Author: Overberg, Regina Ingrid
Title: Breast cancer stories on the internet : improving search facilities to help patients find stories of similar others
Issue Date: 2013-09-10

Breast Cancer Stories on the Internet

Regina Overberg

Improving search facilities to help patients find stories of similar others

BREAST CANCER STORIES ON THE INTERNET

Breast cancer stories on the Internet:
Improving search facilities to help patients find stories of similar others

ISBN/EAN: 978-90-9027748-6

Cover illustration: Tineke van Dijk, bewerkte gouache. www.tinekevandijk.nu
Lay-out: Textcetera, Den Haag
Printed by: Wilco, Amersfoort

The studies presented in this thesis were supported by the Netherlands Organisation for Scientific Research (NWO). The studies are part of the Narrator project (umbrella program ToKeN2000).

Breast cancer stories on the Internet

Improving search facilities to help patients find stories of similar others

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 10 september 2013
klokke 15.00 uur
door

Regina Ingrid Overberg

geboren te Stadskanaal
in 1978

Promotiecommissie

Promotor: Prof. dr. J.H.M. Schonk

Co-promotor: Dr. W. Otten (TNO, Leiden)

Leden: Prof. dr. A.M. van Dulmen (NIVEL, Utrecht/ UMC St Radboud, Nijmegen)
Prof. dr. J.C.J.M. de Haes (AMC, Amsterdam)
Prof. dr. ir. A.M. Stiggelbout

There Is A Light That Never Goes Out
The Smiths

TABLE OF CONTENTS

1. General introduction 9

2. Illness stories on the Internet: Features of websites disclosing breast cancer patients' illness stories in the Dutch language 25
Patient Education and Counseling 2006; 61: 435-442

Addendum to Chapter 2 41

3. Spontaneously published illness stories on a website for young women with breast cancer: do writers and themes reflect the wider population? 53
Informatics for Health and Social Care 2013; 38(1): 54-66

4. Illness stories on the Internet: what do breast cancer patients want at the end of treatment? 73
Psycho-Oncology 2007; 16: 937-944

5. How breast cancer patients want to search for and retrieve information from stories of other patients on the Internet: an online randomised controlled experiment 89
Journal of Medical Internet Research 2010; 12(1): e7

6. Searching for breast cancer stories online: What topics and/or writer characteristics do patients search for? 117
Submitted

7. Searching for breast cancer stories on a website using verbose natural language queries: an exploratory study 135

8. General discussion 153

Summary 169

Samenvatting 173

Appendices 177

Curriculum Vitae 181

List of publications and presentations 183

Dankwoord 185

CHAPTER 1

General introduction

IRMA’S STORY

Irma, 39 years old, lives with her husband and her two children of primary school age. After she was diagnosed with breast cancer, she was treated successfully and is now cancer free. However, since her treatment she has been very tired and her arm remains painful. She would like to know how other patients with similar health issues manage family life and returning to work. She therefore decides to search on the Internet for other people’s stories.

SITUATION IN 2008

Irma types the words ‘breast cancer’ and ‘stories’ into a general Internet search engine. More than 100,000 hits! She does not know where to start. She opens websites and closes them again; they are not what she is looking for. After half an hour, she finds a website that was created by an active group of patients and ex-patients. There are dozens of stories on this website, arranged alphabetically according to the writers’ pseudonyms. She starts to read stories. The writers of these stories are all in different situations. How can she find a story by someone in a similar situation to herself and with similar health issues? She continues scrolling through the list of stories. After one hour she finds a story that matches her own situation. However, while scrolling through the stories, she has also seen several stories by women with metastases. These are stories she would rather not have seen. The one story she did find after an hour may have given her support and information, but the other stories she scrolled through made her anxious.

SITUATION IN 2015?

Using natural language, Irma enters an excerpt of her own story into an Internet search engine: “I am 39 years old, married, two children. After having undergone several treatments, I am now cancer free, but I am still very tired and I have problems using my arm. I’m looking for ways to manage family life and returning to work despite these physical limitations”. After she types this, a number of websites are retrieved. She opens them and discovers that they contain similar stories to her own. Because the stories perfectly match her situation, they immediately give an answer to her question. With little effort she has found stories written by other breast cancer patients that are very helpful to her, without having to read stories that make her anxious.

During the various phases of disease (diagnosis, treatment, recovery) and facing “normal” life again, people require information. Besides the information that one gets, for example, from nurse practitioners about treatment options and the course of the disease, patients often want to know how they are going to experience this process. In such cases, stories written by other patients and published on the Internet can be an important source of information.

Due to the large number of stories on the Internet, it can be difficult for patients to find stories by other patients who are similar to themselves and whose stories therefore may contain information that is relevant and useful to them. If patients cannot find stories by people in a similar situation, they miss out on the opportunity of gaining information and support that matches their situation.

This thesis aims to gain insight into which search facilities for stories facilitate breast cancer patients in finding stories written by other patients who are in a similar situation to their own. Organizations can use the results to implement these search facilities on their websites.

This chapter provides a general introduction to the main themes and presents the aim and outline of the thesis.

Other breast cancer patients as a source of information

Breast cancer is a severe, life-threatening illness with a high prevalence. In 2009, there were more than 93,000 women in the Netherlands who had been diagnosed with breast cancer between 1999-2009. In 2010, about 14,000 Dutch women were newly diagnosed with breast cancer. Breast cancer also occurs in men, although it is uncommon (it is estimated that carcinoma of the male breast accounts for 0.8% of all breast cancers) [1]. Due to an increasing incidence (aging population, early detection of cancer) and improved survival rates (more effective treatments), the number of breast cancer patients will continue to rise [2].

Being diagnosed with breast cancer is a stressful life event that can cause uncertainty [3], severe distress, anxiety, and depression [4]. Patients have to undergo surgery on an organ that plays an important role in body image and sexual functioning [5]. Moreover, the natural sense of continuity is disrupted due to the possibility of dying [6]. Besides the psychological consequences, the often arduous treatments can cause long-lasting or even irreversible physical side effects, such as tiredness, lymphoedema, nerve pain, and early menopause [7]. Patients may also experience problems with social functioning, for example in their role within the family or in returning to work [8].

As a means of coping with their situation, people diagnosed with breast cancer often seek information [9;10]. Health care professionals are an important source of medical information, such as the likelihood of cure, spread of disease, and treatment options [11].

In addition to medical information from health care professionals, other breast cancer patients are also an important source of information during all phases of the disease. The information that other patients provide is essentially different from the information health care professionals provide. Trial and error of the lived experience makes other patients experts in strategies for coping with day-to-day personal health issues and in managing responsibilities and activities associated with family, friends, work, and the home during and after illness [12].

Nowadays, the Internet is an attractive and important source of other patients' experiences, because it can be accessed 24/7 without the need to leave your own house or to disclose your identity [13-15]. Moreover, the Internet contains a large number of patient experiences [16] that can be accessed by many people, since ninety-four percent of all households in the Netherlands have access to the Internet [17].

Breast cancer patients can read about other patients' experiences on the Internet in several ways: a) online support groups [15;18-20]; b) online patient communities, such as PatientsLikeMe [21]; c) social media, such as Facebook, Twitter and YouTube [22-27]; and d) websites with stories written by patients, such as healthtalkonline.org [28], BreastCancerStories.org [29] and de-amazones.nl [30]. The website healthtalkonline.org was developed by researchers, who interviewed a group of patients about their experiences with breast cancer and published these accounts in an edited form on their site. The stories on the websites BreastCancerStories.org and de-amazones.nl are published spontaneously by patients, without any restrictions or editing. This thesis focusses on searching for and reading these spontaneously published breast cancer stories on the Internet.

Spontaneously published breast cancer stories on the Internet

We define spontaneously published breast cancer stories on the Internet as first-person texts that have been written by breast cancer patients, are not part of an interactive forum and were spontaneously published on the Internet at a given moment. Reading these stories differs from reading other types of patient experiences on the Internet in three ways.

Firstly, spontaneously published breast cancer stories tend to be complete stories, meaning that several phases of the disease trajectory are described, such as experiences with treatments and the illness trajectory, feelings about having cancer, and coping with the consequences of being ill. The power of a complete story is that it can integrate events and the writer's thoughts, actions, decisions, and relationships with other persons into a meaningful narrative [31]. In this way, spontaneously published breast cancer stories differ from online support groups and social media where a reader can only read conversations between patients (mostly relatively short fragments in question-answer format) [32], and which, quite often, contain small talk [33].

Secondly, the writers are not selected and their spontaneously published stories are not edited: each patient can publish their own story on the Internet, with self-chosen topics and in self-chosen words. The reader thus has access to the authentic story of another patient, without any modifications and/or additions by professionals/researchers. Frank [34] indicates that such stories that come directly from the patient are important, because they reflect the narrator's truth. In this respect, spontaneously published stories differ from stories published by researchers which are based on patient interviews (e.g. [28]). A particular group of patients may be invited for the interviews and, during the interviews, patients may be asked to describe or write about specific topics to ensure that the set of stories addresses a broad range of topics. Furthermore, the stories will be edited before publication.

Thirdly, the number of spontaneously published stories available on the Internet is not limited and is constantly increasing, since patients can spontaneously publish their story at any time. This is in contrast to stories published by researchers. When researchers have finished interviewing a defined group of patients, they publish these stories at a given moment on their website. As a consequence, the number of stories on researchers' websites will not grow continuously.

The large and increasing number of spontaneously published stories on the Internet will, on the one hand, increase the likelihood that a story that matches the reader's situation becomes available, yet, on the other hand, may makes it difficult to find such a story.

The three above-mentioned characteristics of spontaneously published breast cancer stories show that they are a unique source of information and support. Therefore, this thesis focusses on these spontaneously published stories. To gain insight into the kind of information a reader may find in spontaneously published stories, we performed a content analysis of a set of stories in which we examined the characteristics of breast cancer patients who spontaneously publish their story and the topics they write about (Chapter 3).

Patients who prefer to read complete, authentic stories written by others will choose spontaneously published stories on the Internet as a source. In particular, stories written by others in a similar situation are expected to provide readers with optimal support and information (see next section). However, because there are so many spontaneously published stories on the Internet, it may be difficult for patients to find those stories that are relevant to them. Therefore, an important task for professionals is to make stories written by others in a similar situation findable for readers. To do so, it is important to know the preferences and needs of the patients who want to read stories, and to consider the context and domain in which the information exchange will take place. The issue of making stories findable for patients is therefore not a simple matter of ICT, but should be addressed from a clinical informatics perspective, in constant dialogue with the patient. The field of clinical informatics is ideally suited to address this issue, since it takes the care domain as its starting point instead of information technology, and, therefore, has knowledge and affinity with the domain in which patient stories are exchanged. For this reason, the first step we took was to examine how spontaneously published stories are presented and disclosed on the Internet (Chapter 2). Because the study described in Chapter 2 was already published in 2006, we have added an addendum of the situation in 2013.

In conclusion: spontaneously published stories are a unique source of support and information. Therefore, patients (readers) should not be deprived of this source and should be facilitated in finding stories written by similar others, since these stories are expected to have the greatest positive impact on the readers.

Impact of reading other patients' stories

The possible impact of reading stories written by other patients can be explained from a narrative approach [31] and a social-psychological approach. The most applicable social-psychological theories in this context are social comparison theory [35] and social cognitive theory [36]. This section describes the three approaches/theories and examines their link with reading other patients' stories on the Internet.

The narrative approach suggests that people construct a story (narrative) of their lives to give it meaning [31]. Narratives can give meaning because of their ability to weave together separate events and human actions into a whole (plot), for example by arranging them causally or temporally. Plots are, in turn, part of larger plots, which make individual human lives comprehensible as a whole. In their narratives, people connect their past, present, and future and integrate events, their own actions, thoughts, decisions, expectations, and relationships to others into a meaningful whole, taking into account the limits and opportunities posed by the physical, cultural and personal environments. In this way, a person's narrative is the very process of identity construction. Constructing a narrative is an ongoing process in which new experiences and events are continuously weaved into the narrative and which may retrospectively lead to revisions of previously constructed parts of the narrative. If the narrative is coherent and satisfying, it can serve as a justification for one's present condition and situation, both for the narrator him/herself and for others [31].

Breast cancer diagnosis is an interruption of one's narrative, since becoming a *patient* is a threat to one's identity: due to the cancer and its treatments, a person will experience physical changes, changes in roles, relationships, social circles and activities, and a changed vision of the future. All these changes may cause a curtailed sense of agency. As a consequence of these changes, the pre-diagnosis narrative no longer fits, motivating the reconstruction of this narrative into a new narrative that incorporates the meaning of illness and renegotiates one's identity [37-39]. Constructing a new narrative gives a sense of control, since the patient narrator asserts him/herself as agent: he/she can determine how the events and his/her own thoughts, actions, decisions, etcetera are interwoven [38;39]. Patients can structure their narrative as a restitution narrative, a chaos narrative, a quest narrative, or a combination of these three. A restitution narrative begins with health, followed by sickness and looking to a return to health in the future.

In a chaos narrative, the patient is not (yet) able to form a coherent story: events and thoughts are not woven together into a meaningful whole and the story lacks a clear direction. Finally, a quest narrative shows how illness may be considered useful and can function as an impetus for change [34;40]. The value of constructing an illness narrative is supported by empirical research: extensive research has revealed that forming a narrative about one's experiences in life is associated with improved physical and mental health across a variety of populations [41]. In addition, breast cancer patients may also benefit from reading stories written by other breast cancer patients on the Internet (this thesis). Other patients' stories may inspire readers to construct or reconstruct their own story, since these stories enable readers to see how other patients have woven together events, thoughts and decisions into a meaningful whole and how they have shaped their new identities.

Social comparison theory [35] suggests that people consider one or more people in relation to the self as a source of support and information. In social comparison, one positions oneself relative to another person on some dimension. People may have three motives to engage in social comparison. Firstly, people may try to evaluate their opinions and abilities by comparing themselves with others (self-evaluation), especially when no objective (i.e. non-social) information is available. Secondly, people may have the desire to feel good about themselves and their situation (self-enhancement). They may achieve this by comparing themselves with others who are worse off (downward comparison), thereby providing an indication that their own situation is better. Thirdly, people may have the desire to improve themselves or their situation (self-improvement). They may reach this goal by comparing themselves with others who are doing better (upward comparison), which may provide opportunities for learning and improving their situation. However, downward and upward comparison can also lower self-esteem and cause fear and uncertainty. People who are doing worse suggest that your own situation could take a similar negative turn, and people who are doing better imply that you are not doing as well at the moment. Thus, social comparison is a complex phenomenon. The desired direction (upward/downward) depends on many factors, including a person's personality traits and the comparison dimension.

There are many dimensions on which a person can compare him/herself with another person. Social comparison theory suggests that the comparison on a particular dimension will have more impact when other dimensions (such as gender, age, etc.) between the persons are similar. In the case of breast cancer patients comparing themselves with writers of breast cancer stories on the Internet (this thesis), they may make comparisons on the following dimensions: the degree of support from family and friends, the severity of treatment, speed of recovery, and coping (e.g. [42;43]). Comparison with a writer on a specific dimension (regardless of the direction of the comparison), will have the most impact when other dimensions of the writer, such as the situation of the writer or personal characteristics of the writer, are similar to those of the reader.

Social cognitive theory [36] suggests that people can learn from others by observing their behaviour (vicarious learning). If this behaviour has beneficial effects, people can decide to copy the behaviour (modelling). Learning will most likely occur when the observer is confident that he can successfully execute the behaviour (self-efficacy) and when there is a close identification between the observer and the model. Based on this theory, writers of breast cancer stories on the Internet can serve as models for readers of the stories (this thesis). Readers may learn most when they can identify with the writer/narrator of the story.

Studies have shown that breast cancer patients were more likely to select and read experiences of other patients who were doing better than themselves than those of other patients who were doing worse than themselves [43]. Furthermore, cancer patients seeking social comparison information on the coping dimension have been found to prefer upward comparison [42]. It could be that these upward comparisons offer better opportunities for vicarious learning since stories of patients who are doing better may contain more effective examples of emotion-focused coping and problem-focused coping. Emotion-focused coping is aimed at managing and reducing negative feelings about the disease, for example by seeking distraction. Problem-focused coping is aimed at solving problems caused by the disease, for example wearing a wig to regain a positive self-image. In problem-focused coping, it is important that the patient estimates that his behaviour will lead to the desired outcome (outcome efficacy) and that he is confident that he can successfully execute the behaviour (self-efficacy). Readers of other patients' stories on the Internet (this thesis) could learn from the effective examples of other patients' coping and may decide to apply these forms of coping themselves.

The possible impact of stories outlined above is also acknowledged in other disciplines. Personal stories are, for example, used in patient education, like patient decision aids [44;45] or health promotion interventions [46-48]. Current research is focusing on how to incorporate personal stories and/or information contained in these stories in a way that enhances the effectiveness of the decision aid or intervention [49;50].

In conclusion: the theoretical frameworks suggest that stories by writers who are/were in a similar situation to the reader are most likely to have positive impacts on readers with regard to: a) constructing their own illness story, b) evaluating themselves and their situation, and c) learning effective ways of coping in order to improve themselves or their situation.

Search facilities

To develop search facilities that enable patients to find stories on the Internet written by other patients who are/were in a similar situation to their own, knowledge from the field of information retrieval [51;52] and from research into search engines [53] is used.

First, it is important to note that when searching for spontaneously published patient stories on the Internet, the most important aspect is that the retrieved stories match the situation of the patient who carried out the search (reader) (see previous section). However, it is not important that *all* stories on the Internet that match the reader's situation are retrieved, since the reader will probably find sufficient support and information from just a few stories that match her situation. In the field of information retrieval, this balance is expressed in terms of precision and recall. A search has high precision when the search results consist mainly of documents that are of relevance to the searcher. A search has a high recall when (almost) all documents that are of relevance to the searcher are included in the search results (i.e., when as few as possible relevant documents remain unfound). Thus, when searching the Internet for stories written by other patients, high precision is important, whereas high recall is less important.

Below, various search facilities are described. It is useful to be aware of two important aspects of a search facility, namely:

1) the interface: this enables a user to enter his/her query, i.e. he/she can indicate what he/she is looking for. In addition, the interface displays summary information about the search results and gives the user access to the underlying documents.
2) retrieval of documents: which documents are retrieved depends on how the query is processed and connected to the set of documents that can be searched. The documents that are found are ranked and summary information is derived.

The interface of a general search engine on the Internet usually consists of a search box in which search terms can be typed. A search in such a general search engine, for example Google, using the terms "*borstkanker*" [*breast cancer*] and "*verhalen*" [*stories*] in the search box entitled "Find pages with all these words" yields more than 115,000 hits (search carried out on 11 February 2013). To help the searcher, the search results of a general search engine are sorted (ranking). For the ranking of search results and retrieval of documents (URLs), Google uses algorithms in which various aspects are weighted, such as recentness of information, and language and country of the user [54]. However, the exact ranking measure used by the search engine might differ from the relevance measure expected by the searcher. As a consequence, high-ranked search results will not always be the most relevant results for the user. In our search, the high-ranked search results consisted mainly of general sites about breast cancer, but not of specific stories. A more specific search query, for example, "*verhalen*" [*stories*], "*borstkanker*" [*breast cancer*], "*vrouw*" [*woman*], "*39 jaar*" [*39 years old*], "*armfunctie*" [*arm function*], "*gezinsleven*" [*family life*] and "*werk*" [*work*], yields more specific search results. However, the high-ranked search results seem to contain no complete stories in which all searched topics are present in relation to each other. Moreover, forum posts are also found. Using a general search engine, it is therefore difficult for patients to find complete stories from other patients in the same situation as themselves, without finding stories or forum posts that may not be helpful to them or may even make them anxious.

In terms of internal search facilities provided by websites that contain patient stories, a manual index is often used. To enable document retrieval, the website owner (domain expert) uses the content or meaning of the texts to determine which words are relevant for inclusion in the index, and subsequently assigns these index words to the documents. This method is applied, for example, on the website Healthtalkonline [28]. Here, the interface consists of a full list of topics and subtopics (but not in alphabetical order as in a typical index), which is visible to the user and from which the user can select single items to search for stories. An important limitation of this method is a possible difference in *vocabulary* between the domain expert and the user: to find relevant documents, the user must use the same terms as the domain expert. This limitation may particularly play a role when searching for patient stories, because readers are looking for experiences rather than facts, with facts being easier to divide into pre-defined categories than experiences. Another limitation is *assigning* the index words to the documents. The way the expert assigns these words to the documents may not be what the user expects and this may affect precision and recall of the search results. In addition, stories on the website Healthtalkonline [28] can be retrieved on the basis of a writer's age: the stories are divided into four age categories and each age category is accessible via a button. Yet, it might be difficult for patients to decide which age category to select if they want to search for an age at either the beginning or the end of the pre-defined age categories. Finally, an overarching disadvantage of the manual indexing according to topics and age groups is the amount of time and manpower required to build and maintain such a search facility.

Another possibility to make patient stories findable on a website is an internal search engine based on full-text indexing. In addition to the aforementioned full list of topics and age categories buttons, the Healthtalkonline website [28] also contains such an internal search engine. Retrieval of documents based on full-text indexing is based on words that occur literally (including singulars, plurals and diminutives) in the documents. The interface consists of a search box in which a user can type search terms. Yet, this form of searching can present the user with problems relating to synonyms (different words with the same meaning) and homonyms (same words with different meanings). Another problem is that the user does not know exactly which words appear in the documents and therefore which search terms he/she should use. The Healthtalkonline website supports the user in this respect, since the user is given suggestions through a dropdown menu that appears while typing search terms into the search box. A different kind of problem with a full-text indexing search engine is that the same set of search terms can have different meanings (polysemy). For example, a patient who uses the search terms "*werk*" [*work*] and "*armfunctie*" [*arm function*] may be looking either for stories about how to reintegrate into work despite an impaired arm function or for stories about situations in which reduced arm function can lead to occupational disability. Thus, the overall intention of a text cannot be expressed in individual words; it is the relationship between the words that is of crucial importance. However, one advantage of a full-text indexing search engine is that relatively little time and money is involved in building and maintaining the search engine.

If patients had an interface at their disposal that consisted of a search box in which they could type natural language queries to search for stories, there would no longer be any need for them to translate their query into search terms, thereby making it easier for patients to search for situations similar to their own. Natural language includes syntax and semantics [55] and therefore, to enable document retrieval using a search engine in which users can enter natural language queries, both the query and the documents that can be searched for have to be analysed for syntax and semantics. A parser can be used to analyse the syntax of a query and thus clarify the intention of the user. For example, does the user have a ‘how’ question, a ‘why’ question or an ‘if’ question? Although a complete automatic semantic analysis of texts is not yet available, there have been steps in this direction, such as shallow parsing, part-of-speech tagging, ontologies etcetera. However, these have not yet led to useful results. Nonetheless, some insight into semantics can be gained by Latent Semantic Indexing (LSI) [52]. LSI does not examine the exact semantics, but rather the word combinations in a text, as opposed to single words. In this way, LSI can yield some information about the semantics of a query or of the documents being searched, and could allow a search facility to be developed in which stories can be searched using natural language queries. This method provides a solution to the above-mentioned problems of a full-text search engine (synonyms, homonyms and polysemy). In principle, it is possible to type a natural language query into the search box of a general search engine or a full-text search engine. However, it is important to be aware that a natural language query in many search engines will be processed as a keyword query instead of a natural language query and, as a result, some words in the query will be disregarded and syntax and semantics will not be examined.

In conclusion: to make patient stories findable various retrieval methods can be used, each with their own advantages and disadvantages.

User-centred design

To examine how spontaneously published stories can be made findable for patients (readers), it is important to ask the patients themselves about their preferences. Often, patients will have a good idea of which stories can give them support and which topics and/or writer characteristics they would like to search for to find these stories. Therefore, the method we applied in this thesis was a user-centred design [56;57]. In a user-centred design, the user is involved in the entire process of developing a product. This process consists of the following steps: a) the baseline situation of the product is described; b) the user is asked for his/her preferences with respect to the product; c) these user requirements are implemented in the product and then the product is evaluated by actual users. The development of a product following a user-centred design is an iterative process: the outcomes of step c may lead to repeating steps b and c once, or even multiple, times to eventually achieve at an end product that is optimal for the user. The philosophy behind user-centred design is that it optimizes the product around the user’s preferences and needs, instead of forcing the user to change so as to conform to the product.

Our user-centred approach to investigating search facilities for spontaneously published breast cancer stories included the following steps. Firstly, we described the baseline situation by examining how breast cancer stories are presented and disclosed on websites (step a: Chapter 2). We then interviewed breast cancer patients about which topics and writer characteristics they would like to search for when seeking stories written by other patients (step b: Chapter 4). Subsequently, we translated these patient preferences into search facilities. We based the retrieval of documents (stories) in these search facilities on manual indexing. The interface consisted of a list of topics and/or writer characteristics of which one or more items could be selected by the user. We implemented these search facilities on a study website and recruited breast cancer patients to actually use these search facilities. We examined which search facilities resulted in satisfaction with the search process and with the stories retrieved, as well as the impact of the retrieved stories on self-reported coping with breast cancer (step c: Chapter 5). Furthermore, we developed a search facility in which document retrieval was based on LSI, where the interface consisted of a search box in which natural language queries could be typed. Breast cancer patients used this search facility and their satisfaction with the search process was examined (step c: Chapter 7). Additionally, we analysed the actual search behaviour of breast cancer patients when seeking stories written by other patients (Chapter 6 and 7).

In conclusion: a user-centred design was applied because patients themselves are in the best position to indicate which topics and writer characteristics they feel are necessary in a search facility to enable them to find stories that provide support.

Aim of this thesis

The primary aim of this thesis was to gain insight into which search facilities for spontaneously published stories facilitate breast cancer patients in finding stories written by other patients who are in a similar situation to their own. A related goal was to provide owners of websites or patient decision aids with information about search facilities for stories so that they could apply this information to their website/aid and thereby make it easier for patients to find stories about others in a similar situation.

Another purpose of this thesis was to provide a basis for the further development of search facilities for patient stories. The results of this thesis can serve as a guide for future steps in the iterative process of developing search facilities for patient stories.

The research was conducted from a clinical informatics perspective, using knowledge from the fields of information retrieval, human-computer interaction and psychology.

Outline of the thesis

Chapter 2 describes a study in which the features of websites with breast cancer stories were examined using a theoretical framework for analysing communication processes. The study focussed on how patient stories are presented and disclosed on websites. The *addendum* to Chapter 2 provides an update of the situation in 2013.

Chapter 3 describes the content analysis of a set of 167 spontaneously published breast cancer stories. This analysis examined writer characteristics, themes in the stories, and how these writer characteristics and story themes relate to the wider population of breast cancer patients.

Chapter 4 reports the results of interviews with 25 breast cancer patients about their preferences with regard to the content and appearance of online breast cancer stories, and the search facilities they feel would help them search for these stories.

Chapter 5 describes an online randomized controlled experiment which examined breast cancer patients' opinions and experiences regarding four different search facilities for stories. The search facilities selected for this experiment were based on the results of the interviews in Chapter 4. The experiment had a 2x2 factorial design (story topics search facility 'yes/no' and writer characteristics search facility 'yes/no') which resulted in four groups: 1) no search facility (control group: writers' nicknames in alphabetical order), 2) search facility for writer characteristics, 3) search facility for story topics, and 4) search facility for writer characteristics and story topics (combination group).

An analysis of the log data of the online randomized controlled experiment is described in *Chapter 6*. This analysis gives insight into breast cancer patients' actual search behaviour when looking for stories by other patients.

Chapter 7 describes an exploratory study of breast cancer patients using a search facility for stories written by other patients in which they could enter natural language queries. This search facility was developed using Latent Semantic Indexing (LSI), a technique from the field of Information Retrieval (IR).

Finally, in *Chapter 8* the main results of the above-mentioned studies are brought together and a reflection is provided on the findings and their meaning in a broader context. This results in an overall conclusion and discussion of the thesis, in which both theoretical and practical implications are discussed.

