Journal of Health Informatics and Statistics

Original Article

J Health Info Stat 2021;46(2):171-180 • https://doi.org/10.21032/jhis.2021.46.2.171

pISSN 2465-8014 | eISSN 2465-8022

The Awareness and Usage of Big Data for Cancer in Korea: A Survey Study

Eun Sil Baek[1], Choong-kun Lee[2,3], Jee Suk Chang[4], Jeong Eun Choi[1], Sang Joon Shin[2,3]

[1]Researcher, Songdang Institute for Cancer Research, Yonsei University College of Medicine, Seoul; [2]Professor, Songdang Institute for Cancer Research, Yonsei University College of Medicine, Seoul; [3]Professor, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul; [4]Professor, Department of Radiation Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

Objectives: A growing interest in big data for cancer research and treatment motivated us to investigate the demand for it by healthcare users. **Methods:** The survey was conducted from December 3, 2019 to January 7, 2020, in Korea. Respondents from 18 healthcare organizations participated in the survey. Among the 172 questionnaires received, 164 responses were used for the final analyses. **Results:** The majority of respondents showed a high awareness of big data related to cancer (n=148, 90.2%). However, only about half of the respondents were aware of how big data related to cancer is used (n=85, 51.8%). Among the respondents with experience using big data (n=83, 50.6%), more than half used big data only about once a year (n=43, 51.8%). The majority of respondents had particularly high demand for big data associated with "chemotherapy" (n=154, 94.5%), followed by "cancer type at diagnosis status", "clinical stage," and "recurrence." The main considerations for releasing cancer big data were "trustworthiness of data" (63.2%), "provision of valuable data" (58.9%), and "improvement of data accuracy" (55.8%). **Conclusions:** The study identified that even though respondents have a high awareness of and demand for big data related to cancer, it is not being sufficiently utilized at present. To increase the utilization of big data for cancer research and treatment, it is necessary to consider its purpose and how to make it available in line with the specific requirements of the healthcare industry, hospitals, and academia.

Key words: Big data, Awareness, Cancer, Healthcare, Surveys and questionnaires

INTRODUCTION

Big data in healthcare refers to electronic health datasets so large and complex that they are difficult to manage with traditional software, hardware, or common data management tools and methods. Big data in healthcare is overwhelming not only because of its volume, but also because of the diversity of data types and the speed at which it must be managed [1-3].

Besides structured large-capacity data, the scope of big data has continually expanded in recent years to include unstructured information [1-6]. In particular, due to the development of information and communication technology and the change in the health paradigm that follows new technology, the amount of data in the healthcare field is rapidly increasing. Along with this, there is a growing interest in analyzing and using big data in the field of healthcare.

The potential of big data in healthcare relies on the ability to turn high volumes of data into actionable knowledge for precision medicine and decision-making. Big data analytics in healthcare is evolving into a promising field for the provision of insights from very large data sets while reducing costs [3,7]. Especially in the case of cancer, there is a large amount of data related to the diagnosis, decision-making, treatment, and prognosis of patients. There is high interest and unmet need for the utilization of such big data related to cancer in hospitals, industry, and academia.

Korea has a well-established system for utilizing large amounts of medical data accumulated through electronic medical records, and vari-

Corresponding author: Sang Joon Shin
50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Tel: +82-2-2228-8138, E-mail: ssj338@yuhs.ac

Received: January 27, 2021 **Revised:** March 30, 2021 **Accepted:** April 9, 2021

*This study is supported by a grant from the Big data Center at the National Cancer Center of Korea (Grant number: 2021-data-we06).

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

How to cite this article:
Baek ES, Lee CK, Chang JS, Choi JE, Shin SJ. The awareness and usage of big data for cancer in Korea: a survey study. J Health Info Stat 2021;46(2):171-180. Doi: https://doi.org/10.21032/jhis.2021.46.2.171

ous attempts have been made to utilize this big data [8-12]. However, there have been no discussions on various demands or detailed methods to utilize big data for cancer [13-15]. Therefore, the purpose of this study was to investigate the level of awareness and utilization demand of big data for cancer among workers in the academic, medical, and healthcare industries.

METHODS

Study population and data collection

The target population of this study comprised individuals from tertiary hospitals, research institutes, pharmaceutical companies, contract research organizations, and academia in South Korea. The inclusion criteria were adults aged 19 years or older with the ability to understand a questionnaire, targeting students majoring in statistics, professors, researchers, data scientists, data managers, health information managers, and nurses. Those who refused to participate in this study were excluded. As a result, a total of 300 questionnaires were distributed. Data were collected either through paper surveys in face-to-face interviews or through questionnaires sent via e-mail. The responses were analyzed at the Yonsei Cancer Center, Korea.

