

## Chronologic trends of cancer-related lymph node research in PubMed: informetrics analysis

Hyo-Jung Oh<sup>1,9</sup>, Dongmin Yang<sup>2,9</sup>, Hyo-Won Oh<sup>3,9</sup>, Jae-Gyu Jeon<sup>4,9</sup>, Chonghyuck Kim<sup>5,9</sup>, Jeong-Yong Ahn<sup>6,9</sup>, Se Wung Han<sup>7</sup>, Chan-Young Kim<sup>7,8,9</sup>

<sup>1</sup>Department of Library and Information Science, Jeonbuk National University, Jeonju, Korea

<sup>2</sup>Graduate School of Archives and Records Management, Jeonbuk National University, Jeonju, Korea

<sup>3</sup>Department of Preventive and Public Health Dentistry, College of Dentistry, Wonkwang University, Iksan, Korea

<sup>4</sup>Department of Preventive Dentistry, School of Dentistry, BK21 Program, Jeonbuk National University, Jeonju, Korea

<sup>5</sup>Department of English Language and Literature, Jeonbuk National University, Jeonju, Korea

<sup>6</sup>Department of Statistics, Jeonbuk National University, Jeonju, Korea

<sup>7</sup>Department of Surgery, Jeonbuk National University Medical School, Jeonju, Korea

<sup>8</sup>Research Institute of Clinical Medicine of Jeonbuk National University, Jeonju, Korea

<sup>9</sup>Institute of Medical Information Convergence Research in JBNU, Jeonju, Korea

**Purpose:** Given the long history of investigation into cancer and its relevance to the lymph node (LN), it would be meaningful to plot the trends of research on cancer-related LN.

**Methods:** Queries such as “cancer,” “lymph node,” and “cancer and lymph node” were submitted to PubMed to collect articles on cancer and LN published between 1945 and 2017. The collected articles were then extracted by an automatic web crawler and examined through informetrics and linguistic analysis.

**Results:** The number of articles related to cancer was 2,795,476 and 127,897 articles (4.6%) were found to be relevant to LN. With regard to cancer types, breast cancer was the most studied (37%), followed by gastric cancer (17%). With regard to the subjects in which the surgeon is interested, LN metastasis (57%) was found to be the topic most discussed, followed by LN dissection (22%) and sentinel LN (17%). Publications on LN metastasis gradually increased over time from 1988 to 2017 although those on sentinel LN and LN dissection have stagnated since the early 2000s.

**Conclusion:** Although research on cancer was abundant, only a small portion was dedicated to investigating its relevance to LN. Western countries had led the research on cancer-related LN, but Asian countries began to participate as major players, expanding their contributions. While LN metastasis, one of the major cancer-related LN topics, showed a steady increase, those involved in oncologic surgery such as LN dissection and sentinel LN did not.

[Ann Surg Treat Res 2020;99(5):305-313]

**Key Words:** Informetrics, Lymph nodes, Neoplasms

### INTRODUCTION

Physiologic functions of lymphatics and lymph node (LN) include tissue homeostasis, fluid balance, immune function, absorption of dietary fat, and lipid transport. However, they

are of particular interest to oncologic surgeons because they are also involved in cancer progression. Metastatic initial events of the solid tumors are presumed to occur through the lymphatic system, though the role of metastatic LNs in cancer progression is not yet fully understood [1,2]. Whatever the role

Received April 29, 2020, Revised July 1, 2020, Accepted July 29, 2020

**Corresponding Author: Chan-Young Kim**

Department of Surgery, Jeonbuk National University Medical School, 20 Geonji-ro, Deokjin-gu, Jeonju 54907, Korea

Tel: +82-63-250-2298, Fax: +82-63-231-5298

E-mail: happyhill@jbnu.ac.kr

ORCID: <https://orcid.org/0000-0001-6656-9425>

Copyright © 2020, the Korean Surgical Society

© Annals of Surgical Treatment and Research is an Open Access Journal. All articles are 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.

is, surgeons have concentrated clinical efforts to remove and analyze regional LNs after Halsted performed radical operations involving LNs in breast cancer patients in 1894 [3]. Doubtlessly, research on LNs and their relevance to cancer must have evolved ever since, reflecting various clinical and nonclinical developments that have influenced researchers. However, to the best of our knowledge, nothing has been said about how the research on cancer-related LN has evolved over the years.

Lately, with the advancement of the internet and computerized information, biomedical databases such as PubMed, Scopus, Web of Science, and Google Scholar have been extensively exploited. As Wormell [4] noted, online databases are not only about accessing documents and finding facts but also about tracing trends and developments in society and scientific disciplines. Indeed, recent works by information scientists have analyzed and organized preexisting medical knowledge using informetrics and linguistic analysis proving their usefulness [5-7].

