

Analysis of an Internet Community about Pneumothorax and the Importance of Accurate Information about the Disease

Bong Jun Kim, M.D., Sungsoo Lee, M.D., Ph.D.

Department of Thoracic and Cardiovascular Surgery, Gangnam Severance Hospital, Yonsei University College of Medicine

Background: The huge improvements in the speed of data transmission and the increasing amount of data available as the Internet has expanded have made it easy to obtain information about any disease. Since pneumothorax frequently occurs in young adolescents, patients often search the Internet for information on pneumothorax. **Methods:** This study analyzed an Internet community for exchanging information on pneumothorax, with an emphasis on the importance of accurate information and doctors' role in providing such information. **Results:** This study assessed 599,178 visitors to the Internet community from June 2008 to April 2017. There was an average of 190 visitors, 2.2 posts, and 4.5 replies per day. A total of 6,513 posts were made, and 63.3% of them included questions about the disease. The visitors mostly searched for terms such as 'pneumothorax,' 'recurrent pneumothorax,' 'pneumothorax operation,' and 'obtaining a medical certification of having been diagnosed with pneumothorax.' However, 22% of the pneumothorax-related posts by visitors contained inaccurate information. **Conclusion:** Internet communities can be an important source of information. However, incorrect information about a disease can be harmful for patients. We, as doctors, should try to provide more in-depth information about diseases to patients and to disseminate accurate information about diseases in Internet communities.

Key words: 1. Pneumothorax
2. Database
3. Statistics

Introduction

Technological developments have created new horizons in the 21st century [1]. Our life has changed dramatically because of the popularization of personal computers and the Internet. People can produce a tremendous amount of data using electronic devices and share it via various platforms such as Google, Naver, and YouTube. The development of an online network has allowed us to easily access a wide range of information. Recently, the term 'big data' has appeared in the context of these developments. The

concept of big data includes data with variety, volume, velocity, and/or veracity and value [2,3]. "Variety" includes heterogeneity and independency, and "velocity" refers to real-time data updates. Medical science and healthcare have been deeply influenced by these trends [3]. The amount of information in healthcare systems has grown over the past decades, to the point that healthcare can now be considered to be a domain of big data [3]. In the field of medicine, there are many forms of big data [4], such as the datasets of electronic health records (EHRs) [5], the genetic variants of patients suffering

Received: August 21, 2017, Revised: October 19, 2017, Accepted: October 19, 2017, Published online: April 5, 2018

Corresponding author: Sungsoo Lee, Department of Thoracic and Cardiovascular Surgery, Gangnam Severance Hospital, 211 Eonju-ro, Gangnam-gu, Seoul 06273, Korea
(Tel) 82-2-2019-3381 (Fax) 82-2-3461-8282 (E-mail) Chestlee@yuhs.ac

© The Korean Society for Thoracic and Cardiovascular Surgery. 2018. All right reserved.

© 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.

from cancer or genetic diseases [4], vital signs recorded by mobile health/wearable technologies [1], and doctor-patient and patient-patient interactions through Internet channels. Many recent studies have investigated these types of big data. Several studies have sought to predict readmission rates using pre-admission EHRs [5-7]; others have explored the use of big data in genomics [8], epidemiology [2,9], and personalized medicine [10]; and a recent study analyzed blog data on gastrointestinal diseases within the framework of big data [11]. Furthermore, studies have presented novel approaches to bioinformatics [12] and emergency medicine [13] in the context of big data.

In this study, we focused on patient-patient and patient-doctor communications on Internet channels about pneumothorax in the context of big data. The amount, variety, and accessibility of medical information through online networks are not inferior to the data contained in EHRs. However, studies using Internet data are relatively rare, and no such studies have been published on thoracic conditions. Therefore, we evaluated communications related to pneumothorax, which is the most common thoracic disease among adolescents, in an Internet community to provide suggestions that might be useful for clinicians.

Methods

The largest Internet community dedicated to pneumothorax in the Korean search engine and user-generated content platform Naver was selected for this study. This retrospective study was exempted from institutional review board of Gangnam Severance Hospital, since it was to be conducted on existing anonymous data from previous years. And the informed consent was waived. The study evaluated data from June 2008 to April 2017, which included a total 599,178 visitors. The authors analyzed the age and sex of visitors; the main terms searched by visitors; the number, type, and content of posts; and the content of replies. Intensive analysis of the content of posts and their replies was performed from June 2015 to April 2017. The authors analyzed details about the questions that users asked, the content of posts, and the content and quality of replies.

