Different platforms for different patients’ needs: Automatic content analysis of different online health information platforms

Sanders, R.; Linn, A.J.; Araujo, T.B.; Vliegenthart, R.; Eenbergen, M.C. van; van Weert, J.C.M.

DOI
10.1016/j.ijhcs.2019.102386

Publication date
2020

Document Version
Final published version

Published in
International Journal of Human-Computer Studies

License
Article 25fa Dutch Copyright Act

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Different platforms for different patients’ needs: Automatic content analysis of different online health information platforms

Remco Sanders⁎, Annemiek J. Linn, Theo B. Araujo, Rens Vliegenthart, Mies C. van Eenbergen, Julia C.M. van Weert

Amsterdam School of Communication Research, University of Amsterdam, the Netherlands
Department of Research, Netherlands Comprehensive Cancer Organisation, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Information seeking behavior
Online health information
Information needs
Cancer
Automatic content analysis
Model of patients’ social support needs

ABSTRACT

Prior online health research has mainly focused on the predictors or outcomes of online health information, leaving online health information itself understudied. Therefore, online health information has remained an umbrella term encompassing different platforms (expert- vs. peer-generated). A hybrid method that combines qualitative and computational methods is used to identify different topics discussed on these different platforms, and an initial model of patients’ social support needs was developed and applied to data obtained from the hybrid method. Using topic modeling (Nposts = 52,990), topics on two expert- and two peer-generated platforms were identified. Differences between and within platforms were found. While peer-generated platforms mainly covered interaction on emotional support topics, expert-generated platforms covered informational topics. Within peer-generated platforms, patients used their experiences differently.

1. Introduction

The diagnosis of cancer causes patients and their surroundings a great amount of fear and uncertainty (Leydon et al., 2000; Sanson-Fisher et al., 2000). As a result, cancer patients and their relatives have many questions like “what is cancer?” or “what sort of treatment is available?”. Some of these questions get answered by a medical expert. However, many patients experience unmet cognitive and affective needs during their cancer journey, even after a discussion with a medical expert (hereafter called “social support needs”; Cotten and Gupta, 2004; Eysenbach et al., 2004; Puts et al., 2012; Silence et al., 2006; Iconomou et al., 2001). Acquiring health information online helps patients fulfill these social support needs (Goldsmith, 2004). Currently, up to 70% of chronically ill patients seek online health information (Abdul-Muhsin et al., 2015; Rupert et al., 2016).

Previous research focused either on the predictors or on the outcomes of using online health information (e.g., Cotten and Gupta, 2004; Puts et al., 2012; Sanders and Linn, 2018). However, we lack information on which content, if any, fulfills which social support needs. This information is important to have because online health information has the potential to “set the agenda” for future behavior (Anderson et al., 2003; McCombs and Shaw, 1972; Xiao et al., 2014). A first step is to disentangle the term online health information. This term has been used as an umbrella term to describe multiple platforms on the internet such as online blogs, medical libraries, question-and-answer pages, and discussion forums. These different platforms are likely to contain different topics and are expected to fulfill different social support needs (Anderson et al., 2003; Prestin et al., 2015). While scholars have advocated for a comparison between these platforms, to the best of our knowledge, no previous studies have investigated if and how content differs on these platforms (Coulson et al., 2007; Klemm et al., 2003).

Because online health information is often unreliable or difficult to understand, patients and medical experts should work together to fulfill patients’ social support needs (Rains and Young, 2009; Buiss and Whitten, 2011; Ruiten et al., 2005). Knowledge about the content of different platforms helps in this regard by deciding which platform to visit to fulfill specific support needs.

Using a hybrid method that combines a literature review and a computational social science method, the current study aims to investigate the differences in content between online health information platforms (Shah et al., 2015). A theoretical model is developed to serve as a roadmap to identify emerging topics and gain a better understanding of how online platforms relate to patients’ social support needs.

The current study adds to the literature in multiple ways. On a methodological level, this study bridges the gap between computational
methods and “classic” social science methods. The current study distinguishes itself by using a theoretical model as a starting point. On a theoretical level, this study integrates the diversity of various social support needs and makes a systematic and holistic analysis of the field of online health information research possible. We will demonstrate the usefulness of using a model of social support needs as a basis for unsupervised machine learning. On a practical level, the results allow us to understand which social support needs patients seemingly fulfill online instead of, or in addition to, fulfilling during appointments with a medical expert.

