Journal of Digital Convergence
Vol. 20. No. 5, pp. 681-693, 2022
ISSN 2713-6434 / eISSN 2713-6442
https://doi.org/10.14400/JDC.2022.20.5.681

A Keyword Network Analysis of Standard Medical Terminology for Musculoskeletal System Using Big Data

Byung-Kwan Choi[1], Eun-A Choi[2*], Moon-Hee Nam[2]
[1]Professor, College of Medicine, Pusan National University Hospital
[2]Associate Professor, Dept, of Nursing, Daedong College

빅데이터를 활용한 근골격계 표준의료용어에 대한 키워드 네트워크 분석

최병관[1], 최은아[2*], 남문희[2]
[1]부산대학교 의학전문대학원 정교수, [2]대동대학교 간호학부 부교수

Abstract The purpose of this study is to suggest a plan to utilize atypical data in the health care field by inferring standard medical terms related to the musculoskeletal system through keyword network analysis of medical records of patients hospitalized for musculoskeletal disorders. The analysis target was 145 summaries of discharge with musculoskeletal disorders from 2015 to 2019, and was analyzed using TEXTOM, a big data analysis solution developed by The IMC. The 177 musculoskeletal related terms derived through the primary and secondary refining processes were finally analyzed. As a result of the study, the frequent term was 'Metastasis', the clinical findings were 'Metastasis', the symptoms were 'Weakness', the diagnosis was 'Hepatitis', the treatment was 'Remove', and the body structure was 'Spine' in the analysis results for each medical terminology system. 'Oxycodone' was used the most. Based on these results, we would like to suggest implications for the analysis, utilization, and management of unstructured medical data.

Key Words : Data base(D/B), Medical record, Medical terminology, Keyword network analysis

요 약 본 연구는 근골격계 질환으로 입원한 환자의 의무기록지 키워드 네트워크 분석을 통해 근골격계와 관련된 표준 의료용어를 유추하여 보건의료현장의 비정형화된 데이터 활용 방안을 제시하기 위함이다. 분석 대상은 2010년부터 2019년까지 근골격계 질환 환자의 입퇴원요약지 145부로, 더아이엠씨(The IMC)에서 개발한 빅데이터 분석 솔루션인 TEXTOM을 활용하여 분석하였다. 1차·2차 정제과정을 통해 도출된 177개의 근골격계 관련 용어를 최종 분석하였다. 연구결과 다빈도 용어는 ‘Metastasis’, 의료용어 체계별 분석 결과에서 임상소견은 ‘Metastasis’, 증상은 ‘Weakness’, 진단은 ‘Hepatitis’, 처치는 ‘Remove’, 신체구조는 ‘Spine’, 약물은 ‘Oxycodone’이 가장 많이 사용되었다. 이러한 결과를 바탕으로 정형화되지 않은 의료데이터의 분석과 활용 및 관리 방안에 대한 시사점을 제안하고자 한다.

주제어 : 데이터베이스, 의무기록지. 의료용어, 키워드네트워크 분석

*This research was supported by Korea Evaluation Institute of Industrial Technology.(No. 20002519).
*Corresponding Author : Eun-A Choi(jijin05@daedong..ac.kr)
Received November 26, 2021   Revised December 26, 2021
Accepted May 20, 2022   Published May 28, 2022

1. INTRODUCTION

1.1 The necessity of research

Along with the 4th Industrial Revolution, many new information is being created in various fields with the development of smartphones, mobile Internet, and social media. Economic value of these information is rising because these information makes a big difference in production, consumption, and distribution systems, which can provide important value in health care. Health-related data produced in large quantities in health care can create valuable data related to health care by being collected and big data analysis[1]. In addition, these health care data will continue to develop in the future to provide medical services tailored to individual diverse needs along with the Fourth Industrial Revolution.

Big data refers to data of a size that exceeds the acceptance limits of commonly used data collection, management, and processing software[2]. Such big data analysis includes technologies to collect, analyze, and utilize, and data is collected from the public and private sectors for analysis of health care in domestic. Big data collected in the public sector consists of genetic data, administrative data related to insurance claims, and survey data related to hospitalization, death, and disease[3]. Big data collected from the private sector consists of data collected through smartphones, mobile Internet, and social media, and data collected by medical institutions during patient treatment. Due to the rapid development of IT technology, data with a considerable level of size and diversity are collected and utilized in the public sector, so public medical big data is well utilized, but due to the nature of medical care, there is a limit to data in the public sector alone. In particular, the vast amount of medical data held by medical institutions is significant in terms of patient clinic and treatment, but there are still various problems for meaningful clinical data to be linked and utilized nationwide[4, 5].

In particular, although medical records and discharge records obtained while treating individual patients have a lot of information about the medical situation, there are not any standardization of medical terms that record them, make difficulties in converting them into data. If medical terminology related to medical care is standardized, it will be of great help to patient care by utilizing such big data along with big data analysis to enable epidemiological analysis, efficient treatment improvement, reduction of medical errors, and high decision support.[6]. To this end, many standard terms have been developed at home and abroad, and ISO EN 13606 Part 1 EHR interoperability reference model[7] and HL7 clinical document architecture (CDA)[8] for standardization of expression of clinical problems have been developed. However, this only provides the structural framework of the data to be converted into data on the clinical record and does not define the meaning of the data [9]. Therefore, it will be necessary to standardize medical terminology using big data for health care to improve the quality of treatment and to facilitate research related to health care. Recently, as one of these standardization methods, there is a growing demand to establish standard terms centering on medical terms actually used in clinical practice.