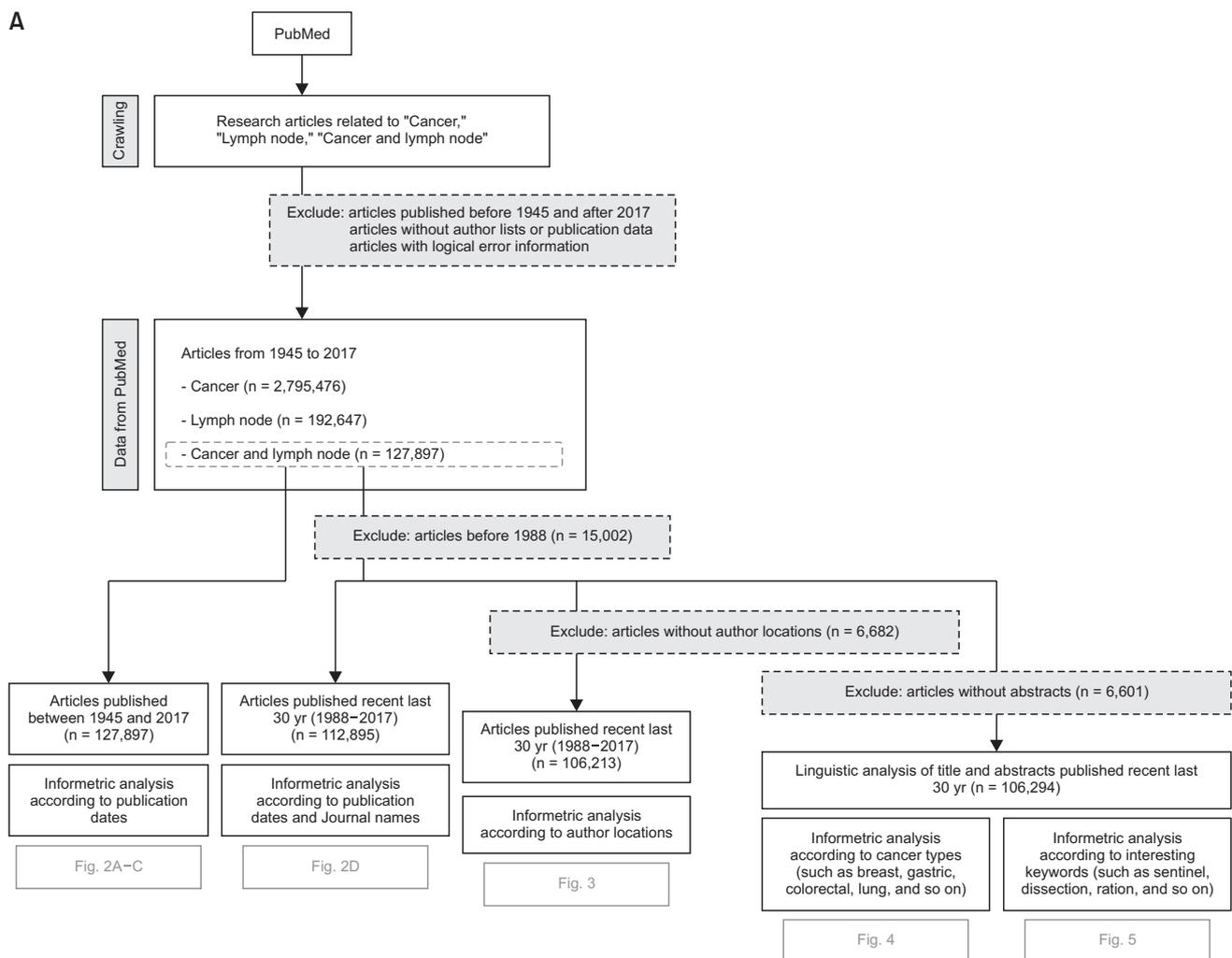
In this article, we attempt to trace the trends of research conducted over the past several decades on cancer-related LN by exploiting the techniques used in informetrics. The results obtained will not only fill the void that has been formed and left untouched since the connection between LN and cancer was made by Halsted, but also provide potentially useful information for researchers and clinicians.

## METHODS

This study was exempted from the ethics approval and written informed consent from the participants from the Institutional Review Board of Jeonbuk National University Hospital (No. 2019-05-067).

### Data source and query selection

Research articles related to cancer and LN published between 1945 and 2017 were searched in PubMed which is



**Fig. 1.** (A) It shows data collection and selection with exclusion criteria for informatics analysis. (B) It is an example showing a method using linguistic analysis.