Ethics

Participation in this study was entirely voluntary, and anonymity was guaranteed. All participants agreed that the results of the survey may be used for research purposes.

The method was approved by the Ethics Committee of the Yonsei University College of Medicine, Seoul, Korea

Questionnaire items

We developed questionnaires to investigate the demand of users for big data related to cancer in healthcare contexts. Survey items were developed with reference to a previous study [16,17].

The questionnaires were originally written and conducted in Korean. The questionnaires consisted of five sections: (1) awareness of big data related to cancer; (2) big data usage status and purpose of use; (3) demand for the release of big data to the public; (4) utilization of big data; and (5) basic characteristics of the respondents. The total number of questions to which the respondents could reply was 18. The number of question responses differed because the questionnaire included some conditional questions (Appendix 1).

Statistical analysis

Categorical data were summarized as frequencies and percentages (%). A chi-square test analysis was performed to test for differences in proportions of continuous and categorical variables between two or more groups. Statistical analysis was performed using the SPSS version 25.0 (IBM Co., Armonk, NY, USA). p-values < 0.05 were considered statistically significant. To ensure the clarity, precision, and accuracy of the results, the cases where respondents did not answer a sufficient number of questions, as well as the nonresponding cases, were excluded from the analyses.

RESULTS

Characteristics and knowledge of respondents

We conducted the survey from December 3, 2019 to January 7, 2020. A total of 300 questionnaires were distributed. A total of 172 questionnaire responses were received, and 164 responses were used for the final analyses. Eight respondents who skipped more than one-third of the

Table 1. Baseline characteristics of the respondents (n=164)

Characteristics	Respondents n (%)
Institution	
Healthcare industry	92 (56.1)
Tertiary hospital	51 (31.1)
Academia	21 (12.8)
Experience (y)	
≥10	81 (49.4)
5-<10	41 (25.0)
1-<5	28 (17.1)
<1	14 (8.5)
Discipline (multiple response question)	
Oncology	113 (76.4)
Bioinformatics	25 (16.9)
Pharmacy	9 (6.1)
Genetics	8 (5.4)
Epidemiology	7 (4.7)
Biochemistry	3 (2.0)
Pathology	2 (1.4)
Cardiology	2 (1.4)
Radiology	2 (1.4)
Pain management	2 (1.4)
Others[1]	7 (4.7)

[1]Includes infectious disease, neurology, chronic diseases, endocrinology, supportive care, surgery, gastroenterology, marketing, and regulatory.

questions were excluded from the final analyses. The majority of survey respondents worked in the healthcare industry (n=92, 56.1%) or in tertiary hospitals (n=51, 31.1%). Most respondents were in the field of oncology (n=113, 76.4%; multiple response question) and had more than 10 years of working experience (n=81, 49.4%) (Table 1).

In addition, we analyzed the results of the baseline characteristics in each group. There was a statistically significant difference in the years of experience of the group participating in the survey ($p<0.001$). However, in discipline comparisons, only oncology and bioinformatics were statistically significant ($p<0.001$ and $p<0.001$, respectively) (Supplementary Table 1).

Awareness of big data related to cancer

The question regarding the awareness of big data was segmented into four levels (Table 2). Most respondents (n=148, 90.2%) reported that they knew about big data for cancer. More than half of the respondents replied that they "know very well" (n=9, 5.5%) or "know a little" (n=83, 50.6%). The awareness of usage of big data in healthcare showed contradictory results, as about half reported "well" (n=81, 49.4%), which was closely followed by those who reported "not very well" (n=74, 45.1%) (Table 2). Most respondents were seen to agree ("strongly agree" (n=107, 65.2%) and "agree" (n=53, 32.3%)) on the need to use big data in healthcare services, business, and research. The utilization of big data in the academia group was the highest with "strongly agree" (n=19, 90.5%). Overall, respondents answered that the use of big data in current healthcare contexts is very necessary. Furthermore, the experience of participants in using big data, in terms of "yes" (n=83, 50.6%) and "no" (n=81, 49.4%), showed similar distributions. However, compared with each group, there was a higher proportion of "no" responses for experience using big data in the tertiary hospital group (n=36, 70.6%).

