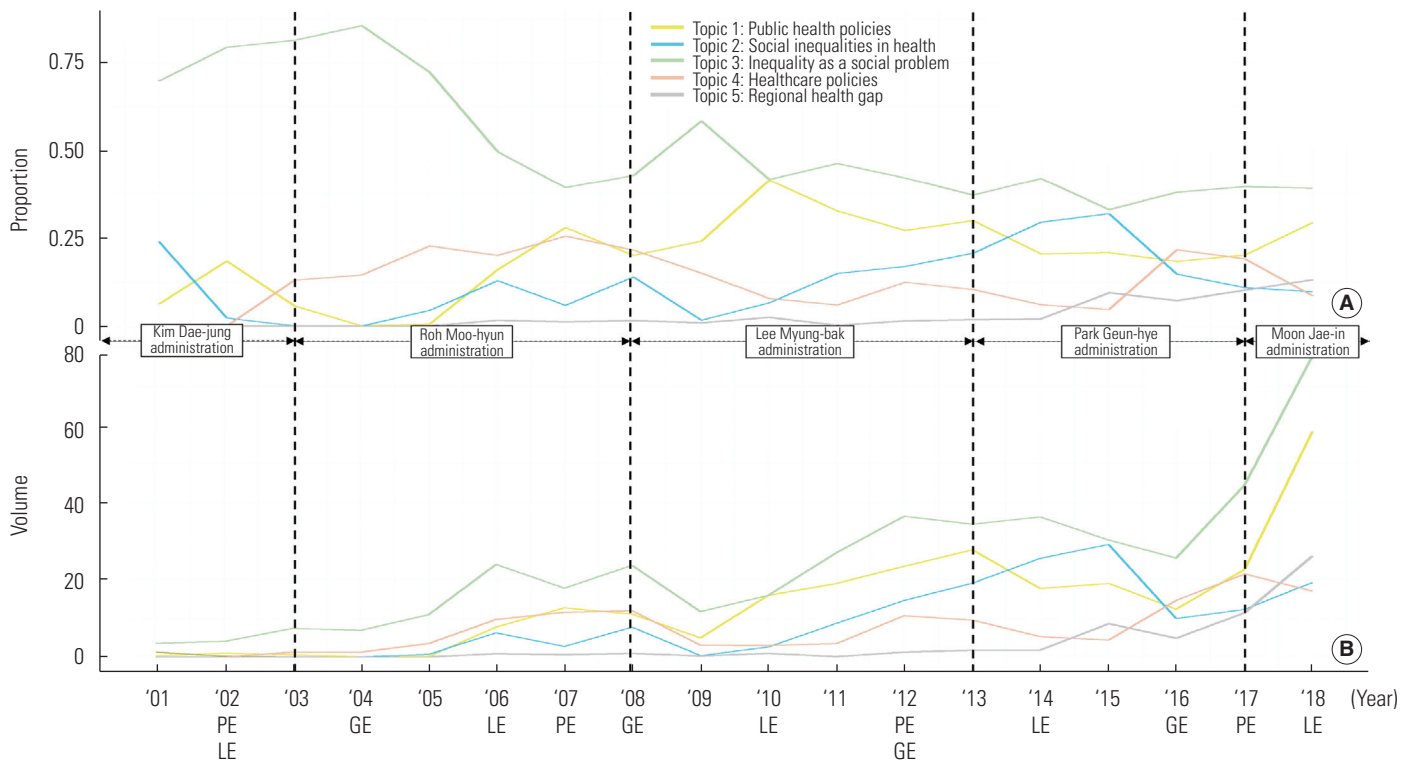


Figure 4. Temporal trends of each topic. (A) Trends of topic proportion, (B) trends of topic volume (volume · proportion). PE, presidential election; GE, general election; LE, local election.

quently in the local press, topics 2 and 3 in the major press, and topic 4 in the specialized press. Topic 1 related to the establishment of new regional public hospitals, while topic 5 related to the efforts of the KSEH, which underscored the status of health inequality as a gap between regions. Topic 4 related to changes in healthcare policies that impact doctors, medial institutions, patients, payment, and health insurance, which could be of particular interest to the specialized press. These findings convincingly demonstrate that the local press is interested in local issues, while the specialized press is interested in issues classically associated with healthcare policy.