References

1. Giordano SH, Buzdar AU, Hortobagyi GN. Breast cancer in men. Ann Intern Med 2002;137(8):678-87.
2. Signaleringscommissie Kanker van KWF Kankerbestrijding. Kanker in Nederland tot 2020: Trends en prognoses. 2011.
3. Shaha M, Cox CL, Talman K, Kelly D. Uncertainty in Breast, Prostate, and Colorectal Cancer: Implications for Supportive Care. Journal of Nursing Scholarship 2008;40(1): 60-7.
4. Spiegel D. Psychosocial aspects of breast cancer treatment. Semin Oncol 1997 Feb;24(1 Suppl 1):S1.
5. Fobair P, Stewart SL, Chang S, D'Onofrio C, Banks PJ, Bloom JR. Body image and sexual problems in young women with breast cancer. Psychooncology 2006 Jul;15(7):579-94.
6. Beatty L, Oxlad M, Koczwara B, Wade TD. The psychosocial concerns and needs of women recently diagnosed with breast cancer: a qualitative study of patient, nurse and volunteer perspectives. Health Expect 2008;11(4):331-42.
7. KWF Kankerbestrijding. Patientenfolder Borstkanker. 2008.
8. Stewart DE, Cheung AM, Duff S, Wong F, McQuestion M, Cheng T, et al. Long-term breast cancer survivors: confidentiality, disclosure, effects on work and insurance. Psychooncology 2001 May;10(3):259-63.
9. Nagler RH, Gray SW, Romantan A, Kelly BJ, DeMichele A, Armstrong K, et al. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. Patient Education & Counseling 2010;81S:S54-S62.
10. Eheman CR, Berkowitz Z, Lee J, Mohile S, Purnell J, Rodriguez EM, et al. Information-seeking styles among cancer patients before and after treatment by demographics and use of information sources. J Health Commun 2009;14(5):487-502.
11. O'Leary KA, Estabrooks CA, Olson K, Cumming C. Information acquisition for women facing surgical treatment for breast cancer: Influencing factors and selected outcomes. Patient Educ Couns 2007;69(1): 5-19.
12. Hartzler A, Pratt W. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians. Journal of Medical Internet Research 2011;13(3):e62.
13. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004 Apr;53(1): 57-64.
14. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564-9.
15. Im EO, Chee W, Lim HJ, Liu Y, Guevara E, Kim KS. Patients' attitudes toward internet cancer support groups. Oncol Nurs Forum 2007 May;34(3):705-12.
16. Keijser WA. Deel VII: Online LotgenotenContact. Online Gezondheidsinformatie en LotgenotenContact: eHealth Serie Deel 1.Utrecht: Wacomed BV; 2005.
17. CBS Statline. Tabel: ICT gebruik van huishoudens naar huishoudkenmerken. 2011.
18. Van Uden-Kraan CF, Drossaert CH, Taal E, Shaw BR, Seydel ER, van de Laar MA. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-17.
19. Hoybye MT, Johansen C, Tjornhoj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. Psychooncology 2005 Mar;14(3):211-20.

20. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003 May;21(3):136-42.
21. Frost JH, Massagli MP. Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data. J Med Internet Res 2008;10(3):e15.
22. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Research Notes 2012;5:699.
23. Farmer AD, Bruckner-Holt CEM, Cook MJ, Hearing SD. Social networking sites: a novel portal for communication. Postgrad Med J 2009;85:455-9.
24. Chou WYS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social Media Use in the United States: Implications for Health Communication. J Med Internet Res 2009;11(4):e48.
25. De la Torre-Diez I, Diaz-Pernas FJ, Anton-Rodriguez M. A Content Analysis of Chronic Diseases Social Groups on Facebook and Twitter. Telemedicine and e-Health 2012;18(6):404-8.
26. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups. J Med Internet Res 2011;13(1):e16.
27. Chou WYS, Hunt Y, Folkers A, Augustson E. Cancer Survivorship in the Age of YouTube and Social Media: A Narrative Analysis. Journal of Medical Internet Research 2011;13(1):e7.
28. Healthtalkonline. DIPEx. Healthtalkonline . 2008. 11-5-2008.
29. BreastCancerStories.org. 2011. 18-12-2011.
30. De Amazones, jonge vrouwen met borstkanker [The Amazones, young women with breast cancer]. De Amazones 2007 [cited 2007 Jan 2];Available from: URL: http://www.de-amazones.nl
31. Polkinghorne DE. Narrative knowing and the human sciences. Albany: State University of New York Press; 1988.
32. Van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Self-Reported Differences in Empowerment Between Lurkers and Posters in Online Patient Support Groups. J Med Internet Res 2008;10(2):e18.
33. Van Uden-Kraan CF, Drossaert CHC, Taal E, Lebrun CEI, Drossaers-Bakker KW, Smit WM, et al. Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur? Computers in Human Behavior 2008;24:309-24.
34. Frank AW. The Wounded Storyteller. Body, Illness, and Ethics. London: The University of Chicago Press; 1995.
35. Wood JV. What is Social Comparison and How Should We Study it? Personality and Social Psychology Bulletin 1996;22(5):520-37.
36. Bandura A. Social foundations of thought and action: A social cognitive theory. New York: 1986.
37. Mathieson CM, Stam HJ. Renegotiating identity: cancer narratives. Sociology of Health & Illness 1995;17(3):283-306.
38. Sakalys JA. Restoring the patient's voice. The therapeutics of illness narratives. J Holist Nurs 2003 Sep;21(3):228-41.
39. Sharf BF, Vanderford ML. Illness Narratives and the Social Construction of Health. In: Thompson TL, Dorsey AM, Miller KI, Parrott R, editors. Handbook of Health Communication. London: LEA Publishers; 2003. p. 9-34.
40. Thomas-MacLean R. Understanding breast cancer stories via Frank's narrative types. Soc Sci Med 2004 May;58(9):1647-57.
41. Pennebaker JW, Seagal JD. Forming a story: the health benefits of narrative. J Clin Psychol 1999 Oct;55(10):1243-54.
42. Bennenbroek FT, Buunk BP, van der Zee KI, Grol B. Social comparison and patient information: what do cancer patients want? Patient Educ Couns 2002 May;47(1):5-12.

43. Van der Zee K, Buunk B, Sanderman R. Neuroticism and reactions to social comparison information among cancer patients. J Pers 1998;66(2):175-94.
44. Molenaar S, Sprangers MA, Rutgers EJ, Luiten EJ, Mulder J, Bossuyt PM, et al. Decision support for patients with early-stage breast cancer: effects of an interactive breast cancer CDROM on treatment decision, satisfaction, and quality of life. J Clin Oncol 2001 Mar 15;19(6):1676-87.
45. Khangura S, Bennett C, Stacey D, O'Connor AM. Personal stories in publicly available patient decision aids. Patient Educ Couns 2008;73(3):456-64.
46. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007;34(5):777-92.
47. Fix GM, Houston TK, Barker AM, Wexler L, Cook N, Volkman JE, et al. A novel process for integrating patient stories into patient education interventions: incorporating lessons from theater arts. Patient Educ Couns 2012;88(3):455-9.
48. Houston TK, Cherrington A, Coley HL, Robinson KM, Trobaugh JA, Williams JH, et al. The art and science of patient storytelling-harnessing narrative communication for behavioral interventions: the ACCE project. J Health Commun 2011;16(7):686-97.
49. Bekker HL, Winterbottom A, Butow P, Dillard A, Feldman-Stewart D, Fowler J, et al. Chapter E: Using personal stories. In: Volk R, Llewellyn-Thomas H, editors. 2012 Update of the International Patient Decision Aids Standards (IPDAS) Collaboration's Background Document. 2012.
50. Shaffer VA, Zikmund-Fisher BJ. All stories are not alike: a purpose-, content-, and valence-based taxonomy of patient narratives in decision AIDS. Med Decis Making 2013;33(1): 4-13.
51. Manning CD, Raghavan P, Schutze H. Introduction to information retrieval. New York, USA: Cambridge University Press; 2008.
52. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 1990;41(6):391-407.
53. Levene M. An Introduction to Search Engines and Web Navigation. Pearson Education; 2006.
54. Google Inside Search. http://www.google.com/insidesearch/howsearchworks/algorithms.html. 2013. 17-3-2013.
55. Dik SC, Kooij JG. Algemene Taalwetenschap. Uitg. Spectrum; 1979.
56. Rosson MB, Carrol JM. Usability Engineering: Scenario-Based Development of Human-Computer Interaction. San Francisco: Morgan Kaufmann; 2002.
57. Neerincx MA, Lindenberg J. Situated cognitive engineering for complex task environments. In: Schraagen JMC, Militello L, Ormerod T, Lipshitz R, editors. Naturalistic Decision Making and Macrocognition.Aldershot, UK: Ashgate Publishing Limited; 2008. p. 373-90.

CHAPTER 2

Illness stories on the Internet: Features of websites disclosing breast cancer patients' illness stories in the Dutch language

Regina Overberg[1]
Pieter Toussaint[1]
Bertie Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden

Patient Education and Counseling 2006; 61: 435-442

Abstract

Objective To examine the way in which breast cancer patients' illness stories in the Dutch language are disclosed on the Internet.
Methods Websites containing Dutch illness stories written by breast cancer patients were selected using a search engine on the Internet. A checklist was developed based on a theoretical framework for analysing communication processes in order to examine the selected websites.
Results None of the websites implements search facilities based on either the content of an illness story, or the personal features of the author. A minority of the websites offers information about the illness stories, such as author's genuineness and editor's review.
Conclusion The Internet's opportunities are not fully utilized with respect to disclosing breast cancer patients' illness stories.
Practice implications In order to support a patient in finding appropriate illness stories in the diverse set of available stories, we are in the process of developing an application that combines symbolic and statistical natural language analysis techniques to facilitate content-based retrieval.

Keywords: Internet; Websites; Breast cancer; Illness stories

1 Introduction

Breast cancer incidence in The Netherlands is among the highest in the world. On the basis of the present incidence rates, about one in every eight to nine women in The Netherlands will develop breast cancer. Although, in general, cancer is most common among the elderly, almost half of the breast cancer diagnosis in the Netherlands in 2000 occurred in women between 20 and 59 years of age [1].

A breast cancer diagnosis and subsequent treatment commonly cause many psychological problems in patients such as distress, anxiety, depression, and fear [2,3]. Patients also experience social isolation [2], a disturbed body image [3], and difficulties with sexual functioning [2]. In some patients, these problems persist even years after diagnosis [2].

One of the ways to cope with these psychological consequences is to exchange illness experiences with fellow patients [4]. During illness experiences exchange patients share experiences that are related to their disease. In this way, they can exchange information, find recognition in each other's experiences, learn from each other how to accept and adapt to the disease, and express their feelings [5].

Illness experiences can be exchanged in different ways. One way is to take part in face-to-face conversations in waiting rooms or during support group meetings. Another possibility is reading books or magazines that compile illness experiences, or watch TV-programmes in which fellow patients tell their story. The Internet is a new but rapidly growing channel for exchanging illness experiences. There are forums, chat rooms and websites that enable patients to exchange illness experiences.

Several studies have examined the effects of face-to-face and web-based illness experience exchange between breast cancer patients. Face-to-face exchange of illness experiences increases breast cancer patients' quality of life [6] and self-esteem [4], and decreases breast cancer patients' anxiety, depression scores [7], and negative feelings [4]. Though in one study on face-to-face social support, some indications of adverse effects on vitality and physical and social functioning were found [8].

Web-based exchange of illness experiences reduced breast cancer patients' scores on depression [9,10], perceived stress, cancer related trauma measures [9], and reactions to pain [10], and is found helpful by breast cancer patients [11]. Yet, also an increase in emotional suppression was found by Lieberman et al. [10]. Klemm et al. [12] conclude in their review that online cancer support groups help people to cope more effectively with their disease. Eysenbach et al. [13] found no robust evidence for positive effects of online peer-to-peer interactions. However, in their opinion the absence of evidence does not mean that virtual communities have no effect [13]. In conclusion, the positive effects of exchanging illness experiences between patients exceed the negative effects.

The Internet is an interesting channel for exchanging illness experiences. In 2004, 65% of the 7 million Dutch households owned a personal computer with Internet access. A good quarter of the 85% information seekers searched for health topics [14]. Among them there will be many patients searching for illness experience exchange on the Internet.

The Internet differs in several ways from the more traditional face-to-face meetings for exchanging illness experiences between patients. An Internet user can stay anonymous, whereas a participant of a face-to-face meeting has to divulge his identity. In addition, a patient who joins a face-to-face meeting has to be present on a specific time at a specific place, while the Internet can be accessed by the patient on a time and at a place he prefers. Also, the diverse possibilities on the Internet for exchanging illness experiences differ from each other in certain ways. A chat session on the Internet is characterised by same time and different place, while forums and websites are characterised by different time and different place.

In this paper, we will focus on the current practice of using the Internet as a communication medium for exchanging illness experiences. Our purpose is to examine the way in which breast cancer patients' illness stories in the Dutch language are disclosed on websites on the Internet. To our knowledge, this was never done before. We focus on websites containing illness stories for two reasons. Firstly, in contrast to forums and chat rooms, websites are easily accessible without needing a password or login name. This implies that also patients who have little knowledge about and experience with using the Internet are able to visit websites. Secondly, fellow patients' illness stories published on websites are more appropriate for restoring the own life story than short fragments of illness experiences are, which are common in forums and chat rooms. The thought behind this is as follows. Every person has an own life story in which he gives meaning to his life, forms his identity, and accounts for his choices. A breast cancer diagnosis disturbs a person's identity and sense of continuity. Someone else's life story who went through the same situation will help to restore the own life story.

2 Methods

2.1 Selection of websites

Before selecting websites three inclusion criteria were formulated: the websites have (1) to be accessible for every internet user, i.e. without password, (2) to be in the Dutch language, (3) to contain at least two illness stories of two different breast cancer patients. Websites were searched for in three ways: (a) by asking breast cancer patients and researchers whether they know any sites, (b) by using the search engine Google on the Internet, and (c) by searching in literature databases. Five websites that were identified by breast cancer patients and researchers met the inclusion criteria [15–19].

Three searches using the search engine Google were conducted (www.google.nl, accessed 30 March 2005) combining the Dutch word for breast cancer (*borstkanker*) with those for stories consisting of experiences (*ervaringsverhalen*; 355 hits), stories (*verhalen*; 6340 hits), or story (*verhaal*; 15,100 hits). The first 100 sites selected by each of the three searches were checked for meeting the inclusion criteria. Links on the websites were not examined, because it was argued that links appear as a hit automatically. The three searches produced, respectively, four [20–23], three [24–26], and one [27] additional website.

In March 2005, literature searches were conducted in the databases Medline, Embase, CINAHL, PsycINFO, PubMed, and Web of Science. Three search concepts were combined: subject (breast cancer patients' illness stories), medium (ICT), and language (Dutch). Also various synonyms for each of these three search concepts were conducted. No websites that fit the inclusion criteria were found using literature databases. Thus, in total 13 relevant websites were found.

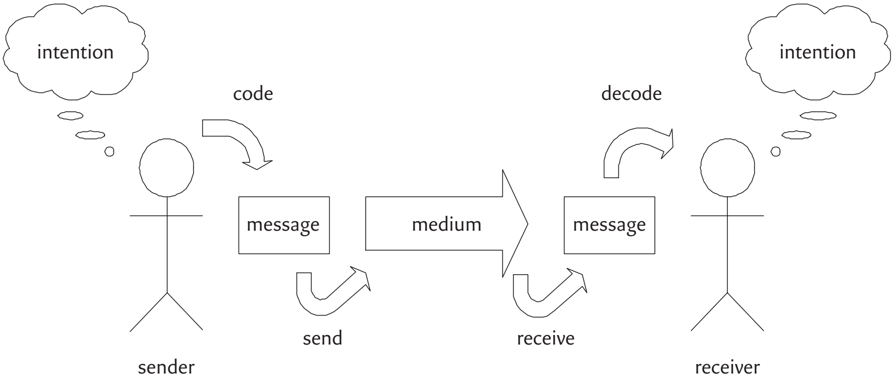


Figure 1 Message exchange between a sender and a receiver

2.2 Evaluation of websites

There are numerous perspectives from which one can look at websites containing illness stories, ranging from technical to functional. We take as our viewpoint that such websites are examples of Computer Mediated Communication applications. Their use is a type of communication practice, and can be best described using features that can also be applied to other communication practices. Communication practices can be described using the message-channel model as basis [28], in which a message is exchanged between a sender and a receiver through a particular medium, as shown in Figure 1. The elements of this model – namely participants, message, medium, and overall situation – are characterised by several features. These features, which are based on the framework presented by Te'eni [29], are listed and explained in Table 1. As depicted in the right column of Table 1, the features' definitions were translated into practice in order to be able to evaluate websites that disclose illness stories in a structural manner. The first author evaluated the selected websites within four weeks of the original search. One site [25] was excluded, because the illness stories were not available at the time this site was evaluated.

3 Results

3.1 Overall features of websites

None of the evaluated websites mentions explicitly the reason why breast cancer patients' illness stories are published on the website. In addition, the time/place context of all the evaluated websites is different time/different place.

Table 1 Features and their definitions for analysing the exchange of illness experiences between patients, and translations of definitions into practice in order to examine websites that disclose illness stories.

Feature		Definition	Translation into website practices
Overall situation	**Goal**	The reason why patients communicate, what they want to achieve with exchanging illness experiences	Which goal is mentioned on the website for publishing the illness stories: to give information, to give recognition, or to express feelings?
	Time/place context	The time and the place at which patients exchange illness experiences	What is the time/place context: same time/same place, same time/different place, different time/same place, or different time/different place?
Participants	**Number**	The number of patients that join the exchange of illness experiences	How many illness stories are available at the web site?
	Anonymity	The state in which the receiving patient can take part in the communication without being identified by the sending patient	Can the visitor of the site (the receiving patient) stay anonymous?
	Authenticity	The degree in which the receiving patient is assured that the sending patient is a genuine patient	Is there a guarantee that the illness stories are from genuine patients?
	Size of common ground	The degree in which personal features (e.g. age, marital status) and disease features (e.g. time since diagnosis, treatment underwent) of a sending patient match with those features of a receiving patient	Can the visitor of the site (the receiving patient) search for illness stories using personal features of the author (the sending patient)?

Table 1 (Continued)

Feature		Definition	Translation into website practices
Message	**Structure**	The degree in which it is beforehand clear to the receiving patient of which parts the illness experience of the sending patient is constructed	Are the illness stories on the site structured by means of a short summary or headings in the text, or are they not structured?
	Accuracy	The degree in which medical facts and spelling in the illness experiences are correct	Does an editorial board review the site including the illness stories?
	Content	The degree in which sending patients' illness experiences fit with the information need of a receiving patient	Can the visitor of the site (the receiving patient) search for stories using content keywords?
Medium	**Capacity**	The degree in which the receiving patient receives non-verbal, additional aspects of a sending patient's illness experience (e.g. sending patient's facial expression, voice/intonation, drawings)	Which modes of information (text, images, video) are used to express the illness stories?
	Evanescence	The degree in which a receiving patient can save or store the sending patient's illness experience	For how long do the illness stories remain available at the site?
	Interactivity	The degree in which a receiving patient can react to a sending patient's illness experiences or can send his/her own illness experience	Can a visitor react on illness stories (react yes/no), or submit his/her own illness story (send yes/no)?

Table 2 Features of websites (n=12) that disclose breast cancer patients' illness stories in the Dutch language.

Feature		Value	Websites % (n)
Participants	**Number**	2-5 illness stories	42 (5)
		6-15 illness stories	25 (3)
		>15 illness stories	33 (4)
	Anonymity	Reading the illness stories without being identified	100 (12)
	Authenticity[a b]	Guarantee that illness stories' authors are genuine patients	8 (1)
	Size of common ground[b]	Search facility for personal features of illness stories' authors	0 (0)
		Brief, separate summary of authors' personal features	8 (1)
Message	**Structure**[ab]	Structure by means of a brief summary of or headings in illness stories	58 (7)
	Accuracy[b]	Review by editor	42 (5)
	Content	Search facility with content keywords	0 (0)
		General search facility with text words	83 (10)
		No search facility, only 'Find (on This Page)' delivered by Internet browser	17 (2)
Medium	**Capacity**[a]	Only text	83 (10)
		Text supported by photos	8 (1)
		Text supported by video fragments with patient's voice	8 (1)
	Evanescence[ab]	Publication dates of illness stories mentioned	50 (6)
	Interactivity	Reacting and sending possible	17 (2)
		Reacting possible	42 (5)
		Sending possible	17 (2)
		No reacting nor sending possible	25 (3)

[a] Feature x is attached to a website if at least half of the illness stories on that website do have feature x. For example, if three out of six illness stories on a website are structured the feature structure is attached to this website.

[b] Does not count till 100 percent (n=12), because only one or a few values of the feature are inserted in this table.

3.2 Participants features of websites

Participants features of the websites are presented in the first part of Table 2. Forty-two percent of the websites contain 5 or fewer illness stories. A quarter of the websites comprise of 6–15 illness stories, while 33% comprise of more than 15 illness stories. The highest number of illness stories on a website is 49 [16], and the highest but one is 40 [21]. Not included in this calculation are: (a) illness stories of cancer patients with a cancer type other than breast cancer, (b) illness stories of male breast cancer patients, (c) stories of persons who are not yet diagnosed with breast cancer, but who have a mutation in the BRCA-1 or BRCA-2 gene, (d) illness stories that are not physically presented on a website, but to which is only referred, and (e) stories that do not contain illness experiences. An example of the latter is a story about fashion for women with only one breast caused by mastectomy [15]. If the same illness story is presented on several websites, this illness story was counted multiple times. One website [21] presents 11 illness stories from another site [15], and two illness stories from yet another website [17]. These thirteen illness stories are counted twice.

A patient who visits the websites in order to read fellow patients' illness stories can stay anonymous in all cases.

None of the evaluated websites guarantee that the authors of the illness stories are genuine breast cancer patients, except one [19]. On this website, the patients/authors of three out of the six illness stories on the site are followed with a camera. Conversations between the patients/authors and specialists are filmed, which points out that the patients are genuine breast cancer patients. On the other 11 websites, no guarantees are given about the genuineness of the patients/authors. However, one of the sites [22] warns visitors that some persons present themselves as a cancer patient without being one.

A search facility based on personal features of authors of illness stories is offered by none of the evaluated websites. In practice, this means that a visitor cannot search for illness stories written by, for example, breast cancer patients between 20 and 30 years of age, or by single breast cancer patients, or by breast cancer patients who underwent radiotherapy but no chemotherapy. In order to find out which illness stories are written by breast cancer patients with a certain personal feature, the illness stories must be read. However, in quite a lot of illness stories personal features of an author, such as age or time since diagnosis, are not mentioned at all. On one website [16], personal features of authors are summarised briefly, separately from the illness stories. Yet, in a few cases the personal features in the brief summary do not fit with the personal features in the illness story. Finally, on a quarter of the websites (n = 3) it is not immediately clear for the visitor which illness stories are from breast cancer patients and which illness stories are from patients with another type of cancer.

3.3 Message features of websites

In the middle part of Table 2 message features of websites are presented. On the majority of the websites (58%) at least half of the illness stories do have structure by means of a short summary beforehand or by means of headings in the text. On the remaining websites (42%), a minority of the illness stories or none of the illness stories do have structure. Headings of the format day-month-year are not seen as structure, because these headings give no information about the content of the text that follows such a heading. If an illness story is not structured it is more difficult for a reader to determine quickly what the story is about. In general, readers are not warned in case an author of an illness story dies.

A majority of the websites (58%) is not reviewed by an editor or nothing is mentioned about editing. A minority of the websites (42%) is reviewed by an editor. Though, it remains unclear whether the illness stories on websites that are reviewed by an editor are checked on correctness of medical facts in the stories.

None of the websites offers a search facility based on content keywords, 10 websites (83%) offer a general search facility with text words, and 2 websites (17%) offer no search facility, only the 'Find (on This Page)' facility delivered by the Internet browser. The general search facility with text words searches the whole website, and the 'Find (on This Page)' facility searches per page. Thus, neither the general search facility with text words nor the 'Find (on This Page)' facility searches exclusively in the illness stories presented on the site. Moreover, both the general search facility with text words as well as the 'Find (on This Page)' facility do not search for synonyms. A search using the Dutch word for radiotherapy that patients often use (*bestraling*) retrieves texts in which the exact word '*bestraling*' occurs; texts in which the Dutch word for radiotherapy that specialists often use '*radiotherapie*' occurs are not retrieved. However, the meanings of '*bestraling*' and '*radiotherapie*' are exactly similar. With a search facility based on content keywords a search for '*radiotherapie*' will also retrieve texts about '*bestraling*'.

3.4 Medium features of websites

The last part of Table 2 shows medium features of evaluated websites. On the majority of the websites (83%), most illness stories are presented in the format of only text. On one website [23] the texts of 10 out of the 12 presented breast cancer patients' illness stories are supported by photos of the patients. On another website [19], three out of six illness stories are presented in the format of video fragments.

On half of the websites information is given about the dates on which the illness stories were published on the websites. The publication dates are ranging from quite long ago till very recently. None of the websites gives information about the period of time the illness stories will be available in future.

A minority of the websites (17%) offers a special button for reacting on illness stories as well as a special button for sending an own story. Forty-two percent of the websites offers only a special button for reacting on illness stories, while 17% of the websites offers only a special button for sending an own story. A quarter of the websites offers neither a button for reacting nor a button for sending. In addition, almost all websites mention an e-mail address to contact the website's owner, but this is in general not meant for reacting on or sending illness stories.

4 Discussion and conclusion

4.1 Discussion

The purpose of this paper was to examine the way in which breast cancer patients' illness stories are disclosed on the Internet. The most striking finding is that none of the websites offers the possibility to search for or within illness stories using either personal features of authors or content keywords. Due to the lack of search facilities, patients have to read the illness stories to find out whether they fit with their own situation and interests. In our opinion, search facilities must be offered to patients looking for illness stories on the Internet, because in that case patients will find stories that fit their needs quicker and easier. A majority of websites offers some kind of structure in the illness stories. This could help patients to get an idea of the content of the illness stories quickly. However, even then a patient has to read all summaries and headings. The medium ICT can be implemented in such a way that searching illness stories is possible, in contrast to, for example, books. It's a pity that the possibilities of Internet are not used at this point.

The results show that the highest number of illness stories on one website is 49. This is quite a lot. Yet, we assume that approximately 10% is relevant for a patient. In addition, we think that the more stories a patient reads/hears, the better she will be able to reconstruct her own life story that was disrupted by a breast cancer diagnosis. ICT has no restrictions at this point, it can contain hundreds or thousands illness stories, in contrast to, for example, a face-to-face meeting at which a limited number of persons can be present. The search facilities become more important if more stories are available to a visitor. If there is no accurate search facility and when a huge number of illness stories is available, a visitor can be overwhelmed. In the case of offering a huge set of illness stories, attention must be paid to the representativeness: stories of patients with diverge personal features (of all ages, marital status, with and without children, long and short time since diagnosis, all kinds of treatment underwent) and with diverge viewpoints (positive, negative, neutral). This diversity will increase the chance that a visitor finds an illness story in which he will recognise himself completely.

Another important aspect which only few websites offer is information about illness stories from which a visitor can distract whether illness stories are reliable or not. Firstly, we saw that only one website guaranteed the genuineness of the authors. Ethically seen it is not justified to offer patients illness stories that are not written by genuine fellow patients. This is not an issue at face-to-face meetings, because it is difficult to fake being a patient when you are not anonymous. However, it is an important issue for ICT. Secondly, only half of the websites mention how old the illness stories are, while this is important information in view of medical developments. For example, a couple of years ago removal of lymph glands was common, nowadays it is first tested if it is necessary to remove them. A patient who reads a story of a couple of years ago in the supposition that it is a recent story, can get very confused. Finally, only a minority of websites is reviewed by an editor whereby it remains unclear whether medical facts in the illness stories are checked. In our opinion, it is very important that the medical facts in illness stories are correct, because otherwise readers can become confused or frightened.

The implementation of the features capacity and interactivity is diverse. Some websites offer a high capacity, others do not. And some websites offer a high interactivity, others do not. We think that patients' preferences with respect to these features differ. Maybe, the best solution is to offer a high capacity and a high interactivity – this is possible with ICT after all – and that patients can switch it off when they do not want to use it.

