



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



What are the main patient safety concerns of healthcare stakeholders: a mixed-method study of Web-based text



Insook Cho^{a,*}, Minyoung Lee^{a,b}, Yeonjin Kim^b

^a Department of Nursing, Inha University, Incheon, South Korea

^b Graduate School, Inha University, Incheon, South Korea

ARTICLE INFO

Keywords:

Patient safety
Web-based text
Healthcare stakeholders
Natural language processing
Topic modeling

ABSTRACT

Objectives: Various healthcare stakeholders define quality of care in different ways. Public policy could advocate all these concerns. This study was conducted to identify the main themes on patient safety of stakeholders expressed before and after the Patient Safety Act was enacted in Korea in 2015.

Design: Longitudinal observational study of the interests of healthcare stakeholders generated between January 2014 and September 2018.

Materials and methods: Text data were collected from 2,487 documents on 18 websites that were identified as representative healthcare stakeholder groups of consumers, providers, government, and researchers. A Korean natural language processing (NLP) package, manual review, and synonym dictionary were used for data pre-processing, and we adopted the unsupervised NLP method of probabilistic topic modeling and latent Dirichlet allocation. A linear trend analysis over time, a qualitative step involving two external experts, and original text reviews were performed to validate the identified topics.

Results: Forty-one topics were identified, and the most common concerns of stakeholders were institutional infection control as triggered by the Middle East respiratory syndrome outbreak in early 2015, and infusion-related infection from late 2017 until the middle of 2018. The other top-three concerns of the stakeholder groups were highly similar, while research topics were limited to the perceptions of providers and the activities and culture of hospitals. Five topics showed statistically significant increasing trends over time, while another five showed decreasing trends (both $P < 0.05$). In the qualitative step, we confirmed 35 themes and revised the other 6.

Conclusions: A common concern among stakeholders was hospital infection control, ranging from nosocomial infections to those brought in by family visiting patients. Government policies and systemic approaches to patient safety were highlighted by different stakeholders. Researchers were focused on hospital sociocultural factors at both the organizational and clinician levels. These identified concerns all should be advocated by the public health policy.

1. Introduction

It is well known that all healthcare systems around the world occasionally unintentionally harm patients who are seeking help. Due to the importance of safety in ensuring a high quality of care, patient safety has become a fundamental aspect of the drive to improve quality, and this is receiving increasing interest from healthcare stakeholders in many countries [1]. Various stakeholders are responsible for ensuring that patient care is safely delivered and that patients are not harmed, including society overall, patients, clinicians including physicians and nurses, educators, administrators, researchers, professional bodies, governments and legislative bodies, and accrediting agencies [2].

Like the governments in other countries, the Korean government recognized that patient safety is an urgent threat to society and should be regarded as the cornerstone of health policy following several serious adverse events. The Ministry of Health and Welfare of the Korean government set out the Patient Safety Act in January 2015 and enacted it in June 2016 [3]. This act focuses on implementing a national reporting and learning system, the development of national healthcare standards and indexes, and formulating a strategic and comprehensive plan for patient safety every 5 years. In accordance with the act, the first national plan for patient safety (covering 2018 to 2022) was released in April 2018, which includes prioritizing R&D and supporting the process under both short-term and long-term national strategies [4].

* Corresponding author at: Department of Nursing, Inha University, 5E301, Inharo 100 Nam-gu, Incheon, 22212, South Korea.
E-mail address: insook.cho@inha.ac.kr (I. Cho).

<https://doi.org/10.1016/j.ijmedinf.2020.104162>

Received 30 January 2020; Received in revised form 20 March 2020; Accepted 28 April 2020

1386-5056/© 2020 Elsevier B.V. All rights reserved.

The design of the Korean national patient safety strategy was based on the classical top-down and essentially rational approach to policy-making. However, little is currently known about the concerns of the various healthcare stakeholders. Many modern theorists have quoted empirical evidence from the field of policy studies to suggest that policies and actions are intertwined [5]. However, national policies should actually be created using a bottom-up approach as much as possible. The report from a study of developing the national plan for patient-safety R&D [6] also emphasized the importance of regularly surveying consumers and providers to explore their concerns and interests and their implications for national strategic planning and decision-making. The various healthcare stakeholders define quality of care in different ways. Traditional survey and content analysis, and consensus panel methods with qualitative interviews are used where evidence is sparse or opinions are diverse. These methods require consultation with a large group from a geographically dispersed population, and their credibility is greatly dependent on the sampling method and panel composition [7].

