KNOWLEDGE GRAPH ANALYSIS FOR CHRONIC DISEASES NURSING BASED ON VISUALIZATION TECHNOLOGY AND LITERATURE BIG DATA

SIYU DUAN AND YANG ZHAO∗

Abstract. The use of knowledge graph analysis for chronic disease nursing based on visualization technology and literature big data is an unexplored area of research in this field of study. To uncover research hotspots and developmental trends in the field of chronic disease nursing, and to provide a scholarly reference, we employed mathematical and statistical methods along with CiteSpace literature visualization analysis software for quantitative analysis of extensive literature data from the Web of Science Core Collection. We examined aspects such as publication trends, journals, author collaborations, research institutions, national and regional distributions, keyword co-occurrence, clustering, time zones, emergence, literature co-citations, and more. These analyses identified the current hotspots and future directions for research. Notably, scholars’ interest in chronic disease nursing exhibited a consistent upward trajectory. In particular, the field of artificial intelligence technology application in nursing yielded 3,610 published papers in 141 journals with more than or equal to 10 published papers on the topic, accounting for 58.41% of the total number of published papers in this field of study. Furthermore, the top three publishers were the “Journal of Clinical Nursing,” “Journal of Advanced Nursing,” and “BMC Health Services Research.” Among authors, Hu, Frank B., Willett, Walter C., and Rimm, Eric B., ranked as the top three, and 12 authors had more than 10 publications. The most active research institutions included Harvard University, Harvard Medical School, Brigham & Women’s Hospital, University of California System, University of London, US Department of Veterans Affairs, Veterans Health Administration (VHA), Harvard T. H. Chan School of Public Health, University of Sydney, and the University of Toronto. The United States, Australia, England, China, Canada, Netherlands, Spain, Italy, Sweden, and Germany emerged as the leading countries in terms of research output, while emerging hotspots encompassed topics such as incidence, rheumatoid arthritis, qualitative research, burnout, kidney transplantation, critical illness, COVID-19, Sars-COV-2, public health, and the well-being of medical staff. These findings present valuable insights for prospective research endeavors.

Key words: Chronic diseases, nursing, literature big data, bibliometric analysis, trends, hotspots

1. Introduction. Chronic non-communicable diseases (NCDs) stand as the world’s foremost cause of death and disability [16], constituting a staggering 73.6% of chronic disease-related deaths, as reported in the “World Health Statistics 2021” by the World Health Organization (WHO) [20]. In 2019, chronic diseases were responsible for nearly 70% of the global disease burden, and they accounted for a staggering 88.5% of deaths in China, with 80.7% attributed to cardiovascular diseases, cancer, and chronic respiratory diseases. Presently, the field of chronic disease prevention and control confronts challenges, underscoring the pressing need to fortify chronic disease nursing practices.

Bibliometrics employs mathematical and statistical methods to scrutinize vast repositories of literature within specific domains and databases, unearthing the current state of research in these fields. This approach predominantly measures documents, authors, and word counts, harnessing mathematical and statistical techniques to conduct a quantitative examination of the knowledge contained within these documents. The insights derived from bibliometrics are further visualized through knowledge graphs, offering a more objective portrayal of the research landscape. Scholars rely on bibliometrics and related literature analysis software to comprehensively dissect the research status and identify hotspots within nursing, ultimately providing invaluable reference points for fellow researchers. For instance, Juan-Jose delved into bibliometric and gender-based analyses of scientific publications within Scopus and Web of Science, offering insights into annual article production, prominent authors, top-cited articles, and thematic keyword analyses [2]. Similarly, Cant conducted a bibliometric exploration of highly-cited virtual simulation nursing education articles, revealing rankings, topic diversity, and authorship patterns [4].

∗School of Health and Wellness, Panzhihua University, Panzhihua 617000, China (zhaoyang@pzhu.edu.cn). This work was sponsored by China Panxi Health Industry Research Center (No. PXXY-YB-202213).
Besides, Hahn engaged in quantitative and statistical analyses of publication trends, prolific authors, highly-cited documents, and keywords about clinical reasoning in nursing [9]. In the same vein, Su, through bibliometrics, explored the trends in high-impact international nursing core competencies research [13]. While Wang employed literature visualization analysis to unveil research hotspots and future directions in Traditional Chinese Medicine (TCM) nursing for insomnia [19]. That said, Yesilbas conducted a literature data analysis to investigate the knowledge structure and developmental process in nursing empowerment [21]. Building on the findings of past studies in this area, Zhang utilized VOSviewer and CiteSpace to scrutinize COVID-19-related nursing research, uncovering the current state and hot topics within this realm [22]. Likewise, Zhao employed CiteSpace and VOSviewer to assess the application of virtual reality technology in nursing studies [23].

Contemporaneously, De Oliveira conducted a bibliometric analysis to discern trends in burnout research among nursing professionals, comparing the contributions of various countries, institutions, journals, authors, keywords, and citations [6]. Intriguingly, Ghamgosar also employed bibliometric analysis to offer insights into global research output on geriatric nursing [7]. Whereas, Huang evaluated the literature on family nursing to identify development trends and research focal points [10]. Correspondingly, Molassiotis conducted a bibliometric exploration of disaster nursing, unveiling global development and trends [12]. Just as Blazun adopted an automated, electronic approach to scrutinize the nursing informatics literature, tracing its historical origins, and analyzing the evolution of topics and themes contribute to the understanding of knowledge development within nursing informatics [17].