Although there are various methods of collecting and analyzing large-scale unstructured data such as health and medical big data, there have been no studies that analyze medical records stored in medical institutions' DBs due to the nature of unstructured data. As an alternative to this, unstructured data used by medical institutions can be similarly collected through smartphones, mobile internet, social media, etc. However, most of the patients with musculoskeletal disorders are the elderly, and more than 90% of the elderly have musculoskeletal disorders that complain of pain[10]. Since most of the patients are elderly, it is difficult to collect information using IT when collecting data for informatization. Therefore, analyzing medical records written by medical personnel will be a very meaningful method of collecting information. There are several types of medical records, such as admission and discharge records, discharge summary papers, progress records, surgery records, consultation diagnosis records, nursing records, and histopathological examination reports. However, it includes many aspects of treatment and nursing. Since it includes data on the diagnosis, chief complaint, past medical history, operation name, treatment process, test results, and prescription drugs, analysis of the admission and discharge summary sheet provides information about the patient's process from hospitalization to discharge. Therefore, this study intends to utilize data stored in the D/B of medical institutions to collect data in the private sector related to the musculoskeletal system. Medical records of medical institutions are composed of unstructured text, and keyword network analysis is attracting attention as a method to analyze big data composed of texts[11, 12]. The keyword network analysis method makes it possible to derive various meanings contained in keywords by examining the relationships between keywords in large-scale unstructured data[13-15].

Therefore, through keyword network analysis of the admission and discharge summary sheet D/B of patients hospitalized for musculoskeletal disorders in P hospital located in B city, this study predicts standard medical terminology related to musculoskeletal disorders by examining medical terms frequently used in clinical settings related to musculoskeletal disorders. Even if there is a huge amount of data in the D/B of medical institutions, it is difficult to analyze and use the data for research purposes due to the unstructured characteristics of medical records. Therefore, this study analyzed medical record sheets using big data for the first time in history to suggest a plan to utilize unstructured medical record sheets in medical institutions. In addition, it is expected that the results of this study will help to reduce the mixed use of unstructured terms occurring in the health care field and medical research fields by predicting and suggesting standard medical terms. Although many efforts are being made at home and abroad to standardize medical terminology, it has not been standardized yet. Therefore, this study intends to serve as basic data so that unstructured data can be used in various ways for the development of medical care.

1.2 Securing research ethics

Since this study analyzed the medical records of patients at a university hospital, it is important to secure research ethics for the protection of patients' personal information. In this study, personal information such as the patient's name, gender, age, and registration number were not used in any analysis.

2. SYSTEM MODEL AND METHODS

2.1 Research design

This study examines frequently used medical terms related to musculoskeletal disorders in clinical settings, and uses atypical data to predict standard medical terms for musculoskeletal disorders using the D/B Keyword network analysis, one of the data analysis methods, was performed [16, 17], and proposed the method to extract field-oriented musculoskeletal standard terms by comparing results with information provided by the medical search engine (KMLE), MeSH and SNOMED CT.

2.2 Data collection target

The data used in this study were collected from the admission and discharge summary sheet D/B of patients with musculoskeletal disease at P Hospital located in city B, and only the data that agreed to participate in the study among the admission and discharge summary sheets focusing on the keyword 'musculoskeletal system' was used as an analysis target. The collection period is 10 years from 2010 to 2019. The collected data are medium-sized data suitable for analysis[18] with 1 case in 2010, 5 cases in 2011, 18 cases in 2012, 13 cases in 2013, 15 cases in 2014, 18 cases in 2015, 20 cases in 2016, 18 cases in 2017, 22 cases in 2018, and 15 cases in 2019(total=145).

2.3 Analysis tool

TEXTOM, a big data analysis solution service developed by The IMC, was used to extract and refine medical terms from the admission and discharge summary paper D/B and derive medical terms most frequently used in the admission and discharge summary record of patient with musculoskeletal disorders[19].

2.4 Data processing

Data purification was conducted for clear interpretation and understanding of the data used in this study [20], and in order to secure content validity, three clinical experts (one nurse, one doctor, one medical recorder) and one big data expert reviewed it. For the primary refinement, text mining was performed centering only on noun-type English keywords in the collected raw data[21], and the EspressoK method was adopted to reflect the proper nouns and complex nouns in the result value. In addition, by using the 'edit right away' function, keywords with low relevance to medical terms related to the musculoskeletal system, keywords with one syllable, and keywords that are difficult to give meaning such as symbols, Chinese characters, Korean characters, and numbers were removed.[22]. In the secondary refinement, the keywords separated by the MeCab method were modified into one keyword, and synonyms and similar words were classified as representative keywords, and keywords with a buzz volume of 10 or less were removed[23].