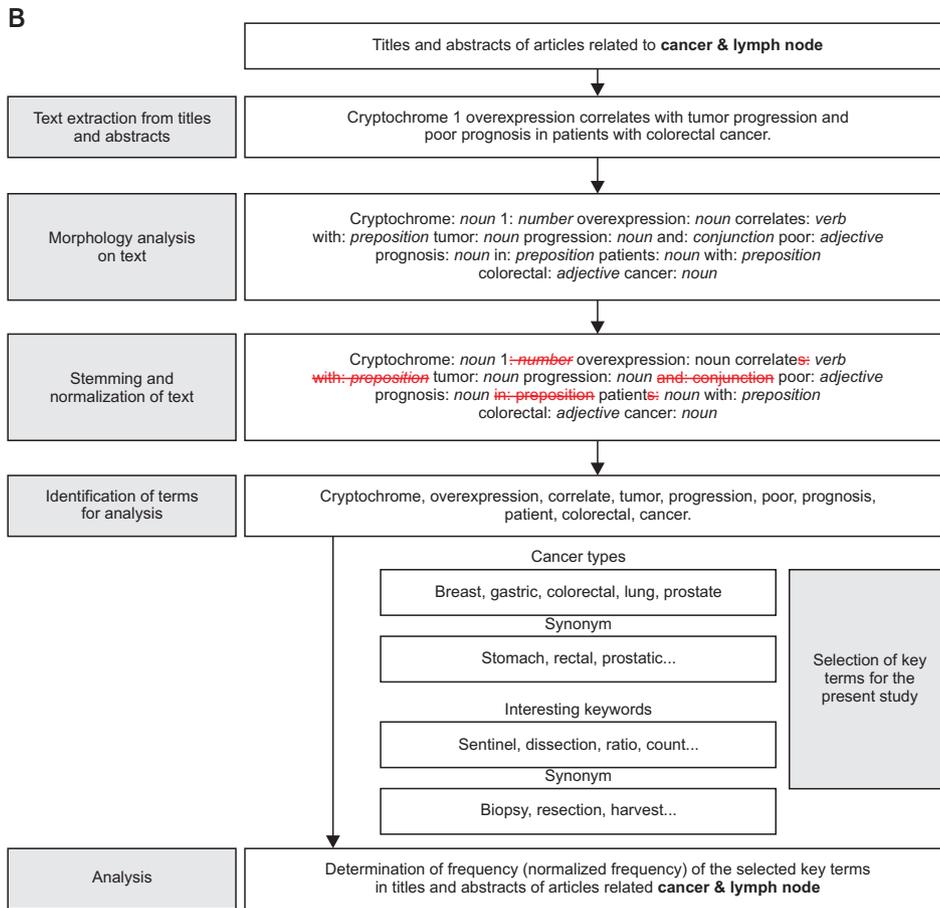


Fig. 1. Continued.

practical—easy to use, fast and useful—as well as authoritative [8]. To search for the articles, queries such as “cancer,” “lymph node,” and “cancer and lymph node” were selected. PubMed translation of each of the queries were as follows: (i) “cancer” PubMed translation: “neoplasms” [MeSH Terms] OR “neoplasms” [All Fields] OR “cancer” [All Fields] AND “English” [Language]; (ii) “lymph node” PubMed translation: “lymph nodes” [MeSH Terms] OR (“lymph” [All Fields] AND “nodes” [All Fields]) OR “lymph nodes” [All Fields] OR (“lymph” [All Fields] AND “node” [All Fields]) OR “lymph node” [All Fields] AND “English” [Language]; (iii) “cancer and lymph node” PubMed translation: (“neoplasms” [MeSH Terms] OR “neoplasms” [All Fields] OR “cancer” [All Fields]) AND (“lymph nodes” [MeSH Terms] OR (“lymph” [All Fields] AND “nodes” [All Fields]) OR “lymph nodes” [All Fields] OR (“lymph” [All Fields] AND “node” [All Fields]) OR “lymph node” [All Fields]) AND “English” [Language].

### Data collection and selection

Once articles on cancer-related LN were generated by PubMed based on our search queries, several pieces of information (their titles, abstracts, publication years, author affiliations, and publication journals) were extracted by an automatic web

crawler that we developed using Eclipse based on Java [9]. The extracted information was then arranged in Extensible Markup Language (XML) format and underwent a round of refinement. Some of the articles collected had to be removed as they were not suited for our analysis (Fig. 1A). The exclusion criteria were as follows: (1) articles published before 1945 or after 2017, those with no author lists or publication dates, and those with logical errors were all excluded; (2) articles published before 1988 were excluded as well when we analyzed their geographic distributions, published journals, and abstracts, because PubMed did not provide relevant information until 1986 (see below for further reasons). After the removal of the uninformative articles, the collected information (titles, abstracts, publication dates, author affiliations, and journal names) was ready for use.

### Informetric analysis

The information gathered (publication dates, author affiliations, and journal names) was informetrically analyzed. We investigated how much the number of articles and journals dedicated to cancer-related LN have grown over time. Along with this, we also plotted the trends that occurred in terms of journal rankings and geographic distribution of authors.

Authors' geographic distribution was determined using the affiliations of the first and corresponding authors. In cases where the affiliations of the first and corresponding authors were different in an article, each of them was counted as having 0.5 value.

### Linguistic analysis

In order to assess the direction of change of the collected articles in their content, we examined the occurrences and distributions of specific keywords in the titles and abstracts. Selected keywords to examine the articles on cancer-related LN are LN metastasis, sentinel LN, LN dissection, LN count, LN ratio, and LN micrometastasis. These keywords were reviewed and selected by 2 oncologic surgeons (CYK and SWH). As for cancer types, breast, gastric, lung, colorectal, gynecologic, prostate, head and neck, thyroid, esophageal, pancreatic, urogenital, liver, and skin were used.

