A bibliometric analysis of the research hotspots in the applications of internet searches in breast cancer patients

Cui Mengyao(1), Jin Feng(1), Jin Zining(1), Yu Xinmiao(1), Cui Lei(2)*

*Correspondence:lcui@cmu.edu.cn

1. Department of Breast Surgery, the 1st Affiliated Hospital, China Medical University, No.155 Nanjing Nan Street, Heping District, Shenyang
2. College of Medical Informatics, China Medical University, 77 Puhe Road, Shenbei Xin District, Shenyang.

Abstract

Background

Women diagnosed with breast cancer need information and support to cope with their illness. The internet has become a powerful way to disseminate health information. The aim of this study is to investigate the research hotspots in the field of the application of the internet in breast cancer patients for information exchange and dissemination.

Material and Methods

Papers published between 1998 and 2016 with major Medical Subject Headings (MeSH) from related citations were downloaded from the PubMed database. Terms appearing in more than 10 papers were kept as high frequency MeSH terms. Terms that appeared in more than 10 papers were considered high frequency MeSH terms and were retained. A co-occurrence matrix was built with co-occurrence times between each pair of high frequency MeSH terms. These terms were then clustered into four groups.
Papers that fell outside of our initial search timeline (published in 2017-2018) were classified into one of the 4 existing clusters of hotspots to predict the emerging trends of the field.

**Results and Analysis**

A total of 459 MeSH terms appeared 1594 times in 313 citations and 26 MeSH terms presented more than 10 times in 809 citations.

We clustered these high frequency MeSH terms into 4 groups. Topics of those 18 new papers include online support group related, diagnosis, treatment and prevention, but showed a significant content expansion of each cluster.

**Conclusions:**

From 1998 to January 2018, internet searches about breast cancer focused on 4 aspects: (1) patients utilized the internet to seek psychological support, (2) patients sought explanations of diagnostic information, (3) breast cancer survivors sought education during their drug therapy duration, and (4) the impact of a web-based patient decision aid for breast cancer treatment options.

Recent studies expanded to specific problems, with features including more social science theories and techniques applied in this field, more clinical applications by physicians, and more social media used by breast cancer survivals and physicians.

**Keywords:** Breast cancer, internet, hotspots, co-word analysis, cluster analysis

**1. Background**

Breast cancer is the most prevalent cancer in women worldwide. According to
data from GLOBOCAN 2012 by the World Health Organization [1], while breast cancer is the most occurring cancer in women, with 1,671,149 current diagnoses, the estimated number of deaths due to breast cancer is 521,907, implying that over 1,000,000 survivors are living all over the world.

An important goal in health care is to sustain the long-term physical health of patients, as well as their mental health and psychological well-being. Given the role of hormones in breast cancer, it is not surprising that nearly 50% of the women with early breast cancer had depression, anxiety, or both in the year after diagnosis, 25% in the second, third, and fourth years, and 15% in the fifth year[2]. Because quality of life of breast cancer patients is of widespread concern, health education of breast cancer patients is indeed a public health problem.

During the course of the illness, women diagnosed with breast cancer seek information and support to cope with the illness. The internet is a powerful way to disseminate health information, popular due to its ease of use, lack of limitations regarding availability, and convenience of access, and lower cost. Therefore, patients are willing to use the internet as a medium to search for and exchange health-related information.

Studies on the application of internet usage among breast cancer patients has increased rapidly. We searched the PubMed database for papers about internet usage and breast cancer and retrieved 334 papers published from 1998-2018. Indeed, excluding the sole paper published in 2018 (as this year was not completed at the time of publication), we see a linear growth in the cumulative number of related citations
by its publishing year (see Figure 1) \(y= 17.874x–35748, R^2 = 0.9565\). However, to date, no known reviews have been published examining the use of the internet by breast cancer patients.

Figure 1: Cumulative number of papers examining breast cancer and internet usage year.

Therefore, the aim of this study is to investigate the research hotspots in the application of internet usage in breast cancer patients for information exchange and disseminations.

2. Materials and Methods

2.1 Materials

We searched the PubMed database with queries “breast neoplasms [majr] AND internet [majr]”, yielding a total of 334 citations (as of January 19, 2018). We
separated these citations into 2 sets; set A includes papers published between 1998-2016, to present research hotspots in this field, and set B includes papers published in 2017 and 2018, to present the recent progress in this field.