2.5 Data processing and analysis

In this study, in order to examine the medical terms frequently used in relation to musculoskeletal disorders in clinical settings, raw data were collected from the admission and discharge summary sheet D/B of patients with musculoskeletal disorders and primary and secondary purification were performed[18]. The detailed data analysis procedure is as follows. First, frequency analysis was conducted based on the data after data purification, and a 177×177 1-mode matrix dataset was created based on the 177 keywords with the highest frequency[24]. Second, keyword network analysis was conducted to examine the relationship between musculoskeletal diseases and related keywords, and to understand the structural properties of the network, we focused on nodes, average connection strength, density, number of components, average connection distance, diameter, and network center[11]. (At this time) Bootstrapping verification was performed to verify the statistical significance of the network[12]. To examine the centrality of keywords, connection centrality (NrmDgree), closeness centrality (Closeness), intermediary centrality (Betweenness), and position centrality (Eigenvector) were performed. Third, in order to examine the actual usage of musculoskeletal-related medical terms according to the medical term system, the medical term system was classified into clinical findings, symptoms, diagnosis, treatment, body structure, and drugs, and frequency analysis and network analysis were conducted for each sub-network. Fourth, an ego network analysis was conducted to examine the relationship between keywords for each medical terminology system. We used word cloud and NetDraw for network visualization[25].

3. RESULTS

3.1 Frequency analysis

Based on the big data provided by the admission and discharge summary record D/B of patients with musculoskeletal disorders, 177 medical terms were extracted in the data purification process for musculoskeletal-related medical terms. The word cloud and frequency analysis results are presented in Fig. 1, Table 1, focusing top 30 keywords in frequency among the selected medical terms.

Fig. 1. A word cloud for musculoskeletal related medical terms(top 30 keywords)

Table 1. Frequency analysis of medical terms related to musculoskeletal system(top 30 keywords)

No	Keyword	N	No	Keyword	N
1	Metastasis	251	16	Insertion	47
2	Hepatitis	106	16	Change	47
3	Cancer	102	18	Control	44
4	Spine	71	18	Acetaminophen	44
5	Removal	70	20	Breast	42
6	State	68	21	Stenosis	41
7	Past History	67	21	Tramadol	41
8	Fusion	66	21	Allergy	41
9	Abdomen	60	21	Carcinoma	41
10	Extension	58	25	Disease	40
11	Lesion	55	26	Flexion	37
12	Oxycodone	51	26	Prostate	37
13	Laminectomy	50	26	Targin	37
14	Compression	49	29	Evaluation	36
15	Weakness	49	29	Medication	36

As a result of frequency analysis of musculoskeletal-related medical terms, ‘Metastasis (251)’ was the highest frequency, followed by ‘Heptasis (106)’, ‘Cancer (102)’, ‘Spine (71)’ and ‘Removal(70)’.

3.2 Overall network analysis

3.2.1 Macro-level overall network analysis

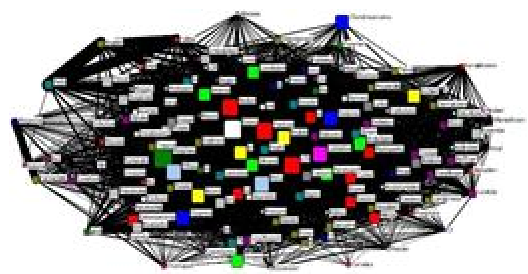

Fig. 2. The overall network structure for musculoskeletal related medical terms(top 177 keywords)

As a result of examining the macro-level overall network attributes for musculoskeletal-related medical terms, nodes were 177 and average connection strength was 61.706, and density was 61.706.351, the number of components was 41, the average connection distance was 1.414, the diameter was 2, and the network center was 1.871%.

3.2.2 Statistical significance test of the entire network

For the statistical significance of the entire network of musculoskeletal-related medical terms, the average sample distribution (M) of the data was 1.8650, and the standard error (SE) was

0.1378 and Z=13.4762. This shows that the probability of the total network statistics appearing larger than |Z| is 0.0002, which rejects the zero hypothesis at the 95% confidence interval, so it was judged appropriate to interpret the analysis results with the entire network used in this study.

3.2.3 Micro-level overall network analysis

Centrality analysis was conducted to examine the overall network properties of micro-level medical terms related to musculoskeletal system, and the results of centrality analysis centering on the top 10 medical terms are presented in Table 2.