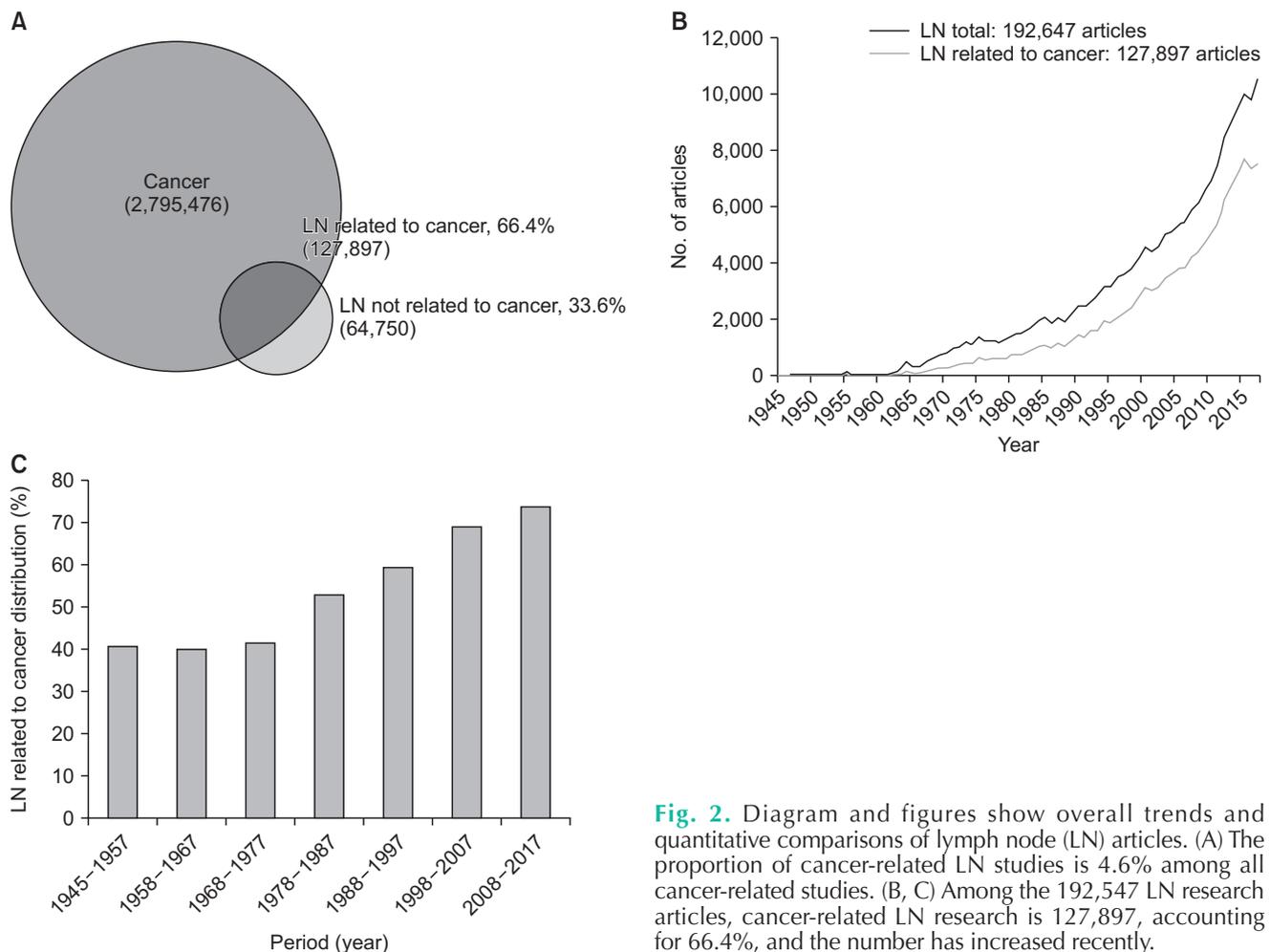
PubMed does not provide abstracts in many of their articles prior to 1986. Thus, only the titles and abstracts of articles published between 1986 and 2017 were eligible for our analysis. In fact, we excluded the articles published before 1988 from our

analysis, as it allows us to present the results of our analysis in decade cycles. Specific procedures we used to analyze the occurrences of the keywords are as follows, which we adopted from Oh et al. [9]: (1) text collection from titles and abstracts in XML format; (2) grammatical analysis of text using morphology; (3) exclusion of articles, prepositions, adverbs, and conjunctions from text; (4) extraction of meaningful terms from text; (5) selection of keywords of cancer-related LN; (6) identification of keywords in the terms extracted; and (7) determination of frequency (or normalized frequency) of occurrence of the selected keywords in the titles and abstracts of articles. A normalized frequency was calculated via the following equation: number of keywords divided by number of articles on cancer-related LN.

## RESULTS

### Chronological trends of articles

In PubMed, the total number of articles on cancer published between 1945 and 2017 was 2,795,476. Among them, 127,897 articles including LN accounted for 4.6% (Fig. 2A). The total



**Fig. 2.** Diagram and figures show overall trends and quantitative comparisons of lymph node (LN) articles. (A) The proportion of cancer-related LN studies is 4.6% among all cancer-related studies. (B, C) Among the 192,547 LN research articles, cancer-related LN research is 127,897, accounting for 66.4%, and the number has increased recently.

number of articles dealing with LN was 192,647. Hence, the majority of the articles concerned with LN were also concerned with cancer (66.4%). These articles dealing with both LN and cancer continued to grow over time, even more dramatically in recent years (Fig. 2B). They began to exceed the number of the articles dealing with just LN in the mid-1970s and have recently reached about 70% (Fig. 2C).