1.1. Towards a model of social support needs

The use of online health information is increasing, and this trend can be placed within a larger societal trend of people taking more control over all aspects of their lives (Beck et al., 1994; Lupton, 1997, 2013). Online health information-seeking behavior can be defined as, “... active efforts to obtain specific information in response to a relevant event” (Niederdeppe et al., 2007, p.154). In the current study, “the relevant event” is operationalized as being diagnosed with or living with (someone with) cancer. The “active effort” is operationalized as searching online for social support. A model will be developed that acknowledges that patients go online because of different social support needs. In this model, it is assumed that due to these different needs, patients search online to find different forms of social support on a variety of topics. To illustrate, users can fulfill their “abstract” cognitive and affective needs (layer I) by going online to find different forms of social support (layer II) which is provided to them via content covering different topics. The layers of the proposed model go from broad to specific. Previous research shows that patients go through similar phases during their health information-seeking process; normally the patient starts with a broad need (e.g., I want to know more about my disease, i.e., cognitive need) and this need becomes more specific through searching and browsing until that specific need is fulfilled (e.g., I want to know more about medicine X; Wilson et al., 2002). See Table 1 for the proposed model of patients’ social support needs.

1.2. Layer I: patients’ needs

The first layer of the framework consists of two basic underlying patients’ needs: cognitive and affective (Bensing and Verhaak, 2004; Rutten and colleagues, 2005). Patients have the need to know and understand (i.e., cognitive needs) as well as the need to feel acknowledged and understood (i.e., affective needs). An example of a cognitive need is the need for information about treatment. An affective need can be the need to receive reassurance. This classification is the first layer of the proposed model.

1.3. Layer II: social support categories

The cognitive and affective needs specified in layer I can be fulfilled by seeking for and receiving online social support, which is the second layer of the framework. The Social Support Behavior Code and the Optimal Matching Model (SSBC; Cutrona, 1990; Cutrona and Russell, 1990; Cutrona and Suhr, 1992; House, 1983) provide a useful categorization of how patients’ needs can be fulfilled by providing and receiving social support in an online context. Within social support, different categories can be distinguished, namely, informational support (fulfilling patients’ cognitive needs) and esteem, network, and emotional support (fulfilling patients’ affective needs). These categories are explained below.

Patients’ cognitive needs in layer I can be fulfilled by providing informational support. Informational support means providing guidance to users on topics such as the disease, symptom interpretation, and illness management (Coulson et al., 2007). Rutten et al. (2005) offer a typology that can be used to further specify the social support categories into social support topics that can be measured using topic modeling. This typology is based on a review clustering the support needs of cancer patients about certain health topics or issues. The typology describes nine cognitive informational support topics: cancer-specific, treatment-related, prognosis, rehabilitation, surveillance and health information, interpersonal/social information, financial/legal information, medical system information, and body image information. Based on other offline literature tangible assistance was added to informational support (Cutrona, 1990). For descriptions of the categories, see Table 1. These informational support categories and topics form the second and third layers of the model.

Rutten et al. (2005) only emotional support topic is coping information. The subcategories of coping information given by Rutten and colleagues (2005) seem to include all social support categories (e.g., emotional reactions, support from other patients, and spiritual support). Although previous studies show a correlation between the various support concepts, we believe there is also a clear theoretical distinction between network and esteem support and to a lesser degree, emotional support. Therefore, we use the definition of Rutten and colleagues (2005) in which the coping information category specifies the three forms of support. The following distinction is made: the subtopic coping and spiritual support reflects esteem support, the subtopic member interaction reflects network support, and the subtopic emotional support reflects emotional support. Esteem support refers to communicating respect and confidence in one’s ability to deal with the disease. Network support entails communicating one’s belonging to a group of people sharing the same concerns or (health) complaints. Emotional support refers to messages expressing empathy, love, and