Table 2. The centrality analysis result(top 10 keywords)

No	NrmDgree		No	Closeness	
1	Extent	2.589	1	Extent	.753
2	Folfiri	2.340	2	Fracture	.747
3	Avastin	2.072	3	c79	.733
4	Nodule	1.927	4	Ultracet	.724
5	Pedicle	1.789	5	Lafutidine	.719
6	Conclusion	1.774	6	ER	.713
7	Fracture	1.756	7	CT	.710
8	MI	1.742	8	Calcium	.707
9	Folfiriq	1.738	9	Fixate	.705
10	Crease	1.676	10	Interval	.699
No	nBetweeness		No	Eigenvector	
1	Extent	.760	1	Folfiri	.453
2	Frature	.670	2	Avastin	.430
3	CT	.552	3	Folfiriq	.382
4	Ultracet	.476	4	Extent	.184
5	Calcium	.465	5	HR	.168
6	Lafutidine	.447	6	L1	.165
7	Crease	.443	7	mcg	.156
8	c79	.442	8	Crease	.146
9	Interval	.433	9	PETt	.139
10	Uptake	.413	10	Interval	.117

As a result of the connection centrality analysis, the keyword with the most relationship among keywords was Extent(2.589), followed by Folfiri(2.340), Avastin(2.072), Nodule (1.927), Pedicle (1.789), and Conclusion (1.774). Fracture(1.756), MI(1.742), Folfiriq(1.738), Crease (1.676).

As a result of proximity centrality analysis, the keyword that forms the closest relationship with other keywords was Extent (.753), followed by Fracture (.747), c79 (.733), Ultracet (.724), and Lafutidine (.719), ER (.713), CT (.710), Calcium (.707), Fixate (.705) and Interval (.699).

As a result of mediation centrality analysis, the keyword that plays the largest mediator role in the relationship between keywords was Extent(.760), followed by Frature (.670), CT(.552), Ultracet(.476), and Calcium(.465), Lafutidine (.447), Crease (.443), c79 (.442), Interval (.433), and Uptake (.413)

As a result of prestige centrality analysis, the keyword most closely related to the keyword with high centrality was Folfiri (.453), followed by Avastin (.430), Folfiriq (.382), Extent (.184), and HR (.168).), L1 (.165), MCG (.156), Crease (.146), PET (.139) and Interval (.117).

3.3 Frequency analysis by medical terminology system

In order to examine the usage status of medical terms related to musculoskeletal system by medical term system, medical terms extracted from admission and discharge summary records of musculoskeletal patients based on big data provided by admission and discharge summary records D/B of musculoskeletal patients were classified according to medical term classification system of SNOMED CT. The medical term classification system includes clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimens. In this study, musculoskeletal-related medical terms were reclassified, focusing on clinical findings, symptoms, diagnoses, procedures, body structures, and pharmaceuticals. In this case, medical terms were classified based on 104 keywords excluding 73 classified as other, and the frequency analysis results for each sub-network are as follows.

3.3.1 Frequency analysis of medical terms related to clinical findings

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 13 medical terms related to clinical findings were presented in in Fig.3, Table 3.

Fig. 3. A word cloud for clinical findings related medical terms(top 13 keywords)

Table 3. Frequency analysis of medical terms related to clinical findings(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Metastasis	251	included
2	Compression	49	included
3	Stenosis	41	included
4	Allergy	41	included
5	Fracture	29	included

As a result of frequency analysis of clinical findings-related medical terms, ‘Metastasis(251)’ showed the highest frequency, followed by ‘Compression(49)’, ‘Stenosis (41)’, ‘Allergy (41)’ and ‘Fracture (29)’.

3.3.2 Frequency analysis of medical terms related to symptoms

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 6 medical terms related to symptoms were presented in Fig. 4, Table 4.

Fig. 4. A word cloud for symptoms related medical terms(top 6 keywords)

Table 4. Frequency analysis of medical terms related to symptoms(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Weakness	49	included
2	Dyspnea	25	included
3	Headache	20	included
4	Tenderness	15	included
5	Fever	15	included

As a result of frequency analysis of symptom-related medical terms, 'weakness (49)' showed the highest frequency, followed by 'dyspnea (25), 'headache (20), 'tenderness (20) and 'fever (15)'.

3.3.3 Frequency analysis of medical terms related to diagnosis

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 19 medical terms related to diagnosis were presented in Fig. 5, Table 5.

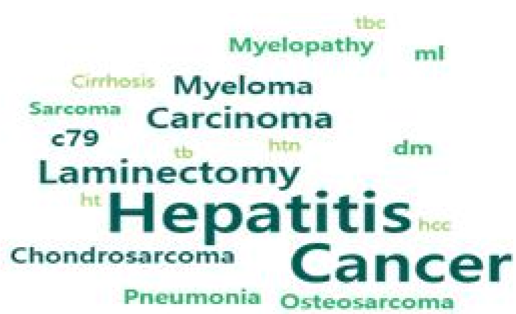

Fig. 5. A word cloud for diagnose related medical terms(top 19 keywords)

Table 5. Frequency analysis of medical terms related to diagnoses(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Hepatitis	106	included
2	Cancer	102	included
3	Laminectomy	50	included
4	Carcinoma	41	included
5	Myeloma	33	included

As a result of frequency analysis for diagnostic-related medical terms, 'Hepatitis (106)' showed the highest frequency, followed by 'Cancer (102)', 'Laminectomy (50)', 'Carcinoma (41)', and 'Myeloma (33)'.

3.3.4 Frequency analysis of medical terms related to procedures

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 6 medical terms related to procedures were presented in Fig. 6, Table 6.