A website that offers more possibilities than the Dutch sites is the English DIPEx website, Database of Individual Patients' Experience [30]. With respect to breast cancer, 40-50 people are interviewed to find out their individual experiences about the illness. The illness experiences are categorised on age of the author, and on subject. Thus, a visitor can find relatively quickly illness experiences in which he recognises himself and in which subjects appear that interest him. Moreover, a visitor can choose to read the illness experiences (texts are presented), to hear them (audio fragments are presented), or to see them (video fragments are presented) [31]. Access to practical and experiential information from other patients was highly valued by patients who visited the DIPEx website. A few features were specifically valued by patients: total privacy, no emotional demand of attendance, 24 h and 7 days a week availability, the possibility to click off distressing or irrelevant information, and no necessity to engage with patients in return [32]. Thus, this study confirms that there is a need for websites that disclose illness stories.

The present study of websites that disclose breast cancer patients' illness stories had a few limitations. Only websites in the Dutch language were examined. Dutch patients also have access to English sites, so why did we restrict to the Dutch language? We argued that it is the best to read/hear illness experiences in your native language, because you can make nuances in your native language and in general it will be more close to you. Furthermore, not every one understands English very well. Misunderstanding can lead to fear. A second limitation is that this study gives only information about the situation on the Internet in March/April 2005. Changes over time are not considered. Existing websites may have been closed down, and new websites may have been started. However, we think this is inherent to the Internet, and unavoidable when studying it. The last limitation is that only one researcher evaluated the websites. Though, factual features instead of interpretable features were checked, so we think that evaluating by one researcher only gives accurate and reliable results.

4.2 Conclusion

In conclusion, quite a lot breast cancer patients' illness stories in the Dutch language are presented on the Internet. Though, websites that disclose illness stories do not offer accurate search facilities. In addition, the reliability of the illness stories is in many cases unclear. Finally, the capacity and interactivity of the Internet are limited with respect to illness stories. Thus, the opportunities of the Internet are not yet fully utilized with respect to disclosing breast cancer patients' illness stories in the Dutch language.

4.3 Practice implications

It is a pity that not all possibilities of the Internet are used with respect to disclosing illness stories. Patients will benefit if illness stories are better disclosed. That is why we are currently in a process of developing a prototype application for storage and retrieval of illness stories in an Internet environment that enables a user to retrieve illness stories:

- using personal features of the author (size of common ground);
- using content keywords (content);
- knowing that the stories are from genuine patients (authenticity);
- knowing that the content has been checked (accuracy).

With respect to the first two requirements our idea is as follows. Patients who use our future application will be able to enter several personal features, such as their age, marital status, time since diagnosis, and treatment underwent. Moreover, they will be able to enter a search query in natural language, for example: ''How to cope with being bald after chemotherapy?'' Illness stories which fit the personal features and the search query will be retrieved from the database and presented to the patient. In order to make retrieval of illness stories based on personal features and search queries in natural language possible, we propose a design that combines symbolic and statistical semantic analysis techniques of illness stories [33].

Acknowledgements

This paper is part of the Narrator project, funded by The Netherlands Organisation for Scientific Research (NWO), addressing the problem of storage and accurate retrieval of illness narratives using Natural Language Processing and Dialogue Management Techniques. The umbrella programme ToKeN2000 (Access to knowledge and its enhancement Netherlands 2000) is an interdisciplinary research programme that focuses on fundamental problems affecting the interaction between human beings on the one hand, and knowledge and information systems on the other hand. ToKeN2000 is concentrated around three fields of application, namely (1) education and culture, (2) law enforcement and the judicial system, and (3) health care.

References

1. Incidence of cancer in the Netherlands 1999/2000. Visser O, Siesling S, Van Dijk JAAM, editors. Eleventh report of the Netherlands cancer registry. Utrecht: Vereniging van Integrale Kankercentra, 2003.
2. Spiegel D. Psychosocial aspects of breast cancer treatment. Semin Oncol 1997;24(S1):36-47.
3. Payne DK, Sullivan MD, Massie MJ. Women's psychological reactions to breast cancer. Semin Oncol 1996;23:89–97.
4. Van Den Borne HW, Pruyn JFA, Van den Heuvel WJA. Effects of contacts between cancer patients on their psychosocial problems. Patient Educ Couns 1987;9:33–51.
5. Isaksen AS, Gjengedal E. The significance of fellow patients for the patient with cancer: what can nurses do? Cancer Nurs 2000;23: 382–91.
6. Ashbury FD, Cameron C, Mercer SL, Fitch M, Nielsen E. One-on-one peer support and quality of life for breast cancer patients. Patient Educ Couns 1998;35:89–100.
7. Montazeri A, Jarvandi S, Haghighat S, Vahdani M, Sajadian A, Ebrahimi M, et al. Anxiety and depression in breast cancer patients before and after participation in a cancer support group. Patient Educ Couns 2001;45:195–8.
8. Helgeson VS, Cohen S, Schulz R, Yasko J. Education and peer discussion group interventions and adjustment to breast cancer. Arch Gen Psychiatry 1999;56:340–7.
9. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003;97:1164–73.
10. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer 2003;97:920–5.
11. Weinberg N, Schmale J, Uken J, Wessel K. Online help: cancer patients participate in a computer-mediated support group. Health Soc Work 1996;21:24–9.
12. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003;21:136–42.
13. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004; 328:1166–70.
14. Database of Statistics Netherlands (Statline – Centraal Bureau voor de Statistiek). Available at: <http://statline.cbs.nl>. Accessed 9 May 2005.
15. Website of the Dutch Patient Organization for Breast Cancer (Website van BorstkankerVereniging Nederland). <http://www.borstkankervereniging.nl>. Accessed 30 March 2005.
16. Website of De Amazones, young women with breast cancer (Website van De Amazones, jonge vrouwen met borstkanker). <http://www.deamazones.nl>. Accessed 30 March 2005.
17. Website of the Dutch Federation of Cancer Patient Organizations (Website van de Nederlandse Federatie van Kankerpatie¨ntorganisaties). <http://www.kankerpatient.nl>. Accessed 30 March 2005.
18. The breast cancer information pages (De borstkanker informatiepagina's). <http://www.borstkanker.net>. Accessed 30 March 2005.
19. Website of the Dutch Cancer Society: information about cancer and combatting cancer (Website van KWF Kankerbestrijding: informatie over kanker en kankerbestrijding). <http://www.kankerbestrijding.nl>. Accessed 30 March 2005.

20. Website of the Foundation Health-of present interest (Website van de Stichting Gezondheid-actueel). <http://www.kanker-actueel.nl>. Accessed 30 March 2005.
21. Breast cancer information pages: contact with fellow patients, chemotherapy, radiotherapy, breast reconstruction, hormone therapy, immune system (Borstkanker informatiepagina's: lotgenotencontact, chemotherapie, bestraling, borstreconstructie, hormoontherapie, immuunsysteem). <http://www.levenmetborstkanker.nl>. Accessed 30 March 2005.
22. Website of KANKER SDK: starting point for cancer patients and their family (Website van KANKER SDK: startpunt voor kankerpatiënten en hun naasten). <http://www.diagnose-kanker.nl>. Accessed 30 March 2005.
23. Website of the Flemish League against Cancer (Website van de Vlaamse Liga tegen Kanker). <http://www.tegenkanker.be>. Accessed 30 March 2005.
24. Breast cancer meeting point: a meeting point for people who feel involved in breast cancer (BorstkankerTrefpunt: een ontmoetingsplek voor mensen die zich betrokken voelen bij borstkanker). <http://www.borstkankertrefpunt.web-log.nl>. Accessed 30 March 2005.
25. Website of a Dutch health care insurer: Achmea Health (Website van een Nederlandse zorgverzekeraar: Achmea Health). <http://www.achmeahealth.nl>. Accessed 30 March 2005.
26. Bosom friends: a website about breast cancer and friendship (Boezemvriendinnen: een website over borstkanker en vriendschap). <http://www.boezemvriendinnen.nl>. Accessed 30 March 2005.
27. Website of the Working party Heredity of the Dutch Patient Organization for Breast Cancer (Website van deWerkgroep Erfelijkheid van de BorstkankerVereniging Nederland). <http://www.brca.nl>. Accessed 30 March 2005.
28. Cragan JF, Shields DC. Understanding communication theory. In: The communicative forces for human action. Boston: Allyn and Bacon, 1998.
29. Te'eni D. Review: a Cognitive-affective model of organizational communication for designing IT. MIS Q 2001;25:1–62.
30. Herxheimer A, McPherson A, Miller R, Shepperd S, Yaphe J, Ziebland S. Database of patients' experiences (DIPEx): a multi-media approach to sharing experiences and information. Lancet 2000;355:1540–3.
31. DIPEx: Personal experiences of health and illness. <http://www.dipex.org>. Accessed 10 May 2005.
32. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004;53:57–64.
33. Wolf L, Hoenkamp E, Overberg R, Reckman H, Toussaint P. Design of the narrator system: processing, storing and retrieving medical narrative data. SDPS Transactions-ISoLA, in press.

ADDENDUM TO CHAPTER 2

1 Introduction

The Internet contains many stories written by breast cancer patients that form an important source of support and information for other breast cancer patients. In 2006, we examined how breast cancer stories were presented and disclosed on the Internet [1]. Since the Internet is dynamic and subject to rapid developments, it is useful to repeat the 2006 study to gain insight into how breast cancer stories are now presented and disclosed (March 2013).

The aim of this addendum was twofold:

1. To provide an update to Chapter 2 that was published in 2006 [1], i.e. to examine how Dutch-language breast cancer stories are now presented and disclosed on the Internet (March 2013).
2. To describe differences and similarities between the presentation and disclosure of breast cancer stories on the Internet in 2006 and in 2013.

2 Methods

To provide an update and to make a comparison between 2006 and 2013, we used the same method as in 2006. This 2006 method is described briefly below.

2.1 Describing the websites

The presentation and disclosure of breast cancer stories on websites was described on the basis of a checklist that had been developed using a theoretical framework for analysing communication processes [2]. This framework distinguishes four elements of a communication process, each of which can be divided into different aspects:

1. overall situation: goal, time/place context
2. participants: number, anonymity, authenticity, extent of common ground
3. message: structure, accuracy, content
4. medium: capacity, evanescence, interactivity.

These aspects were translated into the practical context of websites with breast cancer stories. For each of the aspects, a definition was formulated (for details see Table 1 of Chapter 2).

2.2 Selection of websites

Breast cancer stories were defined as texts or videos which were not part of an interactive forum and which were written/narrated in the first person. Three inclusion criteria for websites were formulated: (1) the stories should be accessible to all web users, i.e. without password; (2) the stories should be written/narrated in the Dutch language; and (3) the websites should contain at least two stories by two different breast cancer patients.

The 12 websites that were described in 2006 were re-examined in 2013 to assess whether these websites still met the inclusion criteria and how the stories on the websites were presented and disclosed in 2013.

In addition, the three Google searches that had been performed in 2006 were repeated: 1) *borstkanker* AND *ervaringsverhalen* (20,800 hits), 2) *borstkanker* AND *verhalen* (131,000 hits), and 3) *borstkanker* AND *verhaal* (281,000 hits) (searches performed on 2 March 2013). The first 100 sites selected by each of the three searches were checked to assess whether they met the inclusion criteria. Links on the websites were not examined, because it was argued that links should automatically also appear as a hit. In comparison to 2006, the three searches produced seven, three and two additional websites, respectively.

3 Results

3.1 Update regarding websites described in 2006

Of the 12 websites with breast cancer stories described in 2006, seven no longer met the inclusion criteria. Three of these websites no longer seemed to exist: www.levenmetborstkanker.nl, www.borstkankertrefpunt.web-log.nl and www.boezemvriendinnen.nl. One website, www.borstkanker.net, contained only one story that had been written by the founder of the website, who had since passed away. The remaining three websites that failed to meet the inclusion criteria no longer presented patient stories. However, visitors to these websites were referred to other websites with patient stories. For example, the website www.nfk.nl (formerly kankerpatient.nl) referred to the website of the Dutch Patient Organisation for Breast Cancer and the websites www.diagnose-kanker.nl and www.tegenkanker.be referred visitors to four and twelve breast cancer patient blogs, respectively. The website www.tegenkanker.be also included all editions of its digital magazine published since 1999, with 26 articles about breast cancer in which patients often recounted their experiences.

Table 1 gives a description of the five remaining websites that still met the inclusion criteria in 2013. Because the website www.brca.nl is part of the website www.borstkanker.nl, these two websites have been considered together in Table 1.

Table 1 Update of the websites described in 2006 that still met the inclusion criteria in 2013.

Framework[a]		www.borstkanker.nl (including www.brca.nl)	www.de-amazones.nl	www.kankerbestrijding.nl (including YouTube-channel)	www.kanker-actueel.nl
Overall situation	Goal	Borstkankervereniging Nederland is a professional association for people with breast cancer, their social environment, and people/families with hereditary breast cancer genes.	Website of Stichting Amazones, founded in 2005 by a group of active breast cancer and ex-breast cancer patients.	KWF Kankerbestrijding is engaged in scientific research, education, patient support and fundraising.	Website containing the latest information about regular and complementary, non-toxic treatments and resources for cancer disease.
Participants	Number	68 (25 'closed' stories, 33 blog stories and 10 stories by patients with hereditary cancer).	Approximately 790 stories / blogs.	4 women are followed in video diaries, 68 women submitted their own story to the website; 2 videos on the YouTube-channel contain breast cancer stories.	13 (text) + 7 (videos) of breast cancer patients; the website also contains experiences of patients with other cancers. Some patients who appear in a video story have also written a text story.
	Authenticity	Unknown. The 25+33 stories were spontaneously published, the patients with hereditary cancer (10 stories) were probably asked to write their story.	Unknown. After a person has created an account (whether someone is a real patient is not checked), he/she can publish his/her story/blog. Persons with an account are called Amazones.	It is likely that the videos follow real patients.	Unknown.
	Extent of common ground	No search facilities for writer characteristics. Sometimes a writer gives a short description of herself at the beginning of her story.	Amazones can be searched according to personal characteristics, medical condition or specific profile/name.	No search facilities for writer characteristics, neither in the videos, nor in the stories.	No search facilities for writer characteristics.

Table 1 (Continued)

Framework[a]		www.borstkanker.nl (including www.brca.nl)	www.de-amazones.nl	www.kankerbestrijding.nl (including YouTube-channel)	www.kanker-actueel.nl
Message	Structure	Some stories contain headings; after opening blog stories, the topics covered are indicated.	Varies.	Varies.	Varies.
	Accuracy	The 33 blog stories are probably not edited. For the 25 'closed' stories, the site indicates that the stories are published after an editorial check.	Anyone can create an account and publish a story/blog; there is no editor.	The videos are directed. Patients can submit their text story using a web form, whereupon an editor decides whether the story will be published, possibly after shortening the story.	The own story can be published after consultation with an editor.
	Content	No search facility for story topics. The part of the website with blog stories has a full text search engine, which often gives no results.	No search facility for story topics. Amazones can be searched for according to personal characteristics, medical condition or specific profile/name. (There is an overall full text search engine for the entire website.)	Both the video diaries and the text stories can be searched for by a full-text search engine. In addition, the video diaries are disclosed with a search tree containing 14 themes and 6 stages in the disease course.	No search facility for story topics (only an overall search engine for the entire website).

Table 1 (Continued)

Framework[a]		www.borstkanker.nl (including www.brca.nl)	www.de-amazones.nl	www.kankerbestrijding.nl (including YouTube-channel)	www.kanker-actueel.nl
Medium	**Capacity**	Mostly stories in text; some blog stories are supported by photos. The brca.nl website contains one video.	Mostly text (supported by smileys).	Text and videos.	Text and videos.
	Evanescence	For the 33 blog stories, dates of publication are specified. For the other stories no publishing dates are specified, but often years are mentioned in the texts.	For the stories as well as the reactions to the stories, the date on which they were published is specified.	For the stories as well as the YouTube-videos, month and year of publication is specified.	For most stories no date of publication is specified.
	Interactivity	Visitors can submit/publish their own (blog) story and/or respond to stories of others. It does not seem possible to submit an own story for publication on the hereditary cancer site.	After creating an account, visitors can publish their own story/blog and can respond to stories/blogs of others.	Patients can submit their text story using a web form, whereupon an editor decides whether the story will be published, possibly after shortening the story.	The own story can be published after consultation with an editor; visitors can respond to stories using a web form.

[a] The aspects time/place context and anonymity are not included in the table, because, for all websites, stories can be read at a self-selected time and from a self-selected location, and readers of stories remain anonymous on all websites.

Table 2 New websites in 2013

Framework[a]		www.pinkribbon.nl	www.borstborst.nl	www.kanker voorbij.com	www.gezondheids plein.nl	www.youtube.com	Theme websites (n=7)
Overall situation	Goal	Foundation that calls for attention to breast cancer.	Foundation engaged in research and treatment plans, patient education, education for professionals, and raising awareness of breast cancer.	Website created by a patient aiming to raise awareness of breast cancer and all aspects of the disease.	Website owned by Medical Media B.V. aiming to provide independent, accessible, reliable and educational information on health.	Website that allows people to discover, watch and share originally-created videos. YouTube provides a forum for people to connect, inform and inspire others.	Websites with a specific goal owned by foundations, pharmaceutical companies and health centres. Goals are, for example, to stimulate sports activities, promote certain treatments or inform about drug side effects.
Participants	Number	6 stories, including one by a man with breast cancer (all in PDF format)	2 stories	24 stories, of which at least 9 stories by breast cancer patients (some stories do not mention the type of cancer)	3 text stories and 4 videos about breast cancer	Searching for ‘borstkanker’ AND ‘verhaal’ on the homepage of YouTube retrieves about 43 videos.	Varies from 3-20 stories per website.

Table 2 (Continued)

Framework[a]		www.pinkribbon.nl	www.borstborst.nl	www.kankervoorbij.com	www.gezondheidsplein.nl	www.youtube.com	Theme websites (n=7)
Participants	Authenticity	It probably involves real patients (stories are not spontaneously published).	Unknown.	Unknown, the stories are spontaneously published.	The 3 text stories are spontaneously published; the videos are edited and probably involve real patients.	Some of these videos are published by organisations, for example by KWF Kankerbestrijding. Other videos are published by individuals whose authenticity is unknown.	Unknown. On mijnmedicijn.nl the stories are spontaneously published; on the other sites stories are probably not published spontaneously.
	Extent of common ground	No search facilities for writer characteristics. A few sentences summarize what the stories are about; in general, age of the writers is mentioned.	Search facility based on age, type of diagnosis, type of surgery, type of lymph nodes, type of medication, type of radiation, type of reconstruction and/or other characteristics.	No search facilities for writer characteristics. The first few sentences of each story are shown in a list. Visitors must click on the title of a story to see the whole story.	No search facilities for writer characteristics. Text stories are arranged alphabetically according to illness/disease.	No search facilities for writer characteristics.	No search facilities for writer characteristics on most of the websites. Sex and age of the writer is mentioned for every story on mijnmedicijn.nl.

Table 2 (Continued)

Framework[a]		www.pinkribbon.nl	www.borstborst.nl	www.kanker voorbij.com	www.gezondheids plein.nl	www.youtube.com	Theme websites (n=7)
Message	Structure	Headings in the stories.	Headings in the stories.	Varies/none.	None.	None.	Varies/ none.
	Accuracy	Edited.	Unknown.	No editor.	Videos are edited.	Videos provided by organisations are edited.	Almost all websites seem to have edited stories.
	Content	No search facility within stories. Ctrl-F on the summary page allows no search within PDFs.	No search facility with content keywords, no search facility with text words (only Ctrl-F).	No search facilities for story topics. The first few sentences of each story are shown in a list. Visitors must click on the title of a story to see the whole story.	No search facility for story topics (only an overall full text search engine for the entire website).	Only overall search engine for the entire website. Within the set of 43 videos, no search facilities are available for story topics.	In general, no search facilities for story topics.

Table 2 (Continued)

Framework[a]		www.pinkribbon.nl	www.borstborst.nl	www.kanker voorbij.com	www.gezondheids plein.nl	www.youtube.com	Theme websites (n=7)
Medium	Capacity	Text supported by photos.	Only text.	Predominantly text; some stories are supported by photos.	Text and videos.	Videos.	Predominantly text; some stories are supported by photos.
	Evanescence	No dates of publication are specified.	No dates of publication are specified.	The reactions to the stories include a date of publication; the stories do not.	No dates of publication are specified.	For the videos as well as the reactions to the videos, month and year of publication is specified.	For some stories, a publication date is specified. (On mijnmedicijn.nl all stories include a date of publication).
	Interactivity	Visitors cannot submit their own story, nor can they respond to other stories.	Visitors cannot submit their own story, nor can they respond to stories by others.	All stories allow for visitor comments and visitors are encouraged to publish their own story.	Visitors can submit their own text story and respond to the text stories by others.	After creating an account, a visitor can publish their own video and can respond to videos by others.	In general, visitors cannot submit their own story, nor can they respond to stories by others (with the exception of mijnmedicijn.nl).

[a] The aspects time/place context and anonymity are not included in the table, because, for all websites, stories can be read at a self-selected time and from a self-selected location, and readers of stories remain anonymous on all websites.

Table 3 Features of websites (n=16) that disclose Dutch-language illness stories written by breast cancer patients: situation in 2013.

Feature		Value	Websites % (n)
Participants	**Number**	2-5 illness stories	38 (6)
		6-15 illness stories	13 (2)
		>15 illness stories	50 (8)
	Anonymity	Reading the illness stories without being identified	100 (16)
	Authenticity	Guarantee that illness story writers are real patients	19 (3)[a]
	Extent of common ground	Search facility for personal features of illness story writers	13 (2)
		Brief, separate summary of writers' personal features	19 (3)[a]
Message	**Structure**	Structure by means of a brief summary of story or headings in story	19 (3)[a]
	Accuracy	Review by editor	38 (6)[a]
	Content	Search facility with content keywords	0 (0)
		General search facility with text words / search engine for entire website	56 (9)
		No search facility (or only 'Find (on This Page)' delivered by Internet browser)	44 (7)
Medium	**Capacity**	Only text	31 (5)
		Text supported by photos	38 (6)
		Text supported by video fragments with patient's voice	31 (5)
	Evanescence	Publication dates of illness stories mentioned	44 (7)[a]
	Interactivity	Responding and submission is possible	44 (7)
		Responding is possible	6 (1)
		Submission is possible	6 (1)
		Responding or submission is not possible	44 (7)

[a] Feature x is associated with a website if at least some of the stories on that website have feature x.

3.2 New websites in 2013

The searches performed in 2013 yielded twelve additional websites. These websites are described in Table 2. The stories on seven websites illustrate a particular theme: www.geefhaareenkans.com, www.tegenkracht.nl, www.mijnmedicijn.nl, www.loopvoorleven.nl, www.hyperthermie.info, www.noadin.nl, and www.breastfriends.nl. These seven websites are considered together in Table 2. The stories on the other five websites include more diverse aspects of breast cancer: www.pinkribbon.nl, www.borstborst.nl, www.kankervoorbij.com, www.gezondheidsplein.nl and www.youtube.com. These five websites are described separately in Table 2.

3.3 The complete picture in 2013

Table 3 summarizes the complete picture in 2013. This table is constructed in the same way as Table 2 in Chapter 2, enabling a clear comparison with 2006.

4 Conclusion

The goal of this addendum was to compare how patient stories are presented and disclosed on websites in 2006 and in 2013. There were three clear results. Firstly, the number of stories per website has increased. While in 2006 the highest number of stories on a website was 49, in 2013 the highest number of stories was about 790. Secondly, the format in which stories appear has changed. In 2013, there is now a trend favouring photos and videos, instead of text only. Thirdly, in 2013 two websites offered a search facility for writer characteristics, while in 2006 none of the websites offered such a search facility. However, in 2013 there were still no websites offering a search option for content keywords.

References

1. Overberg R, Toussaint P, Zwetsloot-Schonk B. Illness stories on the Internet: Features of websites disclosing breast cancer patients' illness stories in the Dutch language. Patient Educ Couns 2006;61:435-42.
2. Te'eni D. Review: A Cognitive-Affective Model of Organizational Communication for Designing IT. MIS Quarterly 2001;25(2):1-62.

CHAPTER 3

Spontaneously published illness stories on a website for young women with breast cancer: do writers and themes reflect the wider population?

Regina Overberg[1]
Andries de Man[1]
Ron Wolterbeek[2]
Wilma Otten[3]
Bertie Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden
[2] Department of Medical Statistics, Leiden University Medical Center, Leiden
[3] Department of Prevention and Care, TNO Quality of Life, Leiden

Informatics for Health and Social Care 2013; 38(1): 54-66

Abstract

This study examined writer characteristics and themes written about in a set of 167 spontaneously published stories on a Dutch website for young women with breast cancer. The stories were coded for six disease characteristics and sixteen themes. Coding results were compared to the characteristics of young women with breast cancer in a hospital cancer register and to the frequency of problems among young breast cancer patients participating in quantitative studies. We found that writer characteristics were diverse. Yet, logistic regression showed that women were more likely to be a writer if they were diagnosed at a younger age (OR 0.82; 95% CI (0.78, 0.85)), underwent a mastectomy (OR 4.63; 95% CI (2.59, 8.26)), or were in the first treatment period (OR 2.83; 95% CI (1.44, 5.58)). All sixteen themes were present in the stories, but some themes were addressed less often than their frequency among participants of quantitative studies suggested. The findings indicate that a set of spontaneously published stories might not completely reflect the characteristics and themes of the wider population of young women with breast cancer. Websites with spontaneously published stories should inform readers about this.

Keywords: Internet, breast cancer, illness stories, writer characteristics, story themes

1 Introduction

Young breast cancer patients can encounter problems with respect to a wide range of issues, such as: body image and sexuality [1]; concerns about pregnancy [2], fertility issues [3,4] and heredity [5,6]; and, communication problems with physicians [7] or with family and friends [8]. For some patients it is helpful to write about their experiences with their illness [9,10]. The process of constructing one's illness story can help with making sense, asserting control, transforming identity and making decisions [11]. Patients can put their stories on the Internet, where others -such as family, friends or other patients- can read them.

For readers of the stories it might be most helpful when a set of stories on a website gives an acceptable impression of the various topics that are relevant for coping with the disease: the set of stories should be a good starting point for a fellow patient, relative or friend to get an impression of what it means to suffer from the disease. Thus, the stories should be about topics that are generally considered important for coping with the disease.

Some of the websites with closed individual stories are provided by research groups, who have carefully selected the set of stories that is presented on the website. The research group who composed the Health Talk Online website (formerly DIPEx) [12], for example, used a method called maximum variation sampling to ensure that a wide range of perspectives and experiences is represented [13]. There is a clear consensus among an international team of patients, practitioners, policy makers, and researchers that such a range of perspectives and experiences should be included [14].

Besides research groups, there are active groups of patients who provide websites with closed individual stories. On many of these websites patients can spontaneously publish their own story. In this way a set of stories develops which is not based on any formal sampling methods. A number of concerns might rise at this point. Butow et al. [15] mention that it might be that only articulate patients or patients who have strong motivations to tell their story spontaneously publish their stories. Moreover, in an earlier study we found that patients wanted to read about themes they experienced problems with [16]. If this also applies to writing, than patients might only spontaneously write about themes they experienced problems with. As a consequence, it might be that a set of spontaneously published stories is written by a specific group of patients and that only specific themes are addressed in the stories.

The present study examines the set of spontaneously published stories on the website of the *Amazones* foundation. This is a Dutch website for and by young women with breast cancer [17]. Our study entailed a content analysis of all 167 spontaneously published stories on this website and encompassed the following questions: 1) What are the characteristics of the patients who spontaneously publish their story on this website?, and 2) Which themes do they address in their stories? In order to answer these questions, we used a coding scheme with six disease characteristics and sixteen themes. Furthermore, we examined whether the writer characteristics reflect the characteristics of the wider population of young women with breast cancer. We used the cancer register of the Leiden University Medical Centre (LUMC) to get an indication of the characteristics of the wider population of young women with breast cancer. Finally, for five clearly defined themes we examined whether the extent to which these themes were written about in the stories reflect the prevalence of these themes among the wider population of young women. We used quantitative studies that examined prevalences of these five themes among young women with breast cancer to get an indication of the extent to which these themes are prevalent in the wider population.

2 Methods

2.1 Main empirical material

In January 2007 we downloaded all 167 stories from the *Amazones* website to analyse them. This was the total number of stories posted since women started to submit anonymously (by using nicknames) their stories for publication on the website in 2004. There were no restrictions for the women as to length, structure and content of their stories. Submitted stories were put on the website by the board of the foundation. The board did not exclude any stories from publication. In this article, the women who posted their story are referred to as the writers.