The recent and continuing rapid growth in user-generated text data shared in Web-based communities, posting boards, and social media has made it possible to study and analyze language at an unprecedented scale. The analysis of user-generated texts related to patient safety could reveal the topics of interest to various stakeholders. Natural language processing (NLP) facilitates the analysis of these data to extract meaning. NLP has been applied in highly diverse disciplines and applications, such as social media, political speeches, and physician discharge summaries [8–10]. As an unsupervised NLP technique, topic-modeling based on latent Dirichlet allocation (LDA) is a statistical approach for discovering or identifying topics associated with certain words or phrases [11]. One study [12] applied the topic-modeling technique to compare narrative reviews of online communities about hospitals to the domain of governmental hospital surveys. Another study [13] applied an opinion-mining technique to social media to examine public opinions.

In this study we applied a topic-modeling technique to explore text data on patient safety collected from Web-based user-generated texts. We were interested in how text data are distributed across the LDA topics, and in particular how this distribution can represent the specific concerns of various stakeholders. We hypothesized that while some LDA topics will show common ground among stakeholders, LDA will reveal new topics or themes that provide different viewpoints among the stakeholders, which may be useful for understanding each stakeholder. A thorough understanding of the interests, motivation, and behaviors of these Web-based healthcare stakeholders could make it possible to obtain voices from a bottom-up perspective.

The next section provides brief background information about the topic-modeling approach and previous studies that have applied topic-modeling techniques in healthcare. This is followed by a description of the study methodology and the main quantitative and qualitative results. We present qualitative results obtained by applying a tool for deep inspections of topic-term relationships, the WHO framework on the international classification of patient safety, and reviews of two external academic experts on patient safety. The paper ends by presenting a discussion and drawing conclusions.

2. Background

2.1. Topic Modeling

Topic modeling is a sophisticated text-mining technique that was appropriate for the present research task—understanding the bottom-up concerns of healthcare stakeholders by identifying topics in the text that appears in Web-based communities. Topic modeling is a statistical method for uncovering abstract topics from a collection of documents [14]. For example, if a document includes ‘law’ as a topic, this document is likely to contain related words such as ‘legislation,’ ‘medical

law,’ ‘health act,’ ‘civic group(s),’ ‘duty,’ ‘inform,’ ‘rule,’ ‘policy,’ and ‘personnel.’ If a document is about the topic of ‘institutional action on patient safety,’ then terms such as ‘action,’ ‘intervention,’ ‘inquiry,’ ‘standards,’ ‘prevention,’ ‘education,’ ‘report,’ ‘regulation,’ and ‘surveillance’ will co-appear with high probabilities. The name of the topic is abstracted and summarized by researchers (e.g., the topics ‘law’ or ‘patient safety activities’) based on the keywords that appear most frequently, because computer algorithms can find statistical clustering patterns of keywords but they cannot summarize the topics that keywords relate to. A document usually has a mixture of different topics, and topic modeling can capture those topics statistically using different algorithms. Topic models have several advantages when working with a large complex text data set where the topic categories might not be clear or easy to discern a priori. These models have been shown to be generally useful in several applications, such as analyzing text-based survey responses, foreign media, and FDA drug labels [8,9,14]. In this study we used the LDA algorithm developed by Blei and colleagues [11,15] to obtain the parameters for every document (i.e., Web-based documents).

2.2. Related Works

Text-mining methods such as topic modeling have many applications in various subjects related to healthcare. Researchers have applied topic models based on LDA to three areas: (i) recognizing specific events from clinical documentation, (ii) reducing the workload when systematically reviewing a large amount of literature, and (iii) understanding or identifying public opinions, concerns, and patterns from social media text data in a wide variety of online communities. Regarding the applications to clinical documentation, one study [10] analyzed the free-text descriptions of a large repository of reports on patient safety events. Those authors were able to automatically re-classify the events that were ambiguously categorized by event reporters. Another study [9] applied topic modeling to drug labeling in order to group drugs with similar safety concerns and/or therapeutic uses together. They used about 800 FDA-approved drug labels and generated 100 topics, each associated with a set of drugs based on their probability analysis.

Regarding the literature-review applications, one study [16] applied topic modeling to perform automatic citation screening for systematic reviews. Those authors reported that term-enriched topics were more informative and less ambiguous for systematic reviews. They also found that utilizing topic modeling dramatically decreased the workload of the screening phase. Another study [17] also used LDA to reduce the burden of screening for systematic reviews. Those authors used the topics as another feature representation for documents having no manually assigned information, such as MeSH terms.