By the same token, Guo harnessed literature mining and information visualization technologies to examine the bibliometric characteristics of cirrhosis nursing articles in the Web of Science spanning from 1986 to 2020. This endeavor aimed to comprehensively depict the present state of this field and furnish essential evidence for enhancing research in nursing and clinical liver cirrhosis within Mainland China [8].

To unveil the current landscape and pressing concerns within chronic disease nursing, this study sought to provide a reference point for further research endeavors by domestic scholars. Employing mathematical and statistical techniques, along with the CiteSpace literature visualization analysis software, we conducted a quantitative exploration of literature big data about chronic disease nursing. This analysis encompassed ten years from 2013 to 2022 and involved key facets such as publication trends, journals, author collaborations, research institutions, national and regional distributions, keyword co-occurrence, clustering, time zone mapping, emergence, literature co-citation, and more. By scrutinizing the interplay and internal correlations within this wealth of information, we aimed to unearth research hotpots and development trends to guide the field of chronic disease nursing.

The initial section of this paper provides an introduction to the research context, status, innovations, and primary contributions. The remaining sections of the paper are arranged as follows: The subsequent section focuses on elucidating the objectives, design, sample, search strategy, inclusion criteria, and statistical methods. The third part delves into a comprehensive discussion based on the results derived from CiteSpace literature visualization analysis, offering insights into pertinent research hotspots and evolving trends. The closing section of the paper presents the conclusions drawn from our research endeavors.

2. Materials and methods.

2.1. Objectives. The objectives of this study encompass the following key aspects: (1) Identifying the principal contributors in the realm of nursing research linked to Virtual Simulation (VS), including countries, institutions, journals, authors, and articles. (2) Analyzing collaborative relationships within this field. (3) Constructing a knowledge network and pinpointing the frontier topics, thus elucidating future directions in this domain.

2.2. Study design. A descriptive bibliometric analysis was conducted on publications within the domain of nursing related to chronic disease research, retrieved from a comprehensive literature database.

2.3. Data source. The data for this research were obtained from the Web of Science™ (WOS) database, specifically the Web of Science™ Core Collection, encompassing SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR, EXPAND, and related indices.
2.4. **Search method.** The search was conducted on April 23rd, 2023, utilizing Web of Science. The search query utilized the following formula: $TS = (\text{chronic disease OR chronic non-communicable disease OR chronic illness or chronic non-communicable illness})$ AND (nursing OR nurse), AND $DOP = (2013–01–01/2022–12–31)$ AND $DT = (\text{Article OR Review})$ was used to screen out publications associated with chronic diseases nursing. This method efficiently filtered publications relating to chronic diseases nursing, specifying that they were either full papers or review articles in the field of nursing. The search yielded a total of 6,180 literature records.

2.5. **Inclusion criteria.** The following criteria were employed for inclusion:

1. Peer-reviewed articles involving VS related to nursing
2. Original articles and review articles
3. Web of Science core collection (WoSCC) literature big database
4. The language of the document is English

2.6. **Statistical analysis.** Bibliometric methods and CiteSpace visualization techniques were employed to analyze the annual volume of articles, authors, institutions, countries or regions, journals, keywords, and literature citations within the scope of Chronic disease nursing research. This approach was undertaken to gauge the influence and attention accorded to each country or region in this field. The process of literature visualization analysis using CiteSpace is depicted in Figure 2.1 [14, 5].

The primary procedural steps, as outlined in reference [5], are as follows:

**Step 1:** Define the research focus, which, in this instance, pertains to chronic disease nursing.

**Step 2:** Gather literature data by formulating a tailored search strategy aligned with the research focus established in Step 1. This strategy may include keyword and topic searches. Additionally, specify the sources of literature, such as WOS, CNKI, CSSCI, SCOPUS, etc., and execute the literature search in the respective databases by the established search strategy. It is crucial to preprocess the retrieved literature, with a particular note that only data from WOS can be directly utilized and analyzed within CiteSpace. Literature data collected from sources like CNKI, CSSCI, SCOPUS, or others necessitate conversion into WOS format for compatibility with CiteSpace.
Leveraging Visualization Technology and Extensive Literature Data for Chronic Disease Knowledge Analysis

**Step 3:** Create an analysis project within CiteSpace. Configure analytical parameters such as time segmentation, network type, and correlation strength. Afterward, initiate the analysis process using CiteSpace.

**Step 4:** Visualize the results. Review the analysis outcomes (detailed analysis content can be found in reference [24], and, as needed, adjust clustering algorithms and relevant parameters. Generate visual displays for various analysis types, including network diagrams, timeline graphs, and temporal zone charts.

**Step 5:** Conduct a visual analysis. Utilize the insights from the analysis results and the designated view types to undertake a comprehensive analysis, integrating domain-specific knowledge to produce the analysis report.

### 2.7. Calculation algorithms for key literature analysis indices.

1. The Ziff’s Law of Co-word Analysis can be expressed as:

\[
\ln C = \ln f + \ln r, 0.1 < C < 1
\]

(2.1)

where \( f \) signifies the frequency of literature occurrence, \( r \) denotes the rank number of literature frequency, and \( C \) represents a constant within the range of \( 0.1 < C < 1 \).

2. Betweenness Centrality calculation algorithm. The calculation algorithm for Betweenness Centrality is given as:

\[
BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}
\]

(2.2)

where \( g_{st} \) denotes the number of shortest paths from node \( s \) to node \( t \), and \( n_{st}^i \) symbolizes the number of shortest paths that traverse nodes within the shortest path between node \( s \) and node \( t \).

