




Mapping of diseases from clinical medicine research—a visualization study

Yilong Chen^{1,2} · Yiting Dong³ · Yu Zeng⁴ · Xiaoyan Yang^{1,2}  · Jiantong Shen⁵ · Lang Zheng⁶ · Jingwen Jiang^{1,2} · Liming Pu⁷ · Qilin Bao⁷

Received: 20 November 2019 / Published online: 12 August 2020
© The Author(s) 2020

Abstract

By employing bibliometric method, this study aimed to visualize the research hotspots and correlations among clinical medicine subjects. Literatures were retrieved from the PubMed database based on MeSH words and free-text phrases and screened based on inclusion and exclusion criteria. The disease themes were manually marked according to ICD-10. Co-word analysis and strategic diagram methods were applied to explore the hot topics and development trends of disease themes. 2551 articles were included after literature screening. The amount of paper showed an increasing trend and reached a peak in 2013. The subjects of adults and the elderly accounted for 45.0% and 27.0% respectively. The United States of America had the most publication, with Massachusetts and California being the most prevalent states, and Harvard University was the most prolific institution. Co-word analysis revealed that research hot topics of diseases were divided into 8 themes, among which the most was “disease of the circulatory system” and “injury, poisoning and certain other consequences of external causes”. The strategic diagram showed that the above two topics were mature but relatively independent, while the “physical fitness” topic was not mature enough but was closely related to the others. There are more and more data-driven studies in the field of medicine and health, while, huge development spaces in the full spectrum of the diseases do exist. Mining the published researches through bibliometrics and visualized methods could come up with valuable results to inform further study.

Keywords Medical data science · Clinical medicine research · Bibliometrics · Co-word analysis · Strategic diagram · Visualization

Yilong Chen, Yiting Dong and Yu Zeng have contributed equally to this work

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11192-020-03646-8>) contains supplementary material, which is available to authorized users.

✉ Xiaoyan Yang
yangxiaoyan@scu.edu.cn

Extended author information available on the last page of the article

Introduction

Medicine is a systematic discipline from the prevention to the treatment of disease, also be divided into many different research directions and research fields (Iserson and Moskop 2007). Disease is the abnormal life process of the body under certain conditions, which is caused by the damage of etiology and is caused by the disorder of self-regulation (Miller and Chappel 1991). Disease category is one of the appropriate ways to distinguish different clinical fields in clinical medicine research.

A method named bibliometric analysis is to make quantitative analysis on appropriate literature sources by showing the growing trends and periodical hot topics of a subject in a visualization way (Shen et al. 2011a, b, c, d). In 2011, Prof. Shen, one of our group members, published a series of articles about visualization studies on evidence-based medicine domain knowledge, which has received widespread attention (Shen et al. 2011a, b, c, d). Meanwhile, medical scholars had adopted the bibliometric method to find theme trends and knowledge structures in different medical fields. For instance, Zhang visualized knowledge domain of patient adherence by co-word analysis and social network analysis (Zhang et al. 2012); Fu utilized bibliometric analysis to research malaria in China during 2004 to 2014 (Fu et al. 2015); Gu, Hsu, and Liao all visualized the big data research in medicine (Gu et al. 2017; Hsu and Li 2019; Liao et al. 2018); Huang analyzed the pelvic organ prolapse during 2007 to 2016 by bibliometric and social network analysis (Huang et al. 2018); Jin and Xin visualized the hotspots and trends of multimedia big data (Jin and Li 2018); Zhao used co-word analysis to study the theme trends and knowledge structure on choroidal neovascularization (Zhao et al. 2018); Saheb adopted bibliometric analysis to exploring IoT big data analytics in the healthcare industry (Saheb and Izadi 2019); Wang applied co-word analysis to investigate the professional-patient relations in the Internet era (Wang et al. 2019); and Yang used the same method to identify the trends in research on Vitamin D (Yang et al. 2019).

However, all those articles were aimed at isolated disease without exploring the relationship among them. Investigating the hotspots and correlations of broad clinical medicine research subjects by bibliometrics method might provide references for clinical scholars and methodologists in their professional research. In this study, we conducted a relatively comprehensive exploration and collation of clinical medicine literature subjects, through manually labeling the disease theme according to the International Classification of Disease, 10th Revision (ICD-10), which was developed by the World Health Organization (WHO) in 1992 (Quan et al. 2005), and, applied bibliometrics and visualized methods to reveal hot topics and relationship among them, so as to inform further study.