Regarding the application to patient-generated text analysis, many studies have used topic modeling to analyze the opinions of online health communities in order to understand the needs and interests of patients. One study [18] used this method to detect patient non-compliance behaviors associated with a drug of interest from message posts on the most-popular French online forums. Other studies [12,19] applied topic-extracting algorithms for LDA to capture what topics health consumers discuss when they are reviewing their health providers online. Ranald et al. [12] compared the content of online narrative reviews of hospitals to the topics in the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey. They found that most topics in online narratives are not measured or reported by HCAHPS. Those authors suggested that policymakers should focus on the measures of hospital quality that matter the most to patients and caregivers. Another study [20] applied automatic topic detection to explore hot topics in online communities about lung cancer, breast cancer, and diabetes. The identified hot topics mainly covered symptoms, examinations, drugs, procedures, and complications.

Few studies have analyzed the content on social media related to

healthcare policy. Kang et al. [13] applied opinion-mining techniques to investigate public opinions about a new school-meals policy for preventing childhood obesity. They classified tweets into three categories: positive, negative, and neutral. Their approach employed several types of lookup table, which was developed specifically for tweets. They found that negative opinions outnumbered positive ones and suggestions for future policy improvements.

3. Methods

3.1. Defining Stakeholders and Web-Based Text as Data Sources

Healthcare stakeholders are generally defined as any person or party with an interest in the financing, implementation, or outcome of a service, practice, process, or decision made by another (e.g., healthcare or health policies). In this study we divided the stakeholders into four groups: consumers (e.g., patients, caregivers, and families), providers (e.g., physicians, nurses, and professional associations), governments (e.g., legislative bodies and accrediting agencies), and researchers. We searched Web-based communities and public websites for posting and news-sharing activities as well as opinions and information on patient safety. The 18 representative websites we found are listed according to stakeholder groups in Table 1: 1 large Korean mobile news portal of NAVER®; 6 websites of representative associations of physicians, nurses, and pharmacists; 10 sites of government organizations that have closely related to patient safety issues; and 1 large database portal website providing search services for academic and domestic research papers.

We collected text documents with timestamps indicating that they had appeared on a posting board or in a newsroom or announcement between January 2014 and September 2018. This period ranged from 1 year before establishment of the Patient Safety Act to the time when data collection was started for this study, and spanned 4 years and 9 months. We retrieved text documents by applying several inclusion keywords [e.g., patient safety, adverse event(s), hospital, healthcare institution, harm, injury, errors, accident, malpractice, misuse, side effects, mistakes, malfunction, and defect] and exclusion keywords (e.g., pet, bid, recruitment, ad, assault, strike, new product, new technology, and development). Fig. 1 shows the construct of documents and words retrieved by stakeholder groups.

For further data cleansing, we reviewed all of the document titles and excluded less-relevant documents manually. This process revealed 2,487 documents with timestamps that were used to extract meaningful words (Table 2). Text data preprocessing was divided into words using

a Korean NLP package, and special characters, pre and post position, and capital letters were treated consistently. For the preprocessed words, two authors (I.C. and M.L.) manually reviewed and filtered less-meaningful words, such as proper nouns and words that are common and either appear in all documents or appear only very rarely (less than 3 times). We also used a synonyms dictionary. This process identified 2,933 words that were used to develop a document-term matrix.

The topics were extracted by implementing LDA using the open-source software package topicmodels in R software [21]. The LDA model was estimated using the maximum likelihood as described by Grün and Hornik [21]. We identified the optimal number of topics by using the Idatuning package of R software and the statistical method of Griffiths and Steyvers [22]. To estimate a posterior distribution, it is using Gibbs sampling as a special case of Markov-chain Monte Carlo methods, which reportedly often yields relatively simple algorithms for approximate inferencing in high-dimensional models such as LDA [23]. According to Phan et al. [23], the estimation accuracy changes slightly with the number of Gibbs sampling iterations, with the iterations being fast and yielding stable results after a burn-in period. We therefore set alpha to 1.2195, and beta was freely estimated by conducting the test with the following setting: 2,000 iterations, burn-in period of 1,000 iterations, and a thinning interval of 500 iterations. (Supplement material A).

For the analysis at the level of topics, we formulated more-sophisticated generative models that incorporated parameters describing the changes in the prevalence of topics over time. The analysis was based on a post-hoc examination of the estimates of topic proportions produced by the model. Being able to identify the hot topics at a particular period is useful when considering the trends in the concerns of stakeholders for performing future strategic planning, by providing quantitative measures of the prevalence of particular kinds of topics. To find topics that consistently increased or decreased in popularity from 2014 to 2018, we conducted a linear trend analysis of a topic proportion on a quarterly basis, using the same single sample as in these analyses.