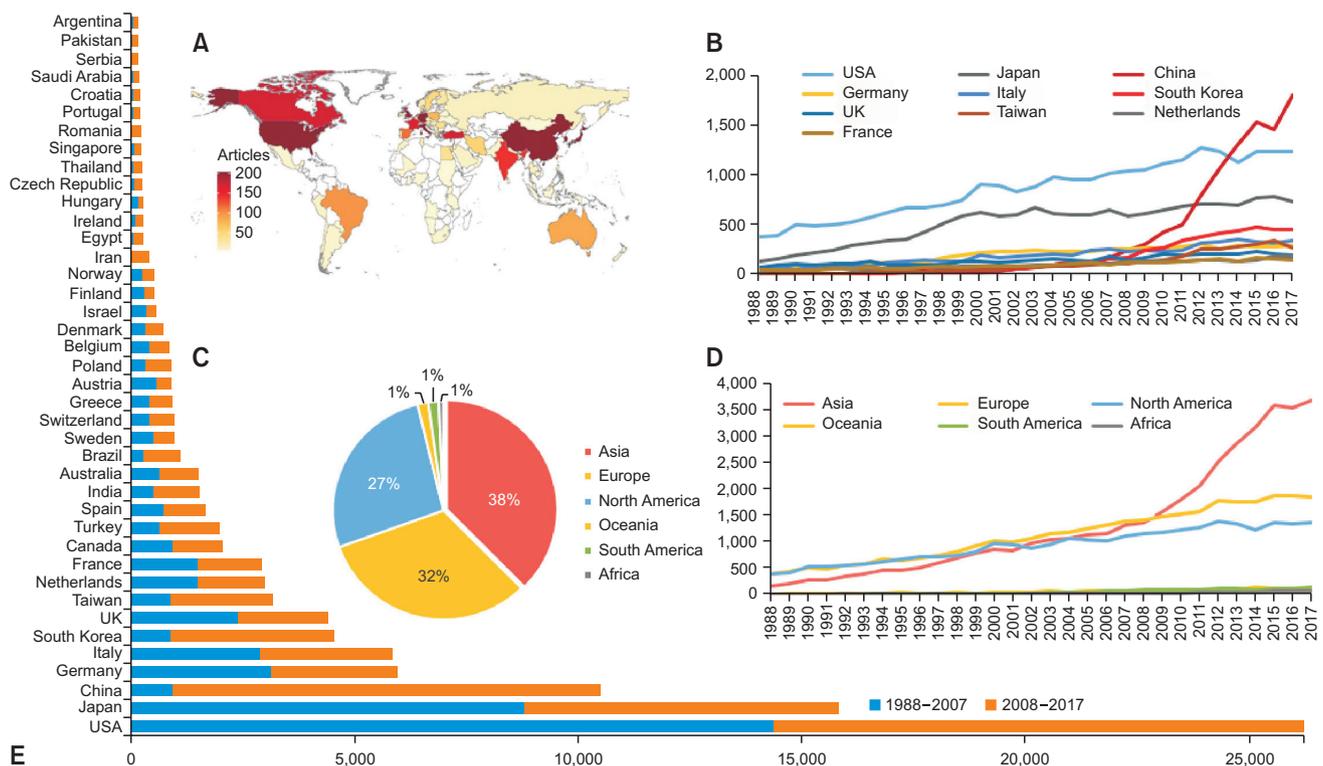
### Author distribution

Turning to the geographical distribution of the articles, the United States emerged as the leading country with 26,386 publications, followed by Japan with 15,918, China with 10,521, Germany with 6,003, Italy with 5,893, South Korea with 4,579, the United Kingdom with 4,450, Taiwan with 3,216, the Netherlands with 3,036, and France with 2,977. Of these top 10 countries, China showed a remarkable performance in recent years, pushing the number of articles from a mere 103 in 2004 to the astonishing number of 1,966 in 2017. In the same

period, the performance of the United States remained more or less the same, with 591 articles in 2004 and 665 articles in 2017. Following China, South Korea saw the second-highest growth rate, increasing the number from 94 articles in 2004 to 455 in 2017 (Fig. 3). Broadening the scope of comparison from individual nations to continents, it was found that the number of articles published in Asia was the highest at 39,933 (37.6%), which was followed by Europe at 33,957 (31.9%), North America at 28,335 (26.7%), South America at 1,666 (1.6%), Oceania 1,588 (1.5%) and Africa 746 (0.7%). The growth rates of Asia were notably high, with their contributions tripling in less than 15 years from 1,076 articles in 2004 to 3,690 in 2017.

### Trend of research on cancer type

Various types of cancer have been studied in relation to LN. Of them, breast cancer was found to be the subject of research that caught the most attention, appearing in 54,556 articles which are 37% of all articles. The second most popular subject



**Fig. 3.** World maps, diagram, and graph show the overall frequency, rate, and trends of the main author's country and continent that carried cancer-related lymph node (LN) articles. (A) Nations' distribution of articles with LN related to cancer. The top 10 nations are the United States, Japan, China, Germany, Italy, South Korea, the United Kingdom, Taiwan, the Netherlands, and France. (B) The trend of the top 10 nations' distribution of articles with LN related to cancer (1988–2017). It is confirmed that China has been publishing articles at a rapid pace since 10 years ago. South Korea is also growing. (C) Continental distribution of articles with LN related to cancer. Asia was 37.6% compared to 31.9% in Europe. (D) The trend of continental distribution of articles with LN related to cancer (1988–2017). Trends are increasing in Asia recently. (E) Distribution of the top 40 nations. The volume of publications in each of the 40 countries is compared to the volume published in the last decade. China's remarkable growth is also confirmed. The United States was the most popular with 26,386 articles, followed by Japan with 15,918 articles. Followed by 10,521 in China, 6,003 in Germany, 5,893 in Italy, 4,579 in South Korea, 4,450 in the United Kingdom, 3,216 in Taiwan, 3,036 in the Netherlands, and 2,977 in France.

of research was gastric cancer, which appeared in 24,290 articles (17%). Other topics of research listed in order of popularity are shown in Fig. 4.