Materials and methods

Data collection

Data sources and searching strategies

The clinical medicine researches based on medical data science with various methods were searched in the PubMed database on March 28th, 2018. When selecting research terms, besides main terms such as “Medicine”, “Health” and “Disease”, we focused on

three topics in fields of medical health. The first topic was “data” because the object of our research aims at medical data science, and the search terms of “big data”, “real-world”, etc. were considered; the second topic was “method”, such as “Data Mining”, “Machine Learning”, “Statistical Models”, “Biostatistics” etc.; and the last topic was “health information” with the search terms being “Health Information Systems”, “Electronic Health Records”, etc.. A total of 25 search terms were used, including Medical Subject Headings (MeSH) terms (which contained vocabulary of biomedical terms and created by U.S. National Library of Medicine) and free-text phrases. Furthermore, we applied the retrieval strategy (Table 1) of the following three steps to obtain literature in initial research: (1) “OR” was used to combine the search terms with the same meaning or topic (#1–#6); (2) the retrieval results from #2 to #6 were connected as #7; (3) “AND” was used to combine #1 and #7 and filtered to remain papers from core clinical journal subsets, which is an optional choice to add journal restriction in the PubMed database searching (https://www.ncbi.nlm.nih.gov/books/NBK3799/#catalog.Searching_for_Journals_in_the_NL).

Inclusion and exclusion criteria

In this study, we included articles published in English (Saheb and Izadi 2019) and studied in human with quantitative data methods, such as statistical models, machine learning, etc., focusing, especially, on clinical diseases, including etiology, mechanism, clinical manifestations, diagnosis, treatment, prevention, prognosis, etc.. Moreover, the following related articles were excluded (Haase et al. 2009; Huang et al. 2018; Wang et al. 2019): (1) Duplicated article; (2) Articles whose abstract cannot be imported into EndNote software (<https://endnote.com>); (3) Articles that just apply descriptive methods; (4) Review articles or systematic review articles; (5) Articles with irrelevant topics. The screening step and results of the literature selection are shown in Fig. 1.

Table 1 Search strategy list

#	Search strategy
1	Medicine OR Health OR Disease
2	big data OR Pragmatic Clinical Trials as Topic OR real world OR real-world
3	Data Mining OR Biostatistics OR Models, Statistical OR Data Interpretation, Statistical
4	Automatic Data Processing OR Artificial Intelligence OR Machine Learning OR Unsupervised Machine Learning OR Supervised Machine Learning
5	Health Information Exchange OR Health Information Management OR Health Information Systems
6	Electronic Health Records OR Health Records, Personal OR Medical Records OR Registries OR electronic medical record* OR EMR
7	2 OR 3 OR 4 OR 5 OR 6
8	1 AND 7 Filters: Core clinical journals

All search strategies were limited in core clinical journals of the PubMed database

#: Search number. Each search number represents a search strategy

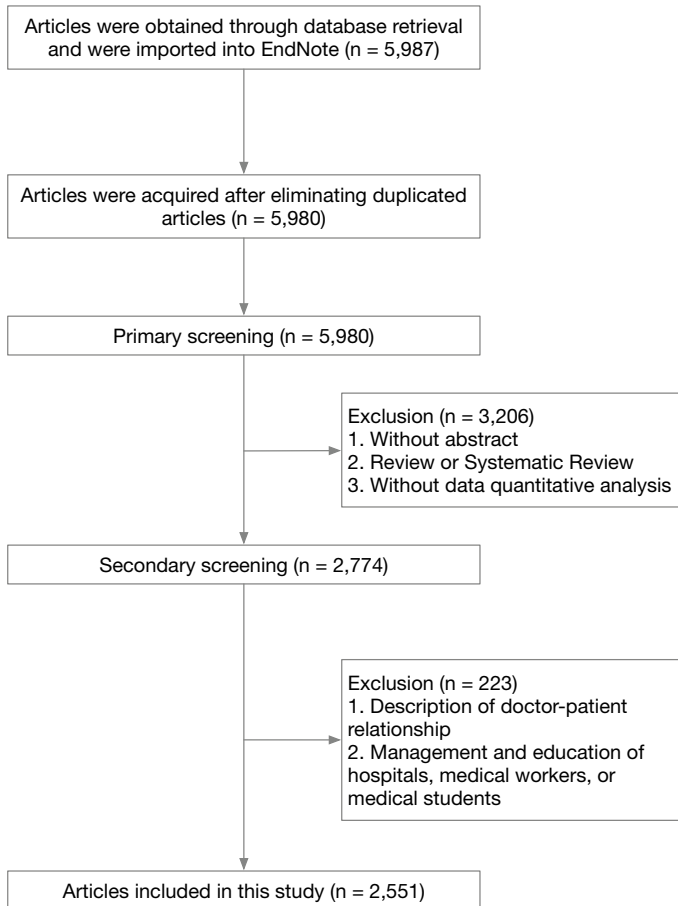


Fig. 1 Screening and result of literature selection

Methods

Paper screening

Based on inclusion and exclusion criteria, two researchers selected the literature back to back. In circumstances of inconsistency, they discussed and decided whether to include, if an agreement could not be reached after discussion, a senior researcher would be asked to make the decision.