3. Network density calculation algorithm. The Network Density is computed using equation (2.3) below:

\[
Density = \frac{m}{C_n^2} = \frac{2m}{n(n-1)}
\]

(2.3)

With \( m \) representing the number of actual network relations, and \( n \) being the number of network nodes. Additionally, equation (2.4) defines the associations between nodes \( i \) and \( j \) in the network.

\[
Q = \frac{1}{2m} \sum_{i,j} (a_{ij} - p_{ij}) \sigma (C_i, C_j)
\]

(2.4)

where \( A = a_{ij} \) is the adjacency matrix of the actual network, \( p_{ij} \) is the expected value of the number of connecting edges between nodes \( i \) and \( j \), and \( C \) represents the associations between nodes \( i \) and \( j \) in the network. If \( C_i, C_j \) are part of the same club, \( \sigma (C_i, C_j) = 1 \). Otherwise \( \sigma (C_i, C_j) = 0 \). The size of \( Q \) relates to node density, and typically, a \( Q \) value greater than 0.3 is preferred.

4. Silhouette calculation algorithm. The Silhouette Calculation Algorithm distinguishes between classes, assigning a value of 0 if \( a(i) \) equals \( b(i) \), 1 if \( a(i) \) exceeds \( b(i) \), and 0 if \( a(i) \) is less than \( b(i) \) according to equation (2.5).

\[
S_i = \begin{cases} 
1 - a(i)/b(i) & \text{if } a(i) < b(i) \\
0 & \text{if } a(i) = b(i) \\
b(i)/a(i) - 1 & \text{if } a(i) > b(i)
\end{cases}
\]

(2.5)

Here, \( a(i) \) signifies the average distance between point \( i \) and other points in the class, while \( b(i) \) denotes the average distance between point \( i \) and all points in the class of the nearest point \( i \). The average silhouette value \( S \) can be used to measure the cluster’s homogeneity, with a higher \( S \) value indicating greater homogeneity within the network. Generally, an \( S \) value greater than 0.5 is indicative of a highly reliable cluster.
3. Results and discussion. In this section, we configured analysis parameters in CiteSpace, including time segmentation, network type, and correlation strength, and subsequently conducted the analysis. We imported a total of 6,180 literature records from the WoSCC database into the CiteSpace software, employing keywords as analysis nodes. The visualization analysis was performed using CiteSpace 6.2r2, with the software set to run from 2013 to 2022. A one-year time slice was applied, and thresholds were established for authors, institutions, countries, keywords $K = 15$, and co-citation $N = 50$. Pruning methods included Pathfinder, Annual pruning, and overall network pruning. We utilized the LLR algorithm to integrate and analyze the data, as well as to visually present the results.

3.1. Annual number of published papers. Analyzing the temporal evolution of the number of published papers in the WoSCC database provides a macroscopic perspective on the research hotspots within the field. Therefore, we documented the annual number of papers related to chronic disease nursing from 2013 to 2022, as illustrated in Figure 3.1.

As depicted in Figure 3.1, the number of papers addressing chronic disease nursing exhibited a consistent uptrend from 2018 to 2021, underscoring the sustained global interest in this subject across various countries and regions. However, it’s important to note that the data for 2022 might not fully represent the total number of papers published on the topic of chronic disease nursing for that year, as certain papers from 2022 may remain unpublished or are yet to be included in the WoSCC.

3.2. Publication journals. A total of 141 journals featured ten or more published papers on the topic of chronic diseases nursing, collectively accounting for 58.41% of the total published papers. Notably, the top three journals with the highest publication counts were “The Journal of Clinical Nursing,” with 218 papers, followed by “The Journal of Advanced Nursing,” with 194 papers, and “The BMC Health Services Research,” with 132 papers. Other notable contributors to the literature included “BMJ Open” with 128 papers, “International Journal of Environmental Research And Public Health” with 99 papers, “Plos One” with 98 papers, “International Journal of Nursing Studies” with 66 papers, “Cochrane Database of Systematic Reviews” with 64 papers, “Journal of The American Association of Nurse Practitioners” with 56 papers. More so, journals publishing over 40 articles are highlighted in Figure 3.2.

3.3. Geographic distribution. The analysis of papers published in the field of chronic disease nursing within the WOS literature database by different countries or regions was conducted using CiteSpace, with the node type set as “Country.” The visual representation of papers published through collaborative efforts between countries or regions is illustrated in Figure 3.3.

Within Figure 3.3, the size of nodes signifies the volume of papers published in the respective country or region, while the connecting lines between nodes indicate the level of collaboration between different countries or regions. Analogously, the line thickness corresponds to the strength of cooperation.

Furthermore, this analysis encompasses a total of 138 nodes and 386 connections, resulting in an overall network density of 0.0408. This density value suggests a significant presence of countries or regions within this
field, underlining the closeness of collaboration between them. Expectedly, the United States emerged as the leading country in terms of publication count, closely followed by Australia and England. The top ten countries or regions with the highest number of published papers can be identified by examining the publication counts across different nations or regions, as presented in Table 3.1. With a frequency of 5,491, these top contributors account for 46.28% of the total. Notably, the United States outpaces all others in terms of published papers, with a significant lead. Besides, it’s worth noting that the number of papers published in the United States surpasses the combined output of the following four countries or regions. From a centrality perspective, countries or regions with higher publication counts generally exhibit more pronounced centrality.

Nevertheless, it is apparent that the publication numbers in countries like China and Canada, despite their volume, do not correlate proportionally with their centrality, indicating less-than-ideal cooperative relationships with other countries or regions.