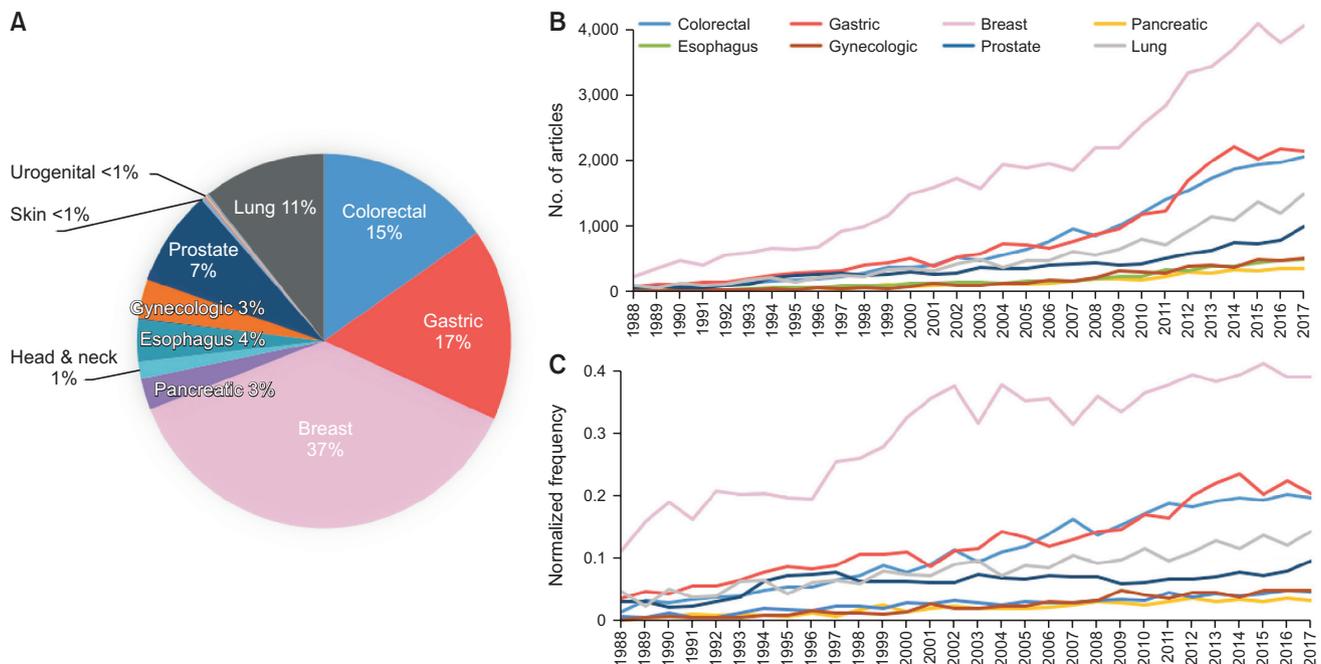
### Trends by keywords: LN metastasis, LN dissection, sentinel LN, LN ratio, LN count, and LN micrometastasis

LN metastasis was the topic discussed in the majority of the articles (57%) followed by LN dissection (22%), sentinel LN (17%), LN ratio (2%), LN count (2%), and LN micrometastasis (1%) (Fig. 5). LN metastasis received steadily increasing attention in the literature. While LN dissection and sentinel LN also appear to get increasing interest over time, their normalized frequency graph shows that the interest in sentinel LN increased from 1996 to 2004, but then declined since, and that the interest in LN dissection increased steadily until 2000, but has since been flat. On the other hand, LN ratio and LN count have not attracted much attention until now. Though recently, 100–200 articles have been published consistently. LN micrometastasis is not of interest and has appeared in less than 30 articles a year.