Manual Labeling

Based on ICD-10, the experienced medical workers conducted a classification supplement of keywords for key diseases in the included literatures. We invited four experienced medical workers to manually label disease themes, the detailed steps and rules are as follows: (1) The medical workers were trained to classify the main research content based on ICD-10 categories by viewing the title, keywords and abstract; (2) To ensuring the label results'

consistency from different medical workers, we randomly chose 50 articles and each one should label the disease theme back to back. If there are two or more different label results from four medical workers in the same article, they would discuss in the group, and the training and test procedure would be carried out again as above, until the label results are consistent. (3) After confirming the label results' consistency, we divided the rest of included articles into four groups and each group was in charge by one experienced medical worker; (4) If there were several different disease themes existed in one article, they would try to classify each article into two or fewer main disease themes; if some articles clearly indicate more than two main disease themes, they would discuss with others to sum up into two or fewer themes, however, more than two disease themes were allowed sometimes if the agreement could be reached after discussion. (5) After finishing marking the articles of one group, another medical worker would double-check them in this group. Let's take a simple example through one group of the articles which the paper Abdelsattar's (Abdelsattar et al. 2017) belonged to. Firstly, we could easily classify the article as "Neoplasms" disease theme based on its title "The impact of health insurance on cancer care in disadvantaged communities" and its keyword "Neoplasms/*epidemiology/mortality", then we had a quick view of its abstract or text to see if there was another disease theme related to, and finally another person did a transvaluation on this article.

After manual labeling, the main classification types of disease themes were set as "Certain infectious and parasitic diseases" (CPD), "Neoplasms" (NPS), "Injury, poisoning and certain other consequences of external causes" (IPC), "Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism" (DBC), "Endocrine, nutritional and metabolic diseases" (ENM), "Mental and behavioral disorders" (MBD), "Diseases of the nervous system" (DNS), "disease of the circulatory system" (DCS), "physical fitness" (PSF), "Diseases of the eye and adnexa" (DEA), "Diseases of the ear and mastoid process" (DEM), "Diseases of the respiratory system" (DRS), "Diseases of the digestive system" (DDS), "Diseases of the genitourinary system" (DGS), and "Pregnancy, childbirth and the puerperium" (PCP).

Descriptive statistics

A histogram chart of the number of publications per year was presented. The proportion of age distribution of research objects was calculated by the frequency data set of keywords indicating age structure, such as "Adult", "Infant", "Newborn", "Child", "Preschool", "Aged", "80 or over", etc. (Shen et al. 2011a, b, c, d). The spacial distribution was presented through filtering out the "country" in the "Author Address" entry from EndNote and counting the word frequency of each "country" for the corresponding author, and the same "country" was merged and sorted by their frequency to obtain national data with significant results. And the same method was applied to acquire the spacial distribution at the state level in the top country, and the affiliation distribution through publication counts of the corresponding author's affiliations in the top three states/regions in the top country. The "maps" and "ggplot2" packages in R software (<https://www.r-project.org>) were used to realize data visualization for heat maps combined with geographic location information.

Association of clinical medicine researches by co-word analysis

Co-word analysis is to reflect the strength of the correlation between keywords by analyzing the co-word of words in literatures. We excluded subject headings like "age", "gender"

and “nationality” from the total keyword list. Based on the keywords of our research, a word co-occurrence matrix representing the number of articles containing two specific keywords was established. The number of articles reflects the correlation strength between the keywords. The more the articles, the closer the connection between the two keywords. In this study, all keywords were extracted from EndNote with the label of 2,551 eligible articles, and we chose the co-occurrence frequency of keywords that is larger than 10 to finish the next steps. According to the word co-occurrence matrix, a common word network composed of those words can be formed, and the corresponding keywords selected in the co-occurrence analysis were treated as nodes, the co-occurrence between the two keywords as connection, and the co-occurrence frequency of keywords as connection weight. We used “igraph” (Csardi and Nepusz 2006) package in R software to visualize the mapping for the association of clinical medicine research, and applying “walktrap.community” function to classify the type of clinical medicine researches (Deitrick et al. 2013). The size of a node was formed by the value of degree of centrality, which is formed by the number of the connection edges that the node possesses (Borgatti 2005; Everett et al. 2005; Valente et al. 2008; Rodrigues 2019). The higher value of degree of centrality, the larger size of the node. The highest value of degree of centrality indicated the most important node with red color labeling in each cluster.

Law et al. put forward the strategic coordinates method in 1988 to describe the correlation between different fields and the correlation within one field (Law et al. 1988). In this study, we introduced a strategic diagram to present the current hotspots and trends of disease themes based on medical data science and to predict future developments. The strategic diagram is composed of centrality and density. Centrality is used to measure the strength of interrelation between a cluster and other clusters, and the high centrality indicates the hotspot of this research has extensive connections with other hotspots (Shen et al. 2011a, b, c, d). Density is used to measure the strength of interrelation between subject terms within the cluster, and the high density indicates that the research subjects are highly correlated and the research tends to be mature (Shen et al. 2011a, b, c, d). The four quadrants indicate the different combination of centrality and density: high centrality and density in quadrant I, high centrality and low density in quadrant II, low centrality and density in quadrant III, and low centrality and high density in quadrant IV.