Fig. 6. A word cloud for procedures related medical terms(top 6 keywords)

Table 6. Frequency analysis of medical terms related to procedures(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Remove	70	included
2	Extension	58	included
3	Insertion	47	included
4	Operate	30	included
5	Biopsy	29	included

As a result of frequency analysis of procedure-related medical terms, "Remove (70)" showed the highest frequency, followed by "Extension (58), "Insertion (47), "Operate (30)" and "Biopsy (29).

3.3.5 Frequency analysis of medical terms related to body structures

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 21 medical terms related to body structures were presented in Fig. 7, Table 7.

Fig. 7. A word cloud for body structures related medical terms(top 21 keywords)

Table 7. Frequency analysis of medical terms related to body structures(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Spine	71	included
2	Abdomen	60	included
3	Breast	42	included
4	Prostate	37	included
5	Artery	24	included

As a result of frequency analysis of medical terms related to body structure, ‘Spine (71)’ showed the highest frequency, followed by ‘Abdomen(60)’, ‘Breast(42)’, ‘Prostate(37)’ and ‘Artery(24)’.

3.3.6 Frequency analysis of medical terms related to pharmaceuticals

Among the 73 medical terms related to the musculoskeletal system, word cloud and frequency analysis results focusing on 39 medical terms related to pharmaceuticals were presented in Fig. 8, Table 8.

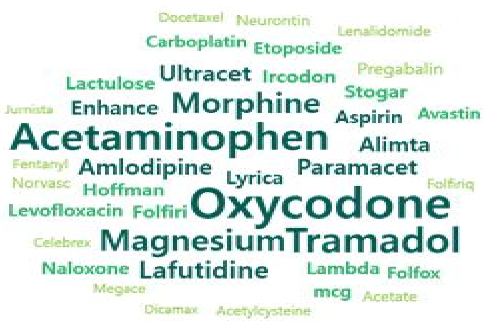

Fig. 8. A word cloud for pharmaceuticals related medical terms(top 39 keywords)

Table 8. Frequency analysis of medical terms related to pharmaceuticals(top 5 keywords)

No	Admission & Discharge Summary Record	N	KMLE/ MeSH/ SNOMED CT
1	Oxycodone	51	included
2	Acetaminophen	44	included
3	Tramadol	41	included
4	Magnesium	33	included
5	Morphine	32	included

As a result of frequency analysis of medical terms related to body structure, ‘Oxycodone(51)’ showed the highest frequency, followed by ‘Acetaminophen(44)’, ‘Tramadol(42)’, ‘Magnesium(33)’, ‘Morphine(32)’.

3.4 Subnetwork Analysis

Based on the frequency analysis results of medical terms related to the musculoskeletal system by medical terminology system, the ego network analysis was conducted for each system, and the results were presented as follows.

3.4.1 Analysis of the ego network for clinical findings

Results of ego network analysis focusing on medical terms related to clinical findings were presented in Fig. 9, Table 9.

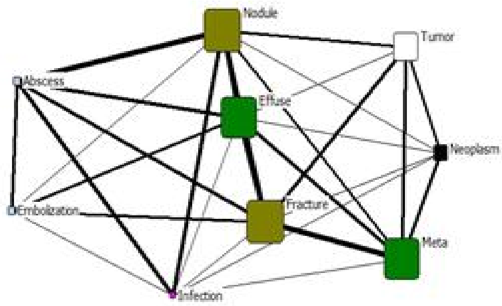

Fig. 9. The ego network structure for clinical findings related medical terms

As a result of examining the connection strength focusing on clinical findings-related medical terms, the connection strength of 'Fracture' was the strongest, and as a result of analyzing the centrality centered on 'Fracture', the highest connection strength with 'Nodule (13)', followed by ‘Meta(11)’, ‘Effuse(9)’, ‘Tumor(7)’, ‘Abscess(6)’, ‘Embolization(4)’, ‘Neoplasm(1)’ and ‘Infection(1)’.

3.4.2 Analysis of the ego network for symptoms

Ego network analysis results focusing on medical terms related to symptoms were presented in Fig. 10.

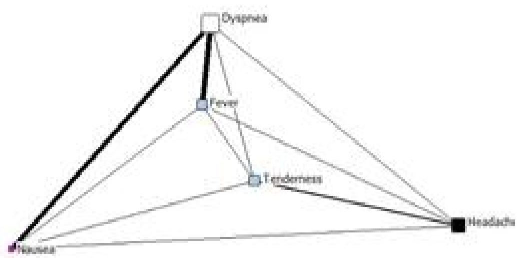

Fig. 10. The ego network structure for symptoms related medical terms

As a result of examining the connection strength focusing on symptom-related medical terms, the connection strength of ‘Dyspnea’ was the strongest, and as a result of analyzing the centrality centered on ‘Dyspnea’ had the highest connection strength with ‘Fever(15)’, followed by ‘Nausea(14)’, ‘Tenderness(4)’ and ‘Headache(3)’.

3.4.3 Analysis of the ego network for diagnosis

Ego network analysis results focusing on medical terms related to diagnosis were presented in Fig. 11.