Because of limitations in the analytical tools pro-

Table 1. Characteristics of community members (N=6,409)

Characteristic	No. of all members (%)
Sex	
Male	4,531 (70.7)
Female	1,878 (29.3)
Age (yr)	
≤ 19	143 (2.2)
20-29	2,895 (45.2)
30-39	1,746 (27.2)
40-49	869 (13.6)
≥ 50	756 (11.8)

vided by this Internet community, we were only able to collect the demographic data of 6,420 community members (Table 1). The personal data of other non-member visitors were not collected. The term analysis and a detailed analysis of the posts covered the specified time range.

A total of 6,513 posts were divided into 2 groups: those made by healthcare workers and those made by others. Posts written by doctors included information about pneumothorax, such as typical symptoms, classification, treatment, and methods of surgical treatment. Posts written by visitors were mainly questions about the symptoms of pneumothorax, recurrent pneumothorax, pneumothorax operations, obtaining a medical certification of having been diagnosed with pneumothorax, and healthcare tips for preventing pneumothorax. Raw data from the Internet community were used and each post was classified according to the category it was published under in the Internet community.

The Internet community provided no tools for categorizing the replies. Therefore, the content of posts and replies from June 2015 to April 2017 was manually analyzed. In the selected period, posts about pneumothorax, recurrent pneumothorax, and operations to treat pneumothorax with at least 1 reply were included in our analysis. Details about the questions, the quality of the replies, and the content of posts expressing distrust toward doctors were analyzed. Posts were categorized as adequate replies, incorrect replies, and replies expressing a distrust of doctors. Posts with at least 1 medically appropriate answer were classified as adequate replies, and posts with only medically inappropriate answers were classified as incorrect replies. Posts that included neg-

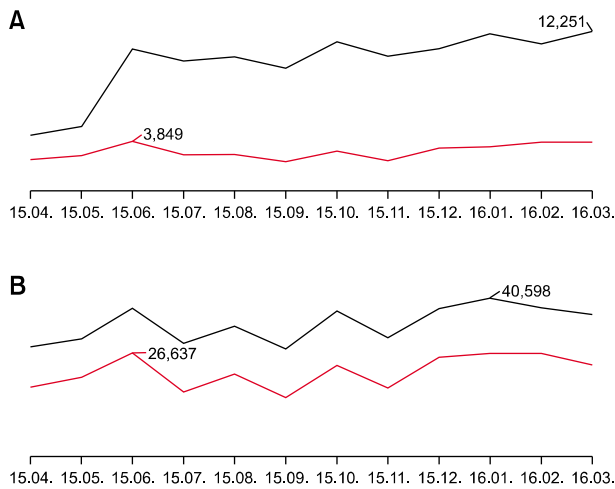


Fig. 1. Statistical trends (April 2015 to March 2016). (A) Trends in visitors, including members (red line) and nonmembers (gray line). (B) Trends in post views, including members (red line) and nonmembers (gray line).

ative references towards doctors were classified as expressing distrust of doctors.

Results

A total of 559,178 people visited the Internet community focusing on pneumothorax between June 2008 and April 2017. The monthly average number of visitors was 5,652 people, and the daily average number of visitors was 190 people. The number of visitors and posts both increased over the analysis period (Fig. 1A, B). Of the members who joined the Internet community, 70.7% were male, 29.3% were female, and 72.4% were young adults (aged 20–40 years). There were relatively few young adolescents, as only 2.2% of the members were younger than 19 years (Table 1). This distribution may be attributed to a requirement of the Korean Internet system. People younger than 19 years of age are unable to undergo network identification without the permission of their parents. Thus, they are restricted from joining Internet communities. However, young adolescents can participate in Internet communities using their parent’s identification, which would result in distribution bias.