3.2. Qualitative Validation of the Extracted Topics

A topic model yields a word list of topics, which may be viewed as keywords of that topic. Texts with a high probability of exhibiting a certain topic may be viewed as key texts of the topic. Although the topics are generated algorithmically by computer software, researchers must assign meanings to them and infer their thematic coherence. This process requires experience and relevant knowledge in a specific domain. Although the present authors already had the required expertise,

Table 1
Data sources retrieved according to different healthcare stakeholder groups

Stakeholder group	Community name	Website URL
Consumer	NAVER® mobile news	https://news.naver.com/
	The Korean Nurses Association news	http://www.nursenews.co.kr/
	The Medical News	http://www.bosa.co.kr/news/
	The Korean Medical Association news	http://www.doctorsnews.co.kr/
	The Korean Pharm Business journal	http://www.pharmnews.com/
	Young Doctor's News	http://www.docdocdoc.co.kr/
Provider	Korean Hospital Association	https://www.kha.or.kr/
	Ministry of Health and Welfare	http://www.mohw.go.kr/
	Korean Institute for Healthcare Accreditation	http://www.koiha.kr/
	Ministry of Food and Drug Safety	http://www.mfds.go.kr/
	Korea Centers for Disease Control and Prevention	http://www.cdc.go.kr/CDC/
	Health Insurance Review and Assessment Service	https://www.hira.or.kr/
	National Health Insurance Service	https://www.nhis.or.kr/
	National Police Agency	http://www.police.go.kr/
	Korea Legal Aid Corporation	https://www.klac.or.kr/
	Korea Medical Dispute Mediation and Arbitration	https://www.k-medi.or.kr/
Researcher	Seoul Metropolitan Government	http://www.seoul.go.kr/
	DBPia®	http://www.dbpia.co.kr/

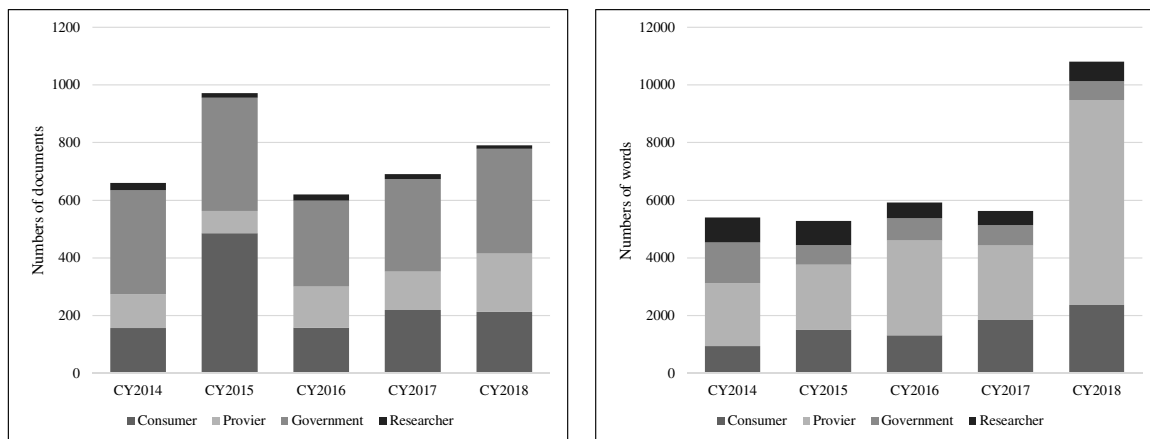


Fig. 1. Numbers of documents and words used by healthcare stakeholder groups over time.

Table 2

Data description used in topic modeling after preprocessing.

Stakeholder group	Number of documents	Number of words
Consumer	216	644
Provider	469	1,246
Government	1,740	679
Researcher	62	364
Total	2,487	2,933

we adopted several methods to stimulate an objective inference process.

First, we used the Web-based interactive visualization tool LDAvis, which enables deep inspections of the topic-term relationships in an LDA model [24]. The LDAvis tool provided a global view of the topics through their prevalence rates and similarities to each other in a compact space. (Supplement material B) Second, we used a list of social events that attracted social attention during the data collection period. This information was useful as cross-reference for determining the social context. For example, when a doctor and the associated hospital

were accused of illegal orthopedic surgeries being performed by an unlicensed person, relevant news and social responses increased with concerns about patient safety. In addition, we referenced the terms of the WHO International Classification of Patient Safety for representing themes [25]. Third, we asked two expert academics working on patient safety to review our inferred themes: an oncology specialist and an emergency medicine professor affiliated to an academic tertiary hospital and university. We explained our research aims, methods, and results, and then asked the two experts to independently score their level of agreement using a 5-point scale [from 1 (strongly disagree) to 5 (strongly agree)] and provide comments and feedback, which were used when confirming and revising the topic themes. After receiving their comments, we conducted text reviews of representative original documents on topics that showed discrepancies between two experts and the research team. In this qualitative step we inspected the properties of the text and disambiguated between word senses to identify the themes that were obscured by the topic word list.