| Table 1 |
| Model of patients’ social support needs. |

<table>
<thead>
<tr>
<th>Layer I: Needs</th>
<th>Layer II: Social support category</th>
<th>Sublayer social support topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive needs</td>
<td>Informational support</td>
<td>Cancer-specific information</td>
<td>Information about (diagnosis of) cancer, physical effects, test, symptoms.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment-related information</td>
<td>Information about available treatments, how treatment works and side effects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prognosis information</td>
<td>Change of cure, survival expectations, expectations for future health.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rehabilitation information</td>
<td>Self care, nutrition, availability medical supplies for at home</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surveillance and health information</td>
<td>Maintaining or restoring physical health after treatment and preventive behaviors.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpersonal/social information</td>
<td>Effect of disease on relations with family, friends, caregiver, employment, and social life in general.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Financial/legal information</td>
<td>Costs associated with disease (e.g., treatment, insurance, writing a testament).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical system information</td>
<td>Health care system, interactions with health care providers, evaluation of physicians.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Body image/sexuality information</td>
<td>Sexuality and physical appearance (i.e., attractiveness).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tangible assistance</td>
<td>Providing, or offering to provide, information, goods, and services.</td>
</tr>
<tr>
<td>Affective needs</td>
<td>Esteem support</td>
<td>Coping and and ss spiritual</td>
<td>Communicating respect and confidence in ones abilities to deal with disease (e.g., compliments and validation).</td>
</tr>
<tr>
<td>Network support</td>
<td>Emotional support</td>
<td>Member interaction</td>
<td>Communicating belonging to a group of persons with similar concerns or experiences</td>
</tr>
<tr>
<td>Emotional support</td>
<td></td>
<td>Emotional support</td>
<td>Communicating love, concern, empathy, expressions of emotions or caring, attentive listening, reassurance</td>
</tr>
</tbody>
</table>
2.1. Data

2. Method

2.1. Towards testing the model of patients’ social support needs

By grouping and combining the previously described layers, the model of patients’ social support needs is proposed (see Table 1). This model allows us to measure latent needs and social support categories via observable social support topics. By combining the outcomes of previous studies, covering multiple sources of information, and targeting different (cancer) populations, the model consistently conceptualizes and identifies different types of social support topics within online health information platforms (Buis and Whitten, 2011). This process will contribute to a greater generalizability of research findings in the future. A first step would be to map the content of these different platforms. The results of this first step can then be applied to the proposed model. The following research question is proposed: What social support topics can be distinguished online on cancer-related platforms, and to what extent do these topics fit within with the model (RQ1)?

3. Analysis

3.1. Topic modeling

Unsupervised computer-assisted content analysis was used to analyze the data without restricting the analysis based on previously defined categories in coding schemes. This technique has several benefits. First, it allows for the analysis of large amounts of data. This method can be used on multiple datasets to compare, unsupervised, the occurrence of topics on different platforms (i.e., without any interference from the researcher). Second, the analysis can be easily replicated by other researchers (Wang et al., 2015).

Latent Dirichlet Allocation (LDA) was used as a topic modeling method using the Gensim package (Rehůřek and Sojka, 2010; Blei et al., 2003). Topic modeling is a computer algorithm that identifies latent constructs by word co-occurrence, using word distribution across documents (e.g., individual posts and questions; Jacobi et al., 2016). Based on these documents, topics were created and displayed as clusters of words. These clusters contain the most-used words for each topic (see preprocessing for more information). For each topic, the latent meaning (topic) can be (subjectively) inferred from these words. To derive reliable and interpretable topics, several steps need to be taken within LDA.

3.2. Preprocessing

The raw data received from any online source, including the sources used in this study, is “noisy”. Therefore, work needs to be performed before analysis can begin (Petz et al., 2014). Data were preprocessed as follows: First, all html mark-up and punctuation was removed (e.g., <b> and nbsp). Second, stop words were removed using a custom made Dutch stop word list (e.g., en [and], te [too] and eens [once/ever]). Third, topic modeling was done. This step resulted in a list of top words, whose latent meanings can be regarded as topics. A list of top words per topic is provided, as well as a score that indicates the weighted presence of the top word per topic (topic loading). Based on these top words and Fig. 1, the outcome of the topic modeling was evaluated on how meaningful, consistent, and discriminative the identified topics were. Fourth, words in top words deemed unnecessary
or uninformative were removed by adjusting the code. For example, based on the first and second cycles, all names were replaced by the word *naam* [name] using a custom made list of popular Dutch names. The same procedure was repeated for all forms of cancer, which were replaced by the word *kankervorm* [type of cancer]. This process was used since an individual name or type of cancer was not needed to answer the current research questions. Fifth, because the results of automatic stemming for the Dutch language was too rudimentary (i.e., resulting in uninterpretable and unreadable words), multiple words were grouped together manually by adding this information to the code (e.g., *hooii* and *hallo* became *hoi* [hi]). To limit the influence of the noninformative difference between multiple conjugations of verbs on the topics, verb conjugations were grouped by the verb stem.

### 3.3. Parameters

To run the LDA model, values for the number of expected topics and the *alpha* parameter needed to be determined. Values of 0.10, 0.05, 0.01, 0.005, and 0.001 were tried. In line with the recommendation of Jacobi et al. (2016), a low alpha value was used to obtain fewer topics per document (i.e., *α* = 0.001).