Status of usage of big data and purpose of use

Specifically, respondents with experience using big data were surveyed on the frequency, area and purpose of big data usage. Overall, responses for "more than once a year" (n=43, 51.8%) was the highest, followed by "at least once a month" (n=18, 21.7%) and "at least once a week" (n=11, 13.3%). Notably, the academia group had the most active big data users, selecting "at least once a day" (n=4, 36.4%) (Table 3). In addition, the utilization of "cancer diseases" big data (n=56, 68.3%) was the highest among all research areas. Regarding the purpose of using big data, responses for "collecting and utilizing data regarding work (service projects, proposals, reports, etc.)" (n=53, 65.4%) was the highest, followed by

Table 2. Awareness of big data related to cancer (n=164)

Variables	Total (n=164)	Healthcare industry (n=92)	Tertiary hospital (n=51)	Academia (n=21)	p[1]
	n (%)	n (%)	n (%)	n (%)	
How much do you know about big data related to cancer?					0.022
Know very well	9 (5.5)	3 (3.3)	4 (7.8)	2 (9.5)	
Know a little	83 (50.6)	47 (51.1)	30 (58.8)	6 (28.6)	
Heard about it	56 (34.1)	34 (37.0)	15 (29.4)	7 (33.3)	
Don't know	16 (9.8)	8 (8.7)	2 (3.9)	6 (28.6)	
Do you think big data is being used in healthcare?					0.057
Very well	4 (2.4)	2 (2.2)	1 (2.0)	1 (4.8)	
Well	81 (49.4)	46 (50.0)	25 (49.0)	10 (47.6)	
Not very well	74 (45.1)	44 (47.8)	20 (39.2)	10 (47.6)	
Not at all	5 (3.0)	0 (0.0)	5 (9.8)	0 (0.0)	
Do you think you need to use big data in your current health care, business or research?					0.053
Strongly agree	107 (65.2)	62 (67.4)	26 (51.0)	19 (90.5)	
Agree	53 (32.3)	28 (30.4)	23 (45.1)	2 (9.5)	
Disagree	3 (1.8)	1 (1.1)	2 (3.9)	0 (0.0)	
Strongly disagree	1 (0.6)	1 (1.1)	0 (0.0)	0 (0.0)	
Do you work with big data?					0.001
Yes	83 (50.6)	57 (62.0)	15 (29.4)	11 (52.4)	
No	81 (49.4)	35 (38.0)	36 (70.6)	10 (47.6)	

[1]p-values for multiple response questionnaires are not applicable.

Table 3. Usage of big data and purpose of use

Experience in big data	In the case of having experience using big data				p^3
	Total (n=164)	Healthcare industry (n=92)	Tertiary hospital (n=51)	Academia (n=21)	
	n (%)	n (%)	n (%)	n (%)	
How often do you use big data in your current business and research? (n=83)					0.002
At least once a day	5 (6.0)	1 (1.8)	0 (0.0)	4 (36.4)	
At least once a week	11 (13.3)	6 (10.5)	3 (20.0)	2 (18.2)	
At least once a month	18 (21.7)	13 (22.8)	3 (20.0)	2 (18.2)	
More than once a year	43 (51.8)	33 (57.9)	7 (46.7)	3 (27.3)	
Don't use it at all	6 (7.2)	4 (7.0)	2 (13.3)	0 (0.0)	
With what type of data do you currently work in your research? (n=82) (multiple response question)					NA
Cancer diseases	56 (68.3)	40 (70.2)	12 (85.7)	4 (36.4)	
Diabetes/Metabolic diseases	12 (14.6)	10 (17.5)	1 (7.1)	1 (9.1)	
Cardio-circulatory system disease	18 (22.0)	16 (28.1)	2 (14.3)	0 (0.0)	
Neurobiology diseases	7 (8.5)	7 (12.3)	0 (0.0)	0 (0.0)	
Digestive system diseases	2 (2.4)	0 (0.0)	2 (14.3)	0 (0.0)	
Respiratory diseases	1 (1.2)	1 (1.8)	0 (0.0)	0 (0.0)	
Infectious disease	7 (8.5)	7 (12.3)	0 (0.0)	0 (0.0)	
Dermatology disease	1 (1.2)	1 (1.8)	0 (0.0)	0 (0.0)	
DNA/RNA/protein sequence	7 (8.5)	4 (7.0)	1 (7.1)	2 (18.2)	
Others[1]	20 (24.4)	13 (22.8)	0 (0.0)	7 (63.6)	
What is the purpose of using big data? (n=81) (multiple response question)					NA
Development of web, application (App), and other services using the cancer data	4 (4.9)	1 (1.8)	2 (14.3)	1 (10.0)	
Collecting and utilizing data regarding work (service projects, proposals, reports, etc.)	53 (65.4)	46 (80.7)	3 (21.4)	4 (40.0)	
Acquiring information for starting companies or finding new business opportunities	9 (11.1)	8 (14.0)	0 (0.0)	1 (10.0)	
Collecting data for academic research	48 (59.3)	25 (43.9)	13 (92.9)	10 (10.0)	
Collecting data for policy research	7 (8.6)	4 (7.0)	3 (21.4)	0 (0.0)	
Collecting data for genome research	7 (8.6)	3 (5.3)	2 (14.3)	2 (20.0)	
Collecting data for new drug development	15 (18.5)	12 (21.1)	1 (7.1)	2 (20.0)	
Others[2]	1 (1.2)	0 (0.0)	1 (7.1)	0 (0.0)	

[1]Includes nonresponse as well as Health Risk Appraisal (HRA), pain data, gynecology, claim data, work-related, chemical compound data, older adults, autoimmune disease, drug safety, real estate data, and credit data.
[2]Includes sales prediction for pharmaceuticals, and feasibility study on drugs.
[3]*p*-values for multiple response questionnaires are not applicable.