Table 3

Top-3 topics and top-10 keywords for each healthcare stakeholder inferred from the LDA modeling.

Stakeholder group	Top-3 topics	Top-10 key words
Consumer	Topic 3: Newborn death at a neonatal ICU	Pediatrics, (bacterial) infection, death, ICU, infusion fluid, aseptic, injection, management, drug, mass media
	Topic 28: Middle East respiratory syndrome and visiting patients in hospitals	Visiting a patient, infection, prevention, litigation, seeing a doctor, ameliorating action, admission, structure, contact, risk
	Topic 30: Illegal surgeries	Clinician, surgery, manager, plastic surgery, suspension of qualification, anesthesia, medical law, commit, explain, protest
Provider	Topic 6: Hospital infection control	Infection, management, case, prevention, explain, risk, comprehensive countermeasure, communication, behavior, report
	Topic 11: Government policy	Government, management, revision bill, development, law, violation, system, regulation, team, errors
	Topic 32: Patient safety precaution alerts	Report, action, alert, errors, precaution, prevention, medication, evaluation case, related organization, recognition
Government	Topic 15: Enactment of legislation	Legislation, medical law, health act, civic group(s), duty, inform, rule, policy, personnel, compensation
	Topic 21: Systems for reducing damage due to adverse drug reactions	Drug, adverse reaction(s), diagnosis, death, system, infection rate, epidemiologic investigation, dosage, compensation, medical records
	Topic 22: Infusion-related infection	Bacteria, injection, hepatitis, prevention, epidemiologic investigation, nutritional fluid, supervision, contamination, treatment, procedure
Researcher	Topic 10: Hospital activities on patient safety	Action, intervention, inquiry, standards, prevention, education, reporting, regulation, surveillance, medical device
	Topic 14: Providers' perceptions of patient safety	Medical practice, provider(s), medical error(s), medical device(s), reporter, education, manager, perception, harm, surveillance
	Topic 38: Communication about patient safety	Acutely ill patient, clinician(s), communication, survey, medical accident, regulation, protect, risk, work-in, working hours

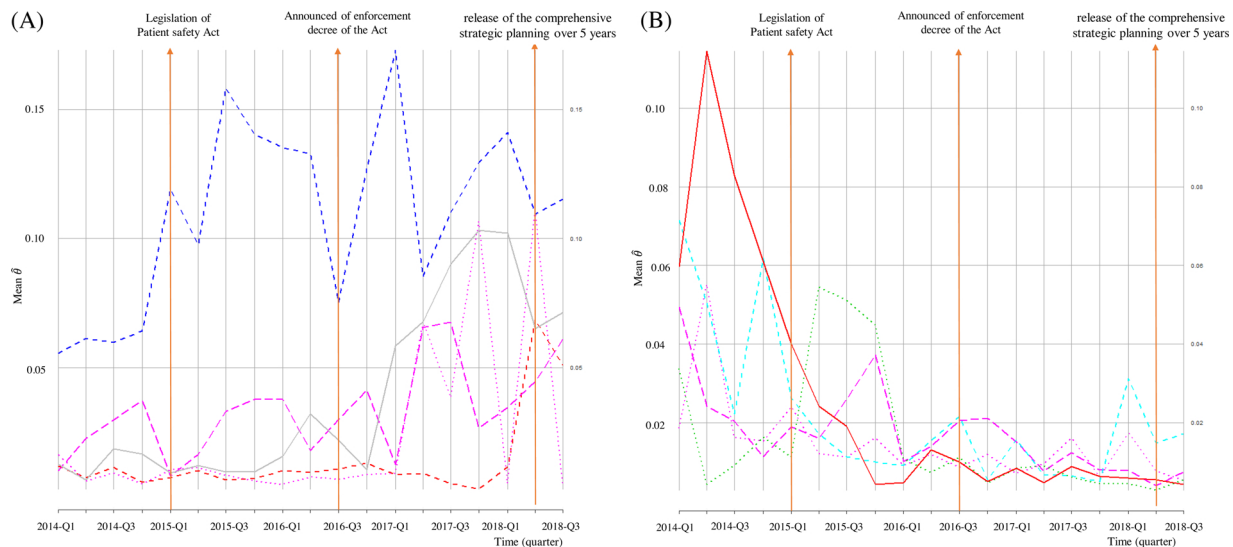


Fig. 2. Plots showing the trends of the five most- and least-popular topics from 2014 to 2018, defined as those topics that showed the strongest positive and negative trends. The most-probable words in those topics are shown below the plots. The orange color vertical bars indicate the time points when the Patient Safety Act was released over, when the act was enforced, and when the comprehensive strategic planning was released over 5 years.