The results of an analysis of keywords used to access the Internet community from January 2017 to April 2017 are presented in Table 2. Similar keywords were grouped together to simplify the data. A

Table 2. Search keywords from January 2017 to April 2017

Search keyword	No. (%)
Pneumothorax	5,295 (30.2)
Recurrent pneumothorax	2,478 (14.1)
Chest tube insertion	1,861 (10.6)
Pneumothorax operation	4,098 (23.3)
Pneumothorax and the army	1,892 (10.8)
Symptoms of pneumothorax	1,420 (8.1)
Chest pain	646 (3.7)
Cough	583 (3.3)
Emphysema	191 (1.1)
Airplane	56 (0.3)
Smoking	452 (2.5)
Total	17,552

total of 17,552 searches were performed, and the main keywords included ‘pneumothorax,’ ‘pneumothorax operation,’ ‘recurrent pneumothorax,’ ‘chest tube insertion,’ and ‘pneumothorax symptoms.’ Additionally, ‘pneumothorax and the army’ was among the main search terms, possibly due to specific aspects of the current political environment in the Republic of Korea.

A total of 6,513 posts were found, but only 21 posts were officially written by healthcare providers. Visitors and community members wrote the remaining 6,492 posts, most of which were questions about pneumothorax (28.5%), questions about recurrent pneumothorax (18.8%), questions about operations to treat pneumothorax (16.0%) and the treatment experiences of actual patients (35.9%) (Table 3). There were 54 informative posts (0.8%) written by visitors. However, some inaccurate information about pneumothorax was posted. Most of the inaccurate information was associated with traditional Korean medicine and treatment methods that have not been scientifically proven (Table 4).

No tools were available to analyze the full content of the replies. Thus, the authors manually analyzed all posts and replies from June 2015 to April 2017. A total of 1,579 posts were made during this period, and 1,254 contained questions about pneumothorax, recurrent pneumothorax, and operations to treat pneumothorax. All these 1,254 posts received at least 1 reply from another visitor, and almost all of them contained a reply by the writer of the original post (Table 5). Compared with active visitor-visitor communications, communications between healthcare

Table 3. Distribution of posts

Post writer	Post category	No. (%)
Healthcare provider	Information and symptoms of pneumothorax	6 (28.6)
	Video and real photographs of surgery	7 (33.3)
	Treatment of pneumothorax	8 (38.1)
Visitors and community members		6,492 (99.7)
	Questions about pneumothorax	1,852 (28.5)
	Pneumothorax operation	1,037 (16.0)
	Recurrent pneumothorax	1,221 (18.8)
	Information about pneumothorax	54 (0.8)
	Treatment experiences	2,328 (35.9)
Total		6,513

Table 4. Classification of information provided by visitors

Post content	No. (%)
Symptoms of pneumothorax	18 (33.3)
Recurrent pneumothorax	8 (14.9)
Treatment of pneumothorax	11 (20.3)
Korean medicine	7 (13.0)
Traditional	3 (5.6)
Pneumothorax and the army	3 (5.6)
Other	4 (7.4)
Total	54

Table 5. Post distribution from June 2015 to April 2017

Post content	No. (%)	Reply by visitor	Reply by writer
Questions about pneumothorax	561 (35.6)	561	560
Questions about recurrent pneumothorax	386 (24.4)	386	378
Questions about pneumothorax operation	307 (19.4)	307	304
Other	325 (20.6)	288	288
Total	1,579		

Table 6. Details of question about pneumothorax

Focus of question	Article	Adequate replies	Wrong replies	Distrust doctor
Pain	168	140	28	41
Symptom	123	97	26	45
Need for operation	103	64	39	22
Pleurodesis	59	31	28	28
Oxygen therapy	56	45	11	8
Exercise	38	35	3	0
Etc.	14	12	2	2
Total (%)	561	424 (75.6)	137 (24.4)	146 (26.0)

Table 7. Details of question about recurrent pneumothorax

Focus of question	Article	Adequate replies	Wrong replies	Distrust doctor
Symptom	258	211	47	35
Need for operation	83	63	20	25
Etc.	45	38	7	7
Total (%)	386	312 (80.8)	74 (19.2)	67 (17.4)

Table 8. Details of question about pneumothorax operation

Focus of question	Article	Adequate replies	Wrong replies	Distrust doctor
Pain	76	67	9	22
Wound	73	55	18	7
Recurrence after operation	61	42	19	31
Bloody sputum	32	31	1	0
Pleurodesis	28	20	8	17
Exercise	15	12	3	0
Etc.	22	13	9	9
Total (%)	307	240 (78.2)	67 (21.8)	86 (28.0)

workers and visitors were rare.