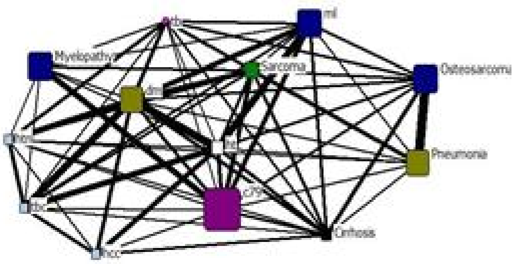

Fig. 11. The ego network structure for diagnoses related medical terms

As a result of examining the connection strength focusing on diagnose-related medical terms, the connection strength of 'c79' was the strongest, and as a result of analyzing the centrality centered on 'c79' had the highest connection strength with 'Myelopathy(11)', followed by 'ml(9)', 'Sarcoma(9)', 'ht(7)', 'dm(5)', 'Osteosarcoma(3)', 'Cirrhosis(3)', 'Pneumonia(2)', 'tbv(2)', 'tb(2)', 'hcc(1)' and 'htn(1)'.

3.4.4 Analysis of the ego network for procedure

Ego network analysis results focusing on medical terms related to procedure were presented in Fig. 12.

Fig. 12. The ego network structure for procedures related medical terms

As a result of examining the connection strength focusing on procedure-related medical terms, the connection strength of 'Biopsy' was the strongest, and as a result of analyzing the centrality centered on 'Biopsy', it was found to be closely related to 'Corpectomy(2)'.

3.4.5 Analysis of the ego network for body structure

Ego network analysis results focusing on medical terms related to body structure were presented in Fig. 13.

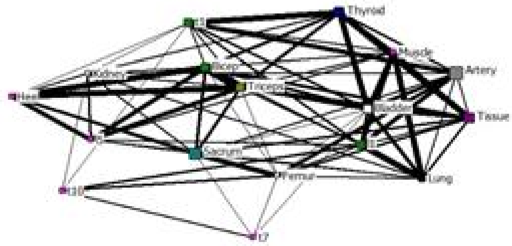

Fig. 13. The ego network structure for body structures related medical terms

As a result of examining the connection strength focusing on medical terms related to body structure, the connection strength of 'Artery' was the strongest, and as a result of analyzing the centrality centered on 'Artery' had the highest connection strength with 'l1(8)' and 't1(8)', followed by 'Bladder(7)', 'Femur(7)', 'Thyroid(4)', 'Sacrum(3)', 'Tissue(2)', 'Lung(2)', 'Kidney(2)' and 'Muscle(1)'.

3.4.6 Analysis of the ego network for pharmaceuticals

Ego network analysis results focusing on medical terms related to pharmaceuticals were presented in Fig. 14.

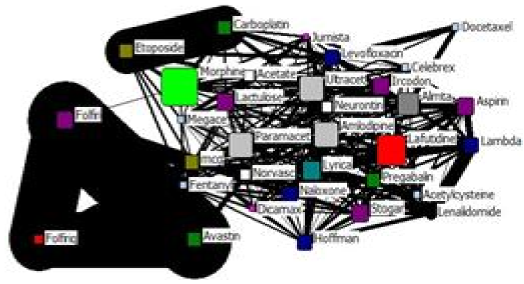

Fig. 14. The ego network structure for pharmaceuticals related medical terms

As a result of examining the connection strength focusing on medical terms related to pharmaceuticals, the connection strength of ‘Morphine’ was the strongest, and as a result of analyzing the centrality centered on ‘Morphine’ had the highest connection strength with ‘Folfox(17)’, followed by ‘Carboplatin(13)’, ‘Etoposide(12)’, ‘Ircodon(5)’, ‘Levofloxacin(5)’, ‘Naloxone(4)’, ‘Acetate(4)’, ‘Neurontin(4)’, ‘Celebrex(4)’, ‘Jurnista(4)’, ‘Amlodipine(3)’, ‘Aspirin(3)’, ‘Lactulose(3)’, ‘Fentanyl(2)’. ‘Megace(2)’, ‘Ultracet(1)’, ‘Paramacet(1)’, ‘Alimta(1)’, ‘Lyrica(1)’, ‘Folfiri(1)’, ‘Norvasc(1)’ and ‘Dicamax(1)’. .

4. DISCUSSION AND CONCLUSIONS

This study was conducted to predict standard medical terms related to musculoskeletal system by examining medical terms frequently used in relation to musculoskeletal diseases in clinical sites through keyword network analysis of D/B on admission and discharge summary of patients hospitalized with musculoskeletal diseases. Meanwhile, the possibility of analyzing health care big data was examined by using D/B on admission and discharge summary of musculoskeletal disorders at P Hospital located in B City. Based on the results shown in this study, the discussion is as follows.

First, as a result of frequency analysis of musculoskeletal-related medical terms, 'Metastasis(251)', 'Hepatitis(106)', 'Cancer(102)', 'Spine(71)', and 'Removal(70)' showed the highest frequency. It is thought that additional research is needed on the reason for the most frequent 'Metastasis' in patients hospitalized for musculoskeletal disorders, and it is also necessary to diversify data collection.