The stories were all in the Dutch language and written in the first person. Most stories followed a chronological order: discovery of breast cancer, treatments and the restitution process. All stories were closed individual stories: each of the stories was published in one time as a whole, i.e. no additions or changes were made after publishing a story online and the stories were not part of an interactive forum. One third of the writers stated their region of origin: more than 90% of them lived in the Netherlands.

2.2 Coding writer characteristics

In order to code the writer profiles a coding scheme with six characteristics was developed. The six characteristics were based on how other websites disclose patient stories [12] and on studies about the aspects patients appreciate in websites with stories [16,18]. The characteristics in the coding scheme were: 1) age at diagnosis, 2) time since diagnosis, 3) partner status, 4) children status, 5) treatment received, and 6) phase in the course of disease.

The continuous variables ‘age at diagnosis’ and ‘time since diagnosis’ were coded with a value (years and months, respectively) or with ‘unknown’. ‘Time since diagnosis’ was defined as the number of months between being diagnosed and writing one’s story. The other writer characteristics were all categorical variables. ‘Partner status’ and ‘children status’ were coded with ‘yes’, ‘no’ or ‘unknown’. ‘Treatment received’ was divided into: a) lumpectomy, b) mastectomy, c) axillary lymph node removal, d) radiation therapy, e) chemotherapy, f) immunotherapy, g) hormonal therapy, and h) breast reconstruction. When writers mentioned some of these treatments, these were coded with ‘yes’ and the ones that were not mentioned, with ‘no’. When none of the treatments were mentioned, all eight were coded with ‘unknown’. For ‘phase in the course of disease’ the categories and code instructions were: 1) in first treatment period (treatments a-f), 2) free of cancer (undetectable cancer or treatments g-h), 3) breast cancer for second time (recurrence of the same breast tumour or a new primary breast tumour), 4) metastasized cancer

(cancer in other parts of the body than the breasts), 5) passed away (obituary at the end of a story written by the website moderators), and 6) unknown.

One coder coded all the stories for writer characteristics, and a second coder coded 15% (n=25). Besides the information in the story itself, the information in the so-called *Amazones* profiles on the website was used: 97 writers (58%) completed such a profile. The inter-rater reliability for each writer characteristic was calculated using the Intraclass Correlation Coefficient (ICC). The ICC's ranged from 0.607 to 1.000 (Table 1). Two ICCs for 'age at diagnosis' were calculated. These indicate the inter-rater reliability for i) assigning a value for age versus assigning 'unknown', and ii) the values that were assigned for age, after excluding the stories with 'age unknown' (n=6). For 'months since diagnosis' one ICC was calculated, since no 'unknown' was assigned for the 25 stories that were coded by two coders.

2.3 Coding story themes

In order to code the themes in the stories a coding scheme with sixteen themes was developed (Table 1). The sixteen themes were divided in four domains: 1) Diagnosis, 2) Treatment, 3) Health care system, 4) Living with it. For the development of the coding scheme, we studied which themes were disclosed in breast cancer stories on two different websites [12,19]. Both were developed by research groups and aim to represent a wide range of perspectives and experiences [13,20]. Two independent coders unfamiliar with the *Amazones* website coded all the stories for themes. The coders were trained using an identical sub-sample (n=13) of the stories under study. A story could receive multiple codes. A code was assigned when women wrote about a theme, irrespective of how they experienced the theme. The inter-rater reliability for each theme was calculated using Cohen's kappa (Table 1).

2.4 LUMC cancer register

To examine how the characteristics of the writers relate to those of the wider population of young women with breast cancer, the characteristics of the writers were compared to those of young women with breast cancer registered at the Leiden University Medical Centre (LUMC). The LUMC is one of eight university hospitals in the Netherlands with a capacity of 882 beds. All cancer patients who consult the LUMC are entered into the LUMC cancer patients register. This register contains information per tumour as to its location, the date of its discovery, the ways in which it was treated, whether it recurred locally/regionally and/or whether it metastasized, and if so when, the age of the patient at diagnosis, and whether or not the patient has died. The register does not include information on partner and children status, and any decisions as to breast reconstruction.

In order to make a useful comparison with the *Amazones* writers, we selected from the register women with breast cancer, diagnosed before the age of 50, who contacted the LUMC between 1997 and 2007. This age range was selected because the *Amazones* present themselves as women younger than 50 [17]. The time range 1997-2007 was selected to ensure that all phases in the course of disease were represented. The selection consisted of n=1160 women with a mean age at diagnosis of 42.4 years (SD=5.6) and a mean time since diagnosis of 56.7 months (SD=36.1). The treatment that most of the women had undergone was radiation therapy (82%; n=952), followed by axillary lymph node removal (73%; n=849), chemotherapy (72%; n=838), mastectomy (48%; n=560), hormonal therapy (43%; n=494) and immunotherapy (6%; n=73). Almost one third of the women was free of cancer (31%; n=359) and one in five had passed away (20%; n=228). The remaining women were in first treatment period (7%; n=76), had cancer for a second time (6%; n=73), had metastasized cancer (6%; n=70) or their phase in the course of disease was unknown (31%; n=354).

2.5 Quantitative studies

In June 2008 PubMed, Embase and PsychInfo were searched for quantitative studies about problems experienced by young women with breast cancer in order to examine how the frequency with which themes were written about spontaneously in the *Amazones* stories relates to the extent to which themes are prevalent in the wider population of young women with breast cancer. In the three literature databases we searched for the themes 'contact with medical staff', 'family and friends', 'body image and sexuality', 'pregnancy issues' and 'concerns about heredity' by combining the search terms *breast cancer* and *young women* successively with *health care system, family (support), social network, social contacts, social behaviour, child(ren), body image, sexuality, pregnancy, heredity, genetic counseling*. All studies that reported prevalences were used for the comparison. Also, for a more useful comparison the theme 'pregnancy issues' was divided into the subthemes 'wanting to become pregnant after treatments' and 'pregnant at diagnosis'.

Table 1 Coding schemes for writer profile and story themes and inter-rater reliability.

Coding scheme writer profile 15% (n=25) of stories were coded by a second coder		Intraclass Correlation Coefficient (ICC)
Age at diagnosis (in years)	Assigning a value for age versus assigning 'unknown'	0.759
	The values that were assigned for age, after excluding the stories with 'age unknown' (n=6)	0.849
Time since diagnosis (in months)	The values that were assigned for time since diagnosis (for none of the 25 stories 'time since diagnosis unknown' was assigned)	0.971
Partner status, n (%)		0.800
Children status, n (%)		1.000

Table 1 (Continued)

Coding scheme writer profile 15% (n=25) of stories were coded by a second coder		Intraclass Correlation Coefficient (ICC)
Treatment received, n (%)	Lumpectomy	0.823
	Mastectomy	0.809
	Axillary lymph node removal	0.914
	Radiation therapy	0.839
	Chemotherapy	1.000
	Immunotherapy	0.657
	Hormonal therapy	0.844
	Breast reconstruction	0.607
Phase in the course of disease, n (%)		0.642
Coding scheme story themes All stories (n=167) were coded by a second coder; the training sub-sample of stories (n=13) was left out when calculating Cohen's kappa's		Cohen's kappa
Diagnosis	Feelings about diagnosis	0.70
Treatment	Decision-making	0.69
	Coping with treatment	0.43
	Side effects	0.79
Health care system	Delay in/errors at diagnosis	0.43
	Waiting for test results	0.68
	Contact with medical staff	0.63
	How treatment was performed	0.39
	Second opinion	0.72
Living with it	Work and insurance	0.57
	Family and friends	0.67
	Body image and sexuality	0.65
	Pregnancy issues	0.88
	Coping with breast cancer	0.57
	Practical advices	0.49
	Concerns about heredity	0.65

2.6 Statistical analysis

The analyses included descriptive statistics of writer characteristics and of the frequency of themes in the stories. Furthermore, a multivariate logistic regression analysis that included all disease characteristics simultaneously was performed to compare the writers with the women in the LUMC cancer register. Group (writers vs. LUMC register women) was included in the model as the dependent variable; disease characteristics were included as independent variables. The two dichotomous variables lumpectomy (yes/no) and mastectomy (yes/no) were combined to form a new dichotomous variable (mastectomy vs lumpectomy only). For the women in the cancer register the time since diagnosis was defined as the period between the date of diagnosis and the date on which the women were selected from the register. For 'phase in the course of disease' a category 'unknown' was included in the model, since for almost one third of the register women this variable was unknown. Confounding was assessed by evaluating changes in the relevant regression coefficients (beta's). Goodness-of-fit was determined by the Hosmer and Lemeshow Test. In the final model non-significant predictors ($p>0.05$) were removed.

2.7 Ethical considerations

The board of the *Amazones* foundation gave us permission to conduct the study. In order to maximize confidentiality, we report our results on an aggregated level without quoting the writers [21]. The LUMC Medical Ethical Committee concluded that our study involved no *medical* intervention and that we could proceed.

3 Results

3.1 Characteristics of the writers

Writer characteristics are shown in Table 2. The mean age at diagnosis was 36 years (SD=6.2). The mean time since diagnosis was 27 months (SD=37.4). The treatments most often undergone were: chemotherapy (77%), mastectomy (68%) and radiation therapy (61%). Few writers received immunotherapy (5%). Most writers were 'free of cancer' (46%) or 'in first treatment period' (34%). In addition, most writers had a partner (76%) and children (74%), and almost one-third of the writers who had received a mastectomy decided to undergo breast reconstruction (31%).

Table 2 Characteristics of the women who published their story on the *Amazones* website; i.e. the *Amazones* writers (n=167).

Characteristics		Amazones writers (n=167)
Age at diagnosis (in years)[a]	Mean (SD)	36 (6.2)
	Minimum	22
	Maximum	52
Time since diagnosis (in months)[b]	Mean (SD)	27 (37.4)
	Minimum	<1
	Maximum	252
Partner, n (%)	Yes	127 (76)
	No	18 (11)
	Unknown	22 (13)
Children, n (%)	Yes	123 (74)
	No	23 (14)
	Unknown	21 (13)
Treatment received, n (%)[c]	Lumpectomy only	43 (26)
	Mastectomy	113 (68)
	Axillary lymph node removal	91 (55)
	Radiation therapy	102 (61)
	Chemotherapy	129 (77)
	Immunotherapy	8 (5)
	Hormonal therapy	67 (40)
	Breast reconstruction	35 (31[d])
Phase in the course of disease, n (%)	In first treatment period	56 (34)
	Free of cancer	76 (46)
	Breast cancer for second time	14 (8)
	Metastasized cancer	11 (7)
	Passed away	5 (3)
	Unknown	5 (3)

[a] age at diagnosis unknown n=33
[b] months since diagnosis unknown n=8
[c] all eight treatments unknown n=6, type of surgery unknown n=1
[d] percentage based on the writers who had a mastectomy (n=113)

Table 3 Final logistic regression model for being a writer on the *Amazones* website.[a]

Disease characteristics		OR[b]	95% CI[c]
Age at diagnosis (years)		0.82	(0.78-0.85)
Time since diagnosis (months)		0.98	(0.97-0.99)
Treatment underwent[d]	Lumpectomy only	1.00	
	Mastectomy	4.63	(2.59-8.26)
	No axillary lymph node removal	1.00	
	Axillary lymph node removal	0.45	(0.26-0.79)
	No hormonal therapy	1.00	
	Hormonal therapy	1.65	(1.00-2.71)
	No immunotherapy	1.00	
	Immunotherapy	0.16	(0.06-0.47)
Phase in the course of disease	Free of cancer	1.00	
	In first treatment period	2.83	(1.44-5.58)
	Cancer for second time	0.64	(0.28-1.48)
	Metastasized cancer	0.85	(0.36-2.01)
	Passed away	0.04	(0.01-0.15)
	Unknown	0.00	(0.00-)

[a] The model includes n=1243 complete cases (n=127 writers and n=1116 cancer register women)
[b] OR=odds ratio; OR's are adjusted for all covariates in the model
[c] CI=confidence interval
[d] Chemotherapy and radiation therapy were not included in the model due to non-significance

3.2 Comparison with the LUMC cancer register women

The results of the logistic regression analysis are shown in Table 3. There was no confounding and goodness-of-fit was satisfactory (Hosmer and Lemeshow Test, p=0.995). The clearest results were that women were more likely to be a writer if they were diagnosed at a younger age (OR=0.82; 95% CI (0.78, 0.85)), underwent a mastectomy (OR=4.63; 95% CI (2.59, 8.26)) and/or were in the first treatment period (OR=2.83; 95% CI (1.44, 5.58)). Radiation therapy and chemotherapy were not significant predictors, nor were three categories of 'phase in the course of disease' (cancer for a second time, metastasized cancer, and unknown).

3.3 Frequency of themes in the stories

All sixteen themes were present in the stories (Figure 1). Many stories contained themes in the domains 'Diagnosis' and 'Treatment'. In 53-61% of the stories women wrote about their feelings on the diagnosis. Treatment decision-making appeared in 47-54% of the stories and side effects were disclosed in 40-45% of the stories. Also, two themes in the domain 'Living with it' were disclosed in many stories: family and friends appeared in 40-47% of the stories, and coping with breast cancer in 41-61% of the stories. The themes least often disclosed were practical advices (written about in 1-4% of the stories) and work and insurance (written about in 4-6% of the stories).

3.4 Comparison with the quantitative studies

The comparison of the prevalences of the themes in the set of stories with the prevalences in the quantitative studies is shown in Table 4. The prevalence of three of the themes was almost equal between the stories and the quantitative studies: 'family and friends' (40-47% vs 13-53%), 'pregnant at diagnosis' (7% vs 4-6%), and 'concerns about heredity' (10-11% vs 7-9%). In the stories the prevalence of the theme 'contact with medical staff' (26-27% vs 37%) was somewhat lower than the prevalence in the quantitative study. The difference in prevalence was substantial for the themes 'body image and sexuality' (20-25% vs 50%) and 'wanting to become pregnant after treatments' (9% vs 54-57%). None of the themes had a greater prevalence in the set of stories than in the quantitative studies.

4 Discussion

We examined the writer characteristics and themes written about in the spontaneously published *Amazones* stories. We found that the stories were written by a diverse group of young breast cancer patients: the group of writers was characterized by a diversity of ages, number of months since diagnosis, treatments and phases in the course of disease. The group of writers slightly deviated from the wider population of young women with breast cancer: in comparison to the women in the LUMC register who were diagnosed with breast cancer before the age of 50 relatively many writers were of a younger age, underwent a mastectomy and were in the first treatment period. With respect to the themes we found that the set of stories contained a broad range of themes: all sixteen themes under study were present in the set of stories. Yet, some themes were present in only a few stories. Furthermore, for two of the five themes we examined closer it was found that these themes did not appear in the set of stories in the extent that could be expected on the basis of quantitative studies among young women with breast cancer.

Table 4 Prevalence of themes in *Amazones* stories (n=167) and in quantitative studies on young women with breast cancer.

Domain	Theme	Stories %	Other studies %	Measurement	Design	Participants	Study
Health care system	**Contact with medical staff**	26-27	37	Reported to have had communication problems with their physicians	In-person interviews, consisting of closed as well as open-ended questions; most questions were adopted from validated questionnaires	185 women diagnosed before the age of 50 who were cancer-free 5 yrs later (median age at the beginning of the study (5 yrs before) was 45 years)	Bloom et al., 2004 [7]
Living with it	**Family and friends**	40-47	22-53	Experienced talking about death/ cancer/ wills and finances/ fears/ future/ feelings with partner as problematic	Structured, postal one-time survey including subscales about relationship with partner and children (adopted from a validated questionnaire)	204 women diagnosed at age 50 or younger who were at least 3 months post-diagnosis (mean current age 43.5 yrs, SD=6.2)	Walsh et al., 2005 [8]
			13-20	Experienced talking about cancer with children/ helping children cope/ caring for children as problematic			
	Body image and sexuality	20-25	50	Experienced ≥2 body image problems some of the time or at least one problem much of the time	Structured, in-person interviews; most questions were adopted from validated questionnaires	549 women aged 22-50 yrs in a stable (un)married relationship within 7 months of diagnosis (no mean age was reported)	Fobair et al., 2006 [1]

Table 4 (Continued)

Domain	Theme	Stories %	Other studies %	Measurement	Design	Participants	Study
Living with it	Pregnancy issues – Wanting to become pregnant after treatments	9	54	Had the question whether becoming pregnant after treatment was possible	Mailed, self-report questionnaire; participants could indicate for a list of fertility- and menopause related questions whether they had had that question or not	228 women diagnosed at age 40 or younger who were 6 to 60 months after diagnosis (mean age at diagnosis 35.5 yrs, SD=4.0)	Thewes et al., 2005 [4]
			57	Recalled substantial concern at diagnosis about becoming infertile with treatment	Structured, e-mailed, one-time survey about fertility issues; questions were as far as possible adopted from validated questionnaires	657 women diagnosed at age 40 or younger who were premenopausal (mean current age 35.8 yrs)	Partridge et al., 2004 [3]
	Pregnancy issues – Pregnant at diagnosis	7	4	Excluded from analysis of the study (see above) due to being pregnant at diagnosis	See above	860 informed consent givers	Partridge et al., 2004 [3]
			6	Was pregnant when breast cancer was diagnosed	Documentation of the clinical presentation and course of breast cancer using data from patient records	407 women aged 20-29 years at diagnosis (median age 28 yrs)	Guinee et al., 1994 [2]
	Concerns about heredity	10-11	7	Was found to carry BRCA1 or BRCA2 deleterious mutations	Screening for BRCA1 and BRCA2 mutations	136 women diagnosed before age 46 years in Spain (no mean age was reported)	De Sanjose et al., 2003 [5]
			9	Was associated with a germline mutation in BRCA1 or BRCA2	Molecular genetic study in which germline mutations in BRCA1 and BRCA2 were investigated	234 women diagnosed under the age of 41 years in southern Sweden (no mean age was reported)	Loman et al., 2001 [6]

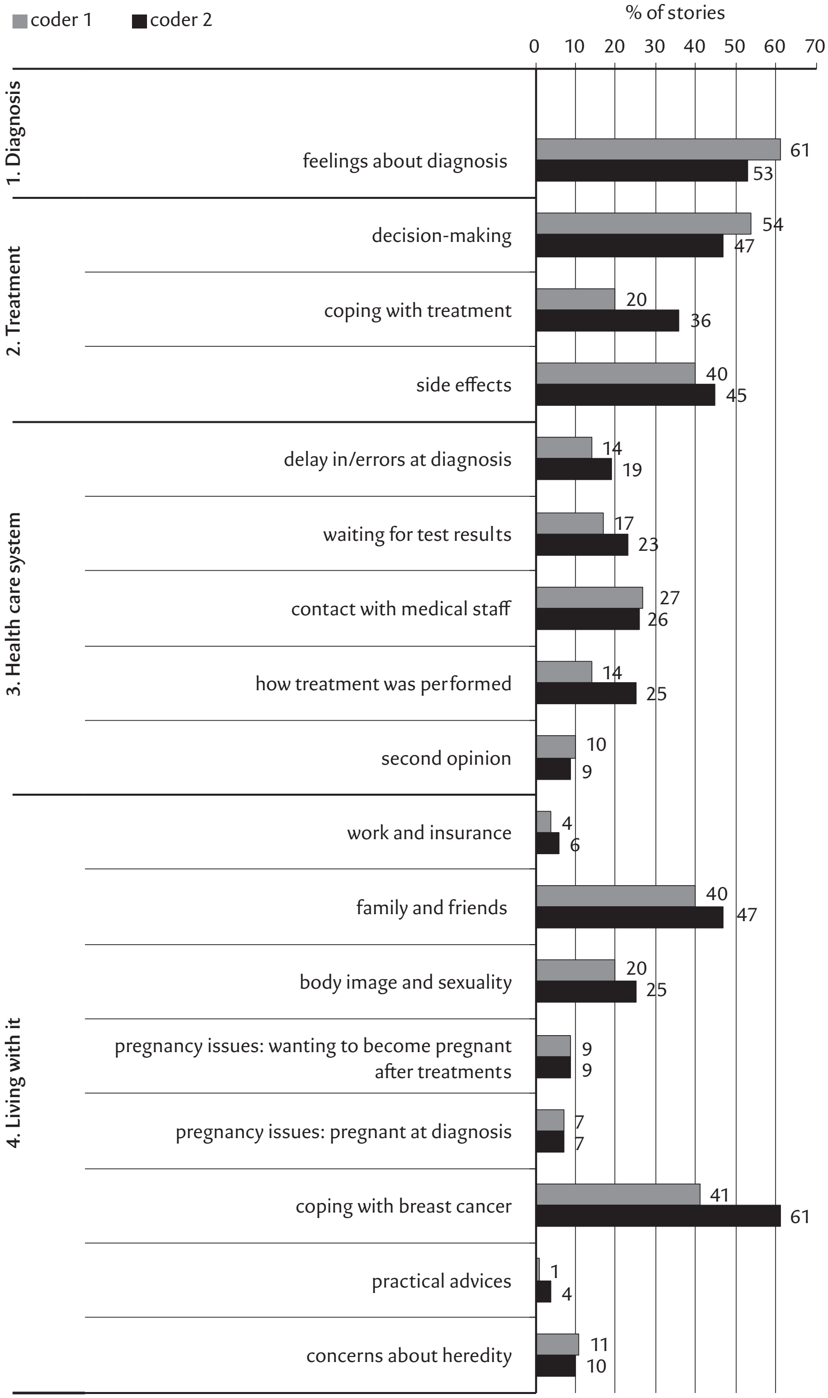


Figure 1 Frequency of themes in the Amazones stories (n=167).

An explanation for our finding that relatively many writers were in the first treatment period might be that the period right after hearing the diagnosis is the most difficult for a patient. In that period many things happen that disturb the normal daily routine. The patient has to find a way to cope with her altered health situation and with the uncertainty this involves. As a result the need to write about it might be relatively high right after hearing the diagnosis.

Patients might predominantly write about the themes that are keeping them busy at the time of writing their story. Since relatively many writers were in the first treatment period, our finding that the themes most frequently written about were the themes directly related to survival (feelings about diagnosis, decision-making about treatment, side effects of treatment) and the effects of illness on daily life (family and friends, coping with breast cancer) are in line with what could be expected. It also may explain our finding that in only few stories themes such as ‘work and insurance’ and ‘practical advices’ were addressed. Such themes may become more important after treatments.

The themes ‘body image and sexuality’ and ‘contact with medical staff’ did not appear in the set of stories in the extent that could be expected on the basis of quantitative studies among young women with breast cancer. Several explanations can be given for this finding. First of all, whether a patient writes about a theme might be dependent on the patient’s demographic and disease characteristics and on the patient’s phase in the course of disease. As mentioned above, a patient who is in the first treatment period will be most concerned with surviving and less with -for example- body image and sexuality. Secondly, it may be that certain themes do not come to a patient’s mind at the time of writing her story, because she has already coped with these themes or these themes are less important to her. For some patients the theme ‘contact with medical staff’ might fall in this category. In this regard the act of writing a personal story is a different type of self-disclosure than filling out a structured questionnaire or answering questions in a structured interview: in a questionnaire or interview a patient is actively asked whether or not he/she experiences problems with a certain theme. A third explanation might be that a certain theme does come to a patient’s mind at the time of writing her story, but that she decides not to write about it because in her opinion the theme is too personal which makes her not feel comfortable to publish her experiences with this theme on a freely accessible website. This explanation might hold for the theme ‘body image and sexuality’. However, when publishing a story on a website the writer can stay anonymous (for example by using a nickname). So, it can also be argued that the Internet is ideally suited to write about personal themes [22].

The question arises whether patients write about those themes other patients want to read about. An earlier study of us showed that breast cancer patients who were asked about what themes they want to read in stories of other patients ranked the themes ‘coping with emotions’ and ‘impact on daily life’ high [16]. The present study showed that the themes ‘feelings about diagnosis’, ‘coping with breast cancer’ and ‘family and friends’ are present in many stories. Thus, with regard to these themes, there is quite a match between the themes patients write about and the themes patients want to read about. Yet, in our earlier study [16] patients mentioned they also wanted to read about ‘practical information’ and ‘work rehabilitation’. The present study showed that these two themes are present in only a few stories.

When a theme is present in only a few stories this can be problematic in the case a website with a set of stories does not have a search facility on story topics: the reader has to scroll through all stories to find the few stories in which the theme is present. Yet, when a website with a set of stories does have a search facility on story topics it may be less problematic that a theme is present in only a few stories: the few stories in which the theme is present will be found relatively quick and easy. However, there are two other possible problems in the case a theme is only present in a few stories. First of all, the range of perspectives and experiences in respect of that theme might be limited. Second, a reader might get the idea that a theme that is written about in only a few stories will also be rare in real life. The latter need not to be true.

Organizations/patients who offer websites with stories should be motivated to describe how the set of stories on their website comes into being and how likely it is that a set of stories contains a broad range of themes and a broad range of perspectives and experiences in respect of these themes [14]. Such a description should certainly be applied to websites with spontaneously published stories, because -as the present study showed- on such websites the range of writer characteristics, perspectives and experiences may not fully reflect the wider population. Yet, when a patient has specific questions to other patients he/she can also consider to pose the question on a forum or to search on a forum for the answer instead of searching for stories. However, Van Uden-Kraan et al. [23] found in their study about health-related Internet use by patients with somatic diseases that only a small proportion of patients had ever read along with an online patient support group or had send postings to such a group.

5 Limitations

The two coders did not completely agree with respect to the writer profiles and the themes. The relatively low inter-rater reliability for 'immunotherapy' might be caused by the fact that few writers underwent immunotherapy. The two coders disagreed about 'immunotherapy' in only one of the 25 stories they both coded for writer characteristics. In addition, the relatively low inter-rater reliability for 'phase in the course of disease' might be caused by the six categories in which phase could be coded: this is quite a large number of categories. The two coders disagreed for 'phase in the course of disease' only in two of the 25 stories. For four of the themes there was some disagreement between the two coders (Cohen's Kappa <0.50). The coding instructions for these themes might not have been completely clear.

In order to examine how the writer characteristics relate to those of the wider population of young women with breast cancer, we compared the writer characteristics to the characteristics of young women with breast cancer in the LUMC cancer register. The LUMC cancer register was the best source we could think of to get insight in the wider population, but it should be noted that a university hospital register might contain a relatively large amount of more complicated cases. In order to make the comparison as useful as possible we selected a group of breast cancer patients from the register based on age, sex and year.

In order to examine how the frequency themes were written about in the stories relates to the frequency of these themes in the wider population, we compared the frequency of themes in the stories with the frequency of problems among young breast cancer patients participating in quantitative studies that use physiological outcome measures (pregnancy, BRCA-genes) or structured validated questionnaires. The background characteristics of participants in some of the quantitative studies differed slightly from those of the writers, which may have biased the comparison. In addition, the quantitative studies focused on difficulties participants had in respect of a particular theme, whilst we focused on whether the writers wrote about a theme at all.

Finally, it should be taken into account that websites are constantly changing. Patients will keep publishing stories on the *Amazones* website, which makes the composition of the set of stories dynamic.

6 Conclusion

This study shows that the set of spontaneously published stories on the *Amazones* website is written by a diverse group of writers and consists of a broad range of themes. Readers of the *Amazones* stories can get an acceptable impression of who may be affected by breast cancer and what it means to suffer from the disease. However, writers and themes on the *Amazones* website do not completely reflect the wider population of young women with breast cancer. Spontaneously published stories on the Internet might be subject to background characteristics of the writers, to main concerns at the time of writing one's story and to the desired degree of self-disclosure. Websites with spontaneously published stories should inform readers about how the set of stories comes into being and that the range of writer characteristics, perspectives and experiences may not fully reflect the wider population.

Acknowledgements

We would like to thank the board of the *Amazones* foundation for giving us permission to conduct this study, Henriët Werij-Wolters and Kristel Schaap for their coding work and Jan Molenaar for retrieving the data from the cancer register.

The Netherlands Organisation for Scientific Research (NWO) financed this study. This study is part of the Narrator project, under the ToKeN2000 umbrella program.