DISCUSSION

This study examined how the media have handled health inequality. As the number of studies on health inequality has increased over the past 2 decades, so too has the number of news articles discussing health inequality. The composition of articles each year changed continually, even as the volume increased overall. The topic modeling technique yielded 5 topics, each of which sheds light on a specific portrayal within the health inequality discourse. Topic 1 implied that regional health policies focused on building new public hospitals. Topic 2 dealt with the way health inequality is measured and recognized according to social gradients. Topic 3 showed that health inequality is a problem embedded in the social structure of inequality. Topic 4 covered typical health insurance coverage expansion policies. Topic 5 was related to the health gaps between regions, and this topic reflected the activity of KSEH.

The trend for each theme over time was associated with events related to its content. One plausible explanation for at least some of the trends observed is that elections affect the probability of a topic appearing. The trend for topic 4, which deals with macro-level healthcare policy, reflected the timing of presidential elections, while regional affairs such as the building of public hospitals and addressing the regional health gap were linked to local elections. In contrast, general elections did not seem to noticeably impact the discourse. The characteristics of each election may explain these results. Presidential candidates need wide support, particularly in the regions where swing votes are most common, and each province or municipality is a unique site of competition in local elections. For general elections, however, cities are divided into several election districts, and rural regions may be grouped together to form a single election district. As a result, general elections do not

create specific accountability for each region.

No previous studies have focused on the relationship between health inequality and media coverage in Korea. However, a study conducted by Kwon et al. [18] did attempt to define the meaning of health inequality. That study identified 4 defining characteristics of health inequality: (1) infringement of basic health rights, (2) unfair utilization of healthcare services, (3) differences in health status between individuals or groups, and (4) social discrimination [18]. While those characteristics align relatively well with the 5 topics identified in this study, they are discordant with the results of interviews of low-income people in health-vulnerable regions, which assessed social determinants of health and related psychological responses [19]. This suggests the existence of a deep discrepancy between experts and people living in vulnerable regions with regard to the understanding of health inequality. This divide partially explains why the efforts of the KSEH to politicize health equality by disclosing the health gaps between regions was not considered very successful.

To the best of the researcher's knowledge, this study is the first to analyze how the Korean press has handled health inequality. The study analyzed newspapers, which are often a neglected route for agenda-setting and politicization. Future studies should go beyond measuring and monitoring health inequalities to produce basic data revealing why health inequality has not been politicized. Another important contribution of the present study is that it introduced an automated content analysis method that is relatively new to the discipline, going beyond previous studies that tracked trends in tweets [20] or identified a network of words that appeared in newspaper articles [21].

However, this study is not without limitations. First, the use of a restricted search term may have limited the comprehensiveness of the analyzed articles. As previously discussed, however, it may be meaningful to analyze articles that include the words "health inequality," which carries its own unique ideological tension, in spite of the information loss resulting from this search term restriction. Second, this study captures reality only after that reality has passed through the lens of the media. This is an important limitation of the research methodology used and requires a separate analysis of how the media frames health inequality. Third, the process of selecting the journals (particularly the specialized journals) and determining the themes of the extracted topics inevitably involved the researcher's subjectivity. However, as the list of

journals was created from the mailing lists of some academic societies in the field of public health, its bias may be minimal. Although the determination of the themes inevitably involved the researcher's subjective decision-making, the extracted themes were assessed for their validity by outside commentators, and the author would argue that the themes presented in this paper were interpreted using generally acceptable logic. Fourth, this study did not include new media, such as social networking services. This, however, is not a serious problem given that most of the content related to health inequality on Twitter are articles from specialized journals or tweets from researchers in this field. Finally, the contents derived from specialized journals may be heterogeneous because they were retrieved from the website of each journal, unlike the content of major and local journals, which were collected through *Big-Kinds*. This is an important limitation that may have impacted the results of the study; however, it was unavoidable, as there was no better way to collect the specialized journal articles.

Given the impact of the media on all steps of the agenda-setting process, more research is needed regarding: (1) how the media frame health inequality (the role of the media), (2) the attitudes of people and policy stakeholders toward health inequality as framed by the media (the interaction between public perception and media framing), and (3) the understanding that people and policy stakeholders have of health inequality (public perception). In addition to these research topics, an analysis of how these patterns of discourse and understandings operate in the dynamic policy process is needed. In summary, this study identified a part of the landscape of health inequality illustrated by the Korean press using newspaper articles. The author hopes that the overall structure of the health inequality discourse will be further understood in the near future through additional studies and through efforts to engage in agenda-setting and policymaking, which are more closely related to the political dimension of health inequality.

SUPPLEMENTAL MATERIALS

Supplemental materials are available at <https://doi.org/10.3961/jpmph.19.221>.