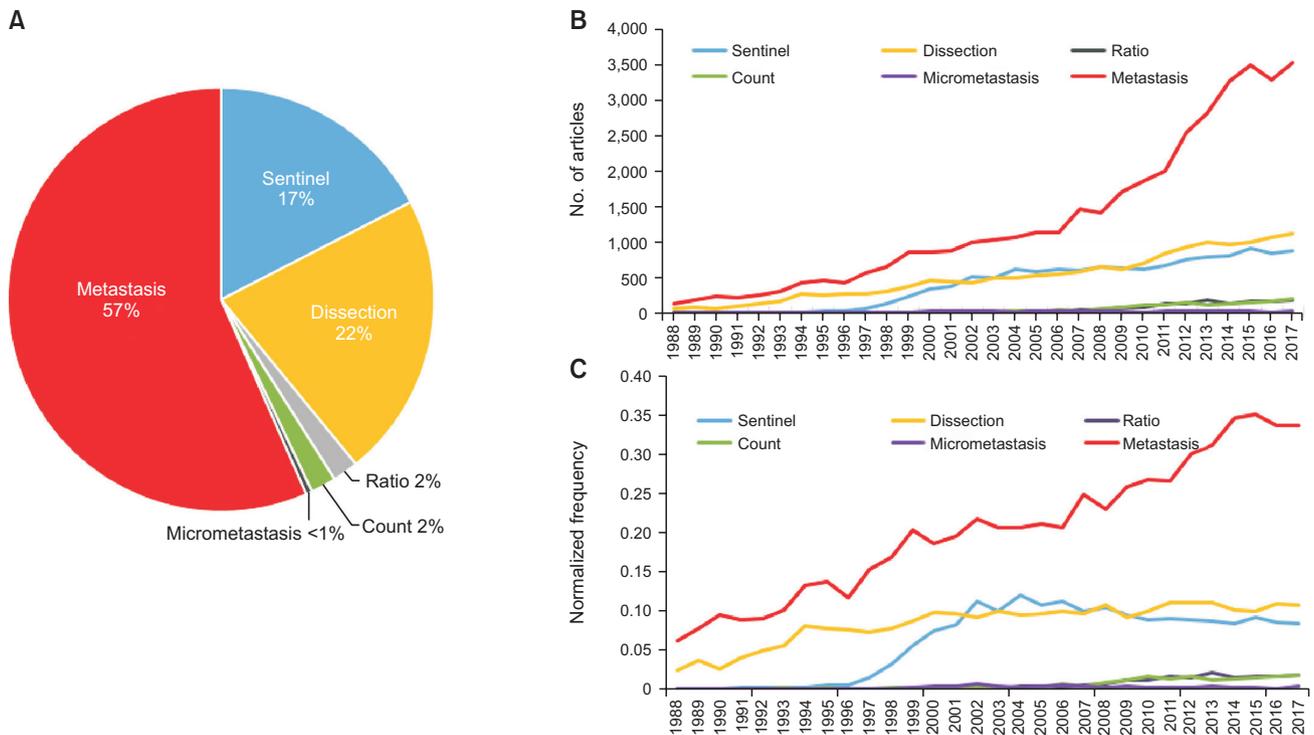
## DISCUSSION

As shown in Fig. 3, the United States produced the largest

number of publications on cancer-related LN, followed by Japan, China, the United Kingdom, and so on. However, if we consider the incremental trends of articles over time, 2 countries stood out; China and South Korea. Their contributions increased significantly since 2007, compared to other countries. Especially, China showed a remarkable growth, publishing more articles than the United States since 2014. This is likely due to China's dramatic increase in investment in research and development (R&D). According to the latest available data on investment in R&D for Organisation for Economic Co-operation and Development (OECD) countries and other major economies published in the OECD's Main Science and Technology Indicators [10], gross domestic expenditure on R&D in China rose by 291% in 2015 compared to 2007. Likewise, South Korea's 180% increase in expenditure on R&D in the same period coincides with the rise in its contribution to the research on cancer-related LN. A rise in the expenditure on R&D appears to be a necessary condition for a rise in publications but not a sufficient condition. This means, in other words, that without a rise in the expenditure on R&D, there would be no rise in publications. This is indeed what happened to the United States. According to the National Cancer Institute report, they reduced funding for research on breast, lung, prostate, and colorectal cancer during the years of 2011–2016 [11]. This cut



**Fig. 4.** Diagrams and graphs show the quantitative proportions and trends of published articles by cancer type. (A) Breast cancer was the most studied cancer type. The total was 54,556, accounting for 37%. Followed by gastric cancer at 24,290 (17%) and colorectal cancer 22,251 at 15%. In order, 15,573 (11%) of lung cancer, 10,932 (7%) of prostate cancer, 5,235 (4%) of esophagus cancer, 4,998 (3%) of gynecologic cancer and 3,975 (3%) of pancreatic cancer. (B) It shows trends by cancer type. Breast cancer has been the most abundant since 1980 and has steadily increased to the present, followed by stomach cancer, colorectal cancer, and lung cancer. (C) A graph that standardizes the amount of published journals to normalized frequency. According to this, breast cancer increased steadily to a normalized frequency of 0.38 until 2002, but then slowly increased to 0.39 by 2017. Gastric cancer, colorectal cancer, and lung cancer have slowly increased until now.



**Fig. 5.** Diagrams and graphs show the proportions and trends of quantities by selectively extracting topics that surgeons may be interested in. The topics selected by surgeon’s opinion are lymph node (LN) metastasis, sentinel LN, LN dissection, LN count, LN ratio, and LN micrometastasis. (A) The cases of 57% were related to LN metastasis, followed by dissection 22%. The sentinel LN accounts for 17% of the topics, but the LN ratio, count, and micrometastasis did not account for much. (B, C) Comparing the amount of articles with the normalized frequency, the LN metastasis is a subject of constant interest, while the dissection is steady without a large increase. In contrast, the sentinel LN showed a slight decline after mid-2000.

in funding directly corresponds to the line, which represents the United States in the graph of Fig. 3B, starting to flatten or decline in the early 2010s.

In addition to outlining the rise and fall of research articles in number with regard to variables such as time, place, and publisher, we investigated how research on cancer-related LN has evolved in their content by tracing the occurrences and distributions of 6 keywords; metastasis, dissection, sentinel LN, LN ratio, LN count, and micrometastasis. As is well known in the literature, it has been always at the center of attention in the literature to understand the process of LN metastasis and to identify the implications of LN dissection. Hence, there are the 2 keywords-metastasis and dissection. Sentinel LN is the first LN drained from a primary carcinoma [12,13], and it is known to have 2 roles, "incubator" and "marker," in cancer metastasis and progression [14]. Even though the 3 keywords of LN ratio, LN count, and micrometastasis are less pervasive terms than the other keywords in the literature, they appear in researches to complement the Union for International Cancer Control TNM stage [15] which is widely used for prognosis prediction.

The amount of articles on LN metastasis was found to be the greatest and steadily increasing over time. The increasing slope of the cancer-related LN articles in Fig. 2B shows a pattern

similar to the slope of LN metastasis. Because LN metastasis is the basis and core of cancer-related research, related articles continue to emerge.

LN dissection is still considered to be the most effective treatment for most solid tumors [14,16]. Despite the steady increment of the amount of research on LN dissection over time, its normalized frequency in Fig. 5C shows that the upward trend leveled off after 2000. This is presumably due to a disagreement among surgeons. While all oncologic surgeons agree that LN dissection is necessary for cancer surgery, they disagree as to the extent to which LN dissection is performed [17]. Sentinel LN attracted explosive interest from the late 1990s but the interest tapered off since the early 2000s. It was once considered as an innovative way to cope with extended LN dissection [18,19]. But since then, the driving force behind the research has weakened because the accuracy and usefulness of sentinel LN in other solid tumors, except for breast cancer, was poor, and clinical application was limited [20].

LN ratio and LN count had hardly been discussed, but after the mid-2000s there was a slight increase in the amount of research on the subjects. By contrast, micrometastasis have still not attracted much attention in the literature. Apparently, the research field of sophisticated prognostic indicators such

as LN ratio, LN count, and micrometastasis needs to gain new momentum to establish itself as one of the popular research fields.