2.2 Methods

(1) Hotspots in 1998-2016 by co-occurrence cluster analysis of high frequency MeSH terms

Co-occurrence cluster analysis of high frequency subject headings is often utilized to discover and present the research hotspots of a scientific field [3-5]. Authors of similar scientific fields tend to frequently use the same subject headings to demonstrate concepts within their areas of interest. Connecting these subject headings by their similarity with co-occurrence strength allows researchers to cluster them into groups and then to present the research hotspots of particular field or topic.

According to this principle, we conducted our study with the following steps:

1. Extracting the subject headings and calculating their frequency

We extracted the major MeSH headings with the subject headings from all citations in set A, counted the frequency of each major MeSH heading/subheading, and defined those terms appearing in more than 10 papers as the high frequency MeSH terms. To easily extract major headings from each article, we developed a package within Visual Basic, entitled Bibliometric Items Co-Occurrence Matrix Builder (BICOMB). These major MeSH/subheadings are the core concepts of the hotspots of the utilization of internet searches for breast cancer, per the existent literature.
2. Building the co-occurrence matrix of high frequency MeSH terms and performing a cluster analysis

Using the BICOMB package, we calculated the co-occurrence of each pair of high frequency MeSH terms and built a co-occurrence matrix. We then entered this matrix into gCLUTO (Graphical CLUstering Toolkit), a platform dedicated to cluster analysis, [6] and clustered these high frequency MeSH terms into 4 groups according to the performance indices of cluster analysis quality measurement.

3. Interpreting the content of each cluster and assigning the cluster tags to these clusters

For each group of high frequency MeSH terms, we manually analyzed the semantic relationship from the core of the cluster tree (see Figure 2). For example, there are 6 MeSH terms in cluster 0, they are “Breast neoplasms/psychology”, “Communication”, “Social Support”, “Self-Help Groups”, “Adaptation/psychological”, and “Internet/organization & administration”. Among these MeSH terms, “Breast neoplasms/psychology” and “Social Support” shared the greatest degree of similarity and therefore comprise the first 2 MeSH terms of the core of this cluster. We can then infer that this cluster is about the social support provided for breast cancer patients’ psychological issues. Then we can expand the content of this cluster by adding the third MeSH term, “Communication”, which now implies that the cluster refers to the social supports to the breast cancer patients’ psychological issues by communication. We also attempted to confirm the cluster content by the representative papers of each cluster. The representative papers are those papers that
contain as many of the MeSH terms from the same cluster. For example, we input a query “Breast neoplasms/psychology [majr] AND Social Support [majr]” into PubMed and retrieved 454 citations, however, by adding the third MeSH term “communication [majr]”, we retrieved 17 citations. Those titles and abstracts retrieved from these searches reflect the semantic relations among these MeSH terms, such as “social networks, social support mechanisms, and quality of life after breast cancer diagnosis”. Therefore, we confirmed our reference about the content of cluster 0. By utilizing the approaches described above, we confirmed four hotspots in this field.

(2) Recent advances by content analyzing papers published in 2017 and January 2018

Content analysis was conducted with a focus on both disease and internet aspects. We then classified all of the papers into one of the four existing clusters of hotspots (as shown in Table 3), to predict the emerging research directions of internet searches and breast cancer.

3. Results and Analysis

3.1 General statistics and MeSH Terms

*High frequency MeSH terms in the field of internet and breast cancer.*

In set A, a total of 459 MeSH terms appear 1,594 times in 313 citations and 26 MeSH terms are presented more than 10 times, distributed throughout 809 citations. This yields a cumulative frequency of 50.75% (809/1594), meaning 5.66% (26/459) of all MeSH terms are involved more than half of papers in this field. Therefore, these 26 MeSH terms should be representative of the research hotspots of this field.
Table 1. List of high frequency MeSH terms extracted from citations on breast cancer and the internet (1998-2016)