"collecting data for academic research" (n=48, 59.3%), and "collecting data for new drug development" (n=15, 18.5%).

Demand for releasing big data related to cancer

We analyzed the results of the questions pertaining to the demand for releasing big data related to cancer. The majority of respondents revealed a particularly high demand for data related to "chemotherapy" (n=154, 94.5%; multiple response question), followed by "cancer type at diagnosis status" (n=139, 85.3%), "clinical stage" (n=138, 84.7%), and "recurrence" (n=136, 83.4%). Overall, respondents wanted information mostly about the treatment and clinical status of patients (Supplementary Table 2).

Regardless of whether they had big data usage experience or not, the respondents said that they required big data in the form of "Excel based file (xls, xlsx, csv)" (n=139, 86.9%), followed by "LOD (Linked Open Data)" (n=19, 10.5%), and "File (json, xml)" (n=11, 6.1%) (Table 4). Most respondents (n=131, 80.4%) revealed a willingness to pay to utilize big data, and the proportion was higher in the group with experience using big data related to cancer (n=70, 84.3%). A higher proportion of academia group respondents with no big data usage experience reported a lack of willingness to pay for big data related to cancer (n=7, 70.0%). The

Table 4. Demand for the release of big data related to cancer

Experience in big data	Total (n=163)		Healthcare industry (n=91)		Tertiary hospital (n=51)		Academia (n=21)		p^2
	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)	
In what data format do you want to receive big data related to cancer? (n=160) (multiple response question)									NA
Excel-based file (xls, xlsx, csv)	72 (88.9)	67 (84.8)	49 (89.1)	29 (87.9)	14 (93.3)	32 (88.9)	9 (81.8)	6 (60.0)	
File (json, xml)	8 (9.9)	3 (3.8)	4 (7.3)	0 (0.0)	1 (6.7)	1 (2.8)	3 (27.3)	2 (20.0)	
Open API	5 (6.2)	1 (1.3)	3 (5.5)	0 (0.0)	0 (0.0)	0 (0.0)	2 (18.2)	1 (10.0)	
LOD	10 (12.3)	9 (11.4)	9 (16.4)	5 (15.2)	0 (0.0)	3 (8.3)	1 (9.1)	1 (10.0)	
Others	5 (6.2)	1 (1.3)	3 (5.5)	0 (0.0)	1 (6.7)	0 (0.0)	1 (9.1)	1 (10.0)	
For statistical data using cancer big data, are you willing to pay the price? (n=163)									<0.001
Yes	70 (84.3)	61 (76.3)	53 (93.0)	31 (91.2)	12 (80.0)	27 (75.0)	5 (45.5)	3 (30.0)	
No	13 (15.7)	19 (23.8)	4 (7.0)	3 (8.8)	3 (20.0)	9 (25.0)	6 (54.5)	7 (70.0)	
How do you mainly intend to use the data? (n=163) (multiple response question)									NA
Development of web, application (App), and other services using the cancer data	10 (12.0)	20 (25.0)	6 (10.5)	8 (23.5)	1 (6.7)	8 (22.2)	3 (27.3)	4 (40.0)	
Collecting and utilizing data regarding work (service projects, proposals, reports, etc.)	54 (65.1)	37 (46.3)	48 (84.2)	22 (64.7)	5 (33.3)	15 (41.7)	1 (9.1)	0 (0.0)	
Acquiring information for starting companies or finding new business opportunities	17 (20.5)	10 (12.5)	13 (22.8)	7 (20.6)	1 (6.7)	1 (2.8)	3 (27.3)	2 (20.0)	
Collecting data for academic research	52 (62.7)	62 (77.5)	30 (52.6)	23 (67.6)	14 (93.3)	29 (80.6)	8 (72.7)	10 (100.0)	
Collecting data for policy research	10 (12.0)	23 (28.8)	9 (15.8)	10 (29.4)	1 (6.7)	10 (27.8)	0 (0.0)	3 (30.0)	
Collecting data for genome research	13 (15.7)	23 (28.8)	4 (7.0)	9 (26.5)	6 (40.0)	13 (36.1)	3 (27.3)	1 (10.0)	
Collecting data for new drug development	24 (28.9)	39 (48.8)	19 (33.3)	21 (61.8)	2 (13.3)	14 (38.9)	3 (273)	4 (40.0)	
Others[1]	1 (1.2)	1 (1.3)	1 (1.8)	1 (2.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
What do you think are important things to consider when releasing cancer big data? (n=163) (multiple response question)									NA
A sufficient amount of data	29 (34.9)	30 (37.5)	19 (33.3)	18 (52.9)	3 (20.0)	6 (16.7)	7 (63.6)	6 (60.0)	
Trustworthiness of data	51 (61.4)	52 (65.0)	35 (61.4)	22 (64.7)	12 (80.0)	25 (69.4)	4 (36.4)	5 (50.0)	
Valuable data	46 (55.4)	50 (62.5)	30 (52.6)	20 (58.8)	9 (60.0)	22 (61.1)	7 (63.6)	8 (80.0)	
Improvement of data accuracy	46 (55.4)	45 (56.3)	35 (61.4)	17 (50.0)	7 (46.7)	21 (58.3)	4 (36.4)	7 (70.0)	
Timely data	21 (25.3)	16 (20.0)	17 (29.8)	6 (17.6)	3 (20.0)	9 (25.0)	1 (9.1)	1 (10.0)	
Up-to-date data	28 (33.7)	20 (25.0)	28 (49.1)	12 (35.3)	0 (0.0)	5 (13.9)	0 (0.0)	3 (30.0)	
Improvement of usability through standardization	34 (41.0)	41 (51.3)	24 (42.1)	18 (52.9)	6 (40.0)	18 (50.0)	4 (36.4)	5 (50.0)	