For our analysis, we used informatics and linguistic filtering methods to articulate the trends. Admittedly, there are some limitations inherent to the methods we adopted, because it only allows us to trace, categorize, and count the words in the titles and abstracts of an article, rather than interpret their meanings in the context in which they are used. These limitations, however, do not undermine our research purpose in any significant way, since our aim here is to trace the general contour of the paths that a large number of articles have taken over a long period.

In conclusion, although research on cancer was abundant, only a small portion was dedicated to investigating its relevance to LN. Western countries had led the research on cancer-related LN, but Asian countries began to participate as major players, expanding their contributions. While LN metastasis, one of the major cancer-related LN topics, showed a steady increase, those involved in oncologic surgery such as LN dissection and sentinel LN did not.

## ACKNOWLEDGEMENTS

### Fund/Grant Support

This paper was supported by Fund of Biomedical Research

Institute, Jeonbuk National University Hospital (No. CUH2018-0023).

### Conflicts of Interest

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

### ORCID iD

Hyo-Jung Oh: <https://orcid.org/0000-0001-8067-2832>

Dongmin Yang: <https://orcid.org/0000-0002-4029-9372>

Hyo-Won Oh: <https://orcid.org/0000-0002-0257-5460>

Jae-Gyu Jeon: <https://orcid.org/0000-0002-8868-1538>

Chonghyuck Kim: <https://orcid.org/0000-0003-2753-4654>

Jeong-Yong Ahn: <https://orcid.org/0000-0002-1007-0550>

Se Wung Han: <https://orcid.org/0000-0001-7252-3976>

Chan-Young Kim: <https://orcid.org/0000-0001-6656-9425>

### Author Contribution

Conceptualization: HJO, JGJ, CYK

Formal Analysis: DY, HWO, SWH

Investigation: HJO, SWH

Methodology: DY, HJO, CK, JYA

Project Administration: HJO, JGJ, CYK

Writing – Original Draft: HJO, CK, CYK

Writing – Review & Editing: DY, HWO, JGJ, JYA, SWH

## REFERENCES

- Stacker SA, Baldwin ME, Achen MG. The role of tumor lymphangiogenesis in metastatic spread. *FASEB J* 2002;16:922-34.
- Pereira ER, Jones D, Jung K, Padera TP. The lymph node microenvironment and its role in the progression of metastatic cancer. *Semin Cell Dev Biol* 2015;38:98-105.
- Halsted WS. I. The results of radical operations for the cure of carcinoma of the breast. *Ann Surg* 1907;46:1-19.
- Wormell I. Informetrics: an emerging subdiscipline in information science. *Asian Libr* 1998;7:257-68.
- Tague-Sutcliffe J. An introduction to informetrics. *Inf Process Manag* 1992;28:1-3.
- Beninger P, Ibara MA. Pharmacovigilance and biomedical informatics: a model for future development. *Clin Ther* 2016;38:2514-25.
- Charidimou A, Fox Z, Werring DJ, Song M. Mapping the landscape of cerebral amyloid angiopathy research: an informetric analysis perspective. *J Neurol Neurosurg Psychiatry* 2016;87:252-9.
- Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2008;22:338-42.
- Oh HJ, Oh HW, Lee DW, Kim CH, Ahn JY, Kim Y, et al. Chronologic trends in studies on fluoride mechanisms of action. *J Dent Res* 2017;96:1353-60.
- Organisation for Economic Co-operation and Development (OECD). Main science and technology indicators (2018) [Internet]. Paris: OECD; 2018 [cited 2019 Feb 28]. Available from: <http://www.oecd.org/sti/msti.htm>.
- National Cancer Institute. Funding for research areas [Internet]. Bethesda, MD: National Cancer Institute; 2018 [cited 2019 Feb 28]. Available: <https://www.cancer.gov/about-nci/budget/fact-book/data/research-funding>.
- Leong SP. Paradigm of metastasis for melanoma and breast cancer based on the sentinel lymph node experience. *Ann Surg Oncol* 2004;11(3 Suppl):192S-197S.
- Hellman S. Karnofsky Memorial Lecture: natural history of small breast cancers. *J Clin Oncol* 1994;12:2229-34.
- Mocellin S, Nitti D. Lymphadenectomy extent and survival of patients with

- gastric carcinoma: a systematic review and meta-analysis of time-to-event data from randomized trials. *Cancer Treat Rev* 2015;41:448-54.
15. Bertero L, Massa F, Metovic J, Zanetti R, Castellano I, Ricardi U, et al. Eighth edition of the UICC classification of malignant tumours: an overview of the changes in the pathological TNM classification criteria: what has changed and why? *Virchows Arch* 2018;472:519-31.
16. Vallbohmer D, Oh DS, Peters JH. The role of lymphadenectomy in the surgical treatment of esophageal and gastric cancer. *Curr Probl Surg* 2012;49:471-515.
17. Mocellin S. The effect of lymph node dissection on the survival of patients with operable gastric carcinoma. *JAMA Oncol* 2016;2:1363-4.
18. Nieweg OE, Uren RF, Thompson JF. The history of sentinel lymph node biopsy. *Cancer J* 2015;21:3-6.
19. Giuliano AE, Gangi A. Sentinel node biopsy and improved patient care. *Breast J* 2015;21:27-31.
20. Read TE, Fleshman JW, Caushaj PF. Sentinel lymph node mapping for adenocarcinoma of the colon does not improve staging accuracy. *Dis Colon Rectum* 2005; 48:80-5.