Declaration of Interest

The authors report no conflicts of interest

References

1. Fobair P, Stewart SL, Chang S, D'Onofrio C, Banks PJ, Bloom JR. Body image and sexual problems in young women with breast cancer. Psychooncology 2006;15:579-594.
2. Guinee VF, Olsson H, Moller T, Hess KR, Taylor SH, Fahey T, Gladikov JV, van den Blink JW, Bonichon F, Dische S, Yates JW, Cleton FJ. Effect of pregnancy on prognosis for young women with breast cancer. Lancet 1994;343:1587-1589.
3. Partridge AH, Gelber S, Peppercorn J, Sampson E, Knudsen K, Laufer M, Rosenberg R, Przypyszny M, Rein A, Winer EP. Web-based survey of fertility issues in young women with breast cancer. J Clin Oncol 2004;22:4174-4183.
4. Thewes B, Meiser B, Taylor A, Phillips KA, Pendlebury S, Capp A, Dalley D, Goldstein D, Baber R, Friedlander ML. Fertility- and menopause-related information needs of younger women with a diagnosis of early breast cancer. J Clin Oncol 2005;23:5155-5165.
5. De Sanjose S, Leone M, Berez V, Izquierdo A, Font R, Brunet JM, Louat T, Vilardell L, Borras J, Viladiu P, Xavier Bosch F, Lenoir GM, Sinilnikova OM. Prevalence of BRCA1 and BRCA2 germline mutations in young breast cancer patients: A population-based study. International Journal of Cancer 2003;106: 588-593.
6. Loman N, Johannsson O, Kristoffersson U, Olsson H, Borg A. Family history of breast and ovarian cancers and BRCA1 and BRCA2 mutations in a population-based series of early-onset breast cancer. Journal of the National Cancer Institute 2001;93:1215-1223.
7. Bloom JR, Stewart SL, Chang S, Banks PJ. Then and now: quality of life of young breast cancer survivors. Psychooncology 2004;13:147-160.
8. Walsh SR, Manuel JC, Avis NE. The impact of breast cancer on younger women's relationships with their partner and children. Families, Systems and Health 2005;23:80-93.
9. Smyth JM, Nazarian D, Arigo D. Expressive Writing in the Clinical Context. In: Vingerhoets AJJM, Nyklícek I, Denollet J, editors. Emotion Regulation. Conceptual and Clinical Issues. New York: Springer; 2008. p 215-33.
10. Pennebaker JW, Seagal JD. Forming a story: the health benefits of narrative. J Clin Psychol 1999;55:1243-1254.
11. Sharf BF, Vanderford ML. Illness Narratives and the Social Construction of Health. In: Thompson TL, Dorsey AM, Miller KI, Parrott R, editors. Handbook of Health Communication. London: LEA Publishers; 2003. p 9-34.
12. Healthtalkonline (formerly DIPEx) [Internet]. Breast Cancer in women [cited 2008 May 11]; Available from: http://www.healthtalkonline.org/Cancer/Breast_Cancer (http://www.webcitation.org/5eAfXVpWN)
13. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Med Educ 2006;40: 405-414.
14. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, Thomson R, Barratt A, Barry M, Bernstein S, Butow P, Clarke A, Entwistle V, Feldman-Stewart D, Holmes-Rovner M, Llewellyn-Thomas H, Moumjid N, Mulley A, Ruland C, Sepucha K, Sykes A, Whelan T. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006;333:417-422.
15. Butow P, Fowler J, Ziebland S. Section E: Using personal Stories. In: O'Connor A, Llewellyn-Thomas H, Stacey D, editors. International Patient Decision Aid Standards (IPDAS) Collaboration Background Document. 2005. p 24-27. Available from: http://ipdas.ohri.ca/IPDAS_Background.pdf

16. Overberg RI, Alpay LL, Verhoef J, Zwetsloot-Schonk JHM. Illness stories on the internet: what do breast cancer patients want at the end of treatment? Psychooncology 2007;16: 937-944.
17. De Amazones, jonge vrouwen met borstkanker/The Amazones, young women with breast cancer [Internet]. Ervaringsverhalen/stories [cited 2007 Jan 2]; Available from: http://www.de-amazones.nl/
18. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004;53:57-64.
19. KWF Kankerbestrijding/Dutch Cancer Society [Internet]. Videodagboek/video diary [cited 2008 Jun 4]; Available from: http://www.kankerbestrijding.nl (http://www.webcitation.org/5eAgM1yAc)
20. Molenaar S, Sprangers MA, Rutgers EJ, Luiten EJ, Mulder J, Bossuyt PM, van Everdingen JJE, Oosterveld P, de Haes HCJM. Decision support for patients with early-stage breast cancer: effects of an interactive breast cancer CDROM on treatment decision, satisfaction, and quality of life. J Clin Oncol 2001;19:1676-1687.
21. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001;323:1103-1105.
22. Joinson AN. Disinhibition and the Internet. In: Gackenbach J, editor. Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications. 2 ed. Burlington/San Diego/London: Academic Press/Elsevier; 2007. p 75-92.
23. Van Uden-Kraan CF, Drossaert CHC, Taal E, Smit WM, Bernelot Moens HJ, Siesling S, Seydel ER, van de Laar MAFJ. Health-related Internet use by patients with somatic diseases: frequency of use and characteristics of users. Informatics for Health & Social Care 2009;34:18-29.

CHAPTER 4

Illness stories on the Internet: what do breast cancer patients want at the end of treatment?

R.I. Overberg[1]
L.L. Alpay[1]
J. Verhoef[2]
J.H.M. Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden
[2] Department of Physical Therapy, Leiden University Medical Center, Leiden

Psycho-Oncology 2007; 16: 937-944

Abstract

The study aims to elicit user requirements for internet-based applications disclosing fellow patients' illness stories for the benefit of breast cancer patients. Twenty-six breast cancer patients, recruited via the Dutch Patient Organization for Breast Cancer, were interviewed about their preferences with regards to content, appearance, and search options concerning fellow patients' illness stories online. The interviews were analysed quantitatively (SPSS) and qualitatively (NVivo). Participants were mainly interested in fellow patients' experiences about how to cope with emotions, the impact of cancer in daily life, and physical discomforts. Most participants preferred a section of an illness story in text format about a specific topic; some of them wanted to be able to click on to the corresponding complete story, comprising of text alone or supported by voice or video clip. A majority of participants wanted to be able to select illness stories on the basis of several authors' features, i.e. treatment underwent, age, presence of metastases, time since diagnosis, and whether or not caring for children. Participants gave arguments for their preferences. The findings of this study will be used for designing an online trial with breast cancer patients aiming at refining the user requirements.

Keywords: Internet; breast cancer; illness stories; user requirements; patients' needs

1 Introduction

> You have received all necessary treatments. It has to be OK. And the world around you says: 'Well, come on, now we are facing normal life again, it's over!' But for yourself, it is not at all over… There is a life before [cancer], and a life after. (39-year-old breast cancer patient)

Cancer has an enormous impact on a patient's physical, psychological and social functioning. After treatments problems often last. Many patients experience long-lasting tiredness, pain, fear of recurrence, uncertainty about the future, difficulties in relationships with others, and problems with work rehabilitation [1]. Having cancer changes a person's activities and thoughts, and the way others treat and view him/her [2].

The life story, which every person constructs by weaving together the variety of events and experiences that make up life into a balanced entity, is disrupted by cancer. It can be repaired by constructing an illness story [3,4], which helps patients to (p. 16) '(a) make sense of health and disease, (b) assert control in the midst of physical and psychological losses, (c) transform their identities and social roles as a result of altered health and disease, and (d) make decisions about their health' [5]. Writing about personally upsetting experiences -like having cancer- can lead to significant health improvements [6,7]. Illness stories are not only beneficial for writers, we think, but also for readers who are in the same situation, fellow patients. They can use writers' illness stories as examples for repairing the life story.

A vast amount of illness stories can be found on the Internet. These stories can anonymously be read when and where readers want [8]. However, having studied features of websites disclosing breast cancer patients' illness stories in the Dutch language, we found neither search facilities based on the content of stories nor on personal features of authors [9]. This complicates both finding stories compiling of relevant topics and finding stories written by comparable fellow patients. Yet, we argue that stories that meet these two criteria are most useful in repairing one's own life story.

In order to develop websites that match patients' needs, the topics breast cancer patients want to read about in fellow patients' stories online and the way they want the stories to be presented to them should be known. To our knowledge no studies have been conducted that address this issue. Studies that have examined breast cancer patients' needs tend to focus on medical information needs [10,11] rather than on the needs concerning fellow patients' illness stories. Moreover, studies that have examined breast cancer patients' listservs [12,13] focused primarily on the topics participants write about and not on the topics they want to read.

Website developers should also be familiar with the personal features breast cancer patients want to know from authors of illness stories in order to identify with them. Bennenbroek [14] studied 'upward' and 'downward' comparison among cancer patients (i.e. comparing with others who are doing better, respectively, worse), but focused not on features such as fellow patient's age or treatment underwent. Rozmovits and Ziebland [15] found that is was positively viewed by their participants to be able to select fellow patients of a particular age, stage of illness, long-term survivors, or someone having opted for similar treatment, when looking for experiences of others. Yet, they did not investigate this further.

The study aim was to elicit user requirements for internet applications that disclose breast cancer patients' illness stories, using individual interviews. We focused on content and appearance of stories, and search options for authors' features.

2 Methods

2.1 Participant recruitment

The Dutch Patient Organization for Breast Cancer [16] published a recruitment advert in their magazine. To recruit a diverse group of participants no limitations were imposed on age, treatment underwent or phase of the breast cancer. When a woman decided to participate an interview appointment was made, her address was asked in order to send written information, and age, time since diagnosis and treatment underwent were noted.

In response to the advert 38 women called, the last seven were noted as reserve participants. Three of the first 31 callers appeared to be not interested in participation, and two women cancelled their appointment due to medical reasons. A number of 26 participants was seen as sufficient for exploring the user requirements.

2.2 Interview questionnaire

The first part of the self-developed questionnaire aimed at letting participants get used to the interview setting and at introducing the main theme by asking about the advantages and disadvantages of several ways to hear about fellow patients' experiences. Parts two to four focused on participants' preferences regarding illness stories' content, appearance, and authors. Part five comprised participants' demographic and health characteristics.

Card-ranking was used to elicit preferred content. This method was selected after test interviews demonstrated that ticking off topics one wants to read about from a list of topics was not useful due to the 'ceiling effect' [17] and that paired comparisons of Thurstone [18] were not applicable due to one's difficulty with choosing and its dullness as a result of the great number of pairs. In the final questionnaire, participants were asked to rank eight cards describing important topics in the scope of coping with breast cancer (identified in [19]) according to the degree they want to read about these in fellow patients' stories. While ranking the cards, participants were asked to think aloud, and to think back to the period they returned home after completion of treatments in hospital. We think that illness stories disclosing internet applications are most beneficial in this period during which many breast cancer patients struggle (e.g. [20]). Participants were also asked whether they want to read about other topics than the eight card topics.

Preferred appearance was also elicited using card-ranking. Participants were asked to rank six cards describing every possible combination of the following two aspects (inspired by [21]): (1) section of an illness story about a specific topic or a complete story about the whole course of illness, and (2) only text, text supported by voice, or text supported by video clip.

In the questions about preferred search options for authors' features clear examples were given to ensure that participants could imagine these accurately. We decided to do this after test interviews demonstrated that questions of the format 'Do you want to know the author's religion?' appeared to be difficult with chance on social desirable answers. In the final questionnaire we therefore used the format: 'We can divide fellow patients' illness stories on the website according to the authors' religion. We can make several buttons, for example: (1) stories of religious women, and (2) stories of women without a religion. Do you want this search option for author's religion?'. Participants were asked for an explanation. After 12 search options were asked for in this way, participants were asked to which three they gave priority and whether they wanted any other author search options.

2.3 Interview procedure

The first author conducted the interviews. An interview started when the interviewee had given permission for audio taping, had signed a consent form, and had no further questions. Interviews took place in booked rooms where interviewer and interviewee could talk without being disturbed in Leiden (LUMC) or in Utrecht (at the patient organization or a colleague's home) and lasted one to one and half hours.

Table 1 Participant characteristics (n=25)[a]

Characteristics		n (%)
Interview age	≤40 yrs	5 (20)
	41-50 yrs	6 (24)
	51-60 yrs	9 (36)
	>60 yrs	5 (20)
Marital status	Single	4 (16)
	Married/ partnered	21 (84)
Education level[b]	Low	2 (8)
	Moderate	9 (36)
	High	12 (48)
Employment at diagnosis	Employed	17 (68)
	Economically inactive	8 (32)
Time since diagnosis	0-24 months	6 (24)
	25-48 months	12 (48)
	>48 months	7 (28)
Diagnosis age (estimated)	≤40 yrs	5 (20)
	41-50 yrs	10 (40)
	51-60 yrs	8 (32)
	>60 yrs	2 (8)

Table 1 (continued)

Characteristics		n (%)
Treatment underwent[c]	Mastectomy	21 (84)
	Breast conserving therapy	10 (40)
	Chemotherapy	17 (68)
	Radiotherapy	12 (48)
	Hormone therapy	14 (56)
Treatment at interview	Chemotherapy and/or radiotherapy	0 (0)
	Hormone therapy and/or check up	25 (100)
Ways to hear about fellow patients' experiences[c]	Face-to-face support groups	17 (68)
	Telephone helplines	14 (56)
	Books and magazines	23 (92)
	Television programmes	22 (88)
	CD-ROMs	3 (12)
	E-mail messages	4 (16)
	Listservs/chat rooms	2 (8)
	Websites with illness stories	18 (72)

[a] One participant was excluded from the analyses due to not being a breast cancer patient, but a preventatively treated BRCA-carrier.

[b] Two missing values.

[c] Does not count till n=25 or 100%, because participants underwent more than one treatment, respectively, used more than one way to hear about fellow patients' experiences.

2.4 Data analysis

The audiotape recordings were transcribed. Quantitative data were analysed in SPSS [22] and qualitative data in NVivo [23]. The data analysis was carried out by the first author, who regularly consulted with the co-authors.

3 Results

3.1 Participant characteristics

One participant was excluded from the analyses due to not being a breast cancer patient, but a preventatively treated BRCA-carrier. The youngest participant was 36 years of age, the oldest 67 (mean= 51; SD= 9.6). Most women were married or living together, higher professional or university educated, and working at the time of diagnosis (Table 1). Time since diagnosis ranged from 7 to 144 months (mean= 41; SD= 28.0). The diagnosis age varied from 32 to 63 years (mean= 48; SD= 9.1) and was estimated by rounding the number of months since diagnosed into years, and subsequently subtracting it from the interview age.

All 25 participants underwent surgery: 15 women (60%) underwent mastectomy; four women (16%) underwent breast conserving therapy; and six women (24%) underwent both. Four women (16%) did not receive any kind of adjuvant therapy, some women received one (n= 8; 32%) or two (n= 4; 16%) kinds. Nine women (36%) received chemotherapy as well as radiotherapy and hormone therapy. None of the women received chemotherapy or radiotherapy at the time of the interview.

All participants were accustomed to hearing about fellow patients' experiences, but they used different ways (Table 1). The mentioned advantages and disadvantages pointed out that participants appreciate: (1) content that is relevant to them (i.e. stories about topics that interest them without extensive elaborations); (2) fellow patients who are comparable regarding treatments underwent; and (3) being able to select stories on the basis of (a) the ending, and/or (b) the writing style.

3.2 Content requirements

Figure 1 shows the results of the card-rank task about preferred content. The topic *coping with emotions* was the most frequently placed high in the order of rank (median= 3; inter-quartile range= 1–3), meaning that the majority of the participants wanted to read about this in fellow patients' stories online when being home again. The topics *impact on daily life* (median= 4; inter-quartile range= 2–4) and *physical discomforts* (median= 4; inter-quartile range= 2–5) were also often placed high. The topic *dealing with physicians* was the most frequently placed low in the order of rank (median= 7; inter-quartile range= 6–8), meaning that there was little need to read about this topic in fellow patients' stories.

The inter-quartile ranges in combination with the highest and lowest place on which a topic was placed indicate the degree of discord between the participants. Participants agreed most about the topics *coping with emotions* and *dealing with physicians*: both topics have an inter-quartile range of only three places (1–3 respectively 6–8) that coincides with an extreme place (1 respectively 8). Most disagreement was about the topic *accepting breast cancer*: the topic was placed on place 1 and 8, and the inter-quartile range comprises exactly the middle of the order of rank (i.e. 3–6).

The think-aloud part pointed out that ranking of the topics was influenced by participants' own experiences and thoughts. They placed a topic high in the order of rank when they had been:

- experiencing long-lasting problems regarding a topic (e.g. long-lasting tiredness or arm lymphedema);
- thinking about a topic over and over (e.g. the question 'Why me?');
- experiencing a topic negatively (e.g. *dealing with physicians*).

Participants placed a topic low in the order of rank when:

- they had not been experiencing a topic at all (e.g. participants who experienced no physical discomforts) or any problems regarding a topic (e.g. participants who dealt very well with physicians);
- they had not been thinking about a topic or already had finished thinking about it (e.g. the question 'Why me?');
- they found a topic only relevant during treatments (e.g. *dealing with physicians*);
- they thought that fellow patients could not give useful comment on a topic (e.g. *the why of cancer*).

Three new topics emerged when participants were asked for any additional topics they wanted to read about in fellow patients' stories:

1. Practical information (e.g. prostheses, wigs, swimming suits, lingerie)
2. Work rehabilitation (returning to profession: how, why, when, etc.).
3. Medical information (treatment and reconstruction options, side effects, heredity).

3.3 Appearance requirements

Due to the combination of two aspects on the appearance cards, seven participants (28%) could not rank them properly. Therefore, instead of analysing the ranking we summarized participants' preferences for the two aspects separately.

As depicted in Figure 2, 15 participants (60%) preferred sections about specific topics, six participants (24%) preferred complete stories, and four participants' preferences (16%) varied. Arguments for preferring sections were: being interested in specific topics; the inconvenience to read long texts on a computer screen; the difficulty to be patient and concentrated when reading complete stories; and, the fact that complete stories deal too much and too long about one particular fellow patient. Eight of the 15 participants who preferred a section wanted to click on to the corresponding complete stories when they got interested. Approaching fellow patients' stories in this way prevents feeling overwhelmed, some participants explained.

Participants who preferred complete stories about the whole course of illness explained wanting to know all in order to gain an accurate view of a fellow patient's situation and to enhance the chance of getting answers to their questions. Participants whose preferences varied explained wanting complete stories just after the diagnosis to read about all the things one has to face, and when gone through all the things (i.e. after treatments) wanting sections to read about specific topics. One participant explained also wanting complete stories when feeling sad to find consolation, but otherwise wanting to read sections.

Figure 3 shows that 13 participants (52%) preferred text only, four participants (16%) preferred text supported by voice, three (12%) preferred text supported by video clip, and five participants' preferences (20%) varied. Arguments for preferring text only were that voice and video clip distract one's attention, do not offer additional value, delay reading, and make it too personal and emotional. Text is least penetrating and quite impersonal and thus appropriate for becoming acquainted with fellow patients' stories, according to some participants. Three of the 13 participants who preferred text wanted to click on to voice or video clip when feeling comfortable with the text.

Participants who preferred text supported by voice explained that voice is easy to take up, gives the idea of really having contact with others, and since it is quite penetrating one cannot deny being a cancer patient. They mentioned turning off the sound when not wanting to hear a fellow patient's voice anymore. Participants who preferred text supported by video clip argued that stories become more human when seeing the author and her facial expressions, and that assessing a fellow patient's personality/type is easier.

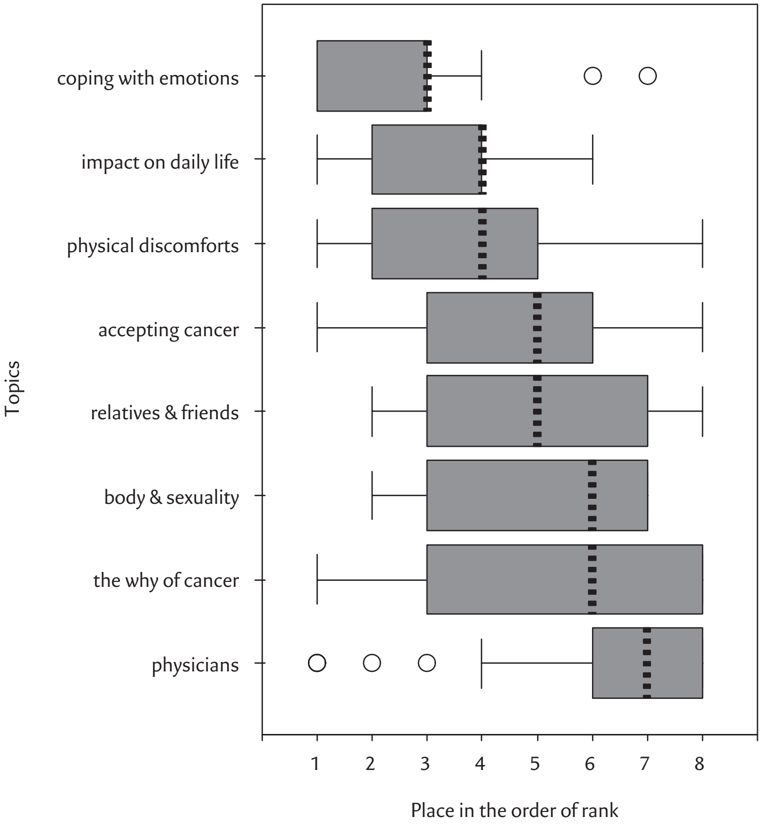


Figure 1 Boxplot of the results of ranking eight topic cards by 25 participants. Place 1= most need, ranging to place 8= least need to read in fellow patients' stories. Per topic is presented: the highest (├) and lowest place (┤) on which the topic is laid, any extreme values (circles), the median (dotted lines), and the middle 50% of observations (i.e. the inter-quartile range; grey rectangles). One participant laid two topics highest in the ranking, to these two the mean of 1.5 is assigned (=(1+2)/2).

Participants with varying preferences explained that their preferences depend on their own emotions, the author's emotions, and the story's length and writing style. When being self very sad or when an author is very emotional reading is preferred above watching video clip. Secondly, video clip during a fragment of a story is fine, but will be too intense during a complete story. Finally, when a story is an enumerative description of all the things an author went through video clip instead of text only is preferred.

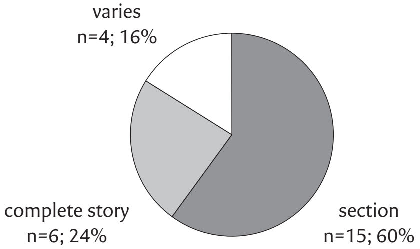


Figure 2 Participants' preferences regarding length and specification of online illness stories. Eight participants who preferred a section wanted to be able to click on to the complete stories.

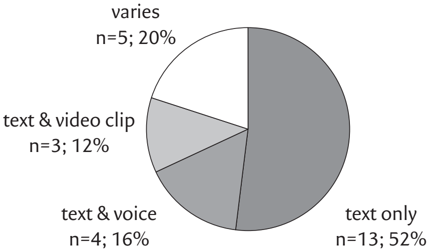


Figure 3 Participants' preferences regarding multimedia support of online illness stories. Three participants who preferred text only wanted to be able to click on to video clip.

3.4 Author requirements

Figure 4 shows participants' preferences concerning search options for authors' features. One search option that was asked for, i.e. future treatments (yes/no), was excluded from the analyses due to being ambiguous: no future treatments can mean both that one has cured from the cancer or that one's cancer has become incurable. The most wanted search options were those for authors' treatment underwent (n= 25; 100%), age (n= 24; 96%), presence of metastases (n= 24; 96%), time since diagnosis (n= 18; 72%), caring for children (n= 17; 68%), and time since last treatment (n= 15; 60%). Less than half of the participants wanted the other five author search options: hobby (n= 12; 48%), marital status (n= 11; 44%), profession (n= 10; 40%), religion (n= 8; 32%), and education level (n= 4; 16%).

Three main reasons emerged for wanting an author search option. Firstly, being able to search for one's own situation and find recognition in the case that different values of a particular feature lead to considerably different stories. For example, participants wanted the age search option because they noted that younger and older patients differ regarding problems in daily life and medical nature of their cancer. Secondly, being able to select or to exclude certain stories. For example, some participants wanted the search option for time since diagnosis to select stories of long-term survivors that can serve as encouraging examples, and some wanted the presence of metastases search option to exclude stories of women with metastatic breast cancer when not wanting that confrontation. Thirdly, being interested in a topic connected to a personal feature. For example, some participants wanted the search option for treatment underwent to read about why some women chose for breast conserving therapy and others for amputation.

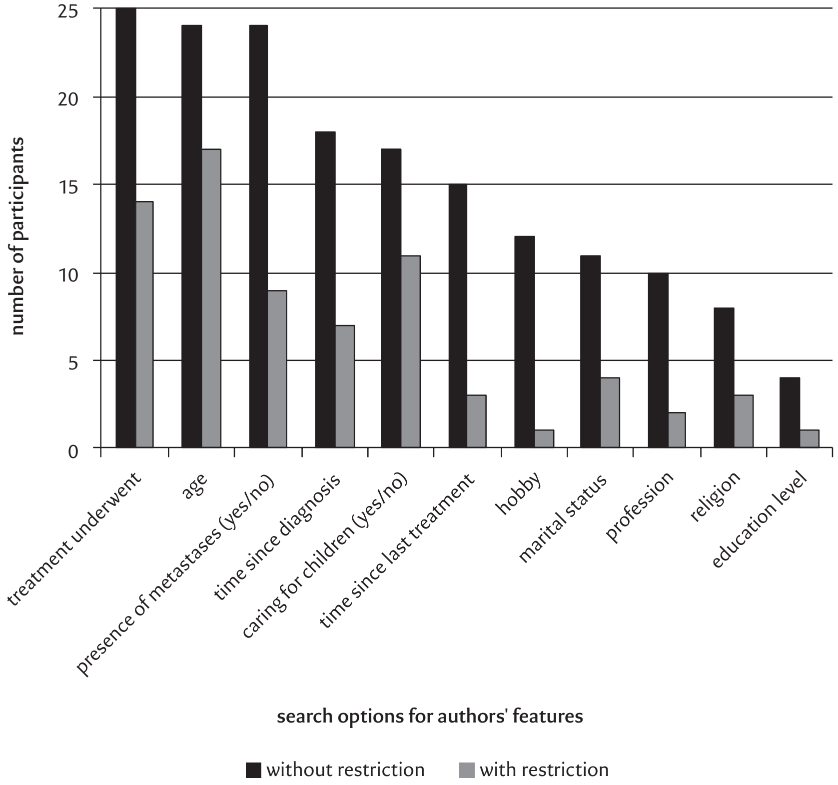


Figure 4 Histogram of preferred search options for authors' features, asked without restriction ('Do you want search option x?') or with restriction ('Which three search options do you give priority?')

Different reasons emerged for not wanting an author search option. When participants thought that different values of a particular feature did not lead to different stories (e.g. hobby), or when according to the participants being ill and having experiences with breast cancer was far more important than the feature (e.g. education level). Some participants did not want a search option because they themselves had a particular value of the feature. For example, some participants who were not religious and therefore not interested in religion did not want the religion search option. Yet, it could be argued that they actually want the search option to exclude stories of religious women (see Discussion). Thirdly, according to some participants separating on certain features is 'unfriendly' (e.g. age), and therefore they did not want these search options.

In addition, one participant expressed her concern about the last treatment search option, because patients do not know whether a last treatment is really the last one. Besides, participants saw a strong relation between the search options for time since diagnosis and time since last treatment, and also, although a weaker one, between the search options for age and caring for children.

When participants were asked which three author search options they gave priority, two were prominent: age (n= 17; 68%) and treatment underwent (n= 14; 56%). These were followed by authors' caring for children (n= 11; 44%), presence of metastases (n= 9; 36%), and time since diagnosis (n= 7; 28%). The necessity of the other search options is limited, ranging from n= 4 (16%; marital status) to n= 1 (4%; hobby and educational level). No other author features emerged when participants were asked for any additional search options they want to have at their disposal.

4 Discussion

This study's purpose was to elicit user requirements for internet-based applications that disclose fellow patients' illness stories for breast cancer patients. Our motivation to this stems from the enormous impact of breast cancer on patients [1,2], the usefulness of hearing and constructing illness stories [3–6], the current lack of search facilities concerning illness stories on the Internet [9], and the fact that applications that match with patients' needs maximize the chance that patients will use the applications and benefit from their use (e.g. [24]).