LOD, linked open data; open API, Open application programming interface.
[1]Includes development of evidence for insurance benefits of new drugs.
[2]*p*-values for multiple response questionnaires are not applicable.

main purpose of big data usage was for "collecting and utilizing data regarding work (service projects, proposals, reports, etc.)" (n=54, 65.1%) among the respondents with big data usage experience. The main considerations for the release/provision of big data related to cancer were: "trustworthiness of data" (n=103, 63.2%), "valuable data" (n=96, 58.9%), and "improvement of data accuracy" (n=91, 55.8%).

DISCUSSION

This survey was conducted to understand the awareness of big data for cancer from the perspective of users in the Korean healthcare environment. The main findings indicate that while most respondents recognize the necessity of using big data related to cancer in healthcare, business, and research, they are not using it frequently in practice. In addition, the majority of respondents indicated a particularly large demand for data on chemotherapeutic agents for treatments and other cancer-specific clinical information.

A higher percentage of people in academia, in comparison to other groups, do not know about big data, while the tertiary hospitals group reported the lowest rate of big data usage experience. Nevertheless, regardless of the institution, most respondents (over 95%) showed a willingness to use big data, especially in academia. The reasons for using big data were different among institutions, and as expected, the purpose of academic research received the highest response for the tertiary hospitals

group (92.9%), with the healthcare industry tending to use big data for projects or reports (80.7%). Most respondents (80.4%) showed a willingness to pay for big data, and this willingness was highest among those in the healthcare industry, followed by tertiary hospitals and academia. The majority of responses were positive, indicating that respondents are aware of big data as an economically valuable resource.

Recently, there have been various attempts [11,16,18-20] to analyze the needs for big data in the field of healthcare and to encourage big data utilization [21,22]. While European countries are supporting initiatives to utilize big data in the field of oncology [7], there has not been much effort in Korea to make legal and institutional improvements to encourage usage of big data for cancer. The term "big data" has become extremely popular globally in recent years and almost every field of research, whether it relates to industry or academics, is generating and analyzing big data for various purposes [23]. Particularly in medicine and healthcare, big data analytics integrates the analysis of several scientific areas such as bioinformatics, medical imaging, as well as medical and health informatics. The application of big data analytics aids the discovery of comprehensive knowledge from the huge amounts of data available [24]. Big data analytics in medicine and healthcare enables analysis of large datasets from thousands of patients, identifying clusters and correlations between datasets, as well as developing predictive models using data mining techniques [25]. The combination of data analysis and artificial intelligence technologies such as machine learning and deep learning allows for innovation in healthcare services such as patient-specific clinical decision support system utilization and precision medicine in real time [26,27]. Additionally, in new drug development, a field that entails enormous time and high investment costs, a partial solution to the cost-efficiency problem is expected through the utilization of accumulated big data on cancer in clinical trials for diagnosis, treatment, results, and prescriptions.