Details of the questions and replies are presented in Tables 6–8. There were 561 questions about pneumothorax, on topics including pain, the symptoms of pneumothorax, whether an operation is necessary, pleurodesis, oxygen therapy, and exercise (Table 6). These questions received 137 inaccurate replies (24.4%), and 146 posts (26.0%) expressed distrust of doctors. Additionally, 386 of the questions about recurrent pneumothorax contained the following 2 questions: ‘Does this symptom indicate a recurrence of pneumothorax?’ and ‘is an operation necessary for

recurrent pneumothorax?’ Seventy-four of these posts (19.2%) contained incorrect replies, and 67 posts (17.4%) contained content expressing distrust of doctors. There were also 307 questions about symptoms after surgery, including pain after the operation, wound healing after the operation, worries about recurrence after the operation, and persistent bloody sputum after the operation. The greatest proportion of questions addressed whether patients can start exercising after surgery. Sixty-seven posts (21.8%) contained incorrect replies and 86 (28.0%) presented suspicions about doctors. There were some inaccuracies in the provided information, but community members played a pivotal role in these interactions by replying to each article.

Discussion

We can easily access large amounts of various types of data through the Internet. Information technology has rapidly developed in the present decade, and we now live in a digital world [1]. The analysis of ‘big data’ is one of the most important approaches in modern academic research and can be readily applied to real-life situations [11]. Hence, analyses of big data have been published in many areas such as economics, industry, social sciences, politics, sports, and even in health and medicine [14]. However, in the Republic of Korea, studies of big data are relatively rare, both in medicine and in other fields [11].

Approximately 67,800 people annually visited the Internet community that we analyzed. Since the incidence of spontaneous pneumothorax is 7.4–18.0 per 100,000 males per year [15] in the Korean population, this number of visitors was not small. There are approximately 50,000,000 people in the Republic of Korea, and if we assume that pneumothorax occurs in 10 per 100,000 people per year, annually 50,000 people suffer from pneumothorax. This would imply that almost all patients with pneumothorax visited the Internet community. This assumption may be valid, because there are no other large pneumothorax-related Internet communities and the visitors came via various routes. Although only 6,409 community members visited regularly, all informative posts and replies were freely accessible to all visitors.

A total of 6,513 posts were made during the study

period, 6,492 of which were written by visitors. Healthcare providers only wrote 21 posts about pneumothorax. Although pneumothorax is a simple disease, patients seemed to have many questions about the disease, even after consulting with a doctor. Accordingly, there were 4,110 question-based posts (63.3% of all posts), and 2,328 posts about personal treatment experiences (35.9% of all posts). Most of the replies and answers to the questions were written by community members, not by doctors or other healthcare providers. Of course, many answers provided accurate information (78%), but inaccurate information could also be found in the replies (22%). This is unavoidable because the community members were not professional doctors. Furthermore, some community members provided information about pneumothorax based on their personal beliefs. Almost all the 54 informative posts were written using information found on other Internet sites; however, those posts included inaccurate information related to Korean medicine and traditional treatment methods. Korean medicine and traditional medicine have not been scientifically proven. Nevertheless, many Koreans readily accept information related to these concepts because of cultural traditions. Incorrect replies to questions and inaccurate information may be harmful for patients. Therefore, doctors should play an active role in addressing this issue.

Doctors should provide more in-depth information about the disease when treating patients experiencing a first attack of pneumothorax. Many patients who posted questions in the community had already been introduced to pneumothorax when they saw a doctor. However, they may not have felt satisfied with the information provided by the doctor, and therefore sought more information about their disease through the Internet community. They repeatedly asked about basic information about the disease, such as pain, symptoms, and the need for an operation, as well as about exercise. Obviously, the majority of doctors may think that they have given a detailed account of the disease to their patients, in addition to performing procedures such as pleurodesis and oxygen therapy. However, this appeared to be insufficient for the patients. As a result, most of the posts contained various repeated questions about pneumothorax. This perceived lack of information

may be explained by several factors. First, most patients experiencing their first attack of pneumothorax arrive in the emergency room feeling confused. Thus, although doctors may provide adequate information about the disease, the patients may not fully understand what they have been told. Second, patients may have found the medical terminology and information to be complicated. Finally, patients may have wanted to confirm the accuracy of the information that they were given. Thus, they looked to other sources of information, such as the Internet. Although patients may be partly responsible for these problems, we should invest the time and effort needed to provide accurate information to patients with pneumothorax, because it is our duty as doctors. Additionally, patient education is one of the most important social roles of doctors. Thus, we should explore various ways to carry out this responsibility.