(A) The five hot topics: blue dashed line is ‘management of regular check-ups’ (topic 1: check-ups, management, ward, medical device, safety education, staff, medical imaging, status report); gray solid line is ‘newborn death at a neonatal ICU’ [topic 3: pediatrics, (bacterial) infection, death, ICU, infusion fluid, aseptic, injection, management, drug, mass media]; pink dashed line is ‘institutional infection control’ (topic 6: infection, management, case, prevention, explain, risk, comprehensive countermeasure, communication, behavior, report); pink dotted line is ‘infusion-related infection’ (topic 22: bacteria, injection, hepatitis, prevention, epidemiologic investigation, nutritional fluid, supervision, contamination, treatment, procedure); red dashed line is ‘patient safety precaution alert’ (topic 32: report, action, alert, errors, precaution, prevention, medication, evaluation case).

(B) The five cold topics: Red solid line is ‘adverse drug reaction(s)’ (topic 5: symptom, drug, disease, medical diagnosis, admission, surgery, emergency, treatment, falls); blue dashed line is ‘government policy’ (topic 11: government, management, revision bill, development, law, violation, system, regulation, team, errors); pink dashed line is ‘nursing service’ (topic 24: work-in, communication, policy, medical law, duty, falls, prevention, mistakes); green dotted line is ‘illegal surgeries’ (topic 30: clinician, surgery, manager, plastic surgery, suspension of qualification, anesthesia, medical law, commit, explain, protest); pink dotted line is ‘fall risk assessment’ (topic 33: evaluation, falls, department safety, accident, prevention, mandatory prescription).

4. Results

4.1. Topic-Modeling Results

Applying the statistical method of Griffiths and Steyvers [22] to calculate harmonic means revealed that the optimal number of topics was 41. Through the exploration and inspection of the topic-term relationships, we identified 9, 20, 8, and 4 topics for the consumer, provider, government, and researcher groups, respectively. Several topics were temporally aligned with specific social events, while other topics with mixed meanings showed weak word–topic relationships. Table 3 presents the three most popular topics translated into English. The title of each topic was summarized by the authors based on the set of keywords returned by the LDA algorithm. We present only 10 of the keywords having higher probabilities under each topic.

Regarding the topics that consistently increased and decreased in popularity, five of the topics showed significant increasing trends while another five showed decreasing trends (both $P = 0.05$). The five hot and cold topics, as assessed by the size of the linear trend test statistics, are shown in Fig. 2. The hot topics were ‘management of regular check-ups,’ ‘newborn death at a neonatal ICU,’ ‘hospital infection control,’ ‘infusion-related infection,’ and ‘patient safety precaution alert.’ The cold topics were ‘adverse drug reaction(s),’ ‘government policy,’ ‘nursing service,’ ‘illegal surgeries,’ and ‘fall risk assessment.’

4.2. Qualitative Validation Results

Two reviewers agreed upon 35 themes that the research team assigned to each topic. One of the reviewers scored the remaining six topics lower than 4 points, and suggested different perspectives, while the other reviewer proposed different themes for four of these six topics. These topics showed characteristics of having no or only a very

small number of salient words in the word distributions and small proportions per document. For example, topic 27 subsumed words that appeared to have weaker relationships with each other. The research team agreed that several perspectives were mixed, and it was difficult to name a single theme: one of the reviewers thought it was about general characteristics of patient safety, while the other reviewer suggested that it was about the safety infrastructure. However, performing a qualitative review of the original text revealed that its main content was about initiating or posting ‘education of medical staff and personnel on patient safety’ at the hospital level and at the collective level of the Korean Hospital Association. The qualitative analysis resulted in the themes of 6 topics being revised and the remaining 35 being confirmed.

5. Discussion

This study explored whether the text data on patient safety collected from Web-based, user-generated documents represent the specific concerns of various healthcare stakeholders. We found that the text data were useful for inferring the latent concerns of consumers, providers, government bodies, and researchers, as well as changes therein over time. In particular, several important pieces of legislation have been introduced and improvements in the social system related to patient safety have been made over the last 5 years in Korea. We found that infection control by medical institutions was the most common concern among the healthcare stakeholders, and there were trends in the changes in concerns about medication errors, adverse drug reactions, and falls between before and after the Patient Safety Act was enacted. We also found that there were discrepancies between the study topics of researchers and the concerns of other stakeholders. These findings and methods could be used as the basis for a bottom-up approach to national strategic planning on patient safety.