### 3.4. Institutional distribution

We harnessed CiteSpace software to visually analyze the dataset within the WOS literature database. Our analysis encompassed the time range from 2013 to 2022, with a yearly breakdown. The $k$ value was set to 25, and Node Types were designated as “Institution.” Pruning was executed
Table 3.1: Top ten countries or regions of published papers

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Country</th>
<th>Frequency</th>
<th>Centrality</th>
<th>Began Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>USA</td>
<td>2089</td>
<td>0.16</td>
<td>2013</td>
</tr>
<tr>
<td>2</td>
<td>AUSTRALIA</td>
<td>677</td>
<td>0.13</td>
<td>2013</td>
</tr>
<tr>
<td>3</td>
<td>ENGLAND</td>
<td>559</td>
<td>0.23</td>
<td>2013</td>
</tr>
<tr>
<td>4</td>
<td>PEOPLES R CHINA</td>
<td>481</td>
<td>0.01</td>
<td>2013</td>
</tr>
<tr>
<td>5</td>
<td>CANADA</td>
<td>443</td>
<td>0.04</td>
<td>2013</td>
</tr>
<tr>
<td>6</td>
<td>NETHERLANDS</td>
<td>301</td>
<td>0.06</td>
<td>2013</td>
</tr>
<tr>
<td>7</td>
<td>SPAIN</td>
<td>276</td>
<td>0.03</td>
<td>2013</td>
</tr>
<tr>
<td>8</td>
<td>ITALY</td>
<td>235</td>
<td>0.07</td>
<td>2013</td>
</tr>
<tr>
<td>9</td>
<td>SWEDEN</td>
<td>216</td>
<td>0.03</td>
<td>2013</td>
</tr>
<tr>
<td>10</td>
<td>GERMANY</td>
<td>214</td>
<td>0.04</td>
<td>2013</td>
</tr>
</tbody>
</table>

Fig. 3.4: Distribution network map of research institutions

using the Pathfinder mode on an annual basis and for the entire network, while other options remained in their default configurations. This process culminated in a visual depiction of research institutions’ distribution, as displayed in Figure 3.4. Moreover, in Figure 3.4, the size of nodes corresponds to the number of papers published by each research institution, while the connecting lines delineate the degree of collaboration between various institutions. Additionally, the color of the lines indicates the collaborative relationships across different periods.

As depicted in Figure 3.4, the network comprises 440 nodes and 1,514 connections, with a network density of 0.0157. This indicates a substantial presence of research institutions, with notable connections among key institutions. The primary institutional collaboration network is notably centered around Harvard University and Harvard Medical School. To gain deeper insights into the accomplishments and collaborative dynamics of these research institutions, we conducted further data analysis in Figure 3.4. This analysis revealed the top ten research institutions with the highest number of publications, as presented in Table 3.2, with a combined frequency of 1,718 times, accounting for 19.29% of the total. Harvard University, Harvard Medical School, and Brigham & Women’s Hospital were the most prolific contributors. Figure 3.4 and Table 3.2 demonstrate the close collaborative network between Harvard University, Harvard Medical School, and Brigham & Women’s
Table 3.2: Top 10 research institutions with the number of publications

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Institutions</th>
<th>Year</th>
<th>Papers</th>
<th>Cooperation Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harvard University</td>
<td>2013</td>
<td>288</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>Harvard Medical School</td>
<td>2013</td>
<td>203</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Brigham &amp; Women’s Hospital</td>
<td>2013</td>
<td>188</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>University of California System</td>
<td>2013</td>
<td>168</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>University of London</td>
<td>2013</td>
<td>161</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>US Department of Veterans Affairs</td>
<td>2013</td>
<td>156</td>
<td>54</td>
</tr>
<tr>
<td>7</td>
<td>Veterans Health Administration (VHA)</td>
<td>2013</td>
<td>154</td>
<td>53</td>
</tr>
<tr>
<td>8</td>
<td>Harvard T. H. Chan School of Public Health</td>
<td>2013</td>
<td>150</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>University of Sydney</td>
<td>2013</td>
<td>140</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>University of Toronto</td>
<td>2013</td>
<td>110</td>
<td>29</td>
</tr>
</tbody>
</table>

Equally, in terms of inter-institutional cooperation, the U.S. Department of Veterans Affairs and Veterans Health Administration (VHA) exhibited a relatively high cooperation density, indicating a concentration of foreign scholars’ research efforts in major institutions. The elevated density of collaboration between these major institutions underscores the maturity of research cooperation within the international community.

Likewise, the Timezone function was applied to assess cooperative institutions through a time series perspective, with the analysis findings presented in Figure 3.5. Over and above that, Figure 3.5 illustrates that the node size corresponds to the number of papers published by each research institution, the connecting lines delineate the intensity of collaboration between different institutions, and the color of the lines signifies the collaborative relationships across distinct periods. Institutions such as Harvard University, Harvard Medical School, and Brigham & Women’s Hospital have a longer history of collaboration.

The peak productivity of institutions was primarily concentrated in the 2013-2014 period, with most of the research institutions represented in yellow, indicating rapid emergence and relatively shorter research duration. These findings align with the shifts in publication trends observed during this timeframe. More recent research institutions in this field include Central South University, Zhejiang University, Yonsei University Health System, and others.

3.5. Authorship distribution. The total number of papers authored by an individual in a journal to some extent signifies the academic standing of that author within the field. The author collaboration network provides a clear depiction of the core author groups and their cooperative relationships in the research domain. For this paper analysis, the analysis node within CiteSpace software was configured as “author,” and the amassed literature data underwent visual analysis. The knowledge graph portraying authors and their collaborative networks is presented in Figure 3.6. Additionally, in Figure 3.6, the font and node size correspond to the number of papers published by each author, while the connections between nodes delineate the cooperative relationships between different authors. Similarly, the line thickness indicates the extent of collaboration.