Korean version is available at <https://doi.org/10.3961/jpmph.19.221>.

CONFLICT OF INTEREST

The author has no conflicts of interest associated with the materials presented in this paper.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Academic Activity Promotion Program for Identifying and Dissolving the Causes of the Inter-regional Health Gap of the Korean Society for Preventive Medicine. I appreciate critical comments from the members of the Critical Health Studies (CHeS).

AUTHOR CONTRIBUTIONS

All work was done by JHK.

ORCID

Jin-Hwan Kim <https://orcid.org/0000-0001-8229-9613>

REFERENCES

1. Cash-Gibson L, Rojas-Gualdrón DF, Pericàs JM, Benach J. Inequalities in global health inequalities research: a 50-year bibliometric analysis (1966-2015). *PLoS One* 2018;13(1): e0191901.
2. Kim C, Kim M, Lee T, Sohn J. Inequality in health: a Korean perspective. Seoul: Seoul National University Press; 2015, p. 136 (Korean).
3. Korean Society for Equity in Health. Health Inequalities in the era of decentralization: what will we do?; 2018 Mar 23 [cited 2019 Sep 27]. Available from: http://www.healthequity.or.kr/bbs/view.php?id=news_notice&page=1&sn1=&divpage=1&sn=off&ss=on&sc=on&select_arrange=headnum&desc=asc&no=157 (Korean).
4. Korea Institute for Health and Social Affairs. Establish of new health plan 2010 [cited 2019 Sep 27]. Available from: <https://khealth.or.kr/board/view?pageNum=2&rowCnt=8&no1=2&linkId=559208&menuId=MENU00829&schType=0&schText=&boardStyle=Gallery&categoryId=&continent=&country=> (Korean).
5. Birkland TA. An introduction to the policy process: theories, concepts, and models of public policy making. 4th ed. New York: Routledge; 2016, p. 203.

6. Choi YJ, Yoon TH, Shin DS. Review of the third Health Plan (2011-2020) in Korea: perspectives on health equity. *J Crit Soc Policy* 2012;37:367-400 (Korean).
7. McCombs ME, Shaw DL. The agenda-setting function of mass media. *Public Opin Q* 1972;36(2):176-187.
8. McCombes M, Lopez-Escobar E, Llamas JP. Setting the agenda of attributes in the 1996 Spanish general election. *J Commun* 2000;50(2):77-92.
9. Whitehead M. Diffusion of ideas on social inequalities in health: a European perspective. *Milbank Q* 1998;76(3):469-492.
10. Kim MH, Lee J. Current status of policy developments in tackling health inequalities and the next steps to be taken in Korea. *J Korean Med Assoc* 2013;56(3):206-212 (Korean).
11. Kawachi I, Subramanian SV, Almeida-Filho N. A glossary for health inequalities. *J Epidemiol Community Health* 2002;56(9):647-652.
12. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health* 2006;27:167-194.
13. Sonia B. The social transformation of health inequities: understanding the discourse on health disparities in the United States [dissertation]. Albuquerque: University of New Mexico; 2013.
14. Kim MH, Kim S, Park Y. Making health inequality policy agenda: where are we now? In: Proceedings of the 2016 Autumn Conference of the Korean Society for Preventive Medicine; 2016 Oct 21; Gyeongju, Korea (Korean).
15. Silge J, Robinson D. Text mining with R: a tidy approach; 2017 [cited 2019 Nov 10]. Available from: <https://www.tidytext-mining.com/>.
16. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3:993-1022.
17. Lee H. The analysis of frames that appear to news reports on Sejong City: focused on the difference between the national media and the local media. *J Polit Sci Commun* 2013;16(1):229-264 (Korean).
18. Kwon JO, Lee EN, Bae SH. Concept analysis of health inequalities. *J Korean Acad Nurs Adm* 2015;21(1):20-31 (Korean).
19. Sim MY, Yeum DM, An SA, Jeong BG. A study on health inequality that low income groups in the health vulnerable regions have recognized. *J Crit Soc Welf* 2012;(37):155-201 (Korean).
20. Kim JY, Park J, Park E, Ji SM. Analysis of issues and trends in cosmetic plastic procedures using tweets from Twitters. *Public Health Aff* 2017;1(1):129-143 (Korean).
21. Min HS, Kim CY. Exploratory study of publicness in healthcare sector through text network analysis. *Health Policy Manag* 2016;26(1):51-62 (Korean).