Recruitment via the Dutch Patient Organization for Breast Cancer [16] may have resulted in a relatively active group of participants. After all, members have put effort in applying for membership and are probably active information seekers. Therefore, the requirements found reflect above all the preferences of active information seekers. Yet, in our opinion, it is precisely this group that uses internet applications most often. Single, lower educated, and economically inactive breast cancer patients were underrepresented and should therefore be involved in future examinations of user requirements. Although the number of participants was limited, we have gained valuable insights from the detailed explanations participants gave for their preferences.

In general, participants wanted to read in fellow patients' illness stories about topics they are themselves suffering from, thinking about, or experiencing as negatively. Thus, participants seemed to take their own story as starting point when searching for fellow patients' stories. This finding supports our thought that patients use fellow patients' stories to further develop and construct their own story into a balanced entity. However, this may not be applicable to all topics. Participants did not want to read about topics of which they thought that fellow patients could not give useful comment on. Since patients' own experiences and thoughts influence what they want to read in fellow patients' stories, we cannot compile an universal list of content requirements. In addition, the fact that *practical information* was often mentioned as additional topic indicates that patients also want to try to make their life more comfortable by following fellow patients' advices.

Since many participants were interested in particular topics and saw text as sufficient, story sections in text format were most preferred. A substantial part of the participants wanted to click on to the corresponding complete story when the section interested them. These findings point out that internet applications that disclose illness stories should: (1) offer a search option using content keywords; and (2) present search results as story sections in text format with the possibility to click on to the complete stories. We realize that not all users will be satisfied with this, since some participants preferred complete stories, voice or video clip, or had varying preferences. However, we think that implementing these two requirements will enable most patients to find quickly and easily what they are looking for, which will be a great improvement compared to the current situation [9].

Most participants wanted the search options for author's treatment underwent and author's age, predominantly to search for their own situation. This finding seems to indicate that patients need these two features in order to be able to identify themselves with a fellow patient. Furthermore, many participants wanted the author search options for presence of metastases and time since diagnosis, among other reasons to search for their possible future situation. This finding suggests that patients are also thinking about their future. These four most wanted search options are in line with the findings of Rozmovits and Ziebland [15], except maybe the slight difference between their 'stage of illness' and our 'presence of metastases'.

The most wanted author search option regarding demographic characteristics -with the exception of age- is caring for children. All other demographic characteristics that were asked for (hobby, marital status, profession, religion, and education level) were only wanted by a few participants as author search options. This finding indicates that patients give priority to authors' health characteristics when searching for fellow patients' illness stories. Furthermore, the question 'Do you want to know the author's feature x?' differs slightly from the final question 'Do you want to search for an author's feature x?', since wanting to know feature x without wanting a search option is possible. Because of this we may have found fewer author requirements, but not incorrect ones, since wanting to search for an author's feature x inherently means wanting to know the author's value for that feature.

The mentioned advantages and disadvantages of several ways to hear about fellow patients' experiences -in the first part of the interview- pointed out that participants liked to hear relevant content from fellow patients who underwent similar treatments as they did. These two wishes were also found in the remainder of the interview as user requirements for internet applications. In theory it will also be possible to implement participants' other wish in an internet application, i.e. being able to select illness stories on the basis of the ending and/or writing style. However, in that case users may only select stories with a positive ending, with the consequence that they will not get a fully representative overview. Moreover, since the writing style of a story is rather subjective, it might in practice be difficult to group stories on the basis of their style.

We would like to make a few comments on our interview questionnaire and on choosing interviewing as method. Firstly, the card-rank task for content consisted of a limited number of topics (eight), which overlapped slightly. For example, *coping with emotions* can influence one's relationship with *relatives and friends*. The card-rank task for appearance was rather difficult, because a combination of two aspects was described on each card. Yet, in our opinion we did overcome these problems by asking also open-ended questions about content and by asking participants to think aloud. Secondly, the explanation of participants with a varying preference for length of a story (i.e. wanting complete stories just after the diagnosis and wanting sections a few years later) indicates that thinking back to the period of returning home after completion of treatments in hospital was difficult. However, in our opinion one cannot interview patients on that emotionally hard moment. Besides, we do not think that we found biased requirements, since we repeatedly reminded participants to think back to that period during the interview. Thirdly, in spite of changing the abstract questions about authors' features into more concrete questions with clear examples, some participants said not wanting particular search options, while their explanation suggested the opposite. Yet, the only feature where this might have led to a bias is author's religion, since for this feature the answers of seven participants seemed not to fit with their explanations. These inconsequent answers may not have occurred, if we had elicited the requirements with the help of a prototype instead of with interviewing. But, with a prototype participants would have been restricted to a particular framework leading them in a particular direction. On a whole, we think that our questionnaire and our choice for interviewing as method were very suitable for a first exploration of user requirements.

This study's findings will be used for designing an online randomized controlled trial aiming at refining the user requirements.

5 Conclusion

Internet applications that disclose fellow patients' illness stories for breast cancer patients should:

- offer search options for stories' content, authors' age and treatment underwent, and authors' presence of metastases and time since diagnosis (or at least mention the authors' values for these features);
- present search results as story sections about the topic that was searched for with the possibility to click on to the corresponding complete stories, both in text format.

This list of universal user requirements we found is rather short, since preferences differed between participants, and sometimes even differed for one and the same participant. In order to make applications valuable to all users, one possible solution would be to offer users an appropriate dialogue manager through which they can make their preferences explicit. Yet, developing such a dialogue manager requires new technical insights and is time-consuming. Meanwhile, applying our list of user requirements will be a big step forward, because applications will then match more with patients' needs which enhance the chance that patients use them and benefit from their use.

Acknowledgements

The authors wish to thank: the participants for their cooperation; the Dutch Patient Organization for Breast Cancer for publishing the advert and for providing an interview room; Prof Dr Jean Pruyn (Institute for Health and Environmental Issues, The Netherlands) for discussions on the topic and for commenting on the draft; and the Netherlands Organisation for Scientific Research (NWO) for financing this study (which is part of the Narrator project, umbrella program ToKeN2000).

The authors state that this manuscript (or parts of it) has not been submitted or published elsewhere and that there is no conflict of interest. They received a declaration of no objection against conducting the reported research from the ethical committee of the Leiden University Medical Centre.

References

1. Sprangers MAG, Smets EMA, Stiegelis H. Gevolgen van de ziekte (Consequences of the disease). In Psychologische patiëntenzorg in de oncologie. Handboek voor de professional (Psychological Patient Care in Oncology. Handbook for the Professional), De Haes JCJM, Gualthe´ rie van Weezel LM, Sanderman R, Van de Wiel HBM (eds). Koninklijke Van Gorcum BV: Assen, 2001; 34–46.
2. Radley A. Making Sense of Illness. SAGE Publications: London, Thousand Oaks, New Delhi, 1994.
3. Kleinman A. The Illness Narratives. Suffering, Healing, and the Human Condition. Basic Books, Inc., Publishers: New York, 1988.
4. Frank AW. The Wounded Storyteller. Body, Illness, and Ethics. The University of Chicago Press, Ltd.: London, 1995.
5. Sharf BF, Vanderford ML. Illness narratives and the social construction of health. In Handbook of Health Communication, Thompson TL, Dorsey AM, Miller KI, Parrott R (eds). Lawrence Erlbaum Associates: NJ, 2003;9–34.
6. Pennebaker JW, Seagal JD. Forming a story: the health benefits of narrative. J Clin Psychol 1999;55(10): 1243–1254.
7. Pennebaker JW, Zech E, Rime´ B. Disclosing and sharing emotion: psychological, social and health consequences. In Handbook of Bereavement Research: Consequences, Coping and Care, Stroebe MS, Hansson RO, Stroebe W, Schut H (eds). American Psychological Association: Washington, DC, 2001; 517–543.
8. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356–371.
9. Overberg R, Toussaint P, Zwetsloot-Schonk B. Illness stories on the Internet: features of websites disclosing breast cancer patients' illness stories in the Dutch language. Patient Educ Couns 2006;61:435–442.
10. Rees CE, Bath PA. The information needs and source preferences of women with breast cancer and their family members: a review of the literature published between 1988 and 1998. J Adv Nurs 2000;31(4):833–841.
11. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). Patient Educ Couns 2005;57(3):250–261.
12. Weinberg N, Schmale J, Uken J, Wessel K. Online help: cancer patients participate in a computer-mediated support group. Health Soc Work 1996;21(1):24–29.
13. Hoybye MT, Johansen C, Tjornhoj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. Psycho-Oncology 2005;14(3): 211–220.
14. Bennenbroek F. Social Comparison and Coping with Radiation Therapy: The Significance of Different Dimensions of Comparison. University of Groningen, Groningen, 2003.

15. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004;53(1):57–64.
16. BVN. Website van BorstkankerVereniging Nederland (Website of the Dutch Patient Organization for Breast Cancer). http://www.borstkanker.nl (15 February 2006).
17. Bilodeau BA, Degner LF. Information needs, sources of information, and decisional roles in women with breast cancer. Oncol Nurs Forum 1996;23(4):691–696.
18. Nooij ATJ. Variabelen en schalen (Variables and scales). In Sociale Methodiek (Social Methodology). Stenfert Kroese: Leiden, 1996; 151–179.
19. Van den Borne B, Jonkers R, Pruyn J, Serail T. Over kanker geschreven: dagboekfragmenten voor lotgenoten (Written About Cancer: Diary Fragments for Fellow Patients). De Toorts: Haarlem, 1987.
20. Ganz PA, Kwan L, Stanton AL et al. Quality of life at the end of primary treatment of breast cancer: first results from the moving beyond cancer randomized trial. J Natl Cancer Inst 2004;96(5):376–387.
21. DIPEx. Personal Experiences of Health and Illness. http://www.dipex.org (10 May 2005).
22. SPSS for Windows. Release 11.5.0 (6 Sep 2002).
23. QSR NVivo Version 2.0.163 (2002).
24. Rosson MB, Carrol JM. Usability Engineering: Scenario-Based Development of Human–Computer Interaction. Morgan Kaufmann: San Francisco, 2002.

CHAPTER 5

How breast cancer patients want to search for and retrieve information from stories of other patients on the Internet: an online randomised controlled experiment

Regina Overberg[1]
Wilma Otten[2]
Andries de Man[1]
Pieter Toussaint[3]
Judith Westenbrink[4]
Bertie Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden
[2] Medical Decision Making, Leiden University Medical Center, Leiden
[3] Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway
[4] The Amazones Foundation, Haarlem

Journal of Medical Internet Research 2010; 12(1): e7

Abstract

Background Other patients' stories on the Internet can give patients information, support, reassurance, and practical advice.

Objectives We examined which search facility for online stories resulted in patients' satisfaction and search success.

Methods This study was a randomized controlled experiment with a 2x2 factorial design conducted online. We facilitated access to 170 stories of breast cancer patients in four ways based on two factors: (1) no versus yes search by story topic, and (2) no versus yes search by writer profile. Dutch speaking women with breast cancer were recruited. Women who gave informed consent were randomly assigned to one of four groups. After searching for stories, women were offered a questionnaire relating to satisfaction with the search facility, the stories retrieved, and impact of the stories on coping with breast cancer. Of 353 enrolled women, 182 (51.6%) completed the questionnaire: control group (n = 37), story topics group (n = 49), writer profile group (n = 51), and combination group (n = 45).

Results Questionnaire completers were evenly distributed over the four groups ($\chi 23 = 3.7$, P = .30). Women who had access to the story topics search facility (yes vs no): were more positive about (mean scores 4.0 vs 3.6, P = .001) and more satisfied with the search facility (mean scores 7.3 vs 6.3, P < .001); were more positive about the number of search options (mean scores 2.3 vs 2.1, P = .04); were better enabled to find desired information (mean scores 3.3 vs 2.8, P = .001); were more likely to recommend the search facility to others or intend to use it themselves (mean scores 4.1 vs 3.5, P < .001); were more positive about how retrieved stories were displayed (mean scores 3.6 vs 3.2, P = .001); retrieved stories that better covered their information needs (mean scores 3.0 vs 2.6, P = .02); were more satisfied with the stories retrieved (mean scores 7.1 vs 6.4, P = .002); and were more likely to report an impact of the stories on coping with breast cancer (mean scores 3.2 vs 2.9, P =. 02). Three main effects were associated with use of the writer profile search (yes vs no): being more positive about (mean scores 3.9 vs 3.6, P = .005) and more satisfied with the search facility (mean scores 7.1 vs 6.5, P =. 01), and being more positive about how retrieved stories were displayed (mean scores 3.8 vs 2.9, P < .001). For satisfaction with the search facility, an interaction effect was found (P = .03): at least one of the two search facilities was needed for satisfaction.

Conclusions Having access to the story topics search facility clearly had the most positive effect on patient satisfaction and search success.

Keywords: Breast cancer; life experiences; social support; Internet; information retrieval; patient satisfaction

1 Introduction

Patients value having access to stories of other patients as it provides them with emotional support, information, reassurance, and practical advice [1]. The Internet is a valuable resource for accessing stories because of its privacy and 24-hour availability without the need to leave one's home [2]. Two well-known examples of web-based applications that include personal stories of patients are The Comprehensive Health Enhancement Support System (CHESS) [3] and The Database of Individual Patients' Experiences of Illness (DIPEx) [4].

Studies of online patient stories have focused on several factors. These include why patients publish their stories online and what this means in a broader sociological context [5-7]. Wise et al [8] found that accessing personal stories in a computer-based patient support system had a positive effect on patients' healthcare participation, which entailed participation preferences, confidence, and communication with their doctor. Little is known, however, about how patients search online for stories of other patients and whether they can find relevant ones.

Some qualitative studies have found that patients appreciate the ability to select stories of other patients of a particular age or who have opted for similar treatment [1,9]. In addition, searching by topics seems also to be of interest [10]. Some websites with patient stories provide a search facility to search for personal characteristics of the story writers and/or for topics written about in the stories [11-13]. However, to our knowledge, patients' satisfaction and search success with these search facilities have not yet been studied.

In the present study, we examined which search facilities for patient stories resulted in satisfaction with the search process and the stories retrieved. We also studied the impact of the stories retrieved on coping with illness. Our expectation was that having a search facility would be an improvement compared with not having a search facility. Moreover, we expected that a combination of search facilities would result in higher satisfaction than a single search facility because a combination of search facilities may result in more opportunities to find a relevant story.

2 Methods

2.1 Design and Procedure

Study Design

We focused our study on patients with breast cancer. We contacted the board of The Amazones Foundation, which was founded by a group of young women with breast cancer to provide their peers with information and support. The Amazones Foundation developed a website for young women with breast cancer [14] that provides information and advice, a calendar of activities, an online support group, and links to other sites. The website also has a section with personal stories. Women can anonymously submit their own story to the site. The stories are presented alphabetically by writers' nicknames. If a writer passes away, her story remains on the site accompanied by an obituary written by the website moderators.

We were granted permission by the board of The Amazones Foundation to conduct our study. In January 2007 we downloaded all 170 stories available at that time on their website for use in our study. We facilitated access to the stories in three ways: (1) with a search facility for story topics, (2) with a search facility for writer profiles, and (3) with a combination of these two search facilities. In addition, a control group could access the stories by means of the original alphabetical listing by story writer. We implemented these four ways of facilitating access to the stories on a separate study website. This resulted in four groups based on two independent factors: (1) no versus yes search by story topics, and (2) no versus yes search by writer profile (Table 1). In each of the four groups the same set of 170 stories could be searched. Figures 1 to 4 show screenshots of the search pages available to the four groups. We requested that the board of the Amazones Foundation not to participate in the study.

The present study is reported in accordance with the CHERRIES checklist, which is a checklist for reporting results of Internet e-surveys [15]. It was not registered as a clinical trial on ClinicalTrials.gov, a registry of clinical trials conducted around the world, because our study does not correspond to the definition of a clinical trial as provided in their glossary.

Table 1 The 2x2 factorial design of the study

		Writer profile search	
		No	Yes
Story topics search	No	control group: See figure 1.	writer profile group: See figure 3.
	Yes	story topics group: See figure 2.	combination group: See figure 4.

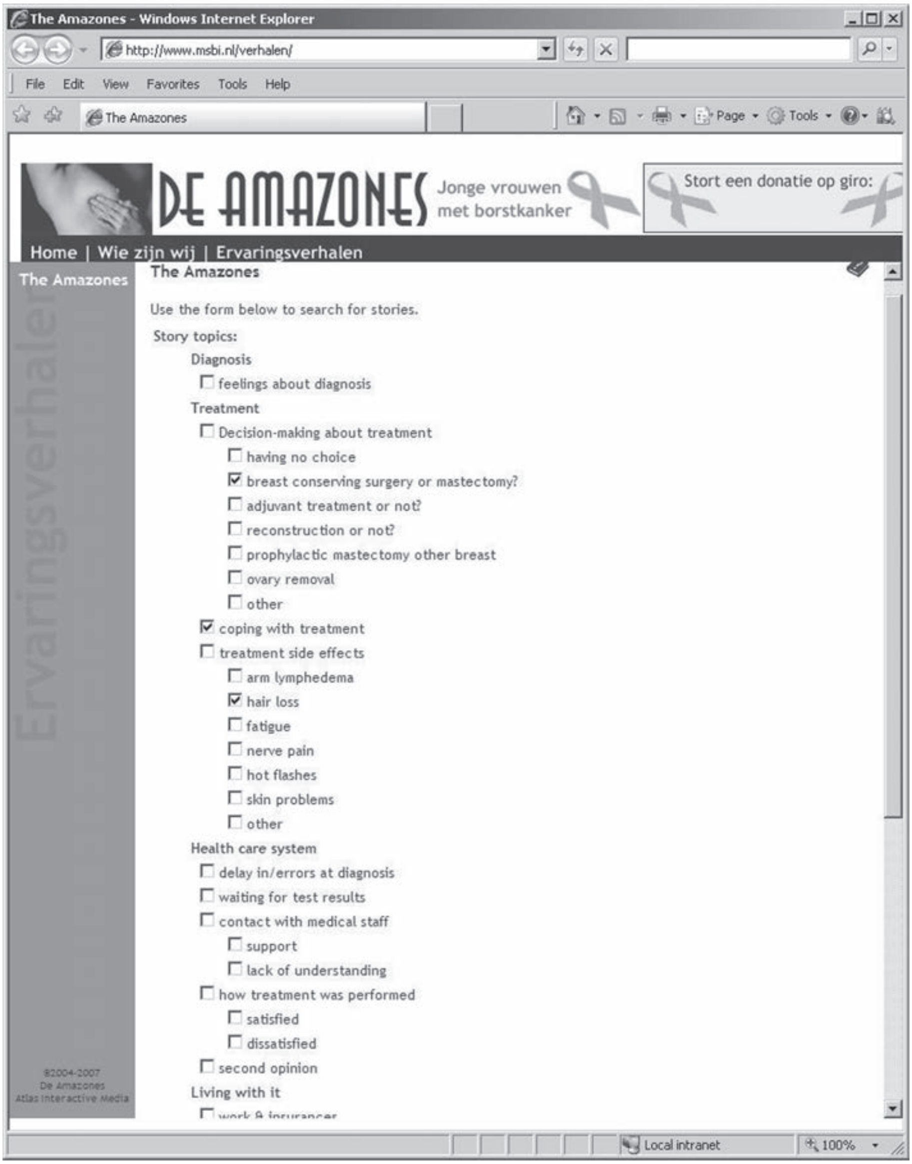
Recruitment Process

Recruitment announcements were disseminated online using banners on the websites of several Dutch patient and health organizations and offline using posters and flyers in waiting rooms of several hospitals. Dutch-speaking women with breast cancer were invited to participate irrespective of other personal characteristics. The offline recruitment announcements gave the URL of the study website, and the online announcements contained a hyperlink to the study website. The study was accessible to each visitor of the site, but only visitors who met the inclusion criteria were further directed to the informed consent page. After finishing the final questionnaire, participants could send a ready-made email message with the URL of the study website to other women who might be interested in participation.

The Amazones

The folders below contain the stories sorted alphabetically by writer name.

Figure 1 Screenshot of the control group search page.



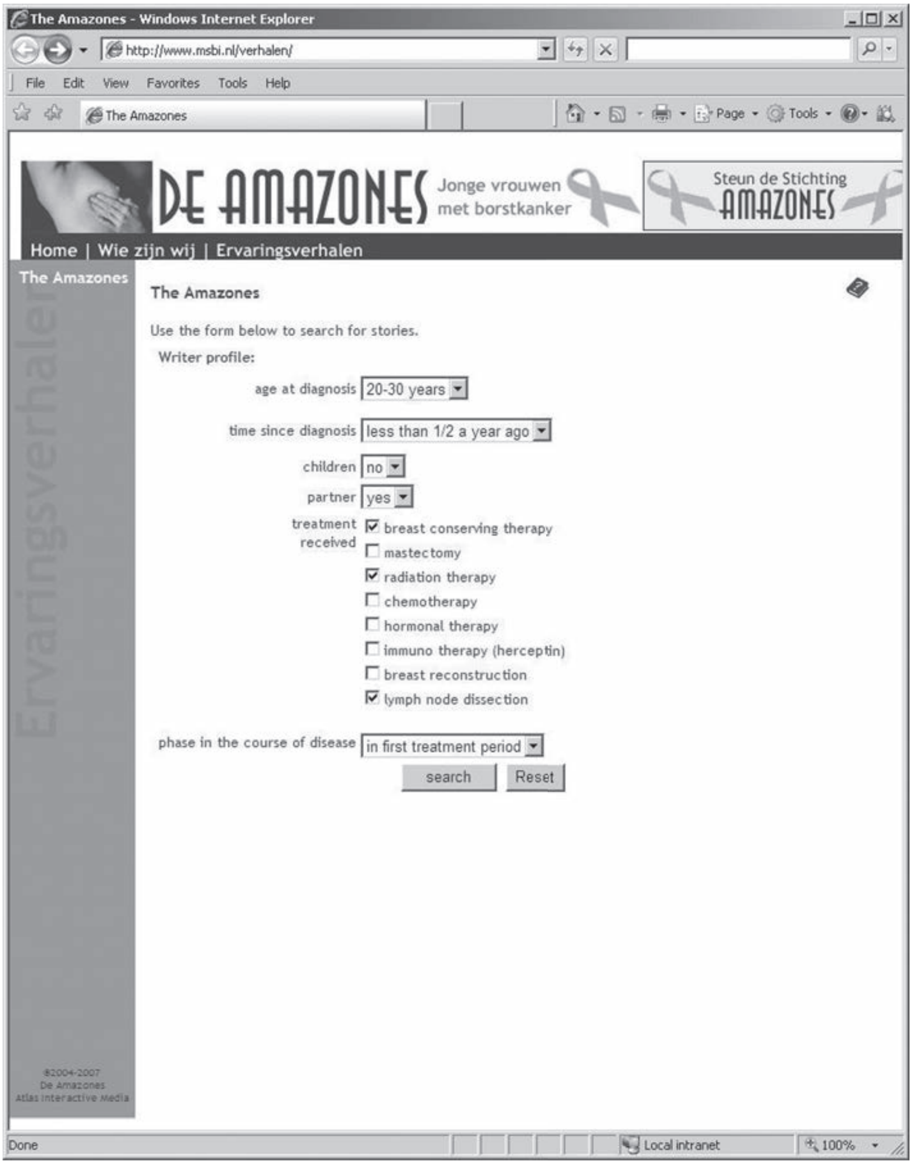
The Amazones

Use the form below to search for stories.

Story topics:

- Diagnosis
  - ☐ feelings about diagnosis
- Treatment
  - ☐ Decision-making about treatment
    - ☐ having no choice
    - ☑ breast conserving surgery or mastectomy?
    - ☐ adjuvant treatment or not?
    - ☐ reconstruction or not?
    - ☐ prophylactic mastectomy other breast
    - ☐ ovary removal
    - ☐ other
  - ☑ coping with treatment
  - ☐ treatment side effects
    - ☐ arm lymphedema
    - ☑ hair loss
    - ☐ fatigue
    - ☐ nerve pain
    - ☐ hot flashes
    - ☐ skin problems
    - ☐ other
- Health care system
  - ☐ delay in/errors at diagnosis
  - ☐ waiting for test results
  - ☐ contact with medical staff
    - ☐ support
    - ☐ lack of understanding
  - ☐ how treatment was performed
    - ☐ satisfied
    - ☐ dissatisfied
  - ☐ second opinion
- Living with it

Figure 2 Screenshot of the story topics group search page.



The Amazones

Use the form below to search for stories.

Writer profile:

age at diagnosis: 20-30 years

time since diagnosis: less than 1/2 a year ago

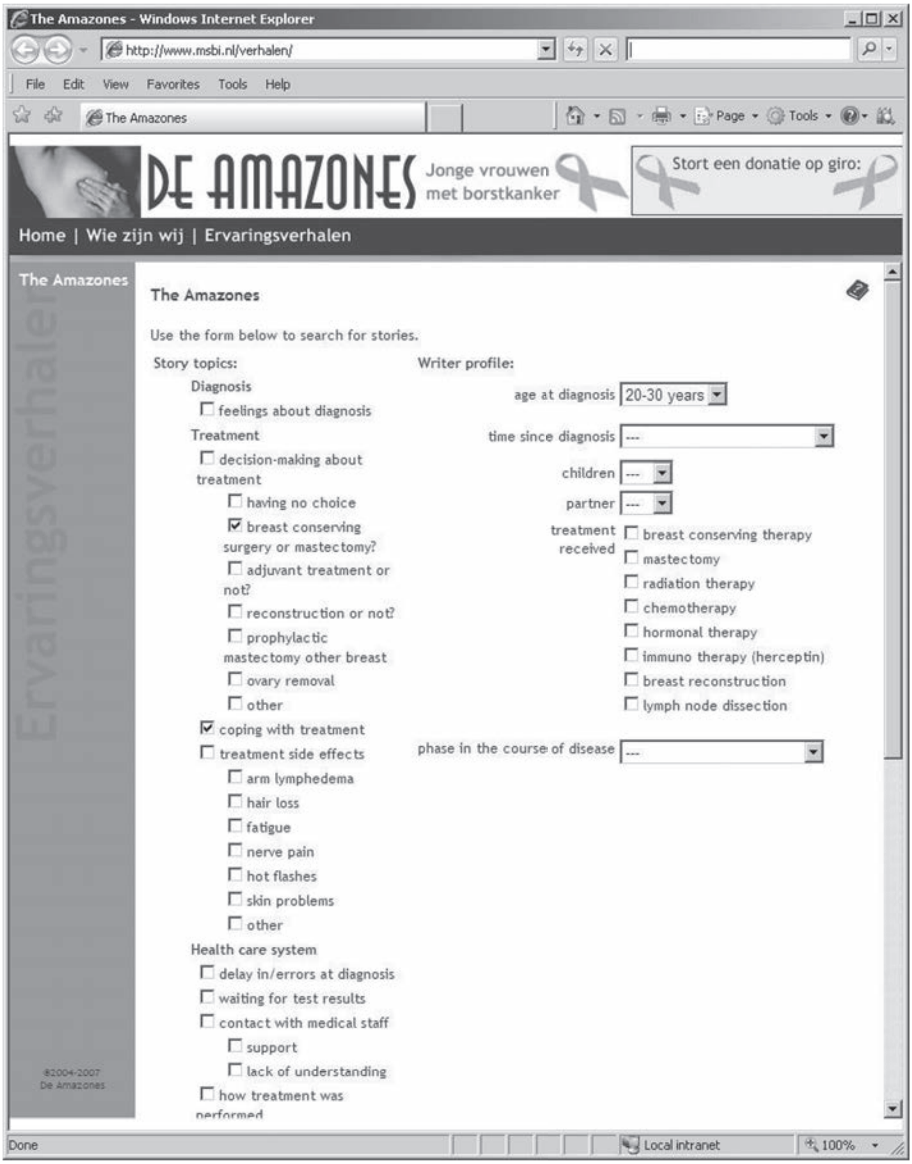
children: no

partner: yes

treatment received:
- ☑ breast conserving therapy
- ☐ mastectomy
- ☑ radiation therapy
- ☐ chemotherapy
- ☐ hormonal therapy
- ☐ immuno therapy (herceptin)
- ☐ breast reconstruction
- ☑ lymph node dissection

phase in the course of disease: in first treatment period

Figure 3 Screenshot of the writer profile group search page.