To establish the number of topics, analyses were performed on 5, 10, 15, 20, 25, 30, and 40 topics. For a maximum of 25 topics, the meaning of each topic could be identified without too much overlap between topics. Next, consensus about the meaning of the 25 topics was reached during meetings with the research team.

### 3.4. Differences between platforms

To answer the second research question, for all 25 topics, linear equations from topic modeling were applied to the sample of posts. This process resulted in a score for all 25 topics per document that measured the extent to which these topics occurred within each document. Then, per platform mean topic scores were calculated for all 25 topics. Finally, MANOVA was used to gain insight into the differences in the mean topic scores between the platforms. The 25 topics were used as dependent variables and the platforms as independent variables. When the Levene’s F test showed significant differences between the four platforms, Bonferroni post hoc tests were used to gain more insights into these differences.

### 4. Results

#### 4.1. Data

The post sample size was 52,995. This sample of documents consisted of all blog posts and replies (n = 22,662), all forum entries (n = 9,573), all expert questions and answers (n = 16,458), and all entries into the medical libraries (n = 4,338). The average number of words per document for each platforms was $M_{blogs} = 78.55$ (SD = 159.89), $M_{forum} = 53.03$ (SD = 64.01), $M_{questions} = 41.62$ (SD = 42.26), and $M_{library} = 163.72$ (SD = 131.47).
<table>
<thead>
<tr>
<th>Subtopic number</th>
<th>Subtopic name</th>
<th>Topwords</th>
<th>Subtopic description</th>
<th>Example post</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sharing current status</td>
<td>Hi, cure, chemo, first day</td>
<td>Sharing previous experiences with cancer is an important part of the treatment process. Patients share their recent experiences to help others navigate their journey.</td>
<td>Hi everyone, today I had the first day of my chemo cure. I'm feeling pretty well, only a little dizzy in my head. (Blog)</td>
</tr>
<tr>
<td>2</td>
<td>Sharing previous status</td>
<td>Hi, cure, chemo, first day</td>
<td>Sharing previous experiences with cancer is an important part of the treatment process. Patients share their recent experiences to help others navigate their journey.</td>
<td>Hi, cure, chemo, first day</td>
</tr>
</tbody>
</table>
4.2. Topic identification

We wanted to gain insight into which social support topics could be distinguished on different cancer-related platforms of a health website and the extent to which these subtopics fit within the model of social support needs (RQ1). Many of the subtopics identified are similar to the social support topics that are part of the model of patients’ social support needs. This similarity means many support needs identified within previous literature on offline information sources exist online. Fig. 1 provides a visual representation of the subtopics. Furthermore, the percentages displayed under Fig. 1 show the relative proportion of a subtopic in the complete sample of posts (i.e., the dominance of a certain subtopic within the content analyzed). Furthermore, Fig. 1 shows several (partly) overlapping subtopics, or subtopics that were similar to each other (i.e., certain words belong to different subtopics).

In Table 2, the integration of the subtopics into the model is summarized, and example posts of each subtopic are given. We elaborate on this process below per layer of the social support model. Which Fyru Furthermore, we discuss the extent to which the predefined topics can be identified in the results of our topic model.

4.2.1. Layer II: social support category—informational support

Based on our theoretical model we expected to find topics ranging from cancer-specific information to tangible assistance that could be grouped under this support category (see Table 3). The analysis revealed that several subtopics were grouped under the social support topic cancer-specific information, such as cancer tests (subtopic 9) and cancer detection (subtopic 14). Additionally, several subtopics were part of the social support topic treatment-related information: cancer-treatment information (subtopic 3), treatment side effects (subtopic 6), other (alternative) treatments (subtopic 19), gastrointestinal treatment (subtopic 21), and treatment consequences (subtopic 23). Thus, when people talk about cancer-specific information, they often use words related to for example detection such as lymphedema, carcinoembryonic antigen test (CEA), contrast fluid, etc. When people talk online about treatment-related information, they use words related to cancer treatment information such as type of cancer or treatment(s).

Subtopics covering rehabilitation were found and grouped under the social support topic rehabilitation information. These subtopics are rehabilitation (subtopic 10) and after care (subtopic 12). In addition, subtopics covering the social support topic surveillance and health information included nutrition (subtopic 17) and hereditary (subtopic 22). We found that the subtopics emotions and cancer (subtopic 4) and impact cancer (relatives) (subtopic 11) were grouped under the social support topic interpersonal/social information. The social support topic financial/legal information included the subtopics insurance/financial (subtopic 15) and employment (subtopic 18). To illustrate, content about rehabilitation information features words like fatigue, recovery, and complaints. However, content included in the subtopic interpersonal/social information featured words such as life, emotions, fear, and feeling.