Software

EndNote was utilized to filter literature and to make manual labels. R software and Microsoft Excel were used to clean and integrate data. All plots were visualized by R software.

Results

The feature of clinical medicine researches

A total of 5,987 articles were obtained after initial search and finally, 2,551 were extracted after screening step by step (Fig. 1).

The number of publications in each year from 1970 to 2017 was counted, it suggests that the overall trend is increasing especially after 1990, with the most researches in 2013 (Fig. 2).

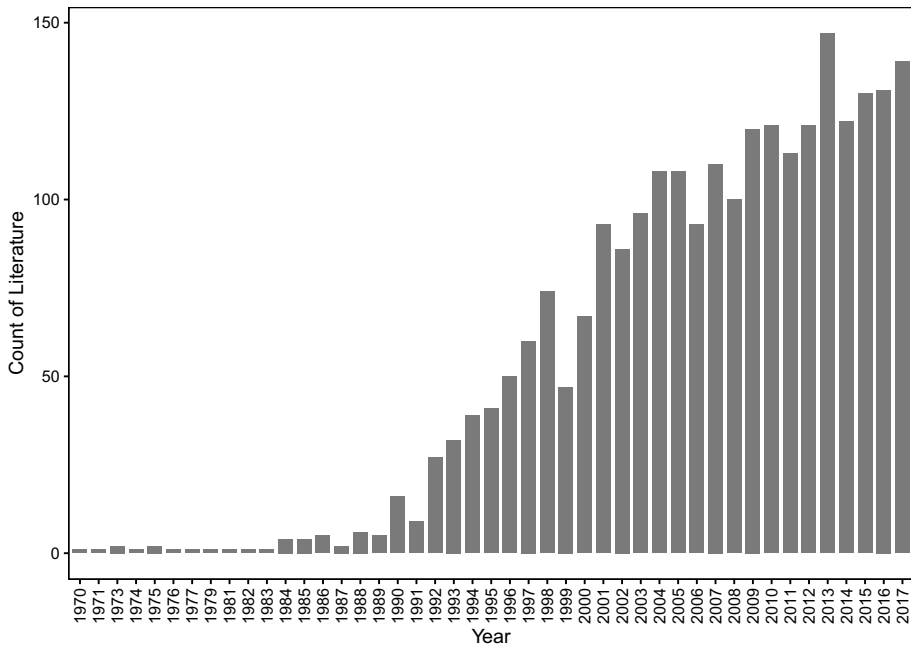


Fig. 2 The publication trend of clinical medicine research from 1970 to 2017. *Note:* it just represents the period from 1970 to 2017, since the literature search is finished on March 28, 2018, and the year 2018 is too short to be compared with other years

Table 2 is depicting the age distribution of the subjects. It shows that the frequency of “adult” reached 1845, accounting for 45.0% of the total, and infants accounted for 6.5%, children for 9.0%, adolescents for 12.5% and aged group for 27.0%.

Mapping of literature publication location (spatial distribution)

The location of published articles in each country is visualized in Fig. 3 marking the top 6 countries which account for 69% (Table S1) of the total, the label color in the heat maps indicates the number of article publication, the deeper color of the label, the larger number of article publications. It is obvious that the United States of America (USA) is the country with the most publication and has been far more than that in other countries. Consequently, we did a further study on papers from the USA, not only counting the number of publications in each state (Table S2), but also exploring each affiliation within the top three states (Fig. 4). It shows that Massachusetts and California have the largest number of published articles compared with the others. And the affiliations with the most publication in Massachusetts, California, and New York were Harvard University, University of California Los Angeles, and Columbia University, respectively (Fig. 4).

Mapping of association between clinical medicine researches

Figure 5 represents the association between disease themes. In this figure, it was classified into eight disease themes which are DCS, IPC, NPS, MBD, PSF, PCP, ENM, and DRS.

Table 2 Age distribution in clinical medicine researches from 1970 to March 2018 ($N=4106$)

Age group	Frequency	Percentage (%)
Infant	267	6.5
Infant, newborn	132	3.2
Other ^a	135	3.3
Child	369	9.0
Child, preschool	262	6.4
Other ^b	107	2.6
Adolescent	515	12.5
Adult	1845	45.0
Young adult	226	5.5
Middle-aged	1280	31.2
Undefined ^c	339	8.3
Aged	1110	27.0
Aged, 80 or over	386	9.4
Other ^d	724	17.6

^aIndicates infant except for newborn infant

^bRepresents child except for the child in preschool

^cIndicates adults that don't clearly distinguish young adult or middle-aged

^dMeans aged people who were under 80 or undefined age

The largest cluster is DCS, with IPC as the second, suggesting these two are popular diseases that investigators would like to explore. Additionally, we observed that “Stroke” and DNS were included in the DCS cluster, which means that they are highly correlated with each other, and they would, normally, be investigated together. Furthermore, in Fig. 5, it is obvious to see that DRS, NPS, and PCP are three independent clusters, they do not have any edge between other clusters, meaning that these three disease themes are more independent. PSF also has a small cluster, but it has multiple edges with other clusters, such as DCS, MBD, and ENM, suggesting that PSF research is not mature, but it is often researched with other disease themes together.