As observed in Figure 3.6, the network encompasses 492 nodes and 566 connections, with an overall network density of 0.0047, signifying robust collaborative ties among authors in the research domain. The most extensive collaborative author network within this field includes figures such as Hu, Frank B., Willett, Walter C., Rimm, Eric B., Rexrode, Kathryn M., and others. Notably, the outer circle of authors, including Willett, Walter C., Rexrode, Kathryn M., and Chan, Andrew T., is depicted in red, indicating recent article contributions. In terms of the number of articles published by these authors, Hu, Frank B., Willett, Walter C., and Rimm, Eric B. secured the top three positions. There were also 12 authors with ten or more articles published. The top ten authors with the highest number of publications are detailed in Table 3.3, accounting for 7.7% of the total. Considering the cooperative degree among research authors, it is evident that the primary authors maintain a relatively high level of cooperation. This signifies the establishment of a close-knit and mature cooperative network within this field. Generally, a substantial correlation exists between highly prolific authors and the density and intensity of their collaborative networks, thereby fostering denser cooperation networks.

To explore author relationships from a time series perspective, this paper employs the Timezone (Time
Fig. 3.5: Time zone diagram of Institutions

Table 3.3: Top ten authors with several publications

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Author</th>
<th>Year</th>
<th>Numbers</th>
<th>Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hu, Frank B</td>
<td>2013</td>
<td>39</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>Willett, Walter C</td>
<td>2013</td>
<td>36</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>Rimm, Eric B</td>
<td>2015</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>Bommer, Ann</td>
<td>2016</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Rexrode, Kathryn M</td>
<td>2017</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>Chan, Andrew T</td>
<td>2016</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>Kubzansky, Laura D</td>
<td>2017</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>Halcomb, Elizabeth</td>
<td>2015</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Manson, JoAnn E</td>
<td>2014</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Missmer, Stacey A</td>
<td>2016</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

The function in CiteSpace, depicting author relationships along a coordinate axis with time as the horizontal parameter, as displayed in Figure 3.7. In this time zone diagram, the node size represents the frequency of an author’s presence, the year associated with each node denotes the author’s initial appearance, and the color of the lines connecting nodes signifies the timing of the author’s co-appearances.

As depicted in Figure 3.7, Hu and Frank B emerge as nodes with the highest number of publications in the related literature, commencing their publication year in 2013. These nodes exhibit extensive connections and a prolonged timespan, highlighting the significant academic status and reference value associated with this author and their work in this field. Over time, the number of authors contributing to related studies increased, and other prolific authors have an extended publication history, indicating the field’s sustainability. Remarkably, authors such as Willett, Walter C., Chan, Andrew T., Bommer, Ann, Kubzansky, and Laura D., among others, have both a high publication output and recent contributions, making the exploration of their research trajectories potentially valuable to this field.
3.6. High-frequency keywords. Co-word analysis primarily involves extracting title information, such as keywords and abstracts, from citations and forming an informative knowledge map through statistical analysis. Research into high-frequency keywords can elucidate the prevailing trends in the field of chronic disease nursing over a specific period. The software’s operating timeframe was set as “2013-2022,” with a threshold of $K = 15$, YearPerSlice configured as “1,” and pruning performed annually and across the entire network. This
facilitated visual analysis, leading to the creation of a co-occurrence map of frequently used keywords in the literature, showcased in Figure 3.7. Within Figure 3.8, 332 high-frequency keywords were identified, forming 428 connections. The node size and text denote keyword frequencies, while the lines connecting nodes signify associations established during various periods. The thickness and density of these lines reflect the intensity of keyword co-occurrence. Notably, “nursing” emerges as the largest node, followed by “chronic disease” and “primary health care.” In terms of historical presence, keywords like nursing, chronic disease, primary health care, and chronic obstructive pulmonary disease have appeared early. More recently, terms like burnout, kidney transplantation, critical illness, quality of health care, and coronavirus have surfaced, potentially signifying new research directions in chronic disease nursing.

On top of that, the mediating centrality of keywords serves as a pivotal metric for evaluating research hotspots and scholars’ primary areas of interest in this field. Analyzing the mediation centrality index, which represents nodes’ facilitating influence (as shown in Table 3.4), reveals that “incidence,” “rheumatoid arthritis,” and “qualitative” exhibit strong connectivity with other prominent keywords. This suggests that these keywords often lie within the communication path with other terms, actively contributing to the mutual citation relationships among literature.

Methodologically, the use of keywords encapsulates the essential content of scholarly work, and by conducting a co-occurrence analysis of high-frequency keywords, we can pinpoint research hotspots within the domain of chronic disease nursing. The inter-mediation centrality value offers insight into the significance and impact of keywords, with higher values indicating greater mediating influence. Table 3.4 showcases the occurrence frequency and inter-mediation centrality values (Centrality ≥ 0) of keywords in the field of chronic disease nursing. As revealed by the centrality values in Table 3.4, “incidence” boasts the highest centrality value (Centrality ≥ 0.34) and exhibits the closest associations with other keywords. Notably, “rheumatoid arthritis,” “qualitative,” and other keywords also display substantial intermediate centrality values (Centrality ≥ 0.3). Considering both keyword occurrence frequency and centrality values, it becomes evident that the primary research foci in chronic disease nursing revolve around “incidence,” “rheumatoid arthritis,” and “qualitative.”