Within the social support topic medical system information, the following subtopics were found: cancer organizations (subtopic 13), cancer research (subtopic 16), and medical trials (subtopic 20). The social support topic referral to other information sources included the topics: information sources (subtopic 24) and external organizations (subtopic 25). In summary, the medical system information contained content around the words such as Dutch Cancer Society [kwf], trial, research, and scientific.

Based on our theoretical model we expected the presence of the social support topics tangible assistance, prognosis information, and body image/sexuality information. However, the results of our topic model included no such subtopics.

4.2.2. Layer II: social support—esteem, network and emotional support

Also for this layer, we investigate the extent to which and how the
under the social support topic named on the status of relatives. Because the subtopics are better grouped about their current status and reactively posted their previous status or contain similar top words). Here, users proactively posted status updates support from others and groups topic 7). Furthermore, the library than it is in all other platforms, including expert questions information cancer specific 13, 16, 20, 24, and 25). To illustrate, within the topic cancer specific information, the subtopic cancer tests (subtopic 9) is covered more in the library than it is in all other platforms, including expert questions (Mlibrary = 13.95, Mblog = 2.38**), Mforum = 1.67*, and Mquestions = 1.83*, see Table 3). Within the topic treatment-related information, the subtopic cancer treatment information is most dominant, i.e., this subtopic was covered more often in the library than in all other platforms (see subtopic 3 in Table 3). The same holds true for the following topics within informational support: rehabilitation information (rehabilitation; subtopic 10), surveillance and health information (nutrition; subtopic 17), financial and legal information (employment; subtopic 18), medical system information (cancer organizations; subtopic 13), and referral to other sources (information sources; subtopic 24). The topic interpersonal/social information showed mixed findings. While emotions and cancer (subtopic 4) were covered more within the library than they were in all other platforms, impact cancer (on relatives) (subtopic 11) was covered more on blogs than it was in all other platforms. In sum, the library showed the highest coverage in topics concerning general cancer-related information regarding the cancer patients’ whole journey, from diagnosis through after care.

4.2.3. Different platforms for different needs

We explored how peer- and expert-generated cancer-related platforms differ from each other in terms of the social support provided. Overall, the MANOVA shows significant differences between the platforms of all 25 subtopics (see Table 3 for the complete table). While the social support topics were not strictly divided between the platforms, the platforms differed in the degree to which social support topics were present in the different platforms.

5. Informational support

Informational support subtopics were found most in the library platform of the website (9, 14, 3, 6, 19, 21, 23, 10, 12, 17, 22, 4, 15, 18, 13, 16, 20, 24, and 25). To illustrate, within the topic cancer specific information, the subtopic cancer tests (subtopic 9) is covered more in the library than it is in all other platforms, including expert questions (Mlibrary = 13.95, Mblog = 2.38**), Mforum = 1.67*, and Mquestions = 1.83*, see Table 3). Within the topic treatment-related information, the subtopic cancer treatment information is most dominant, i.e., this subtopic was covered more often in the library than in all other platforms (see subtopic 3 in Table 3). The same holds true for the following topics within informational support: rehabilitation information (rehabilitation; subtopic 10), surveillance and health information (nutrition; subtopic 17), financial and legal information (employment; subtopic 18), medical system information (cancer organizations; subtopic 13), and referral to other sources (information sources; subtopic 24). The topic interpersonal/social information showed mixed findings. While emotions and cancer (subtopic 4) were covered more within the library than they were in all other platforms, impact cancer (on relatives) (subtopic 11) was covered more on blogs than it was in all other platforms. In sum, the library showed the highest coverage in topics concerning general cancer-related information regarding the cancer patients’ whole journey, from diagnosis through after care.