We used the density and centrality value of each disease theme (Table 3) to visualize a strategic diagram. (Fig. 6). It shows that IPC possesses the largest density value, which indicates the nodes within the cluster are strongly related to one another and it is mature in the clinical medicine research subjects. While, it is obvious to see that the centrality value of IPC is much smaller than that of PSF, presenting that PSF has more ability to connect with other disease themes. Additionally, the disease themes of NPS and DRS are both located in Quadrant III, suggesting that these two disease themes were neither popular nor mature in clinical medicine research. It's worth noting that there is no disease theme in Quadrant I, indicating none disease theme falls into the development period of both mature in itself and as the center of the others (that is, extensively connected with other disease themes). So, there are still huge development spaces in medical data science in terms of disease domain research.

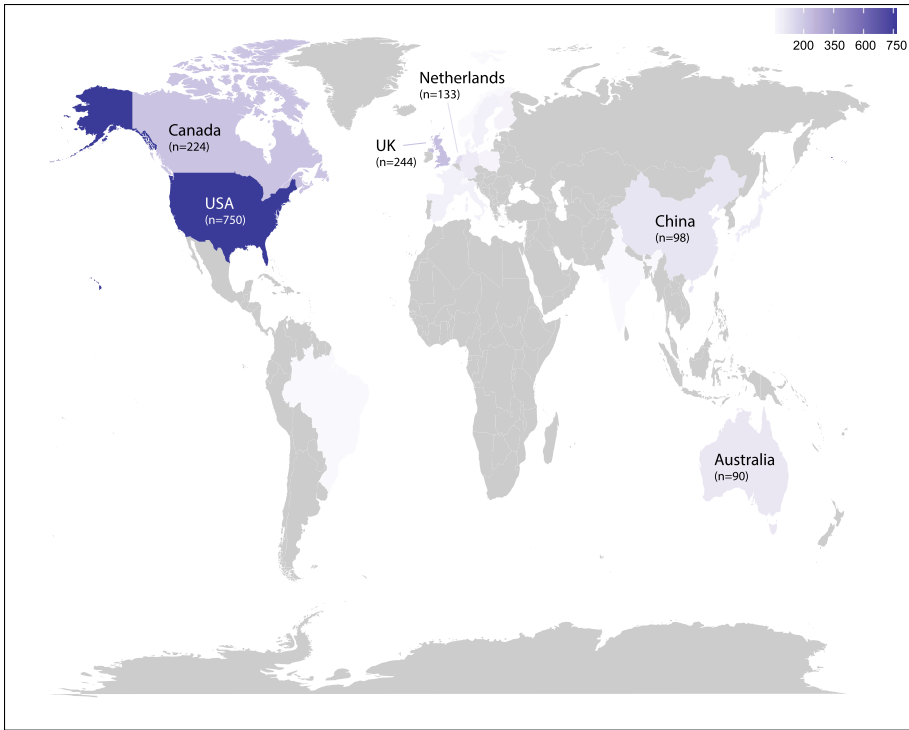


Fig. 3 The situation of clinical medicine research publication in the world from 1970 to March 2018. *n* means the number count of clinical medicine research publication. The label color in the heat maps indicates the number of clinical medicine research publications, the deeper color of the label, the larger number of clinical medicine research publications

Discussion

With the development of medical data, more and more researchers are applying them in various fields, and most of the researches could be divided into two kinds: those of methods and those of applications. To inform clinical scholars and relevant methodologists, this study explored the hot spots and relationships among different disease themes to stimulate further applicational and methodological studies in clinical medicine.

In our study, we clustered into eight disease subjects and it is found that DCS was the most frequently studied topic, followed by IPC, which informs us that these two topics attracted wide attentions and had great importance than the others in clinical data-driven researches, and further, the conclusions from these two disease themes might be reached through a variety of techniques, which is worth investigating in the perspective of methodology.

Results of the strategic diagram graph show that, in terms of the development relationships of the clusters, PSF has the highest centrality, suggesting its center role of the research network; and IPC has the highest density, suggesting its maturity itself among the research network. There are no cluster falls in Quadrant I area, that is, none disease theme is both central and developed, suggesting great improvement opportunities in the future.