The Amazones

Use the form below to search for stories.

Story topics:

Diagnosis
- ☐ feelings about diagnosis

Treatment
- ☐ decision-making about treatment
  - ☐ having no choice
  - ☑ breast conserving surgery or mastectomy?
  - ☐ adjuvant treatment or not?
  - ☐ reconstruction or not?
  - ☐ prophylactic mastectomy other breast
  - ☐ ovary removal
  - ☐ other
- ☑ coping with treatment
- ☐ treatment side effects
  - ☐ arm lymphedema
  - ☐ hair loss
  - ☐ fatigue
  - ☐ nerve pain
  - ☐ hot flashes
  - ☐ skin problems
  - ☐ other

Health care system
- ☐ delay in/errors at diagnosis
- ☐ waiting for test results
- ☐ contact with medical staff
  - ☐ support
  - ☐ lack of understanding
- ☐ how treatment was performed

Writer profile:

age at diagnosis: 20-30 years

time since diagnosis: ---

children: ---

partner: ---

treatment received:
- ☐ breast conserving therapy
- ☐ mastectomy
- ☐ radiation therapy
- ☐ chemotherapy
- ☐ hormonal therapy
- ☐ immuno therapy (herceptin)
- ☐ breast reconstruction
- ☐ lymph node dissection

phase in the course of disease: ---

Figure 4 Screenshot of the combination group search page.

The study website was accessible in the period June through November 2007. During this time period, women could choose for themselves on what day and time they wanted to participate. No incentives were offered for participation.

Study Website

The first page of the study website provided the following information about the study: study objective, information about the researchers, study inclusion criteria, details about participation, expected time within which to complete the study, and contact details. When a website visitor chose to participate by clicking the button "I would like to participate", two questions were presented to check whether the visitor met the inclusion criteria (i.e., being female and having been diagnosed with breast cancer). If this was the case, the visitor was asked to read the informed consent statement. By agreeing with the informed consent statement, the visitor declared that her participation was voluntary, that she understood what participating entailed, and that she was aware of what data would be recorded. To agree with the informed consent statement, the visitor had to check the box "I agree" and then click on the button marked "Next". After this the visitor was asked whether she was certain that she agreed with the informed consent statement. In this way, we assured that women did not automatically agree to participate. Before the participant was randomly assigned to one of the four groups, she was asked to provide a short description of the information she wanted to search for in the stories. Random assignment of each participant to a study group was carried out by an algorithm that was part of the study website. The chance of being assigned to a group was equal for all four groups, that is, 1 in 4 (except when more than one session was conducted from the same IP address; see the section below labeled "Technical Aspects"). Once assigned to a group, a participant could search for and read the stories as long as she liked. When finished with searching and reading she was asked to complete a final questionnaire posted on the study website about satisfaction with the search process and the stories retrieved and the stories' impact on coping with breast cancer.

Ethical Aspects

Participants remained anonymous since no log-in, name, or address were required. In order to minimize traces on each client's computer, no cookies were used. Recording of log data did not start until participants had agreed to the informed consent statement. Questionnaire responses were not saved until participants confirmed at the end of the final questionnaire that they agreed to submit their responses. Data were saved in a password protected SQL database only accessible to the researchers. Participants could stop at any time without receiving pop-ups or text when leaving the study website.

Before the start of the study, our research proposal (see Multimedia Appendix 1) was presented to the Ethical Committee of the Leiden University Medical Centre (archive number 06/43). The Committee concluded that our study involved no medical intervention and that we could proceed. Our intervention consisted of providing access to the stories already available on the website of the Amazones Foundation in several new ways.

2.2 Development of Intervention Groups

Search Facilities

To develop the search facilities, all 170 stories were coded according to a coding scheme for story topics and a coding scheme for writer profile (Table 2). The topics and personal characteristics in the coding schemes were chosen because they had been used on other websites that contained breast cancer stories [11-13] or by other authors of studies in this field [1,9,10]. For the characteristic "phase in the course of disease," the category "passed away" was assigned to stories that contained an obituary.

Table 2 Coding schemes for story topics and writer profiles.

Search Facility	Coding Scheme
Story topics (domains)	
Diagnosis	Feelings about diagnosis
Treatment	Decision-making about treatment: (1) having no choice; (2) breast conserving surgery or mastectomy?; (3) adjuvant treatment or not?; (4) reconstruction or not?; (5) prophylactic mastectomy other breast; (6) ovary removal; (7) other
	Coping with treatment
	Treatment side effects: (1) arm lymphedema; (2) hair loss; (3) fatigue; (4) nerve pain; (5) hot flashes; (6) skin problems; (7) other
Health care system	Delay in/errors at diagnosis
	Waiting for test results
	Contact with medical staff: (1) support; (2) lack of understanding
	How treatment was performed: (1) satisfied; (2) dissatisfied
	Second opinion
Living with it	Work and insurances
	Family and friends: (1) support; (2) lack of understanding; (3) talking with and worrying about
	Body image and sexuality: (1) (partly) missing a breast; (2) partner's reaction
	Pregnancy issues: (1) pregnant at diagnosis; (2) wanting to become pregnant after treatments
	Coping with breast cancer: (1) thinking (emotional-focused coping); (2) doing (problem-focused coping)
	Practical advices
	Concerns about heredity
	Coping with metastasized breast cancer

Table 2 (Continued)

Search Facility	Coding Scheme
Writer profile (personal characteristics)	
Age at diagnosis	Number of years
Time since diagnosis	Number of months
Partner	(1) No, (2) yes
Children	(1) No, (2) yes
Treatment received	(1) Breast conserving therapy, (2) mastectomy, (3) radiation therapy, (4) chemotherapy, (5) hormonal therapy, (6) immuno therapy (herceptin), (7) breast reconstruction, (8) lymph node dissection
Phase in the course of disease	(1) in first treatment period, (2) free of cancer, (3) cancer for second time, (4) metastasized cancer, (5) passed away

Participants could search for age using the categories: 20-30 years, 30-40 years, 40-50 years, and over 50 years. Participants could search for time since diagnosis using the categories: less than half a year ago, ½ – 1 year ago, 1-3 years ago, 3-5 years ago, and more than 5 years ago. To ensure that participants were aware of all search facilities, the search button was placed at the bottom of the page. In the groups with access to a single search facility, it was possible to search for more than one topic or more than one writer characteristic. In the combination group, participants could choose whether they wanted to search for story topics only, for writer characteristics only, or for both. Searching for more items simultaneously was based on the OR Boolean operator.

Weight Assignment in Story Retrieval

For every search performed by the participants, a weight between 0 and 1 was assigned to each of the 170 stories in the database. If a story matched exactly with the search objectives, it received a weight of 1. Story weights were calculated with every new search. Therefore, the weight assigned to a story could change with every search.

In the story topics group, weights were calculated by dividing the number of topics found in a story by the number of topics that were searched for. For example, when a participant searched for four topics, all stories containing one of these four topics received a weight of ¼ (0.25).

In the writer profile group, a weight was assigned to each of the personal characteristics that a participant searched for. These weights were then multiplied with each other to calculate the weight of a story as a whole.

If the age of a writer fell in the age category that the participant was searching for, then “age” received a weight equal to 1. The more the age of the writer deviated from the age category that the participant was searching for, the lower the weight that “age” received. In a similar way, weights for “time since diagnosis” were assigned.

If the partner status of a writer exactly matched the partner status that the participant was searching for, then “partner” received a weight equal to 1. If the partner status of a writer is unknown, then this characteristic received a weight of 0.5 irrespective of the partner status that the participant was searching for. If the partner status of a writer was the opposite of the partner status that a participant was searching for, then this characteristic received a weight of 0.2. We did not assign a weight of 0 for the latter case because then the weight for the whole story would be 0. In similar ways, weights for the other categorical variables were assigned.

In the combination group, weight assignment was similar to that of the previous two groups or a multiplication of these two, depending on whether a participant searched for story topics only, for writer profile only, or for both.

Number of Stories Retrieved

It was decided to present participants with at least ten stories after each search. The total number of stories presented after a search depended on the distribution of the weights assigned to the stories. All stories with the same weight as the tenth story were presented because we saw no valid reason for presenting only a portion of the stories with that weight. For example, when five stories matched exactly with the search objectives, that is, weight equal to 1, 20 stories received a weight of 0.80, 45 stories received a weight of 0.60, and 100 stories received a weight of 0.40, then 25 stories (5 + 20) would be presented. Accordingly, if no stories exactly matched the search objectives, still at least ten stories were presented. The list of retrieved stories showed the extent to which the stories matched the search objectives. In this way we tried to present participants with neither too few nor too many stories. When participants did not fill in the search page, no stories were presented to them because we wanted to ensure that participants were aware of the search facility.

Story Display and Sequence

The retrieved stories were displayed as a list giving for each story the writer's nickname and the story's weight. Weights were represented as a number of pink ribbons. In addition, the search criteria that were fulfilled were given in each group, that is, the topics found, the writer's characteristics, or both (see Multimedia Appendix 2). The list of stories was sorted by weight with the story with the heighest weight at the top. Stories with the same weight in the story topics group were displayed as follows. For each story, the percentage of text of the story relating to the topics the participant searched for was calculated. Stories with the highest percentage were ranked first. In the writer profile group, stories with the same weight were sorted by the age of the writers, and if age of writers was equal by time since diagnosis. Clicking on a story title from the list displayed the complete story.

2.3 Final Questionnaire

Demographic and Disease Characteristics

Participants were asked to provide demographic information such as age, marital status, children, religion, education, and employment status. They were also asked to report characteristics of their cancer, such as time since diagnosis, type of diagnosis, metastases in axillary lymph nodes or other parts of the body, treatment undergone, and prognosis.

Use of the Internet and the Amazones Website (Before Study Participation)

The participants were asked to indicate their frequency of Internet use, the type of activities in which they engage on the Internet, and whether they had read stories of other patients on the Internet before. Moreover, they were asked to report how often they had visited the Amazones website before, how familiar they were with this website, and how many of the stories on this website they read before.

Satisfaction With the Search Process, the Stories Retrieved, and the Stories' Impact on Coping With Breast Cancer

The constructs listed below were used to measure the three outcomes. Cronbach alphas were calculated using SPSS version 16.0 (SPSS Inc, Chicago, IL, USA) by conducting reliability analyses. Reverse phrased items were recoded. We found the internal consistency for each construct to be good or satisfactory (Cronbach alpha = 0.71 to 0.88). The items "overall satisfaction with the search facility" and "overall satisfaction with the stories retrieved" were answered using 10-point Likert scales; all other items were answered using 5-point Likert scales. For an overview of the items belonging to all the below mentioned constructs, see the Multimedia Appendix 3.

To measure satisfaction with the search process, 13 items were formulated (partially based on [16,17]). "Opinion about the search facility" was measured using 5 items (Cronbach alpha = .88). To measure "the extent to which the search options enabled finding information one was looking for," 4 items were formulated (Cronbach alpha = .75). "Recommendation to others and future own use" was measured with 2 items (Cronbach alpha = .82). "Opinion about the number of search options" and "overall satisfaction with the search facility" were measured with 1 item each.

To measure satisfaction with the stories retrieved, 18 items were formulated (partially based on [16,17]). "Opinion about the stories retrieved" was measured with 6 items (Cronbach alpha = .71). To measure "opinion about the list of stories displayed after a search" 4 items were formulated (Cronbach alpha = .76). "The extent to which the stories retrieved covered one's information need" was measured with 4 items (Cronbach alpha = .82) and "recommendation to others and future own reading" with 2 items (Cronbach alpha = .77). "Opinion about the number of stories retrieved" and "overall satisfaction with the stories retrieved" were measured with 1 item each.

"The stories' impact on coping with breast cancer" was measured with 6 items (Cronbach alpha = .85) which were based on an extensive literature on coping [18-21]. Two of the items were formulated to measure problem-focused coping ("By reading the stories I have learnt things" and "By reading the stories I know what to do"), another two items were formulated to measure emotion-focused coping ("By reading the stories I am more able to understand my feelings" and "By reading the stories I can see that certain emotions are part of learning to live with breast cancer"), one item was formulated to measure reappraisal ("By reading the stories I view things in a different way") and one item was formulated to measure social comparison ("By reading the stories I see that others have experienced the same things"). The existing validated coping scales, such as the Ways of Coping checklist [18,19] and the COPE inventory [21], could not be used because they were too general for our research question.

2.4 Technical Aspects

We tested the usability and technical functionality of the study website, including the final questionnaire, multiple times, and we solved all appearing errors. During participants' search processes, log data recorded how long participants surfed on the study website, how many searches they performed, how many stories they accessed, and how long the text of the stories was displayed on the screen. In the control group, clicking on a folder (A to Z) was regarded as performing a search, and subsequently clicking on a name was seen as accessing a story. Also, the time participants needed to fill in the final questionnaire was recorded.

Participants who were searching for or reading the stories were reminded to fill in the final questionnaire by a yellow figure on the left side of the screen with the text "Do not forget to complete the questionnaire," which was highlighted every five minutes. Adaptive questioning was used to reduce the number and complexity of the questions. Questions were not randomized or alternated. The final questionnaire was distributed over five pages in the following sequence:(1) the search process, (2) the stories retrieved, including the stories' impact on coping with breast cancer, (3) use of the Internet and the Amazones website, (4) disease characteristics, and (5) demographic characteristics. When participants clicked on the "Next" button at the end of a page, JavaScript was used to check for completeness. Unanswered questions were highlighted, and participants were asked to answer these. Yet, answering was not enforced, since by clicking on the "Next" button again, the next page was reached. Participants were not able to review and change their answers in previous parts to prevent a possible influence of questions asked later in the questionnaire.

Log data and questionnaire responses were saved automatically in an SQL database. In preparation for data analysis, sessions from the same IP address with a time interval of less than 20 minutes were merged, and those with a time interval of greater than 20 minutes were kept as two separate sessions. We assumed that in the former case the sessions were from the same participant and, in the latter, from different participants. Applying the first rule resulted in 23 merged sessions; the latter rule was applied to 6 pairs of sessions. Merging was possible because in all cases the questionnaire was filled out only once. Participants were only distinguished by IP address. A particular IP address was always assigned to the same intervention group. This was done to prevent women from participating multiple times when trying to get in another study group.

2.5 Data Analysis

The data were imported into SPSS version 16.0. Differences in the log data between questionnaire completers and noncompleters were assessed using Mann-Whitney tests. The noncompleters were excluded from further analyses, since no questionnaire responses for this group were available. For the completers, there was no time frame for filling in the questionnaire. Differences between the four groups in baseline characteristics were assessed using Chi square tests, 1-way ANOVA, or Kruskal-Wallis tests (depending on variable type and skewness).

Kruskal-Wallis tests were performed to assess differences between the four groups in search behaviour (ie, the log data). Significant differences were examined further by performing post hoc tests. We chose to use Mann-Whitney tests with a Bonferroni correction, and as as the critical level of significance we used .05/6 = .008 because with four groups six comparisons were performed.

For each construct of the three outcome measures (satisfaction with the search process, the stories retrieved, and the stories' impact on coping with breast cancer) a mean total score was calculated. A higher mean indicated a higher satisfaction or impact respectively. The effects of the search facilities on the constructs of the three outcome measures were examined using ANOVA with two independent factors (search facility for story topics yes/no; search facility for writer profiles yes/no) to assess possible main and interaction effects. This analytical approach was chosen in order to examine the effects of the two search facilities both independently and in combination. P values above .05 were considered not significant.

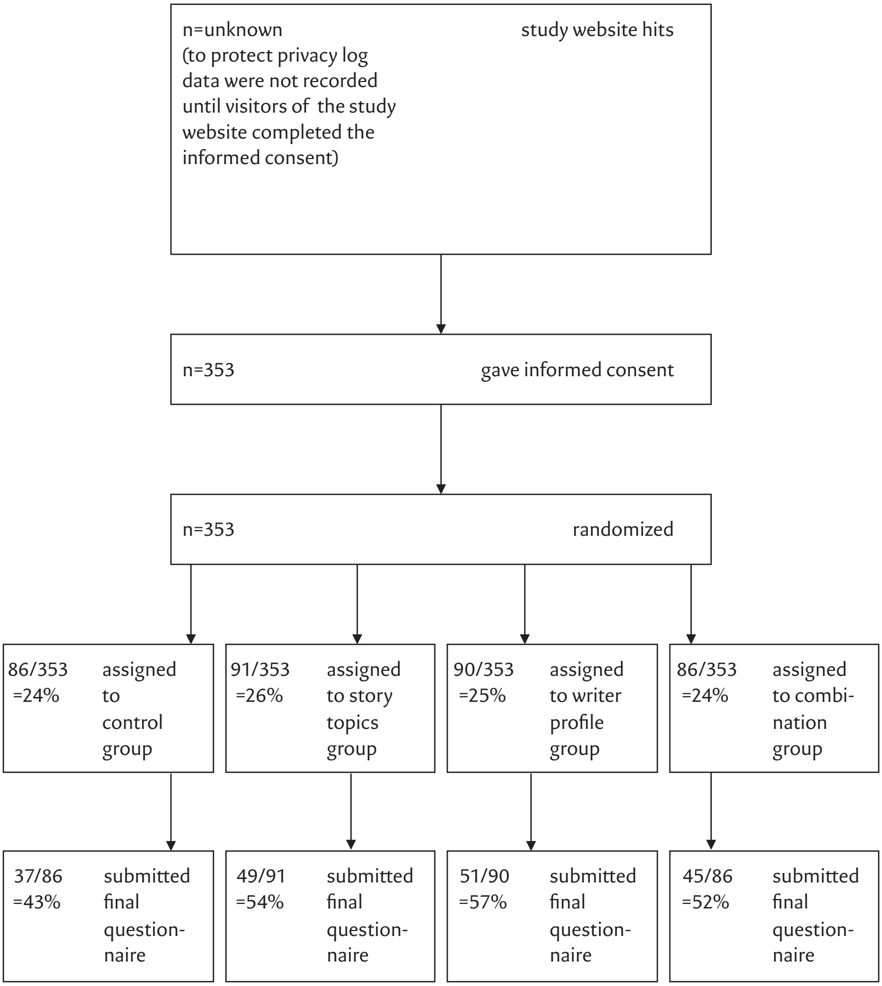


Figure 5 Flow of study participants.

3 Results

3.1 Participant Statistics

Informed consent was given by 353 people, of whom 182 (51.6%) completed the final questionnaire (Figure 5). No significant difference was found between the four groups in percentage questionnaire completers (χ^2_3= 3.7, P = .30). The mean time that participants needed to fill in the final questionnaire was 15.3 minutes (SD = 12.7; min = 5.0, max = 138.4). In comparison with questionnaire noncompleters, questionnaire completers spent less time visiting the study website (mean = 809.1 seconds vs 928.0 seconds, P < .001), but completers performed more searches (mean = 2.1 vs 1.7, P < .001), accessed more stories (mean = 6.6 vs 3.4, P < .001), and their mean reading time per story was longer (mean = 92.8 seconds vs 63.5 seconds, P < .001).

Table 3 shows the baseline characteristics of the questionnaire completers. No significant differences between the four groups were found for demographic and disease characteristics and use of the Amazones websites. With respect to use of the Internet, the writer profile group was less familiar with accessing fellow patients' stories on the Internet.

3.2 Search Behaviour

Table 4 shows that there were no differences between the four groups in time spent on the study website or in the number of searches performed. However, we found differences between the four groups in the number of stories accessed and in the mean reading time per participant per story. Post hoc tests (with a critical level of significance of P = .008 due to Bonferroni correction) showed that compared with the control group, fewer stories tended to be accessed in both the writer profile group (P = .01) and the combination group (P = .02). In addition, in the control group, the mean reading time per participant per story was shorter than in the writer profile group (P = .007) and tended to be shorter compared with the story topics group (P = .02) and the combination group (P = .009).

3.3 Satisfaction With the Search Process

Table 5 shows that having access to the story topics search facility resulted in a more positive opinion about the search facility (1a), in a more positive opinion about the number of search options (1b), in being better enabled to find the information one was looking for (1c), in being more inclined to recommend it to others or to use it more often themselves in future (1d), and in a higher overall satisfaction with the search facility (1e), compared with not having access to this search facility (all comparisons were significant at P < .05). Having access to the writer profile search facility compared with not having access to this search facility resulted in a significantly more positive opinion about the search facility (1a), and in a significantly higher overall satisfaction score (1e).

An interaction effect was found for the overall satisfaction score (1e). When participants could search using the story topics, they were satisfied with this search facility regardless of whether (mean = 7.3, SD = 1.5) or not (mean = 7.2, SD = 1.4) they could also search with the writer profile. The effect of having access to the writer profile search facility when also having access to the story topics search facility was not significant (P = .90). However, when participants could not use the story topics to search the stories, they were more satisfied with having access to the writer profile as a search facility (mean = 6.8, SD = 1.6) compared with not having access to any search facility (mean = 5.7, SD = 2.3). The effect on satisfaction of having access to the writer profile search facility when not having access to the story topics search facility was significant (P = .009).

Table 3 Baseline characteristics of the four groups.

			Control group (n=37)[a]		Story topics group (n=49)[a]		Writer profile group (n=51)[a]		Combination group (n=45)[a]		P-value[b]
			n	%	n	%	n	%	n	%	
Demographic characteristics	**Age in years (mean, SD)**		49.1 (7.5)		49.9 (8.3)		49.5 (9.4)		45.9 (9.9)		.12[c]
	Married or living together	**Yes**	29	78.4	38	77.6	40	78.4	34	75.6	.99
		No	8	21.6	11	22.4	11	21.6	11	24.4	
	Children	**Yes**	29	78.4	40	81.6	41	80.4	31	68.9	.45
		No	8	21.6	9	18.4	10	19.6	14	31.1	
	Religious	**Yes**	16	43.2	19	38.8	19	37.3	12	26.7	.44
		No	21	56.8	30	61.2	32	62.7	33	73.7	
	Higher professional education or university degree	**Yes**	11	29.7	18	36.7	23	45.1	15	33.3	.47
		No	26	70.3	31	63.3	28	54.9	30	66.7	
	Employed	**Yes**	22	59.5	24	49.0	31	60.8	26	57.8	.64
		No	15	40.5	25	51.0	20	39.2	19	42.2	

Table 3 (Continued)

			Control group (n=37)[a]		Story topics group (n=49)[a]		Writer profile group (n=51)[a]		Combination group (n=45)[a]		P-value[b]
			n	%	n	%	n	%	n	%	
Disease characteristics	**Time since diagnosis in months (mean, SD)**		36.8 (45.8)		34.9 (41.4)		42.0 (41.8)		34.0 (37.3)		.49[d]
	Diagnosed with one tumour	**Yes**	23	62.2	35	72.9	37	72.5	32	71.1	.69
		No	14	37.8	13	27.1	14	27.5	13	28.9	
	Size of tumour	**<2 cm**	9	25.0	15	32.6	18	36.7	20	44.4	.32
		≥2 cm	27	75.0	31	67.4	31	63.3	25	55.6	
	Cancer in axillary lymph nodes at diagnosis	**Yes**	15	41.7	27	57.4	24	49.0	23	51.1	.56
		No	21	58.3	20	42.6	25	51.0	22	48.9	
	Metastases to other parts of the body	**Yes**	6	16.7	4	8.5	5	10.2	6	13.3	.68
		No	30	83.3	43	91.5	44	89.8	39	86.7	
	Breast conserving surgery	**Yes**	13	35.1	21	42.9	18	35.3	20	44.4	.71
		No	24	64.9	28	57.1	33	64.7	25	55.6	
	Mastectomy	**Yes**	19	51.4	24	49.0	34	66.7	30	66.7	.15
		No	18	48.6	25	51.0	17	33.3	15	33.3	
	Radiation therapy	**Yes**	22	59.5	28	57.1	25	49.0	20	44.4	.47
		No	15	40.5	21	42.9	26	51.0	25	55.6	
	Chemotherapy	**Yes**	24	64.9	35	71.4	31	60.8	26	57.8	.54
		No	13	35.1	14	28.6	20	39.2	19	42.2	
	Hormonal therapy	**Yes**	21	56.8	23	46.9	23	45.1	18	40.0	.50
		No	16	43.2	26	53.1	28	54.9	27	60.0	
	Cancer free	**Yes**	26	70.3	30	61.2	32	62.7	28	62.2	.83
		No	11	29.7	19	38.8	19	37.3	17	37.8	

Table 3 (Continued)

			Control group (n=37)[a]		Story topics group (n=49)[a]		Writer profile group (n=51)[a]		Combina-tion group (n=45)[a]		P-value[b]
			n	%	n	%	n	%	n	%	
Use of the Internet and the Amazones website	**Daily Internet use**	**Yes**	31	83.8	43	87.8	47	92.2	35	77.8	.23
		No	6	16.2	6	12.2	4	7.8	10	22.2	
	Familiar with searching online for specific information	**Yes**	36	97.3	46	93.9	48	94.1	43	95.6	.88
		No	1	2.7	3	6.1	3	5.9	2	4.4	
	Familiar with accessing fellow patients' stories on the Internet	**Yes**	30	81.1	44	89.8	33	64.7	40	88.9	.005
		No	7	18.9	5	10.2	18	35.3	5	11.1	
	Visited the Amazones website at least once a month before participation	**Yes**	11	29.7	18	36.7	15	29.4	19	42.2	.52
		No	26	70.3	31	63.3	36	70.6	26	57.8	
	'Rather well' or 'well' familiar with Amazones website	**Yes**	7	38.9	18	52.9	15	50.0	13	38.2	.56[e]
		No	11	61.1	16	47.1	15	50.0	21	61.8	
	Read half or more of the Amazones stories before	**Yes**	5	27.8	15	44.1	13	43.3	12	35.3	.62[e]
		No	13	72.2	19	55.9	17	56.7	22	64.7	

[a] N (%) is shown unless noted otherwise.

[b] *P*-values are for chi-square tests comparing the four groups unless noted otherwise.

[c] *P*-value for 1-way ANOVA test to compare the four groups with respect to age.

[d] *P*-value for Kruskal-Wallis test to compare the four groups with respect to time since diagnosis.

[e] Percentages and tests based on the number of participants who had previously visited the Amazones website: control group (n=18), story topics group (n=34), writer profile group (n=30), combination group (n=34).

Table 4 Comparison of the four groups for the search behaviour measures recorded by the log data.

		Control group (n=37)	Story topics group (n=49)	Writer profile group (n=51)	Combination group (n=45)	P-value[a]
Time spent on the study website in seconds	**mean (SD)**	754.00 (966.33)	984.55 (1278.94)	595.39 (630.04)	905.49 (1054.71)	.45
	median	496.00	634.00	389.00	636.00	
Number of searches	**mean (SD)**	3.89 (4.71)	1.88 (2.32)	1.53 (1.59)	1.69 (1.58)	.07
	median	2.50	1.23	1.24	1.28	
Number of stories accessed	**mean (SD)**	13.19 (18.98)	6.73 (6.36)	4.18 (4.53)	3.93 (3.61)	.01
	median	5.67	5.56	3.00	3.29	
Reading time per participant per story in seconds	**mean (SD)**	49.24 (54.65)	94.16 (94.90)	99.72 (128.27)	119.26 (113.00)	.02
	median	28.11	71.33	67.00	89.00	

[a] *P*-value for Kruskal-Wallis tests comparing the four groups with respect to the four search behaviour measures.

3.4 Satisfaction With the Stories Retrieved

Having access to the story topics search facility resulted in a more positive opinion about the list of stories displayed after a search (2c), a greater extent to which the stories retrieved covered one's information need (2d), and a higher overall satisfaction score with the stories retrieved (2f) compared with not having access to this search facility (Table 5). Having access to the writer profile search facility compared with not having access to this search facility resulted in a more positive opinion about the list of stories displayed after a search (2c).

There were no interaction effects observed in satisfaction with the stories retrieved.

3.5 The Stories' Impact on Coping With Breast Cancer

Table 5 shows that the stories retrieved using the story topics search facility had a greater impact on coping with breast cancer (3a). When analysing each of the six coping items individually, we observed that having access to the story topics search facility resulted in a significantly higher score for having learned things (3a.1).