Second, as a result of the centrality analysis of musculoskeletal-related medical terms, various drugs such as ‘Extent’ and ‘Fracture’ were related, of which ‘Folfiri’ and ‘Avastin’ were mainly used as chemotherapy agents for cancer patients. These results also reflect the situation in which data collection is limited to university hospitals, mainly hospitalization for chemotherapy. Centrality analysis is a meaningful analysis method in network analysis, and in the future, it is considered necessary to subdivide it according to the purpose of research when analyzing data.

Third, in the results of frequency analysis by medical term system, the most commonly used word among clinical findings is 'Metastasis', which is 'the transfer of disease from one institution or part to another that is not directly connected to it'. In this case, metastasis of disease may be carried by pathogens such as Mycobacterium tuberculosis or by cells within the malignant tumor. Metastasis is a characteristic of all malignancies[26], meaning that patients linked to musculoskeletal diseases were often related to ‘Metastasis’. 'Weakness', 'Dyspnea', and 'Headache' were the most frequent major symptoms of patients with musculoskeletal disorders at the time of admission and discharge. The reason for these results is that the data collection institution is a university hospital, so it is presumed that there were more patients with complex disease or metastasis to diseases such as cancer than patients with simple musculoskeletal system. Therefore, different data collection institutions such as university hospitals, small and medium-sized hospitals, and private hospitals can produce different results, and these results can be said to suggest the direction of future research methods. According to the results of this study, 'Hepatitis' and 'Cancer' were found to be the most relevant diagnoses for patients with musculoskeletal system. It is considered that there has been no study on the relationship between patients with musculoskeletal system and 'Hepatitis' so far, so further research is needed. 'Oxycodone', 'Acetaminophen', and 'Tramadol' were found to be the most prescribed drugs at the time of discharge related to musculoskeletal disorders. These are drugs mainly used as analgesics, and it is presumed that these results were obtained because the analysis data used admission and discharge summary D/B.

Fourth, lastly, it is true that medical terms currently used in the medical field are difficult to utilize as research data, even though there is a vast amount of medical information, due to the musculoskeletal system-related medical terms that are not standardized not only in Korea but also worldwide. Therefore, this study was meaningful in that it proposed the necessity of converting medical information into data and the research method for it by analyzing the D/B big data of the musculoskeletal admission and discharge summary records of P hospital located in B city, converted into non-standardized medical terms

Based on these discussions, suggestions for follow-up research are as follows.

First, there is a limit to generalizing the research results in that the analysis target of this study is data limited to P hospital located in B city. Therefore, in the follow-up study, it will be necessary to expand in investigating the subjects focusing on tertiary general hospitals that serve as base hospitals across the country and musculoskeletal specialist hospitals.

Second, the medical terms currently used in the medical field were difficult to utilize as data for research despite a vast amount of medical information due to the musculoskeletal system-related medical terms that are not standardized not only in Korea but also worldwide. Therefore, it is expected that standardization studies on musculoskeletal related medical terms will be conducted in follow-up studies.

Third, various health care-related big data are currently provided through the public health and medical big data open system and public data portal. However, while standard health insurance-related glossaries such as health insurance glossaries are provided, dictionaries of various standard medical terms, including musculoskeletal-related terms, are not provided. Therefore, it is expected that a systematic medical term D/B management system will be developed and established for efficient management of standardized data in follow-up studies.

REFERENCES

[1] J. U. Kim, S. Y. Sin, H. S. Cha, & S. J. An. (2015). Smart health data based on smart medical technology standards. *Communications of the Korean Institute of Information Scientists and Engineers, 33(3),* 60-66.

[2] C. Snijders, U. Matzat, & U. D. Reips. (2012). "Big Data": Big gaps of knowledge in the field of Internet, *International Journal of Internet Science, 7(1),* 1-5.

[3] M. K. Kim, & Y. B. Cho. (2019). An Analysis of Impact on the Quality of Life for Chronic Patients based Big Data. *Journal of the Korea Institute of Information and Communication Engineering, 23(11),* 1351-1356. DOI : 10.6109/jkiice.2019.23.11.1351

[4] M. K. Kim, & Y. B. Cho. (2018). An Analysis of Factors Affecting Quality of Life through the Analysis of Public Health Big Data. *Journal of the Korea Institute of Information and Communication Engineering, 22(6),* 835-841. DOI : 10.6109/jkiice.2018.22.6.835

[5] M. H. Park, Y. B. Cho, S. Y. Kim, J. B. Park, & J. H. Park. (2018). Analysis of Factors for Korean Women's Cancer Screening through Hadoop-Based Public Medical Information Big Data Analysis. *Journal of the Korea Institute of Information and Communication Engineering, 22(10),* 1277-1286. DOI : 10.6109/jkiice.2018.22.10.1277

[6] B. K. Choi, E. A. Choi, & M. H. Nam. (2019). Development of Clinical Terminology System for Human Body: Convergence Research of SNOMED-CT. *Journal of Digital Convergence, 17(4),* 177-185. DOI : 10.14400/JDC.2019.17.4.177

[7] D. Kalra. (2006). Electronic health record standards. *Yearbook of medical informatics, 15(1),* 136-144.

[8] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. Shabo, & A. Shvo. (2006). HL7 Clinical Document Architecture Release 2. *Journal of the American Medical Informatics Association, 13(1),* 30-39.