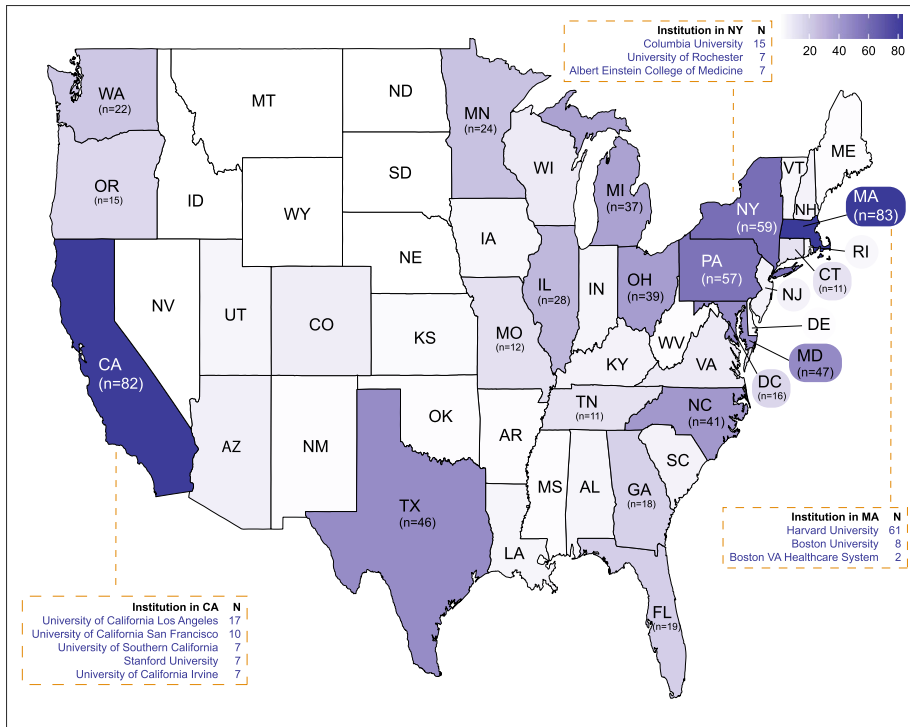


Fig. 4 Clinical medicine research publications in the states of USA and in the institutions of top states. In the top three publication states, the institutions with the top three (including the tie) publications in the state were listed. *N* or *n* means the number of clinical medicine research publications. The label color in the heat maps indicates the number of clinical medicine research publications, the deeper color of the label, the larger number of clinical medicine research publications

Recently, some studies focus on research in healthcare informatics based on big data. For example, Gu introduced the contribution of countries in healthcare informatics and the knowledge structure of healthcare big data research (Gu et al. 2017); Liao explored the situation of publication, country, institution, published journal of the medical big data research (Liao et al. 2018); Hsu investigated the research trends and relationship between knowledge clusters in medical big data (Hsu and Li 2019). All of them concluded the trend and correlation between knowledge clusters in healthcare big data research. Here, our research focuses on using the bibliometric analysis to find the hotspots and relationship of disease themes in data-driven clinical medicine researches, with a wide variety of the data volume. In addition, it is not found yet that there are other studies exploring relationships among disease themes labeled manually according to ICD-10, which might contribute a lot to inform scholars in this field.

It is found that the USA was the top publication country in the world, which is the same as other researches. For example, Fu pointed out that the USA is the top one country on the number of papers and citations in the 19 complementary and alternative medicine journals from the 1980 to 2000 s (Fu et al. 2011); Filser found that the USA is the most cited country in the literature on Lean Healthcare (Filser et al. 2017); Tran discovered that the USA is the most prolific country in Artificial Intelligence in Health/Medicine research (Tran et al.

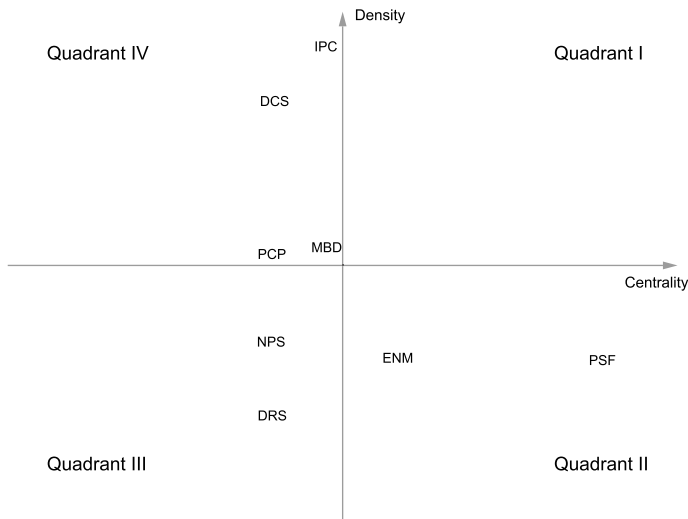


Fig. 6 Strategic diagram for disease themes. Centrality feedbacks the degree of connection to other disease themes, the higher value of centrality, the stronger connection ability; Density feedbacks the degree of correlation between each node in this disease theme, the higher value of density, the more mature disease theme