<table>
<thead>
<tr>
<th>Rank</th>
<th>MeSH/subject headings</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Internet</td>
<td>222</td>
</tr>
<tr>
<td>2</td>
<td>Breast Neoplasms / Psychology</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>Breast Neoplasms</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Internet / Utilization</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>Breast Neoplasms / Therapy</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>Social Support</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>Breast Neoplasms / Prevention &amp; Control</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>Self-Help Groups</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>Breast Neoplasms / Diagnosis</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>Patient Education as Topic</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>Patient Education as Topic / Methods</td>
<td>19</td>
</tr>
<tr>
<td>12</td>
<td>Breast Neoplasms / Surgery</td>
<td>17</td>
</tr>
<tr>
<td>13</td>
<td>Internet / Standards</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>Mammography</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>Breast Neoplasms / Diagnostic Imaging</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>Breast Neoplasms / Genetics</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>Social Media</td>
<td>13</td>
</tr>
<tr>
<td>18</td>
<td>Decision Making</td>
<td>12</td>
</tr>
<tr>
<td>19</td>
<td>Information Seeking Behavior</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>Breast Neoplasms / Drug Therapy</td>
<td>12</td>
</tr>
<tr>
<td>21</td>
<td>Internet / Organization &amp; Administration</td>
<td>12</td>
</tr>
<tr>
<td>22</td>
<td>Adaptation, Psychological</td>
<td>12</td>
</tr>
<tr>
<td>23</td>
<td>Communication</td>
<td>11</td>
</tr>
<tr>
<td>24</td>
<td>Health Knowledge, Attitudes, Practice</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>Survivors / Psychology</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>Breast Neoplasms / Epidemiology</td>
<td>10</td>
</tr>
</tbody>
</table>

3.2 Co-occurrence clustering analysis of high frequency MeSH terms from 1998-2016

Figure 2 shows one of the results of the two direction clustering by gCLUTO, entitled heat map. This figure shows high frequency MeSH terms on the right side (vertically), and their corresponding hierarchical tree on the left side is their corresponding hierarchical tree. On the bottom of the map there are the IDs (the PMID in PubMed) of papers containing those high frequency MeSH terms, and on the top of the map there is a hierarchical tree of these papers.
To clearly illustrate the clustering results, we summarized each high frequency MeSH term and the corresponding titles in Table 2.

It is important to consider that there are two aspects in this topic: a disease (breast cancer), and a platform/tool used for communication and information gathering (the internet). Therefore, we discussed the research hotspots represented in each cluster in two dimensions.

1. On the disease dimension, cluster 0 relates to the psychological problems of breast cancer patients (breast neoplasms/psychology), however, on the internet dimension, it is about the organization and administration of the internet (internet/organization & administration). The outstanding problem in the combination of these two aspects is that breast cancer patients utilize the internet to seek psychological support, with particular differences in the form of online support groups compared to searches of how to improve their health.

2. Cluster 1 is comprised of the diagnostic imaging of breast cancer (breast neoplasms/diagnosis).
neoplasms/diagnosis, breast neoplasms/diagnostic imaging) and the quality of information on the internet (internet/standards). After breast cancer patients are informed of the diagnosis by mammography, health educators seek information on how to educate these patients via the internet; they investigate the informational characteristics such as relevance, comprehensiveness, and personalized on websites and social media.

Table 2: List of clusters of high frequency MeSH terms and the titles of the representative papers
<table>
<thead>
<tr>
<th>C</th>
<th>MeSH terms</th>
<th>Titles of Representative papers</th>
</tr>
</thead>
</table>
| 0 | Breast Neoplasms/psychology, Social Support, Self-Help Groups, Internet/organization & administration, Adaptation, Psychological Communication | 1. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. PMID: 18235163  
2. Giving and receiving peer advice in an online breast cancer support group. PMID: 23659724  
3. Effects of prayer and religious expression within computer support groups on women with breast cancer. PMID: 17131348  
4. Emotional coping differences among breast cancer patients from an online support group: a cross-sectional study. PMID: 24499687 |
| 1 | Breast Neoplasms, Breast Neoplasms/diagnosis, Patient Education as Topic, Internet/standards, Mammography, Breast Neoplasms/diagnostic imaging | 1. Mammography Patient Information at Hospital Websites: Most Neither Comprehensible Nor Guideline Supported. PMID: 27610583  
2. Designing a tailored Web-based educational mammography program. PMID: 21099540  
3. Characteristics of YouTube™ Videos Related to Mammography. PMID: 25502853  
4. Quality of health information on the Internet. PMID: 11694141 |
| 2 | Internet/utilization, Breast Neoplasms/therapy, Social Media, Information Seeking Behavior, Breast Neoplasms/drug therapy, Survivors/psychology | 1. Online discussion of drug side effects and discontinuation among breast cancer survivors. PMID: 23322591  
2. Understanding topics and sentiment in an online cancer survivor community. PMID: 24395991  
3. Why do patients seek an alternative channel? The effects of unmet needs on patients' health-related Internet use. PMID: 20390984  
4. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. PMID: 24345206 |
2. Use of Online Communication by Patients With Newly Diagnosed Breast Cancer During the Treatment Decision Process. PMID: 27468161  
3. Efficacy of a web-based intelligent tutoring system for communicating genetic risk of breast cancer: a fuzzy-trace theory approach. PMID: 24829276  
4. Less scarring or more symmetry? Reconstruction following metachronous bilateral breast cancer. PMID: 22560454 |