5.1. Esteem, network, and emotional support

Within the social support topic coping and spiritual, the subtopic affective support (subtopic 7) is prominent. This topic was covered more on blogs than it was on all other platforms (Mblog = 4.97, Mforum = 2.70*, Mlibrary = 3.44*, and Mquestions = 2.40*). Within the topic support from others and groups, member interaction (subtopic 5) was found, which was covered more on forums than it was on all other platforms. For sharing status, a nuanced image appears: sharing current status (subtopic 1) and sharing status relatives (subtopic 8) were covered more on blogs than they were on discussion forums and other platforms (Mblog_subtopic1 = 19.28, Mforum_subtopic1 = 5.05*, Mlibrary_subtopic1 = 1.40*, Mquestions_subtopic1 = 3.05**), Mblog_subtopic8 = 4.85, Mforum_subtopic8 = 2.64*, Mlibrary_subtopic8 = 2.52*, and Mquestions_subtopic8 = 2.02*). Sharing previous experience, however, was found equally on blogs and discussion forums, where it was found more often than it was on the library and question-and-answers platforms.

Table 3

Mean topic scores comparison per platform.

<table>
<thead>
<tr>
<th>Social support category</th>
<th>Social support topic</th>
<th>#</th>
<th>Subtopic</th>
<th>Bloga</th>
<th>Forumb</th>
<th>Libraryc</th>
<th>Questionsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informational support</td>
<td>Cancer-specific information</td>
<td>9</td>
<td>Cancer tests</td>
<td>2.38bcd</td>
<td>1.67bc</td>
<td>13.95bcd</td>
<td>1.83**</td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Cancer-detection</td>
<td>14</td>
<td>1.26bcd</td>
<td>0.59bc</td>
<td>7.73bcd</td>
<td>0.44bc</td>
<td></td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Cancer-treatment information</td>
<td>3</td>
<td>2.58bcd</td>
<td>4.56bc</td>
<td>50.54bcd</td>
<td>4.32**</td>
<td></td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Treatment side effects</td>
<td>6</td>
<td>3.51bcd</td>
<td>2.45bc</td>
<td>9.74bcd</td>
<td>2.48bc</td>
<td></td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Other (alternative) treatments</td>
<td>19</td>
<td>0.53bc</td>
<td>0.62abcd</td>
<td>1.92abcd</td>
<td>0.49bc</td>
<td></td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Colorectal cancer</td>
<td>21</td>
<td>0.48c</td>
<td>0.43c</td>
<td>2.50bcd</td>
<td>0.41c</td>
<td></td>
</tr>
<tr>
<td>Treatment-related information</td>
<td>Treatment consequences</td>
<td>23</td>
<td>0.61bcd</td>
<td>0.47bc</td>
<td>0.74bcd</td>
<td>0.46c</td>
<td></td>
</tr>
<tr>
<td>Rehabilitation information</td>
<td>Rehabilitation</td>
<td>10</td>
<td>2.11bcd</td>
<td>2.09bcd</td>
<td>14.60bcd</td>
<td>1.53bc</td>
<td></td>
</tr>
<tr>
<td>Rehabilitation information</td>
<td>After care</td>
<td>12</td>
<td>1.32bcd</td>
<td>1.79bcd</td>
<td>8.66bcd</td>
<td>0.73bcd</td>
<td></td>
</tr>
<tr>
<td>Surveillance and health information</td>
<td>Nutrition</td>
<td>17</td>
<td>0.90bcd</td>
<td>0.70c</td>
<td>2.85bcd</td>
<td>0.54c</td>
<td></td>
</tr>
<tr>
<td>Interpersonal / social information</td>
<td>Hereditary</td>
<td>22</td>
<td>0.65bcd</td>
<td>0.41bcd</td>
<td>1.84bcd</td>
<td>0.29bcd</td>
<td></td>
</tr>
<tr>
<td>Financial / legal information</td>
<td>Impact cancer (on relatives)</td>
<td>11</td>
<td>4.98bcd</td>
<td>1.22bcd</td>
<td>0.95c</td>
<td>0.72bc</td>
<td></td>
</tr>
<tr>
<td>Financial / legal information</td>
<td>Emotions and cancer</td>
<td>4</td>
<td>6.80bcd</td>
<td>4.32bcd</td>
<td>18.08bcd</td>
<td>2.11bc</td>
<td></td>
</tr>
<tr>
<td>Medical system information</td>
<td>Insurance/financial</td>
<td>15</td>
<td>1.40bcd</td>
<td>0.97bcd</td>
<td>1.84bcd</td>
<td>0.50bcd</td>
<td></td>
</tr>
<tr>
<td>Medical system information</td>
<td>Employment</td>
<td>18</td>
<td>1.00bcd</td>
<td>0.74abcd</td>
<td>2.43abcd</td>
<td>0.40abcd</td>
<td></td>
</tr>
<tr>
<td>Medical system information</td>
<td>Cancer organisations</td>
<td>13</td>
<td>0.60bc</td>
<td>0.95abcd</td>
<td>10.42bcd</td>
<td>0.57bc</td>
<td></td>
</tr>
<tr>
<td>Medical system information</td>
<td>Cancer research</td>
<td>16</td>
<td>0.97bcd</td>
<td>1.14abcd</td>
<td>3.33abcd</td>
<td>0.56abcd</td>
<td></td>
</tr>
<tr>
<td>Medical system information</td>
<td>Medical trails</td>
<td>20</td>
<td>0.78bcd</td>
<td>0.44abcd</td>
<td>1.55abcd</td>
<td>0.31abcd</td>
<td></td>
</tr>
<tr>
<td>Referral other information sources</td>
<td>Information sources</td>
<td>24</td>
<td>0.40bcd</td>
<td>0.29bc</td>
<td>0.97abcd</td>
<td>0.23bcd</td>
<td></td>
</tr>
<tr>
<td>Referral other information sources</td>
<td>External organisations</td>
<td>25</td>
<td>0.36bcd</td>
<td>0.21abcd</td>
<td>0.90bcd</td>
<td>0.16abcd</td>
<td></td>
</tr>
<tr>
<td>Esteem support</td>
<td>Coping and spiritual</td>
<td>7</td>
<td>Affective support</td>
<td>4.97bcd</td>
<td>2.79abcd</td>
<td>0.34abcd</td>
<td>2.40abcd</td>
</tr>
<tr>
<td>Network support</td>
<td>Support from others and groups</td>
<td>5</td>
<td>Membership interaction</td>
<td>5.28bcd</td>
<td>6.09bcd</td>
<td>1.12abcd</td>
<td>5.54abcd</td>
</tr>
<tr>
<td>Emotional support</td>
<td>Sharing status</td>
<td>1</td>
<td>Sharing current status</td>
<td>19.53abcd</td>
<td>5.05abcd</td>
<td>1.40abcd</td>
<td>3.05abcd</td>
</tr>
<tr>
<td>Emotional support</td>
<td>Sharing previous status</td>
<td>2</td>
<td>10.20bcd</td>
<td>10.51abcd</td>
<td>2.79abcd</td>
<td>9.48abcd</td>
<td></td>
</tr>
<tr>
<td>Emotional support</td>
<td>Sharing status relatives</td>
<td>8</td>
<td>4.85bcd</td>
<td>2.64bcd</td>
<td>2.52ab</td>
<td>2.02abcd</td>
<td></td>
</tr>
</tbody>
</table>