Korea has recently started to promote the use of accumulated big data in cancer research and treatment along with enhancement of privacy and utilization through the enactment of relevant laws. The release of generated big data relating to cancer is a current trend creating added value. It is very important to consider the requirements of various stakeholders in big data usage and related analyses.

This study has several limitations. First, there are important differences that may limit the generalizability of the study findings to the Korean population and may not reflect the opinion of all survey respondents. Second, responses may differ depending on the public policies or legal frameworks in other countries. Third, the respondents in this study were primarily from the healthcare industry and tertiary hospitals, and notably, the respondents from academia were few. Moreover, those who were not familiar with big data did not participate in the survey, and thus, the participants willing to answer the survey could be those who have more knowledge or awareness about the topic and might bias the results. Lastly, the cross-validation of the same perception among institutions was not compared, and thus requires further investigation. Hence, the results of this survey may only reflect the current situation in South Korea. Despite these limitations, this study forms an important baseline for future studies.

The study was able to identify the high awareness and demand for big data related to cancer among the respondents through the survey. However, compared to the high demand indicated in the survey responses, big data is not being well utilized. There will likely be more demand for big data utilization in the "new normal" era following the coronavirus disease (COVID-19) pandemic. To increase the utilization of big data for cancer, it is necessary to consider ways to release the information in accordance with the purpose and the finer details necessary for using such data in the healthcare industry, hospitals, and academia.

Furthermore, it is necessary to lay the foundation for an environment that can enhance consumer-centered data accessibility and establish detailed policies regarding the scope of using such data, and the methods and procedures associated with rapid and secure release of cancer related big data.

ORCID

Eun Sil Baek *https://orcid.org/0000-0003-4127-1328*
Choong-kun Lee *https://orcid.org/0000-0001-5151-5096*
Jee Suk Chang *https://orcid.org/0000-0001-7685-3382*
Jeong Eun Choi *https://orcid.org/0000-0001-8859-7347*
Sang Joon Shin *https://orcid.org/0000-0001-5350-7241*

REFERENCES

1. Frost, Sullivan. Drowning in big data? reducing information technology complexities and costs for healthcare organizations, 2015.
2. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al.

Big data: The next frontier for innovation, competition, and productivity. UK: McKinsey Global Institute; 2011.
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3. DOI: 10.1186/2047-2501-2-3
4. Kim J, Kim H, Son K, Song Y, Yoon J, Lim H, et al. Medical utilization of big data. Inf Sci Manage 2014;32(3):18-26 (Korean).
5. Lee J, Jae M, Jo M, Son H. Big data utilization trends in the healthcare. J Korea Inst Electronic Commun Sci 2014;2(1):63-75 (Korean).
6. Popovic JR. Distributed data networks: a blueprint for Big Data sharing and healthcare analytics. Ann N Y Acad Sci 2017;1387(1):105-11. DOI: 10.1111/nyas.13287
7. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health 2019;29(Supplement_3): 23-27. DOI: 10.1093/eurpub/ckz168
8. Park YT, Han D. Current status of electronic medical record systems in hospitals and clinics in Korea. Healthc Inform Res 2017;23(3):189-198 (Korean). DOI: 10.4258/hir.2017.23.3.189
9. Park YT, Kim YS, Yi BK, Kim SM. Clinical decision support functions and digitalization of clinical documents of electronic medical record systems. Healthc Inform Res 2019;25(2):115-23 (Korean). DOI: 10.4258/hir.2019.25.2.115
10. Seong SC, Kim YY, Khang YH, Heon Park J, Kang HJ, Lee H, et al. Data resource profile: the National Health Information Database of the National Health Insurance Service in South Korea. Int J Epidemiol 2017;46(3):799-800. DOI: 10.1093/ije/dyw253
11. Kim HH, Kim B, Joo S, Shin SY, Cha HS, Park YR. Why do data users say health care data are difficult to use? A cross-sectional survey study. J Med Internet Res 2019;21(8):e14126. DOI: 10.2196/14126
12. Yu HW, Choi JY, Park YS, Park HS, Choi Y, Ahn SH, et al. Implementation of a resident night float system in a surgery department in Korea for 6 months: electronic medical record-based big data analysis and medical staff survey. Ann Surg Treat Res 2019;96(5):209-215. DOI: 10.4174/astr.2019.96.5.209
13. Willems SM, Abeln S, Feenstra KA. The potential use of big data in oncology. Oral Oncol 2019;98:8-12. DOI: 10.1016/j.oraloncology.2019.09.003
14. Major A, Cox SM, Volchenboum SL. Using big data in pediatric oncology: current applications and future directions. Semin Oncol 2020;47(1):56-64. DOI: 10.1053/j.seminoncol.2020.02.006
15. Schlick CJR, Castle JP, Bentrem DJ. Utilizing big data in cancer care. Surg Oncol Clin N Am 2018;27(4):641-652. DOI: 10.1016/j.soc.2018.05.005
16. Barone L, Williams J, Micklos D. Unmet needs for analyzing biological big data: a survey of 704 NSF principal investigators. PLoS Comput Biol 2017;13(10):e1005755. DOI: 10.1371/journal.pcbi.1005755
17. Dillman DA, Smyth JD, Christian LM. Internet, mail, and mixed-mode surveys: The tailored design method (3rd ed.). Hoboken, NJ: John Wiley & Sons; 2009.
18. Brennan PF, Bakken S. Nursing needs big data and big data needs nursing. J Nurs Scholarsh 2015;47(5):477-484. DOI: 10.1111/jnu.12159
19. McNutt TR, Moore KL, Quon H. Needs and challenges for big data in radiation oncology. Int J Radiat Oncol Biol Phys 2016;95(3):909-915. DOI: 10.1016/j.ijrobp.2015.11.032
20. Chen B, Butte AJ. Leveraging big data to transform target selection and drug discovery. Clin Pharmacol Ther 2016;99(3):285-297. DOI: 10.1002/cpt.318
21. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care?. J Arthroplasty 2018;33(8):2358-2361. DOI: 10.1016/j.arth.2018.02.067
22. Capobianco E. Data-driven clinical decision processes: it's time. J Transl Med 2019;17(1):44. DOI: 10.1186/s12967-019-1795-5
23. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019;6:54. DOI: 10.1186/s40537-019-0217-0
24. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integr Bioinform 2018;15(3):20170030. DOI: 10.1515/jib-2017-0030
25. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. IEEE J Biomed Health Inform 2015;19(4): 1209-1215. DOI: 10.1109/JBHI.2015.2406883
26. Song TM, Ryu S. Big data analysis framework for healthcare and social sectors in Korea. Healthc Inform Res 2015;21(1):3-9 (Korean). DOI: 10.4258/hir.2015.21.1.3
27. Wang YC, Hajli N. Exploring the path to big data analytics success in healthcare. J Bus Res 2017;70:287-299. DOI: 10.1016/j.jbusres.2016.08.002