3.7. Keywords clustering. To intuitively visualize the research hot topics within the papers found in the WOS literature big database, we employed CiteSpace software along with the LLR algorithm for keyword co-occurrence cluster analysis. The resulting keyword clustering view is depicted in Figure 3.9, where color blocks delineate distinct clusters, each containing associated keywords. The analysis encompasses N = 332 keywords,
Table 3.4: Top ten centrality of keywords

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Keywords</th>
<th>Frequency</th>
<th>Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Incidence</td>
<td>15</td>
<td>0.34</td>
</tr>
<tr>
<td>2</td>
<td>Rheumatoid arthritis</td>
<td>10</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td>Qualitative</td>
<td>55</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>Chronic heart failure</td>
<td>43</td>
<td>0.28</td>
</tr>
<tr>
<td>5</td>
<td>Cardiovascular disease</td>
<td>67</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>Adherence</td>
<td>48</td>
<td>0.24</td>
</tr>
<tr>
<td>7</td>
<td>Nurse practitioners</td>
<td>27</td>
<td>0.21</td>
</tr>
<tr>
<td>8</td>
<td>Chronic</td>
<td>23</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>Prevalence</td>
<td>31</td>
<td>0.18</td>
</tr>
<tr>
<td>10</td>
<td>Treatment</td>
<td>16</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Fig. 3.9: Spectrum of keywords clustering

$E = 1110$ connections, and a network density of $0.0202$. Also, the size of module $Q$, a measure related to node density, plays a crucial role in scientific cluster analysis, with a larger $Q$ indicating a more effective clustering result. The average silhouette value $S$ gauges the homogeneity of clusters, with higher values indicating greater credibility. In this context, Figure 3.9 reveals a $Q$ of 0.54227, signifying a well-structured network with a favorable clustering effect. The associated $S$ value of 0.7154 underscores the high homogeneity of clusters, showcasing a clear distinction between different clusters. This figure showcases ten clusters, spearheaded by “Chronic Illness,” “self-management,” and “nursing home.” That said, the primary clusters have an average inception around 2014-2016, signifying a period of maturity in related studies. The largest cluster, “Chronic Illness,” with an initiation year of 2013, comprises 50 keywords, with key terms such as nursing, chronic disease, qualitative research, and caregivers, among others. The main keywords for each cluster are summarized in Table 3.5.

3.8. Keywords time zone analysis. To further explore the development and evolution of research over time, the analysis of keywords within the papers from the WOS literature big database was carried out using the time zone map feature in CiteSpace. As illustrated in Figure 3.8, the size of each node signifies the
### Table 3.5: Keywords time zone analysis

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Clustering name</th>
<th>Numbers</th>
<th>S</th>
<th>Year</th>
<th>Main keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Chronic illness</td>
<td>50</td>
<td>0.673</td>
<td>2016</td>
<td>Chronic illness (171.43, 1.0E-4); nursing (156.72, 1.0E-4); qualitative research (97.66, 1.0E-4); caregivers (87.04, 1.0E-4)</td>
</tr>
<tr>
<td>1</td>
<td>Self-management</td>
<td>46</td>
<td>0.699</td>
<td>2015</td>
<td>Self-management (144.95, 1.0E-4); self-care (113.15, 1.0E-4); self-efficacy (99.87, 1.0E-4)</td>
</tr>
<tr>
<td>2</td>
<td>Nursing home</td>
<td>39</td>
<td>0.674</td>
<td>2014</td>
<td>Nursing home (129.62, 1.0E-4); dementia (128.91, 1.0E-4); elderly (105.91, 1.0E-4)</td>
</tr>
<tr>
<td>3</td>
<td>Chronic kidney disease</td>
<td>33</td>
<td>0.704</td>
<td>2016</td>
<td>Chronic kidney disease (264.35, 1.0E-4); hemodialysis (89.86, 1.0E-4); peritoneal dialysis (85.15, 1.0E-4); education (82.7, 1.0E-4)</td>
</tr>
<tr>
<td>4</td>
<td>Depression</td>
<td>33</td>
<td>0.663</td>
<td>2016</td>
<td>Depression (227.75, 1.0E-4); anxiety (154.97, 1.0E-4); quality of life (95.92, 1.0E-4); physical activity (80.33, 1.0E-4); mental health (68.92, 1.0E-4)</td>
</tr>
<tr>
<td>5</td>
<td>Qualitive</td>
<td>27</td>
<td>0.711</td>
<td>2016</td>
<td>Qualitative (60.29, 1.0E-4); chronic (37.67, 1.0E-4); health care (27.02, 1.0E-4); implementation (24.85, 1.0E-4)</td>
</tr>
<tr>
<td>6</td>
<td>Mortality</td>
<td>27</td>
<td>0.758</td>
<td>2016</td>
<td>Mortality (90.25, 1.0E-4); risk factors (87.43, 1.0E-4); epidemiology (85.53, 1.0E-4); pulmonary rehabilitation (78.92, 1.0E-4); copd (74.92, 1.0E-4)</td>
</tr>
<tr>
<td>7</td>
<td>Primary care</td>
<td>26</td>
<td>0.733</td>
<td>2016</td>
<td>Primary care (238.24, 1.0E-4); general practice (174.01, 1.0E-4); nurse practitioner (75.35, 1.0E-4); primary health care (72.55, 1.0E-4)</td>
</tr>
<tr>
<td>8</td>
<td>Telemedicine</td>
<td>21</td>
<td>0.817</td>
<td>2015</td>
<td>Telemedicine (171.13, 1.0E-4); telehealth (132.73, 1.0E-4); mhealth (72.38, 1.0E-4); hypertension (56.84, 1.0E-4); chesl (52.95, 1.0E-4)</td>
</tr>
<tr>
<td>9</td>
<td>COVID-19</td>
<td>15</td>
<td>0.824</td>
<td>2020</td>
<td>COVID-19 (173.91, 1.0E-4); Sars-COV-2 (111.21, 1.0E-4); pandemic (55.38, 1.0E-4); healthcare workers (49.17, 1.0E-4); Coronavirus (49.03, 1.0E-4)</td>
</tr>
</tbody>
</table>