Note. All letters, unless otherwise specified, indicate p < .001. For example: bc indicates the mean score for that topic differ significantly from the mean score for that topic on the platforms forum, library, and questions with p < 0.001.

* p < .05, ** p < .01.
6. Discussion

This study investigated the different social support needs and topics on online health information platforms. A theoretical model integrating social support needs and topics was developed (i.e., the model of patients’ social support needs) and applied to different online platforms using computer-assisted content analysis. Results showed differences between expert- and peer-generated platforms and differences within these types of platforms. Expert-generated platforms generally included informational support and related social support topics focusing on cognitive needs, whereas peer-generated platforms generally included affective support and related social support topics focusing on affective needs. When looking at the differences within one type of platform we found that users talk differently about their experiences on different peer-generated platforms (blogs and forums). These results demonstrate that the platforms fundamentally differ in terms of content. As these platforms may influence the fulfillment of social support needs we can expect that these platforms also affect patient outcomes differently. Future research should therefore explicitly distinguish which platforms are being studied.

On a methodological level, the current study showed the benefits of using a hybrid method, in which a computational social science method is combined with a theoretical a priori–defined model. It allowed us to use a theoretical model to aid interpretation of the results but at the same time to remain open for unexpected findings. For example, many of the subtopics identified were similar to the social support topics of the theoretical model. However, the results indicated that the original model should be extended with the additional social support topic named sharing status. To provide informational support, users shared (previous) status updates about themselves or relatives with other users. Although not included in previous typologies of patient needs, previous research highlighted the patient’s need to provide their knowledge, quote their experiences, and share their progress online (Eysenbach et al., 2004; Gooden and Winefield, 2007; Coulson, 2005; Wang et al., 2015). If a supervised machine learning approach was used instead of an unsupervised machine learning approach this topic would have remained missing in the proposed model. In sum, the current approach allowed for additional topics to emerge from the data.