국문초록

한국의 암 빅데이터 인식과 활용: 설문 조사

백은실[1] · 이충근[2,3] · 장지석[4] · 최정은[1] · 신상준[2,3]

[1]연세대학교 의과대학 송당암연구센터 연구원, [2]연세대학교 의과대학 송당암연구센터 교수, [3]연세대학교 의과대학 내과학교실 교수, [4]연세대학교 의과대학 강남세브란스병원 방사선종양학교실 교수

목적: 암 연구 및 치료를 위한 빅데이터에 대한 관심이 높아짐에 따라 헬스케어 관련 종사자들의 암 빅데이터 수요도를 조사하였다.

방법: 본 연구는 2019년 12월 3일부터 2020년 1월 7일까지 한국에서 실시되었다. 18개 헬스케어 관련 기관이 설문조사에 참여하였다. 회신 받은 172건의 설문지 중 164건의 설문지가 최종 분석에 활용되었다.

결과: 응답자의 대부분은 암 빅데이터에 대한 높은 인지도를 보였다(n = 148, 90.2%). 다만 응답자의 절반 정도만이 암 빅데이터가 잘 활용되고 있다고(n = 85, 51.8%) 답변했다. 빅데이터 사용 경험이 있는 응답자(n = 83, 50.6%) 중 절반 이상은 1년에 한 번 정도만 빅데이터를 사용한 것으로 나타났다(n = 43, 51.8%). 응답자들은 '항암제'(n = 154, 94.5%) 정보 개방을 가장 필요로 하였으며, '암종별 진단 상태', '임상 병기', '재발 정보' 등의 순서로 응답했다.

결론: 본 연구는 설문 조사를 통해 암 빅데이터에 대한 응답자들의 높은 인지도와 수요도를 확인할 수 있었다. 하지만 그에 비해 빅데이터는 잘 활용되지 못하는 것으로 확인하였다. 암 연구와 치료를 위한 빅데이터 활용도를 높이기 위해서는 의료 산업, 병원, 학계의 구체적인 요구사항에 적용할 수 있는 활용방안을 고려할 필요가 있다.