[9] H. A. Park, H. Y. Kim, & Y. H. Min. (2012). Use of clinical terminology for semantic interoperability of electronic health records. *Journal of the Korean*

Medical Association, 55(8), 720-728.

[10] H. S. Jeong, & Y. J. Lee. (2018). Factors Affecting Quality of life in the Elderly with Chronic Musculoskeletal Pain. *Journal of Muscle and Joint Health, 25(2),* 92-103.
DOI : 10.5953/JMJH.2018.25.2.92

[11] B. M. Kim, & K. H. Lee. (2020). An exploration of a directions for preparing early childhood finance policy to promote common good based on complexity: Focusing on social discourse and competitive discourse analysis. *Journal of Learner-Centered Curriculum and Instruction, 20(10),* 407-438.
DOI : 10.22251/jlcci.2020.10.407

[12] K. H. Lee. (2019) An exploration of a direction to promote early childhood playfulness based on big data. *Journal of Educational Innovation Research, 29(3),* 257-289.
DOI: 10.22251/jlcci.2019.19.15.891

[13] H. I. Hwang, K. H. Lee, & H. S. Je. (2018). A network analysis of young children' self-esteem based on big data. *Korean Journal of Early Childhood Education, 38(4),* 213-235.

[14] W. S. Suh, K. H. Lee, & S. W. Kim. (2019). A keyword network analysis for early childhood education finance: Focus on big data. *The Journal of Learner-Centered Curriculum and Instruction, 19(12),* 97-124.
DOI : 10.22251/jlcci.2019.19.12.97

[15] H. I. Hwang, S. H. Kim, & K. H. Lee. (2018). A network analysis of mothers' happiness concepts based on big data. *Journal of Future Early Childhood Education, 25(2),* 45-66.
DOI: 10.22155/jfece.25.2.45.66

[16] B. M. Kim, J. J. Youn, S. J. Jeong, & K. H. Lee. (2018). Keyword network analysis for creativity and personality of early childhood teachers: Focus on big data. *Indian Journal of Public Health Research & Development, 9(11),* 850-856.
DOI:10.5958/0976-5506.2018.01566.8.

[17] B. M. Kim, & K. H. Lee. (2020). Keyword network analysis on the integrated research trends of early childhood education and childcare. *International Journal of Innovation Creativity and Change, 13(3),* 595-607.

[18] S. S. Lee. (2012). *Network analysis methods.* Seoul, Korea: Nonhyung.

[19] TEXTOM. Big data analysis solution [Internet]. Available: http://www.textom.co.kr.

[20] B. M. Kim, & K. H. Lee. (2021). A study on exploring directions for early childhood safety education based on big data: Focus on semantic network analysis. *Review of International Geographical Education Online, 11(2),* 636-647.
DOI : 10.33403/rigeo. 800461

[21] K. H. Lee, & H. S. Je. (2018). A keyword network analysis of young children's creativity and personality based on big data. *Journal of Future Early Childhood Education, 25(4),* 121-144.
DOI: 10.22155/jfece.25.4.121.144

[22] K. H. Lee, & H. I. Hwang. (2020). A discourse analysis of Nuri-curriculum evaluation based on big data. *Journal of Future Early Childhood Education, 27(4),* 77-109.
DOI: 10.22155/jfece.27.4.77.109

[23] K. H. Lee, & H. I. Hwang. (2020). A keyword network analysis of '2019 revised Nuri-curriculum evaluation' based on bis data. *The Journal of Korea Open Association for Early Childhood Education, 25(4),* 79-109.
DOI : 10.20437/KOAECE25-4-04

[24] S. P. Borgatti, M. G. Evertt, & J. C. Johnson. (2013) *Analyzing social networks.* Thousand Oaks, CA: Sage Publications.
DOI : 10.1080/10705422.2016.1209400

[25] L. C. Freeman. (1992). *Research methods in social network analysis.* New York, America: Routledge.
DOI : 10.4324/9781315128511

최 병 관(Choi, Byung Kwan) [정회원]

· 2003년 2월 : 부산대학교 의학박사
· 2007년 9월 ~ : 부산대학교 신경외과 교수
· 2012년 8월 ~ : 부산대학교 병원 의료정보 센터장
· 2015년 7월 ~ : 부산대학교 병원 해양의료연구 센터장
· 관심분야 : 신경외과, 병원의료 정보, 해양의료
· E-Mail : spine@pusan..ac.kr

최 은 아(Choi, Eun A) [정회원]

· 2016년 8월 : 고신대학교 간호학 박사
· 2015년 3월 ~ : 대동대학교 간호학부 부교수
· 관심분야 : 간호, 간호정보, 아동간호
· E-Mail : jijin05@daedong.ac.kr

남 문 희(Nam, Moon Hee) [정회원]

· 2010년 8월 : 인제대학교 보건학박사
· 2010년 8월 : 가야대학교
· 2014년 3월 ~ 현재 : 대동대학교 간호학부 부교수
· 관심분야 : 간호, 간호정보, 간호관리학
· E-Mail : ny386@hanmail.net