Considering the common concern of the stakeholders, the Middle East respiratory syndrome outbreak in early 2015 [26] and infusion-related infections from late 2017 thorough to the middle of 2018 received considerable attention as a topic related to institutional infection control. Healthcare facilities increased their efforts in infection control and the government implemented a patient safety precaution alert system in late 2017 and comprehensive measures for preventing healthcare-associated infections, such as by releasing guidelines, training, and performing epidemic investigations and surveillance. However, the research topics did not reflect these concerns. One possible reason is that hospital infections represent strictly confidential information for each hospital, and so both the size and types of problems are unknown. Most of the cases are treated internally, and the information becomes public only if several patients got unexpected infection problems or died simultaneously. Other unusual situations, such as the increased cases of hepatitis C virus at a clinic in 2015 [27], become public after they are investigated by public health authorities. Such cases are often due to the lack of routine measurements and patient monitoring, and reviewing and improving everyday healthcare practices.

Regarding changes in stakeholder concerns between before and after the Patient Safety Act was enacted in January 2015, topics of adverse drug reactions, medication errors, and falls surged during 2014 and in early 2015, but these topics typically decreased after the legislation. This finding could be explained by two ways: (i) that the act and subsequent measures were effective in decreasing the occurrence of harmful events in healthcare settings, although this is less likely over a short period, but also there is no evidence for this, and (ii) the presence of the interests tilting phenomena, but this is still vague and needs further follow-up and investigation.

We did not find any trends in the most- or least-popular research topics. Unlike other stakeholders, the research topics were similar and showed little changes over time. The study topics were a hospital's patient safety activities and the perceptions, attitude, and awareness of clinicians and how they communicated patient safety issues. These results are similar to those of Cho et al. [28], who performed a systematic review of domestic papers on patient safety 6 years ago. That review found that research topics were limited to descriptive surveys of safety cultures and communication issue in a hospital and the perceptions of clinicians. The results of that study showed support the results obtained in our topic-modeling approach.

Finally, this study adopted the topic-modeling method to explore the heterogeneous concerns of healthcare stakeholders. Topic models have several advantages when working with a large complex text data set where the topic categories might not be clear or easy to discern a priori, and they have been shown to be generally useful in several applications. However, a few studies have used topic modeling and social media to capture public opinions about healthcare policy from different user groups. Traditional ways to obtain a consensus from stakeholders including using group panels with the RAND Appropriate Method [29] or the Delphi technique [30]. Different stakeholders have diverse views about the quality of care that are influenced by their past experiences, expectations, and definitions of quality of care, as well as the perceived power relationship between stakeholders [7]. These differences need to be captured and translated into policy. We believe that a text-mining technique such as topic modeling that we have applied has great potential in obtaining useful clues and insights hidden in large text data sets.

6. Conclusions

A common concern among stakeholders was hospital infection control, ranging from nosocomial infections to those brought in by visiting patients. Government policies and systemic approaches to patient safety were highlighted by different stakeholders. Researchers were focused on hospital sociocultural factors at both the

organizational and clinician levels. The new understanding about stakeholders will enhance the ability to harmonize policies related to patient safety, by incorporating factors such as culture, organizational structures, and modes of operation as well as wider political processes. In addition, our research methods and the results obtained could be used when designing a national patient safety strategy as a policy created using a bottom-up approach as well as when setting research priorities for academia.

Contributors

I.C. conceived the study. M.L. and Y.J.K. extracted and processed the data, and conducted the analyses. I.C. and M.L. conducted the qualitative review and validated the results, and drafted the first version of the manuscript. All of the authors commented on the manuscript drafts and then gave their approval for the final version to be published. I.C. acts as the guarantor.

Funding

This study was supported by grants from the National Research Foundation of Korea (numbers NRF-2016R1D1A1A09919502 and NRF-2019R1A2C2007583) and the Ministry of Trade, Industry and Energy of Korea (number 20004861).

Declaration of Competing Interest

None to declare.