3.9. Timeline of keywords analysis. A two-dimensional timeline, referred to as the “Timeline graph,” was utilized to display literature keyword clustering, providing researchers with insights into the evolving processes and cutting-edge trends within topic clusters. The Timezone function in CiteSpace was used to analyze the keywords, shedding light on the development and evolution of research within the WOS literature big database. The size of each node in Figure 3.11 reflects the frequency of the keyword, while the year associated with each node indicates when the keyword was first used. The connecting lines between nodes reveal instances where different keywords appeared together in the same articles, representing relationships of inheritance and keyword’s frequency, and the year assigned to each node marks the keyword’s initial appearance. Connecting lines between nodes represent instances where different keywords appear in the same articles simultaneously, revealing relationships of inheritance and evolution across different periods. This approach not only helps identify primary areas of research focus during hot periods but also elucidates the stages of development within the field. As seen in Figure 3.10, the most significant node is “nursing,” introduced in 2013. In the early studies, high-frequency keywords included chronic disease, primary health care, chronic obstructive pulmonary disease, and elderly care, among others. These related concepts have spanned a significant timeframe and had a substantial impact. Research has continued to this day, introducing fresh concepts such as burnout, kidney transplantation, and critical illness in recent studies.
Leveraging Visualization Technology and Extensive Literature Data for Chronic Disease Knowledge Analysis

Fig. 3.10: Time zone diagram of keywords

Fig. 3.11: Timeline diagram of keywords
evolution over time. This approach combines the number of publications over the years, providing insights into primary research areas during hot periods and illustrating the field’s developmental stages. Figure 3.11 reveals that the most prominent node is “nursing,” introduced in 2013. Early studies featured high-frequency keywords such as chronic disease, primary health care, chronic obstructive pulmonary disease, and elderly care, among others. These concepts have exerted significant and lasting influence. Recent studies have introduced novel concepts like burnout, kidney transplantation, and critical illness, signifying new directions in the field.

Additionally, the Timeline graph, featured in Figure 3.11, presents the clustering of literature keywords on a two-dimensional timeline. This provides researchers with valuable insights into the evolution and cutting-edge trends of specific topic clusters and the mutual relationships between these hot topics. The graph displays different color-coded clusters, each representing a set of important keywords within the same cluster. The top 10 clusters included #0 chronic illness, #1 self-management, #2 nursing home, #3 chronic kidney disease, #4 depression, #5 qualitative, #6 mortality, #7 primary care, #8 telemedicine, and #9 COVID-19. Figure 3.11 shows that the largest cluster in related literature was “chronic illness,” consisting of 50 keywords, with an average year of 2016. Key terms in this cluster included nursing, chronic disease, qualitative research, caregivers, and more. Over time, additional keywords such as medication management and intensive care units made their appearance. The cluster report generated by the system highlighted Facchinetti, G. (2019) and their study titled “Discharge of Older Patients with Chronic Diseases: What Nurses Do and What They Record. An observational study.”

3.10. Keywords burstness. The table in section 3.10 (Table 3.6) reveals emergent keywords in the research field over the past decade. The “beginning year” indicates when a particular keyword’s frequency began to surge, and the “end year” represents when that keyword’s frequency started stabilizing. The intensity of emergence reflects the degree of a sudden increase in a keyword’s frequency during its emergence period, often correlated with its research popularity. Keywords with red bars signify their relevance in specific durations.
The table encompasses 30 emergent keywords, and by considering their commencement times, “randomized controlled trial,” “chronic disease management,” and “malnutrition” emerge as early research focal points. From a duration perspective, keywords like “randomized controlled trial,” “screening,” “evidence-based practice,” and “readmission” have maintained their relevance over an extended period, indicating their prolonged status as research hotspots. Regarding the strength of emergent keywords, “COVID-19” (Strength = 35.66), “phenomenology” (Strength = 7.65), “Sars-COV-2” (Strength = 6.98), and “humans” (Strength = 6.95) exhibit substantial sudden intensity, signifying significant changes in their frequency of occurrence. In summary, “COVID-19,” “Sars-COV-2,” “public health,” and “medical staff” not only boast high emergence intensity but
also closely align with the current timeline, suggesting that they represent the latest emerging research hotspots.

In general, as time progresses and society evolves, along with shifts in the external environment, the research content and hotspots within Chronic disease nursing continue to change. This dynamic landscape underscores the enduring research value of Chronic disease nursing from a different perspective.

3.11. Literature co-citation analysis. The analysis of literature co-citation serves to identify the interconnections between co-cited works within a specific research field, shedding light on influential literature that has a substantial impact on both the field itself and related disciplines. The quantity of co-citations directly correlates with the strength of associations between works and the significance of high-level literature. Illustrated in Figure 3.12, this co-citation network features 346 nodes, 455 connections, and a network density of 0.0076, highlighting several prominent co-citation relationships. Notably, the works by Braun V. (2006), Tong A. (2007), and Wagner E.H. (2001) emerge as key figures, boasting relatively high citation frequencies. For instance, Tong, A.; Sainsbury, P.; Craig, J.’s extensive search of various sources for existing checklists used to assess qualitative studies, including systematic reviews and major medical journals, has garnered a remarkable 15,637 citations [15]. Braun, V.; and Clarke, V., known for their outline of thematic analysis and its relation to other qualitative analytic methods, have accumulated 4,406 citations [3]. Wagner E.H., whose work focuses on the challenges faced by those with chronic illnesses in accessing appropriate medical care, particularly in systems designed for acute illnesses, has amassed 4,127 citations [18].