Accurate and up-to-date information is important, especially when multiple sources of information are available [1]. Everyone can access the wealth of information on the Internet, a veritable sea of information, through devices such as smartphones. Of course, it is difficult to determine whether a piece of information is trustworthy [16]. Patients can easily search for information about their disease, and sometimes they trust blog posts more than doctors' recommendations. In this study, approximately 24% of patients believed incorrect replies rather than doctors' explanations, and they showed a distrust of doctors. In order to change this, the most crucial task for future doctors will be to provide detailed explanations about the disease and to establish a close rapport with patients in order to build trust. In comparison to previous decades, doctors can now communicate with patients more comfortably and easily through Internet communities and smart devices. We should embrace these technological changes and strive to provide accurate information to patients.

There are several limitations of this study. First, we were unable to access the personal data of all visitors because of limitations in the available analytical tools. We analyzed the data of community members instead of all visitors, making selection bias inevitable. Second, we were unable to trace the behavior of each visitor; thus, subgroup analysis was impossible. An analysis of personal data and in-

dividual behaviors would improve our interpretations of the data. Nevertheless, the protection of personal information has emerged as an important consideration in recent studies involving big data.

This study aimed to analyze an Internet community for the exchange of information about pneumothorax in the context of big data. In conclusion, the number of visitors and page views suggests that Internet communities can function as communication platforms for patients with pneumothorax. However, the information found in Internet communities can be inaccurate; thus, healthcare providers should attempt to provide accurate information about a disease and establish close rapport with patients so that they are more likely to trust a doctor's recommendation. In addition, doctors should provide accurate and in-depth information to patients experiencing their first attack of pneumothorax when they are hospitalized. Future studies will examine changes associated with the efforts of doctors, potentially underscoring the crucial role of doctors in public-facing healthcare.

Conflict of interest

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

Acknowledgments

This study was supported by a Grant of the Samsung Vein Clinic Network (Daejeon, Anyang, Cheongju, Cheonan; Fund no. KTCS04-098).

References

1. Elenko E, Underwood L, Zohar D. *Defining digital medicine*. Nat Biotechnol 2015;33:456-61.
2. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. *Clinical epidemiology in the era of big data: new opportunities, familiar challenges*. Clin Epidemiol 2017;9:245-50.
3. De la Torre Diez I, Cosgaya HM, Garcia-Zapirain B, Lopez-Coronado M. *Big data in health: a literature review from the year 2005*. J Med Syst 2016;40:209.
4. Psaty BM, Breckenridge AM. *Mini-Sentinel and regulatory science: big data rendered fit and functional*. N Engl J Med 2014;370:2165-7.
5. Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. *Predicting 30-day readmissions with pre-*

- admission electronic health record data. Med Care* 2015; 53:283-9.
6. Amarasingham R, Moore BJ, Tabak YP, et al. *An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care* 2010;48:981-8.
 7. Watson AJ, O'Rourke J, Jethwani K, et al. *Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure. Psychosomatics* 2011;52:319-27.
 8. O'Driscoll A, Daugelaite J, Sleator RD. *'Big data', Hadoop and cloud computing in genomics. J Biomed Inform* 2013;46:774-81.
 9. Young SD. *A "big data" approach to HIV epidemiology and prevention. Prev Med* 2015;70:17-8.
 10. Alyass A, Turcotte M, Meyre D. *From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics* 2015;8:33.
 11. Choi J, Park H, Lee CH. *Analysis of a blog for gastrointestinal disease in the view point of the big data: a single institutional study. Korean J Gastroenterol* 2014;63: 361-5.
 12. Merelli I, Perez-Sanchez H, Gesing S, D'Agostino D. *Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. Biomed Res Int* 2014;2014:134023.
 13. Wong HT, Yin Q, Guo YQ, Murray K, Zhou DH, Slade D, et al. *Big data as a new approach in emergency medicine research. J Acute Dis* 2015;4:178-9.
 14. Simpao AF, Ahumada LM, Galvez JA, Rehman MA. *A review of analytics and clinical informatics in health care. J Med Syst* 2014;38:45.
 15. Luh SP. *Review: diagnosis and treatment of primary spontaneous pneumothorax. J Zhejiang Univ Sci B* 2010;11:735-44.
 16. Moskowitz A, McSparron J, Stone DJ, Celi LA. *Preparing a new generation of clinicians for the era of big data. Harv Med Stud Rev* 2015;2:24-7.