Table 3 Density and centrality values in the strategic diagram

	NPS	MBD	PCP	ENM	PSF	DRS	IPC	DCS
Density	- 13.21	3.16	1.99	- 16.01	- 16.41	- 26.01	37.99	28.51
Centricity	- 5.81	- 1.31	- 5.81	4.53	21.39	- 5.81	- 1.38	- 5.81

NPS neoplasms, *MBD* mental and behavioral disorders, *PCP* pregnancy, childbirth, and the puerperium, *ENM* endocrine, nutritional and metabolic diseases, *PSF* physical fitness, *DRS* diseases of the respiratory system, *IPC* injury, poisoning and certain other consequences of external causes, *DCS* diseases of the circulatory system

study included articles in the core journal of the PubMed database and did not expand the methodological aspect, resulting that the papers included are a subset of the whole; with the recent popularity of artificial intelligence, there are more and more scholars applying machine learning methods in clinical medicine research (Beam and Kohane 2018; Tran et al. 2019), and as the methodology system is getting more and more complete in the medical field, we are carrying out further research targeted specially on the methodological aspect, based on experiences of this study. Secondly, we counted the frequency of keywords (such as “Adult”, “Infant”, “Newborn”, “Child”, “Preschool”, “Aged”, “80 or over”, etc.) to show the age distribution, by doing so, we are not able to reveal the whole situation, but trying to provide as much information as we could, because, as a matter of fact, not all the studies revealed the objects’ age, even in the whole paper.

In conclusion, there are more and more data-driven studies in the field of medicine and health, while, huge development spaces in the full spectrum of the diseases do exist. Mining the published researches through bibliometrics and visualized methods could come up with valuable results to inform further study.

Acknowledgements This study was funded by the project of Science & Technology Department of Sichuan Province (grant number 2016SZ0072), Key Research and Development Project of Sichuan Provincial Science and Technology Department (2017FZ0104) and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (grant number ZYJC18010).

Conflict of interest The authors declare that they have no conflict of interest.


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdelsattar, Z. M., Hendren, S., & Wong, S. L. (2017). The impact of health insurance on cancer care in disadvantaged communities. *Cancer*, (2016, 11/20 ed., Vol. 123, pp. 1219–1227.
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Borgatti, S. P. (2005). Centrality and network flow. *Journal of Social Networks*, 27(1), 55–71.
- Churruca, K., Pomare, C., Ellis, L. A., Long, J. C., & Braithwaite, J. (2019). The influence of complexity: A bibliometric analysis of complexity science in healthcare. *BMJ Open*, 9(3), e027308.
- Csardi G, Nepusz TJ (2006) The igraph software package for complex network research. *International Complex Systems*, 1695(5), 1–9.
- Deitrick, W., Valyou, B., Jones, W., Timian, J., & Hu, W. (2013). Enhancing sentiment analysis on twitter using community detection. *Scientific Research*, 5(3), 6.
- Everett, M. G., & Borgatti, S. P. J. M. (2005). Extending centrality. *Models and Methods in Social Network Analysis*, 35(1), 57–76.
- Filser, L. D., da Silva, F. F., & de Oliveira, O. J. J. S. (2017). State of research and future research tendencies in lean healthcare: A bibliometric analysis. *Scientometrics*, 112(2), 799–816.
- Fu, H., Hu, T., Wang, J., Feng, D., Fang, H., Wang, M., & Feng, Z. (2015). A bibliometric analysis of malaria research in China during 2004–2014. *Malaria Journal*, 14, 195–201. <https://doi.org/10.1186/s12936-015-0715-2>
- Fu, J. Y., Zhang, X., Zhao, Y. H., Huang, M. H., Chen, D. Z. (2011). Bibliometric analysis of complementary and alternative medicine research over three decades. *Scientometrics* 88(2), 617–626.
- Gu, D., Li, J., Li, X., & Liang, C. (2017). Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *The International Journal of Medical Informatics*, 98, 22–32. <https://doi.org/10.1016/j.ijmedinf.2016.11.006>
- Haase, M., Bellomo, R., Devarajan, P., Schlattmann, P., & Haase-Fielitz, A. (2009). Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: A systematic review and meta-analysis. *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, 54(6), 1012–1024. <https://doi.org/10.1053/j.ajkd.2009.07.020>
- Hsu, W.-C., & Li, J.-H. (2019). Visualising and mapping the intellectual structure of medical big data. *Journal of Information Science*, 45(2), 239–258. doi:<https://doi.org/10.1177/0165551518782824>.
- Huang, F., Zhou, Q., Leng, B. J., Mao, Q. L., Zheng, L. M., & Zuo, M. Z. (2018). A bibliometric and social network analysis of pelvic organ prolapse during 2007–2016. *Journal of Chinese Medical Association*, 81(5), 450–457. <https://doi.org/10.1016/j.jcma.2017.08.012>
- Iserson, K. V., & Moskop, J. C. J. (2007). Triage in medicine, part I: Concept, history, and types. *Annals of emergency medicine*, 49(3), 275–281.
- Jin, Y., & Li, X. (2018). Visualizing the hotspots and emerging trends of multimedia big data through scientometrics. *Multimedia Tools and Applications*, 78(2), 1289–1313. <https://doi.org/10.1007/s11042-018-6172-5>