주제어: 빅데이터, 인식, 암, 헬스케어, 설문조사

Appendix 1. Questionnaire on utilization of cancer-related big data

Awareness of big data related to cancer

1. How much do you know about big data related to cancer?
   ① Know very well
   ② Know a little
   ③ Heard about it
   ④ Don't know

2. Do you think big data is being used in healthcare?
   ① Very well
   ② Well
   ③ Not very well
   ④ Not at all

3. Do you think you need to use big data in your current health care, business or research?
   ① Strongly agree
   ② Agree
   ③ Disagree
   ④ Strongly disagree

Usage of big data and satisfaction

4. Do you work with big data?
   ① Yes
   ② No (-> No.9)

5. How often do you use big data in your current business and research?
   ① At least once a day
   ② At least once a week
   ③ At least once a month
   ④ More than once a year
   ⑤ Don't use it at all.

6. With what type of data do you currently work in your research? (Select all that apply.)
   ① Cancer Diseases
   ② Diabetes / Metabolic Diseases
   ③ Cardio- Circulatory System Disease
   ④ Neurobiology Diseases
   ⑤ Digestive system diseases
   ⑥ Respiratory Diseases
   ⑦ Infectious Disease
   ⑧ Dermatology Disease
   ⑨ DNA/RNA/protein sequence
   ⑩ Other (please specify)

7. Please select the source and satisfaction of the big data you have used.

	satisfaction				
	Very satisfied	Satisfied	Natural	Dissatisfied	very dissatisfied
☐ In-house data	①	②	③	④	⑤
☐ Commercial data	①	②	③	④	⑤
Public data (domestic)					
☐ Ministry of Food and Drug Safety	①	②	③	④	⑤
☐ Health Insurance Review and Assessment Service	①	②	③	④	⑤
☐ National Health Insurance Service	①	②	③	④	⑤
☐ Public data (overseas)	①	②	③	④	⑤
☐ Open Source ware treated data: news release	①	②	③	④	⑤
☐ others	①	②	③	④	⑤

8. What is the purpose of using big data? (Select all that apply.)
   ① Development of web, application (App), and other services using the cancer data
   ② Collecting and utilizing data regarding work (service projects, proposals, reports, etc.)
   ③ Acquiring information for starting companies or finding new business opportunities
   ④ Collecting data for academic research
   ⑤ Collecting data for policy research
   ⑥ Collecting data for genome research
   ⑦ Collecting data for new drug development
   ⑧ Other (please specify)

Demands for releasing cancer big data to the public

9. Please select the items you wish to provide.

No.	Classification		check
1	Diagnosis	Cancer Type at Diagnosis status	☐
2		Initial Stage	☐
3		Metastasis	☐
4		Recurrence	☐
5		Physical Measurement	☐
6	Test	Laboratory Test	☐
7		Image Test	☐
8		Cancer Genetic Test	☐
9	Pathology	Surgical Pathology	☐
10		Immunopathology	☐
11		Molecular Pathology	☐

No.	Classification		check
12	Operation	Operation Information	☐
13		Operation Finding	☐
14		Pre-Operation Clinical findings	☐
15		Post-Operation Complications	☐
16	Treatments	Chemotherapy	☐
17		Hormone Therapy	☐
18		Radiation therapy	☐
19		Other treatments	☐

10. Feel free to use any cancer big data that you want to release in addition to the items selected above.

11. What data format do you want to receive big data related to cancer?
① File(xls, xlsx, csv)
② File(json, xml)
③ Open QPI
④ LOD(Linked Open Data)
⑤ Other(please specify)

Demands regarding the release of big data related to cancer

12. For statistical data using cancer big data, are you willing to pay the price?
① Yes
② No

13. How do you mainly intend to use the data? (Select all that apply.)
① Development of web, application (App), and other services using the cancer data
② Collecting and utilizing data regarding work (service projects, proposals, reports, etc.)
③ Acquiring information for starting companies or finding new business opportunities
④ Collecting data for academic research
⑤ Collecting data for policy research
⑥ Collecting data for genome research
⑦ Collecting data for new drug development
⑧ Other (please specify)

14. What do you think are important things to consider when releasing the cancer big data?
① A *sufficient amount of data*
② Faithful details of data
③ *Providing* valuable data
④ Improving data accuracy
⑤ *Providing timely data*
⑥ *Up-to-date data*
⑦ Improvement of usability through standardization
⑧ Other (please specify)

15. Please feel free to write down any suggestions for successful release and use of cancer big data.

Basics for Statistical Analysis

16. What position(s) do you currently hold?
① Sole Proprietorship / Start-up company
② SMEs
③ Major company
④ University Institution (Professor / Researcher)
⑤ Public Institution
⑥ Other (please specify

17. How many years experience do you have in your field?
① < 1 year
② 1 year to < 5 years
③ 5 years to < 10 years
④ ≥ 10 years

18. In what academic discipline is your research? (Select all that apply.)
① Oncology
② Pathology
③ Genetics
④ Biochemistry
⑤ Bioinformatics
⑥ Other (please specify)

Thank you for your participation.