(3) Cluster 2 relates to the drug therapy of breast cancer on disease dimension
(breast neoplasms/ drug therapy), and relates to concerns on the utilization of the internet (internet/utilization) on the internet dimension. This research direction focuses how information is being transmitted and how patients seek relevant information during drug therapy duration.

(4) The last cluster contains only one MeSh that represents internet, namely “internet”. However, there are multiple disease related MeSH terms, including “breast neoplasms/ prevention & control”, “breast neoplasms/ surgery”, “breast neoplasms/genetics” and “breast neoplasms/ epidemiology”. While it seems complicated to find the semantic relations among them, we may refer to one special case could help us to link all of them together. Angelina Jolie, a famous actress, underwent a preventive double mastectomy following the discovery of BRCA gene, a gene known to be linked to an increased risk of breast cancer[7]. The three remaining MeSh terms “patient education as topic/methods”, “decision making” and “health knowledge, attitudes, practice” complete the MeSH terms. Studies with these terms include the impact of a web-based patient decision aid for breast cancer treatment options, specifically for patients who are informed of a genetic risk of breast cancer.

After summarization above clusters, we found some characteristics about the research hotspots in this field:

(1) What events would evoke patients’ information needs, driving them to seek for relative information?

Those are key points to supply them tailored information. Soon after they obtain diagnostic information (ie. mammography), they usually want to know the meaning of
the picture and data. During their treatment process, they would need information on drug side effects and psychological support. Finally, when they are warned for a high genetic risk of breast cancer, they desire information to explain related evidence based practices and to support them making a decision on a preventive mastectomy.

(2) What is the appropriate channel by which one should supply patients with information?

For any psychological problems that develop during the course of the disease, online support groups on the internet are easy to use and they provide feedback instantly, so it is a convenient platform for patients to communicate everyday issues about breast cancer. Official web pages monitored by health organizations usually provide a one direction communication channel (information is only disseminated from web to patients), but they are expected to provide more authority and accurate guidelines for diagnosis and treatment. Users of social media are permitted to respond to the content on the web pages and even upload multiple media data (pictures or video) in order to discuss the problems they face. Given the interactive nature of this medium, it may have potential as an information exchanging tool.

3) How do you ensure qualified information is delivered to the patients?

The unregulated nature of the internet allows the information on it to have great variations in quality. More professionals should be involved in the provision of health related knowledge in order to provide authority and accurate explanations about the diagnostic and genetic screening test results.
### 3.3 Content analysis of papers published 2017-January 2018

Among all of the 18 papers published between 2017 and January of 2018, 3 of them are related to online support groups, 5 are about diagnosis, 6 are about treatment (not limited in drug therapy), and 4 are about prevention. However, there is a significant content expansion of each cluster, so classification of an appropriate cluster was difficult.

<table>
<thead>
<tr>
<th>Cluster No. and MeSH terms of Disease</th>
<th>Titles and PMIDs of papers</th>
</tr>
</thead>
</table>
| 0 Breast neoplasms/psychology        | Factors Contributing to Dropping-out in an Online Health Community: Static and Longitudinal Analyses. PMID: 28269969  
Online support groups for young women with breast cancer: a proof-of-concept study. PMID: 28281051  
Development of Trust in an Online Breast Cancer Forum: A Qualitative Study. PMID: 28536093 |
| 1 Breast neoplasms/diagnosis, Breast neoplasms/diagnostic imaging | Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. PMID: 28248621  
How do different delivery schedules of tailored web-based physical activity advice for breast cancer survivors influence intervention use and efficacy? PMID: 27498099  
Frequencies of Private Mentions and Sharing of Mammography and Breast Cancer Terms on Facebook: A Pilot Study. PMID: 28600279  
Online Educational Tool to Promote Bone Health in Cancer Survivors. PMID: 28922062  
Clinical usefulness of the free web-based image analysis application ImmunoRatio for assessment of Ki-67 labelling index in breast cancer. PMID: 28298390 |
| 2 Breast neoplasms/drug therapy      | "I just googled and read everything": Exploring breast cancer survivors' use of the internet to find information on complementary medicine. PMID: 28735830  
Efficacy of Internet-Based Cognitive Behavioral Therapy in Improving Sexual Functioning of Breast Cancer Survivors: Results of a Randomized Controlled Trial. PMID: 28240966  
Pilot Testing a Web-Based System for the Assessment and Management of Chemotherapy-Induced Peripheral Neuropathy. PMID: 28002115 |
By content analysis of the recently published papers in 2017-January 2018 we concluded there are new emerging topics based on the existing 4 hot spots:

(1) More social science theories and techniques applied in this field.

As the research progresses, researchers are seeking theoretical foundations. For example, Hummel et al. evaluated the effect of internet-based cognitive behavioral therapy (CBT) on sexual functioning in breast cancer survivors [8]. Similarly, Abrahams et al. investigated the efficacy of internet-based CBT for severely fatigued survivors of breast cancer [9] and Diddi and Lundy discussed the application of the theoretical parameters of the Health Belief Model (HBM) in Organizational Twitter [10]. Zhang and Elhadad studied the factors contributing to dropping out of an online health community [11] and Lovatt, Bath, and Ellis studied development of trust in an online breast cancer forum [12].
(2) More clinical applications by physicians.

As shown in the results, the internet has been widely used among patients for online self-support and by the organizations for patient education, but in recent years it has started to be used by physicians, not only for communicating with their peers to discuss clinical problems [12], but also applied directly to aid their clinical practice, especially in their diagnosis and treatment plan. For example, Yeo et al. developed and evaluated a free web-based image analysis application for assessment of Ki-67 labelling index in breast cancer [13]. Additionally, Knoerl et al. developed a web-based system for the assessment and management of chemotherapy-induced peripheral neuropathy [14]. El Hage et al. reviewed literature of online prognostication tools, and concluded that online tools are valuable in guiding adjuvant treatment, especially in resource constrained countries. [15].

(3) More social media used by breast cancer survivals and physicians.

As social media has increased in popularity, it has simultaneously attracted more interest to investigate the characters of the utilization of this platform for communication. For example, Huesch et al. investigated the frequencies of private mentions and sharing of mammography and breast cancer terms in Facebook [16] and Diddi and Lundy conducted a content analysis of organizational tweets (via Twitter) during a breast cancer propaganda period [10], however, to date, there is no application study about the use of mobile phones presented in these 18 papers.

1. Conclusions
From 1998 to January 2018, the internet has been applied in breast cancer widely and tends to be focused into 4 aspects:

(1) Breast cancer patients utilize the internet to seek psychological support, particularly examining the effects of online support groups of breast cancer and how to improve patients’ health by different management methods.

(2) The education of patients by the internet on the explanation of diagnostic information, particularly on mammography.

(3) Education of breast cancer survivors during their drug therapy duration, how the information being transmitted, or how these patients seek relevant information.

(4) The impact of a web-based patient decision aid for breast cancer treatment options, especially while patients are informed of their genetic risk of breast cancer.

Recent studies expanded to specific problems, including more social science theories and techniques applied in this field, more clinical applications by physicians, and more social media used by breast cancer survivors and physicians.

1. List of abbreviations

   MeSH: Medical Subject Headings

   BICOMB: Co-Occurrence Matrix Builder

   gCLUTO: Graphical CLUstering TOolkit

2. Declarations
Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The software Co-Occurrence Matrix Builder (BICOMB) is available at 202.118.40.8/bc

The software Graphical CLUstering Toolkit (gCLUTO) is available at: http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/download

GLOBOCAN 2012. Graph production: global cancer observatory is available at: http://globocan.iarc.fr/

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CM and JF conceived and designed the study. CL carried out the searches and downloaded the related citations from PubMed. CM and JZ performed statistical analyses, including cluster analysis. JF interpreted the content of clusters. CM drafted the manuscript. All authors have participated in critically revising the manuscript, have read the final draft, have given final approval of the version to be published, and agree to be accountable for all aspects of the work.