Furthermore, Bodenheimer T., in their exploration of the potential of the computer revolution in improving primary care, particularly through systems aimed at enhancing physician performance and patient outcomes, has garnered 243 citations [1]. Meanwhile, Lorig, K.R.; Holman, H.R.’s work on self-management tasks and skills has accumulated 2,296 citations, emphasizing its profound influence within the field [11]. In addition, the centrality of literature co-citation identifies works like Barlow J. (2002), Higgins J.P.T. (2003), and Bodenheimer T. (2002) as frequently cited classical references within the literature. A comprehensive list of the top ten cited papers is available in Table 3.7.

4. Conclusions. To gain insights into research hotspots and development trends within the domain of chronic disease nursing, this study employed visual analysis techniques on literature within the extensive WOS literature database, examining parameters such as the annual number of published papers, geographical and institutional distributions, authorship, high-frequency keywords, keyword clustering, keyword time zone analysis, keyword burstness, and co-citation relationships. Notably, several papers related to chronic disease nursing-
hibited a steady increase from 2013 to 2021, signaling sustained global scholarly interest. Within this research, 141 journals published a significant number of papers on the application of artificial intelligence technology in nursing, totaling 3,610 published papers, constituting 58.41% of all publications. The top three journals with the highest publication volume were “Journal of Clinical Nursing,” “Journal of Advanced Nursing,” and “BMC Health Services Research.” Among the authors, twelve individuals authored ten or more articles, with Hu, Frank B., Willett, Walter C., and Rimm, Eric B. leading the way, collectively contributing to 110 papers or 7.7% of the total. Their prolific output suggests a strong cooperative network among key researchers in this field.

Additionally, the research uncovered the top ten research institutions by publication frequency, including...
Harvard University, Harvard Medical School, Brigham & Women’s Hospital, and others, amounting to 19.29% of publications. Harvard University, Harvard Medical School, and Brigham & Women’s Hospital featured prominently, indicating a closely-knit cooperation network among these core institutions. Furthermore, the analysis of countries or regions revealed the USA, Australia, and England as the leading contributors, publishing 5,491 papers and accounting for 46.28% of the total. The centrality of these publications correlated positively with their quantity, while countries such as China and Canada exhibited disproportionality between their publication volumes and centrality, suggesting room for improved international collaboration. Similarly, cooperation degrees in the USA, Australia, and England stood at 0.16, 0.13, and 0.23 respectively. Interestingly, certain papers such as those authored by Braun V. (2006), Tong A. (2007), Wagner E.H. (2001), Bodenheimer T. (2002), Lorig K.R. (2003), Barlow J. (2002), Charlson M.E. (1987), Folstein M.F. (1975), Zignod A.S. (1983), Graneheim U.H. (2004) garnered widespread citations, underscoring their substantial impact on the field. Current research hotspots encompass incidence, rheumatoid arthritis, and qualitative aspects, as well as emerging areas like burnout, kidney transplantation, critical illness, COVID-19, Sars-COV-2, public health, and medical staff. These findings reveal the evolving nature of research in chronic disease nursing, underscoring its enduring research value in a dynamically changing landscape. Data supporting this study’s findings are available from the corresponding author upon request. The authors declare no known financial or personal conflicts of interest that could have influenced the reported work.

In conclusion, this study has provided a comprehensive analysis of the research landscape within the field of chronic disease nursing, revealing key trends and hotspots. The increasing number of papers published over the years underscores the enduring relevance of this field to scholars across the globe. The substantial prevalence of publications related to artificial intelligence applications in nursing further highlights the field’s dynamism. Prolific authors and core institutions with high cooperation degrees exemplify a closely-knit scholarly community actively contributing to this domain. Additionally, influential papers with widespread citations illustrate the pivotal role of certain research in shaping the discourse. As research evolves, new emerging hotspots, such as burnout, kidney transplantation, and COVID-19, indicate the field’s responsiveness to evolving societal and environmental changes. This study underscores the continued value of chronic disease nursing research and its ability to adapt to the evolving healthcare landscape.

Based on the findings of this study, it is recommended that healthcare professionals, researchers, and policymakers continue to invest in and support research in the field of chronic disease nursing, with a particular focus on emerging areas such as burnout, kidney transplantation, and COVID-19. This will enable a more comprehensive understanding of the evolving healthcare landscape and help address the challenges associated with chronic disease management. One limitation of this study is that it primarily relies on data from the WoS literature database, which may not encompass all relevant research. Additionally, the analysis is based on quantitative metrics and may benefit from qualitative insights and interdisciplinary perspectives to provide a more holistic understanding of the field. Future research could explore the qualitative aspects of chronic disease nursing, including the experiences of patients, healthcare providers, and caregivers. Lastly, interdisciplinary collaborations and mixed-methods approaches could provide a richer understanding of the field’s dynamics and the impact of healthcare policies on chronic disease management.

Data availability. The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgment. The authors thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

Leveraging Visualization Technology and Extensive Literature Data for Chronic Disease Knowledge Analysis


Edited by: Zhigao Zheng
Special issue on: Graphi Powered Big Aerospace Data Processing
Received: Nov 4, 2023
Accepted: Jan 8, 2024