- Law, J., Bauin, S., Courtial, J.-P., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics*, *14*, 251–264.
- Liang, H.-N. (2010). *Overview of the health informatics research field: A bibliometric approach*. Paper presented at the IMIA/IFIP Joint Symposium on E-Health.
- Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X.-J. (2018). A bibliometric analysis and visualization of medical big data research. *Sustainability*, *10*(2), 166. <https://doi.org/10.3390/su10010166>
- Miller, N. S., & Chappel, J. N. J. P. A. (1991). History of the disease concept. *Psychiatric Annals*, *21*(4), 196–205.
- Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., et al. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, *43*(11), 1130–1139.
- Rodrigues, F. A. (2019). Network centrality: An introduction. In *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems* (pp. 177–196). Springer.
- Saheb, T., & Izadi, L. (2019). Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends. *Telematics and Informatics*, *41*, 70–85. doi:<https://doi.org/10.1016/j.tele.2019.03.005>.
- Shen, J., Yao, L., Li, Y., Clarke, M., Gan, Q., Fan, Y., & Wang, L. J. (2011a). Visualization studies on evidence-based medicine domain knowledge (series 2): Structural diagrams of author networks. *Journal of Evidence-Based Medicine*, *4* (2), 85–95.
- Shen, J., Yao, L., Li, Y., Clarke, M., Gan, Q., Fan, Y., & Wang, L. (2011b). Visualization studies on evidence-based medicine domain knowledge (series 1): Mapping of evidence-based medicine research subjects. *Journal of Evidence-Based Medicine*, *4*(2), 73–84. <https://doi.org/10.1111/j.1756-5391.2011.01126.x>
- Shen, J., Yao, L., Li, Y., Clarke, M., Gan, Q., Li, Y., & Wang, L. J. (2011c). Visualization studies on evidence-based medicine domain knowledge (series 3): Visualization for dissemination of evidence based medicine information. *Journal of Evidence-Based Medicine*, *4*(2), 96–105.
- Shen, J., Yao, L., Li, Y., Clarks, M., Gan, Q., Fan, Y., & Wang, L. (2011d). Visualization studies on evidence-based medicine domain knowledge (series 1): Mapping of evidence-based medicine research subjects. *Journal of Evidence-Based Medicine*, *4*, 73–84.
- Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q.-H., Ho, M.-T., Vuong, T.-T., & Nguyen, H. L. (2019). Global evolution of research in artificial intelligence in health and medicine: A bibliometric study. *Journal of Clinical Medicine*, *8*(3), 360.
- Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connect (Tor)*, *28*(1), 16–26.
- Wang, Z., Deng, Z., & Wu, X. (2019). Status quo of professional(-)patient relations in the internet era: Bibliometric and co-word analyses. *International & Journal of Environmental Research and Public Health*, *16*(7), 1183–1201. <https://doi.org/10.3390/ijerph16071183>
- Yang, A., Lv, Q., Chen, F., Wang, D., Liu, Y., & Shi, W. (2019). Identification of recent trends in research on vitamin D: A quantitative and co-word analysis. *Medical & Science Monitor*, *25*, 643–655. <https://doi.org/10.12659/MSM.913026>
- Zhang, J., Xie, J., Hou, W., Tu, X., Xu, J., Song, F., & Lu, Z. (2012). Mapping the knowledge structure of research on patient adherence: Knowledge domain visualization based co-word analysis and social network analysis. *PLoS ONE*, *7*(4), e34497. <https://doi.org/10.1371/journal.pone.0034497>
- Zhao, F., Shi, B., Liu, R., Zhou, W., Shi, D., & Zhang, J. (2018). Theme trends and knowledge structure on choroidal neovascularization: A quantitative and co-word analysis. *BMC Ophthalmology*, *18*(1), 86–96. <https://doi.org/10.1186/s12886-018-0752-z>

Affiliations

Yilong Chen^{1,2} · Yiting Dong³ · Yu Zeng⁴ · Xiaoyan Yang^{1,2}  · Jiantong Shen⁵ · Lang Zheng⁶ · Jingwen Jiang^{1,2} · Liming Pu⁷ · Qilin Bao⁷

¹ West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, No. 37 Guoxue Xiang, Chengdu 610041, Sichuan, China

² Medical Big Data Center, Sichuan University, Chengdu, China

³ West China School of Medicine, Sichuan University, Chengdu, China

- ⁴ Ministry of Education Virtual Research Center for Evidence-based Medicine, West China Hospital/West China School of Medicine, Sichuan University, Chengdu, China
- ⁵ Medical School of Huzhou University, Huzhou, China
- ⁶ School of Finance, Central University of Finance and Economics, Beijing, China
- ⁷ West China School of Public Health, Sichuan University, Chengdu, China