While most of the social support topics as identified in the theoretical model were found, three topics were not identified (i.e., prognosis information, body image, and tangible assistance). This finding does not necessarily mean that these topics were not covered in the online context per se. We chose to conduct our analysis with 25 topics. While 25 topics made interpretation of the topics possible, this number might not have been sufficiently discriminative to differentiate between very specific topics. Another explanation can be found in the definition used. For example, the definition used to identify tangible assistance was maybe difficult to find online (Coulson et al., 2007). Words in this definition such as “concrete” and “physical action” were previously mentioned concerning offline tangible assistance (Coulson et al., 2007). Future research should redefine the definition appropriate to the online context. The definition should cover examples such as “I’ll email you details of this diet ... I hope it helps!” or “I am happy to help any way I can ... just let me know” (Coulson, 2005, p.583). While most of these examples do not provide tangible evidence of support, the willingness to help is evident in this definition.

To further specify and expand the proposed model, the current study should be replicated using different analyses and perspectives. As an illustration, our model can be used as a starting point for future research for supervised machine learning on different online platforms. By applying this method, a more nuanced picture and more specific topics can be identified. For example, using annotated data on how certain topics are discussed would offer a deeper understanding of the nature of patients’ needs. While the current study included multiple, highly-used platforms, these platforms were all part of one source, a Dutch website, that focused on one illness, i.e., cancer. Therefore, generalizability to other countries, platforms, illnesses, and cultures is unknown. For instance, popular social media platforms such as Facebook, Twitter, and YouTube contain “countless discussions of health, illness, and medicine” (Lupton, 2016, p.55). These platforms offer different experiences to the user (e.g., more direct contact, closer-knit groups, and videos) and receive scarce scholarly attention. Analyzing these platforms using the current theoretical model could replicate and validate our findings as well as further improve our understanding of online health information platforms. The model can be further expanded by including more types of diseases across more countries and cultures. Especially when the diseases are considered taboo. In this latter case, online health information might be even more important to these patients. Similar differences in reliance on different sources of health information could be found in other countries and cultures in which medical experts’ expertise is more highly valued. Possibly this limits the dependence on online health information. Future research in other countries showing a higher difference in power between individuals (e.g., China, Japan, and Brazil; Hofstede, 2011), and thus more reliance on experts for information could broaden our understanding of the role online health information in the context of other sources.

Last, researchers are encouraged to consider group interaction on blogs and discussion forums. Due to the current individual focus in our model, the kinds of needs fulfilled by the online interaction between users remain unclear. Given the large degree to which these responses are given, clearly, some needs are fulfilled. Furthermore, due to the nature of the analysis applied, only the needs and content of posters are analyzed. However, previous research showed most users do not actively write on these platforms but merely read them (i.e., lurkers; Coulson, 2005). Future research should therefore apply different techniques (e.g., pop-up online surveys or focus groups) to identify the needs and motives of this group of users.

Technological advances are changing the way patients search for information, manage their health and the patient-provider relationship (Asan and Carayon, 2017). Technological and health care advances provide the means to more effectively and less expansively manage chronic diseases (Stephanidis et al., 2019). With these technological advances and the availability of online health information platforms, more content (i.e., data) is created (Shah et al., 2015; Topol, 2015; Sillence et al., 2006). How this data is gathered, shared and, used can further improve our understanding of how different stakeholders use this information to manage health (Victorelli et al., 2019; Branford, 2019). As a result of this ever-increasing amount of data, the need to better order and search content online becomes a priority. Insights originating from the current study could influence practice in this regard. Two main suggestions that will help aid and improve content ordering, patients’ health information-seeking behavior, and in turn, patient-provider interaction will be discussed. On the one hand, design factors in the interface could help users to navigate and match content to their needs. The model of patients’ social support needs can be used as a blueprint for ordering online content. By ordering the content by the proposed model, the content will follow the same route from very “abstract” social support needs towards increasingly precise needs and topics (Wilson et al., 2002). Second, using the algorithms to annotate the data based on patients’ social support needs, the content can automatically be accompanied by visuals that indicate to the users which topic is discussed. For example, platforms can include visual cues such as an icon of a brain for cognitive content or an icon of a heart for affective content. Alternatively, parts of texts that match the social support needs of the user can be highlighted or include an exclamation mark that would signal patients to the relevancy of those parts based on their indicated social support needs. This enables patients to intuitively browse and more easily find content based on their needs. A well-designed website, with a focus on information seeking (ordering) and integration (visual cues), could benefit patients by limiting the possibility of encountering misinformation and/or reducing cognitive...
Declaration of Competing Interest

None.

References