References

- [1] Rq Lewis, M. Fletcher, Implementing a national strategy for patient safety: lessons from the National Health Service in England, *Qual Saf Health Care* 14 (2) (2005) 135–139.
- [2] M. Dixon-Woods, C. Tarrant, J. Willars, A. Suokas, How will it work? A qualitative study of strategic stakeholders' accounts of a patient safety initiative, *Qual Saf Health Care* 19 (1) (2010) 74–78.
- [3] Si. Lee, Patient safety and health technology assessment in Korea, *J Health Tech Assess* 4 (1) (2016) 5–7.
- [4] Ministry of Health and Welfare, Republic of Korea, The First Version of Comprehensive Planning on Patient Safety (2018~2022), Ministry of Health and Welfare, Seoul, 2018, p. 91.
- [5] S. Barrett, C. Fudge, Policy and action: essays on the implementation of public policy, Methuen young books, 1981 ISBN-13: 978-0416306705.
- [6] Korea Institute for Healthcare Accreditation, Developing the National Plan for Patient Safety R&D, Ministry of Health & Welfare Republic of Korea, Seoul, 2017 Contract No.: TRKO201800036984.
- [7] S. Campbell, T. Shield, A. Rogers, L. Gask, How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *BMJ Quality & Safety* 13 (6) (2004) 428–434.
- [8] M.E. Roberts, B.M. Stewart, D. Tingley, et al., Structural topic models for open-ended survey responses, *Am J Polit Sci* 58 (4) (2014) 1064–1082.
- [9] H. Bisgin, Z. Liu, H. Fang, et al., Mining FDA drug labels using an unsupervised learning technique—topic modeling, *BMC Bioinformatics* 12 (Suppl 10) (2011) S11.
- [10] A. Fong, R. Ratwani, An evaluation of patient safety event report categories using unsupervised topic modeling, *Methods Inf Med* 54 (4) (2015) 338–345.
- [11] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, *J Mach Learn Res* 3 (2003) 993–1022.
- [12] Bl Ranard, Rm Werner, T. Antanavicius, Ha Schwartz, Rj Smith, Zf Meisel, Da Asch, Lh Ungar, Rm. Merchant, Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care, *Health Affairs* 35 (4) (2016) 697–705.
- [13] Y. Kang, Y. Wang, D. Zhang, L. Zhou, The public's opinions on a new school meals policy for childhood obesity prevention in the US: A social media analytics approach, *Int J Med Inf* 103 (2017) 83–88.
- [14] M.E. Roberts, B.M. Stewart, D. Tingley, et al., The structural topic model and applied social science, *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*, Harrahs and Harveys, Lake Tahoe, 2013.
- [15] D.M. Blei, Probabilistic topic models, *Commun ACM* 55 (4) (2012) 77–84.
- [16] Y. Mo, G. Ktonatsios, S. Ananiadou, Supporting systematic reviews using LDA-based document representations, *Systematic reviews* 4 (1) (2015) 172.
- [17] M. Miwa, J. Thomas, A. O'Mara-Eves, S. Ananiadou, Reducing systematic review workload through certainty-based screening, *J Biomed Inf* 51 (2014) 242–253.
- [18] R. Abdellaoui, N. Foulquié, N. Texier, C. Faviez, A. Burgun, S. Schück, Detection of cases of noncompliance to drug treatment in patient forum posts: topic model

- approach, *J Med Internet Res* 20 (3) (2018) e85.
- [19] H. Hao, K. Zhang, The voice of Chinese health consumers: a text mining approach to web-based physician reviews, *J Med Internet Res* 18 (5) (2016) e108.
- [20] Y. Lu, P. Zhang, J. Liu, J. Li, S. Deng, Health-related hot topic detection in online communities using text clustering, *Plos one* 8 (2) (2013) e56221.
- [21] B. Grün, Hornik K. topicmodels: an R package for fitting topic models, *J Stat Softw* 40 (13) (2011) 1–30.
- [22] T.L. Griffiths, M. Steyvers, Finding scientific topics, *Proceedings of the National academy of Sciences* 101 (suppl 1) (2004) 5228–5235.
- [23] X.H. Phan, L.M. Nguyen, S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections, *Proceedings of the 17th international conference on World Wide Web, ACM* (2008) 91–100.
- [24] C. Sievert, K. Shirley, LDAvis: A method for visualizing and interpreting topics, *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (2014) 63–70.
- [25] World Health Organization and WHO Patient Safety, The Conceptual Framework for the International Classification for Patient Safety Version 1.1, World Health Organization, 2010 Contract No.: WHO/IER/PSP/2010.2.
- [26] M.D. Oh, W.B. Park, S.W. Park, et al., Middle East respiratory syndrome: what we learned from the 2015 outbreak in the Republic of Korea, *Korean J Intern Med* 33 (2) (2018) 233–246.
- [27] Y.S. Chung, J.Y. Choi, M.G. Han, et al., A large healthcare-associated outbreak of hepatitis C virus genotype 1a in a clinic in Korea, *J Clin Virol* 106 (2018) 53–57.
- [28] Ha Cho, Hs. Shin, A systematic review of published studies on patient safety in Korea, *J Korean Acad Dent Admin* 2 (1) (2013) 61–82.
- [29] R.H. Brook, M.R. Chassin, A. Fink, D.H. Solomon, J. Kosecoff, R.E. Park, A method for the detailed assessment of the appropriateness of medical technologies, *Int J Tech Assess in health care* 2 (1) (1986) 53–63.
- [30] F. Hasson, S. Keeney, H. McKenna, Research guidelines for the Delphi survey technique, *J Adv Nur* 32 (4) (2000) 1008–1015.