4 Discussion

4.1 Principal Findings

To our knowledge, this study is the first randomized controlled experiment with a 2x2 factorial design that examined search facilities for accessing online patient stories. We observed that the story topics search factor had a strong impact on patient satisfaction and search success: participants were the most satisfied with this search facility and the stories retrieved. Also, the stories retrieved had a greater impact on coping with breast cancer. The effect of the writer profile search factor was limited. This search facility resulted only in a few effects, predominantly on satisfaction with the search process. The two search factors combined generally had no amplified effect on patient satisfaction or search success as we only had one significant interaction.

Table 5 Means (SD) of the constructs of the three outcome measures asked for in the final questionnaire by search factor.

		Story topics[a]			Writer profile[a]			Interaction[a]
		Yes (n=94)	No (n=88)	P-value[b]	Yes (n=96)	No (n=86)	P-value[c]	P-value[d]
1. Satisfaction with the search process	**a. opinion about the search facility (range 1-5)**	4.0 (0.7)	3.6 (1.1)	.001	3.9 (0.9)	3.6 (1.0)	.005	.21
	b. opinion about the number of search options (range 1-3)[e]	2.3 (0.6)	2.1 (0.8)	.04	2.3 (0.7)	2.2 (0.8)	.29	.23
	c. the extent to which search options enable finding information one was looking for (range 1-5)	3.3 (1.0)	2.8 (1.0)	.001	3.1 (1.0)	3.0 (1.0)	.27	.59
	d. recommendation to others and future own use (range 1-5)	4.1 (1.0)	3.5 (1.2)	< .001	3.9 (1.1)	3.8 (1.2)	.29	.13
	e. overall satisfaction with the search facility (range 1-10)	7.3 (1.4)	6.3 (2.0)	< .001	7.1 (1.6)	6.5 (2.0)	.01	.03

Table 5 (Continued)

		Story topics[a]			Writer profile[a]			Interaction[a]
		Yes (n=94)	No (n=88)	P-value[b]	Yes (n=96)	No (n=86)	P-value[c]	P-value[d]
2. Satisfaction with (the information in) the stories retrieved	**a. opinion about the stories retrieved (range 1-5)**	3.5 (0.6)	3.4 (0.7)	.54	3.5 (0.6)	3.4 (0.7)	.36	.18
	b. opinion about the number of stories retrieved (range 1-3)[e]	2.3 (0.7)	2.1 (0.7)	.27	2.1 (0.7)	2.3 (0.7)	.18	.17
	c. opinion about the list of stories displayed after a search (range 1-5)	3.6 (0.9)	3.2 (1.2)	.001	3.8 (0.9)	2.9 (1.1)	< .001	.06
	d. the extent to which the stories retrieved covered one's information need (range 1-5)	3.0 (1.0)	2.6 (1.0)	.02	2.7 (1.1)	2.9 (1.0)	.56	.91
	e. recommendation to others and future own reading (range 1-5)	4.1 (1.0)	3.8 (1.0)	.08	4.0 (1.0)	4.0 (1.0)	.67	.71
	f. overall satisfaction with the stories retrieved (range 1-10)	7.1 (1.5)	6.4 (1.7)	.002	6.7 (1.7)	6.9 (1.6)	.80	.35

Table 5 (Continued)

		Story topics[a]			Writer profile[a]			Interaction[a]
		Yes (n=94)	No (n=88)	P-value[b]	Yes (n=96)	No (n=86)	P-value[c]	P-value[d]
3. The stories' impact on coping with breast cancer	**a. the stories' impact on coping with breast cancer (range 1-5)**	3.2 (0.9)	2.9 (1.0)	.02	3.0 (1.0)	3.1 (1.0)	.53	.71
	a.1. By reading the stories I have learnt things (range 1-5)	3.0 (1.4)	2.5 (1.3)	.007	2.6 (1.4)	2.9 (1.3)	.24	.53
	a.2. By reading the stories I know what to do (range 1-5)	2.7 (1.2)	2.4 (1.2)	.14	2.5 (1.3)	2.6 (1.1)	.71	.70
	a.3. By reading the stories I am more able to understand my feelings (range 1-5)	3.0 (1.3)	2.6 (1.4)	.07	2.7 (1.4)	2.9 (1.3)	.70	.37
	a.4. By reading the stories I can see that certain emotions are part of learning to live with breast cancer (range 1-5)	3.8 (1.2)	3.5 (1.5)	.17	3.7 (1.4)	3.6 (1.3)	.82	.76
	a.5. By reading the stories I view things in a different way (range 1-5)	2.7 (1.3)	2.4 (1.3)	.16	2.4 (1.3)	2.7 (1.3)	.12	.16
	a.6. By reading the stories I see that others have experienced the same things (range 1-5)	4.1 (1.0)	3.9 (1.3)	.16	4.1 (1.1)	3.9 (1.1)	.28	.31

[a] ANOVA with two independent factors (search facility for story topics yes/no; search facility for writer profiles yes/no). Higher means indicate better outcomes.

[b] *P*-value for possible main effect of story topics search

[c] *P*-value for possible main effect of writer profile search

[d] *P*-value for possible interaction effect between story topics search and writer profile search

[e] Asked on a 5-point scale, but for analysis recoded into 3-points (see also Multimedia Appendix 3)

These findings are contrary to our expectation, which was that the combined search facilities (the interaction) would outperform a single search facility because this combination is more complete and differentiated resulting in greater opportunities to find a relevant story. Apparently, this quantity argument seems to be less important than the type of the search facility (quality). In line with our expectation was that a single search facility was an improvement compared with the alphabetically listed stories in the control group.

Participants in the three search facility groups accessed fewer stories and read longer per accessed story compared with the control group. An explanation for this might be that the stories retrieved in the search facility groups were more relevant to the participants. A search facility probably increases the proportion of the documents retrieved relevant to the user's information need [22].

The story topics search facility resulted not only in participants being more satisfied with the search process, but also in participants retrieving stories that better covered their information needs and retrieving stories from which they learned more. Patients might use online stories predominantly for information, and, therefore, the topics described in the stories might be more important for them than the writer's profile. Patients' profiles might be more important when seeking face-to-face contact. This difference between seeking information and seeking contact has also been noted by Bennenbroek et al [23] in their research on social comparison.

Our observation that the writer profile search facility compared with not having this search facility resulted in a more positive opinion about the search facility and in a higher overall satisfaction with the search facility is in line with the results of Rozmovits and Ziebland [1]. They found in interviews that patients positively evaluated the ability to select other patients of a particular age, stage of illness, or patients who were long-term survivors or who had opted for similar treatment. However, although our study also showed that participants were more satisfied with this search facility, they were not more satisfied with the stories retrieved using this facility.

4.2 Limitations of the Present Study

A considerable number of participants performed searches but did not complete the questionnaire. Compared with completers, non-completers spent more time on the study website while they performed fewer searches, accessed fewer stories, and spent less time reading per story. Non-completers might not have been sure about how to use the search facilities, or they might not have been as interested. Yet, we could not empirically evaluate these hypotheses nor perform any statistical analyses, since we had no further information about non-completers. The number of completers and non-completers was evenly distributed over the four study groups. Therefore, we believe that potential bias equally affected all four groups. In addition, the direction of the bias is probably twofold: dissatisfied participants might have stopped or they might have completed the questionnaire to express their annoyance.

More than half of the participants who completed the questionnaire (63.7%) had previously visited the Amazones website. This could introduce bias because participants familiar with the original disclosure of stories might be especially satisfied with the new search facilities. However, frequency of visiting the Amazones site, knowing the site “rather well” or “well,” and the number of Amazones stories read before, were all evenly distributed over the four study groups. Therefore, we do not think this previous experience with the Amazones website affected the results.

Since the experiment was conducted completely online [24], we cannot verify that all participants indeed had (or had had) breast cancer. However, we targeted this group for recruitment and asked relevant questions before randomization. We assume that all participants were sincere, because overall they spent 15 minutes filling in the final questionnaire. This suggests that participants were interested in the subject matter.

A limitation of the design was a possible confounding between type of search facility (story topics, writer profile) and number of search options (17 topics, 6 personal characteristics). The story topics search might have been more appreciated because it was more extensive than the writer profile search. However, an argument against this reasoning is that the most extensive search facility (i.e., story topics in combination with writer profile) was not the most favourite.

In addition, one could question the content of the search facilities. Were the most appropriate topics and personal characteristics included in the facilities? Yet, the topics and characteristics we used were chosen based on other websites containing breast cancer stories [11-13] and other studies in this field [1,9,10].

A final limitation is that participants may have been annoyed when stories were presented that did not exactly match their search objectives. However, a search resulting in no stories could also be a cause of annoyance. This is why we chose to present at least ten stories after each search. In order to ease interpretation of the resulting list of stories, weights (as pictured in the form of pink ribbons) were used to indicate to what extent a story matched with the search objective.

4.3 Conclusions and Practical Implications

Earlier studies have shown that patients can benefit from stories of other patients, and that the Internet is an important source of these stories. Our current study suggests that a story topics search facility would be most helpful to patients. With a story topics search facility, participants were better enabled to find the information they were looking for. Also, they retrieved stories that more closely covered their information needs and they learned more from the stories retrieved.

Thus, patient organisations or website developers that offer patient stories on their websites can best provide access to them using a story topics search facility. However, constructing such a search facility is very time consuming and labour intensive since stories have to be coded for content. An efficient method might be to use a system analogous to social bookmarking/tagging [25] in which story readers assign keywords or tags to the stories, and the keywords or tags that are most often assigned are seen as most important in describing the content. Another possibility is to construct a list of items from which writers can compose descriptions of their stories. Finally, stories could also be classified by automatic full text indexing or clustering. This will be the subject of our next study.

Acknowledgments

The authors wish to thank the participants for their time and interest. We also wish to thank the organizations that disseminated the call for participation on their websites or in their waiting rooms.

The paper is part of the Narrator project addressing the problem of storage and accurate retrieval of illness narratives. This project is conducted under the umbrella program ToKeN2000 and is funded by the Netherlands Organization for Scientific Research (NWO).

Conflicts of Interest

None declared

Multimedia Appendices

1. Research proposal (in Dutch) as presented to the ethical committee before the start of the study
2. Screenshots illustrating the search page and the list of retrieved stories for each of the 4 search conditions (i.e., control, story topics, writer profile, and combination condition)
3. An overview of the constructs and items belonging to the three main outcome measures

References

1. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004 Apr;53(1):57-64.
2. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564.
3. McTavish FM, Gustafson DH, Owens BH, Hawkins RP, Pingree S, Wise M, et al. CHESS (Comprehensive Health Enhancement Support System): an interactive computer system for women with breast cancer piloted with an underserved population. J Ambul Care Manage 1995 Jul;18(3):35-41.
4. Herxheimer A, McPherson A, Miller R, Shepperd S, Yaphe J, Ziebland S. Database of patients' experiences (DIPEx): a multi-media approach to sharing experiences and information. Lancet 2000 Apr 29;355(9214):1540-1543.
5. Pitts V. Illness and Internet empowerment: writing and reading breast cancer in cyberspace. Health (London) 2004 Jan;8(1):33-59.
6. Hardey M. 'The story of my illness': personal accounts of illness on the Internet. Health (London) 2002;6:31-46.
7. McLellan MF. The Electronic Narrative of Illness [dissertation]. Galveston: University of Texas Medical Branch; 1997.
8. Wise M, Han JY, Shaw B, McTavish F, Gustafson DH. Effects of using online narrative and didactic information on healthcare participation for breast cancer patients. Patient Educ Couns 2008 Mar;70(3):348-356.
9. Overberg RI, Alpay LL, Verhoef J, Zwetsloot-Schonk JHM. Illness stories on the internet: what do breast cancer patients want at the end of treatment? Psychooncology 2007 Oct;16(10):937-944.
10. Williamson K. Where one size does not fit all: understanding the needs of potential users of a portal to breast cancer knowledge online. J Health Commun 2005 Sep;10(6):567-580.
11. Healthtalkonline (formerly DIPEx). Health Experiences Research Group URL: http://www.healthtalkonline.org/Cancer/Breast_Cancer [WebCite Cache]
12. BreastCancerStories.com. URL: http://www.BreastCancerStories.com [WebCite Cache]
13. KWF Kankerbestrijding (Dutch Cancer Society). URL: http://www.kankerbestrijding.nl [WebCite Cache]
14. De Amazones, jonge vrouwen met borstkanker (The Amazones, young women with breast cancer). URL: http://www.de-amazones.nl [WebCite Cache]
15. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34.
16. Spink A. A user-centered approach to evaluating human interaction with Web search engines: an exploratory study. Info Process Manag 2002;38(3):401-426.
17. Zeng QT, Kogan S, Plovnick RM, Crowell J, Lacroix EM, Greenes RA. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. Int J Med Inform 2004 Feb;73(1):45-55.
18. Folkman S, Lazarus RS. An analysis of coping in a middle-aged community sample. J Health Soc Behav 1980 Sep;21(3):219-239.
19. Folkman S, Lazarus RS. If it changes it must be a process: study of emotion and coping during three stages of a college examination. J Pers Soc Psychol 1985 Jan;48(1):150-170.
20. McCrae RR, Costa PT. Personality, coping, and coping effectiveness in an adult sample. Journal of Personality 1986;54:385-405.

21. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989 Feb;56(2):267-283.
22. Van Rijsbergen CJ. Information Retrieval. 2nd edition. London, United Kingdom: Butterworths; 1979.
23. Bennenbroek FTC, Buunk BP, van der Zee KI, Grol B. Social comparison and patient information: what do cancer patients want? Patient Educ Couns 2002 May;47(1):5-12.
24. Nosek BA, Banaji MR, Greenwald AG. E-Research: Ethics, security, design, and control in psychological research on the Internet. J Soc Issues 2002;58:161-176.
25. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007;24:2-23.

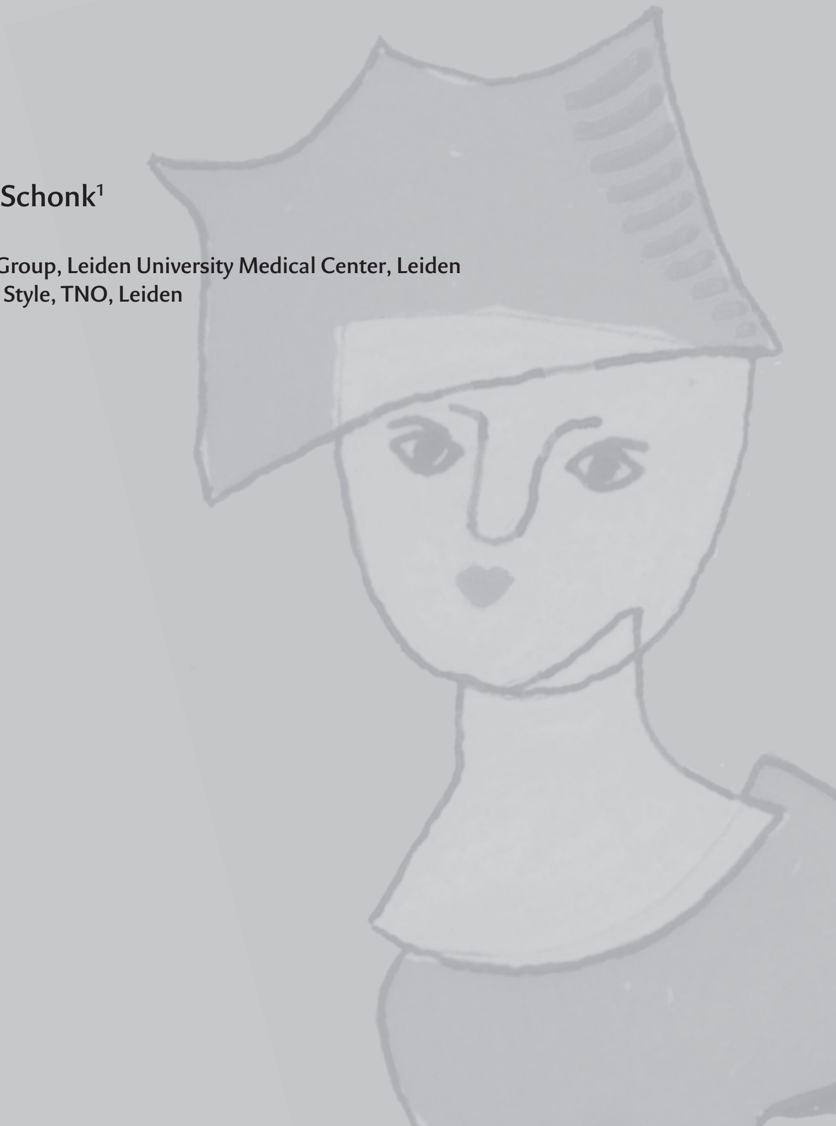
CHAPTER 6

Searching for breast cancer stories online: What topics and/or writer characteristics do patients search for?

Regina Overberg[1]
Wilma Otten[2]
Andries de Man[1]
Bertie Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden
[2] Expertise Centre Life Style, TNO, Leiden

Submitted



Abstract

Online patient stories can provide other patients with support and information. This study explores breast cancer patients' information need from stories and use of social comparison at the level of writer characteristics. Log data of 122 patients who participated in an online randomized controlled experiment on search facilities for stories were analysed. Many participants searched for topics in the domains 'Treatment' and 'Living with it', and for the writer characteristics 'treatment received', 'age at diagnosis' and 'time since diagnosis'. Per search patients used more writer characteristics than topics. They preferred to search for others with similar characteristics as their own.

1 Introduction

Breast cancer has a major impact on physical, psychological and social functioning. On the Internet there are many stories of breast cancer patients in which they describe what happened to them -from the suspicion of cancer, the diagnosis, the treatments and onwards- and how they cope with cancer (e.g. [1,2]). These stories can provide other patients with support and information [3]. In the context of online patient stories it has been studied how giving and receiving social support is experienced [4] and what the effect is of online patient stories on health care participation [5] and making health care decisions [6].

Yet, to date, little is known about the specific information breast cancer patients actually search for in stories of other patients. In an earlier study we interviewed breast cancer patients and asked them what they wanted with respect to stories of others [7]. These interviews showed that they preferred to know a writer's age and treatment undergone and to read how other patients coped with emotions and physical discomforts and how having cancer impacted their daily life. To our knowledge, no other studies focused on the specific information need from stories. Yet, there are studies on general information need about breast cancer showing that younger patients are interested in information about physical and sexual attractiveness [8], breast conserving therapy and metastases [9], and recently diagnosed patients in topics such as likelihood of cure, the spread/stage of the disease and treatment options [10-12]. In one of these studies, patients actually selected information from a CD-ROM [9], the other studies used surveys to examine information need. Yet, patients' general information need may differ from their information need from stories, since the information in stories is from fellow patients who describe their real life experiences in everyday language.

In reading stories of other patients social comparison plays a role [13]: when someone is uncertain (especially when there is no objective information) about how to think, feel or behave, then thoughts and behaviours of other persons who are in a similar situation can be used as a source of information. Comparison with someone who is as close as possible to oneselves on several characteristics (age, treatment received and so on) is assumed to provide information with the most impact (parallel comparison). Bellizzi et al. [14] found, for example, that in autobiographical books cancer survivors used predominantly these parallel comparisons. However, patients can also choose to compare themselves with others who are worse off (downward comparison) in order to feel good about themselves, or with others who are doing better (upward comparison) to engender hope and optimism [15]. As far as we know, it is not known how patients who search online for stories of others apply social comparison at the level of writer characteristics: do they search for writers with similar characteristics as their own? Research from the group behind the healthtalkonline website (formerly Dipex) [3] showed that patients viewed it positively to have access to stories of others who had opted for similar treatment as they had opted. Yet, our interviews revealed that patients wanted a search facility for time since diagnosis to be able to select stories of long-term survivors that can serve as encouraging examples [7].

In an earlier study –an online randomized controlled experiment- we examined breast cancer patients' satisfaction and search success when using a story topics and/or a writer characteristic search facility to search for other patients' stories [16]. The experiment yielded log data of the actual searches of patients. The present article presents the results of an extensive analysis of these log data. Besides our interest in the information need from stories and in social comparison at the level of writer characteristics as described above, we were interested in the specificity of searches (i.e. the number of topics and the number of writer characteristics that is combined per search). To our knowledge no earlier studies were performed about specificity of searches in the domain of breast cancer stories. In the present article we have addressed the following three research questions:

1. Which topics and/or writer characteristics do breast cancer patients search for when they have a search facility for story topics and/or writer characteristics? (information need)
2. Do breast cancer patients search for writer characteristics that are similar to their own? (social comparison)
3. What is the specificity of the searches breast cancer patients perform (i.e. how many topics and how many writer characteristics do patients combine per search)?

2 Methods

2.1 Empirical material

We used log data of an earlier performed online randomised controlled experiment that examined breast cancer patients' satisfaction with the search process and the stories retrieved of four different search facilities for other patients' stories [16]. In this earlier experiment, Dutch-speaking women who were diagnosed with breast cancer and who gave informed consent were randomly assigned to one of four search facilities: 1) alphabetical listing by story writers nicknames (control group), 2) a search facility for story topics, 3) a search facility for writer characteristics, and 4) a combined search facility for story topics and writer characteristics. In the three latter groups participants could check the items they wanted to search for and then click the search button. By clicking on the search button, stories were retrieved and presented to the participants. Participants could search for stories of *De Amazones*, a group of young women with breast cancer who founded a website with stories and information in order to provide their peers with information and support [17]. After the participants searched for stories as long as they liked, they filled out a questionnaire including questions about demographic and disease characteristics. Log data (for example, the number of searches participants performed and the items they searched for in their searches) and questionnaire responses were automatically saved in a database. See [16] for more details. In the present study we analysed the log data of the participants who used the story topics and/or writer characteristics search facility. The group with the alphabetical list of stories is not considered in the present study.

2.2 Search facilities

Figure 1 shows a screenshot of the search page of the combined search facility. The search page of the story topics search facility consisted of only the left side of this screenshot and the search page of the writer characteristics search facility consisted of only the right side.

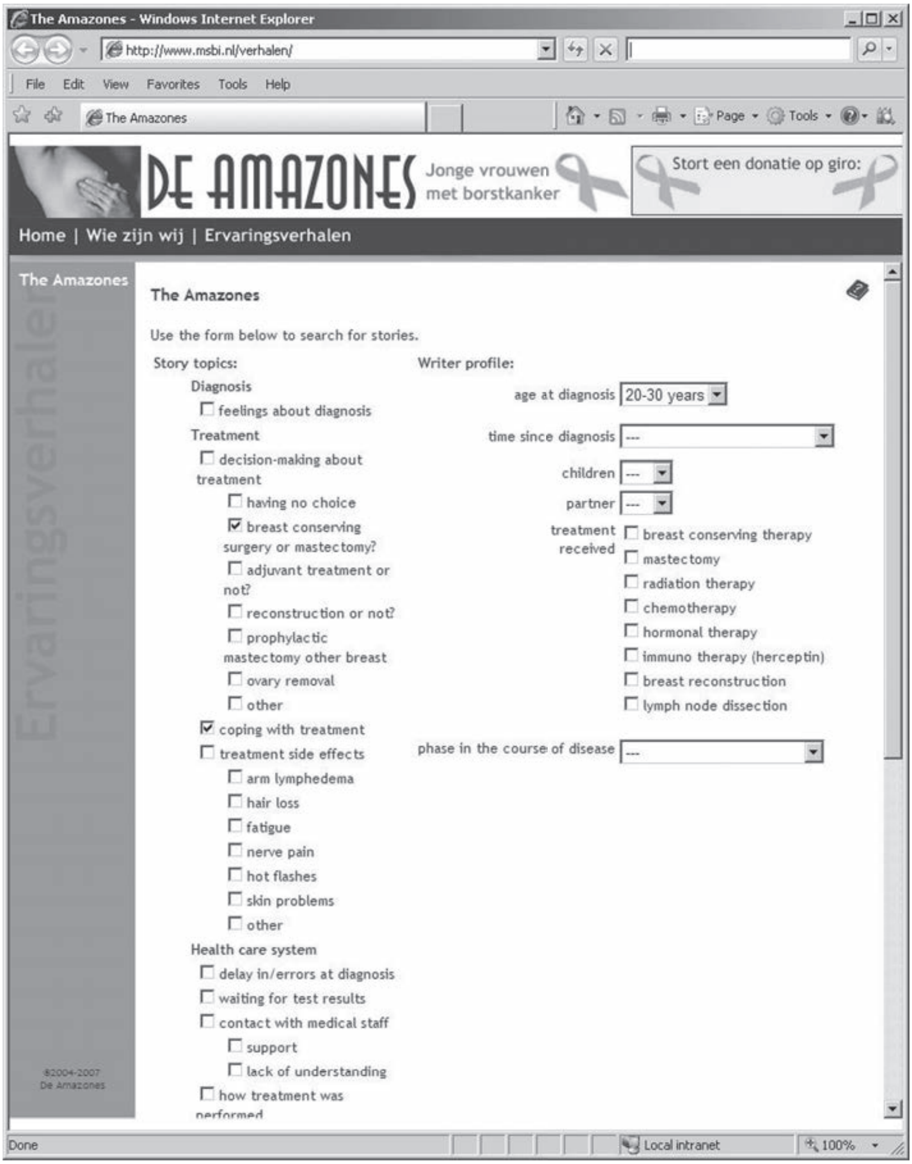
With the story topics search facility one could search for topics in four domains: I) Diagnosis, II) Treatment, III) Health care system, and IV) Living with it. The four domains contained a total of 17 main category topics, of which 9 were not divided in subcategory topics and 8 were divided in subcategory topics. In total, there were 27 subcategory topics (Table 1). In the remainder of this article we will use the general term 'topic' to refer to: a) a main category topic without subcategory topics, b) a main category topic with subcategory topics, or c) a subcategory topic. For each of the topics the search page contained a tick box (Figure 1). One could search for stories by checking as many tick boxes as preferred. When the tick box of a main category topic with subcategory topics was checked, all the subcategory topics were automatically included in the search.

With the writer characteristics search facility one could search for six writer characteristics: 1) age at diagnosis, 2) time since diagnosis, 3) children, 4) partner, 5) treatment received, and 6) phase in the course of disease (Table 2). 'Treatment received' was divided in eight treatments, each with a tick box (Figure 1). To search for a treatment the tick box of the treatment could be checked. The other five writer characteristics could be searched for by drop down menus: from each drop down menu one category could be selected in a search (Figure 1). For 'age at diagnosis' the drop down menu consisted of four categories: 20-30 years, 30-40 years, 40-50 years, and over 50 years. The drop down menu of 'time since diagnosis' consisted of five categories: less than half a year ago, ½-1 year ago, 1-2 years ago, 3-5 years ago, and more than 5 years ago. Both the drop down menu of 'children' and the drop down menu of 'partner' consisted of two categories: yes and no. For 'phase in the course of disease' the drop down menu consisted of five categories: in first treatment period, free of cancer, cancer for second time, metastasised cancer, and passed away (stories of deceased writers remained on the website with an obituary added by the moderators). All dropdown menus had a "neutral" option, which was the default setting. When a participant did not change the default setting, the concerning writer characteristic was not searched for.

With the combined search facility one could search for story topics only, for writer characteristics only, or for a combination of these two.

2.3 Data analysis

The statistical data analysis was performed with SPSS version 17.0. Log data of participants who completed the final questionnaire and who performed at least one search were analysed. These were the log data of 122 participants: 40 participants in the story topics search group, 43 participants in the writer characteristics search group and 39 participants in the combination search group. Descriptive analyses were performed to describe the participants' demographic and disease characteristics and their number of searches performed. A search was defined as clicking the search button after one had checked the tick boxes and/or selected the categories from the drop down menus of the items one wanted to search for.



The Amazones

Use the form below to search for stories.

Story topics:

Diagnosis

☐ feelings about diagnosis

Treatment

☐ decision-making about treatment
  ☐ having no choice
  ☑ breast conserving surgery or mastectomy?
  ☐ adjuvant treatment or not?
  ☐ reconstruction or not?
  ☐ prophylactic mastectomy other breast
  ☐ ovary removal
  ☐ other

☑ coping with treatment

☐ treatment side effects
  ☐ arm lymphedema
  ☐ hair loss
  ☐ fatigue
  ☐ nerve pain
  ☐ hot flashes
  ☐ skin problems
  ☐ other

Health care system

☐ delay in/errors at diagnosis

☐ waiting for test results

☐ contact with medical staff
  ☐ support
  ☐ lack of understanding

☐ how treatment was performed

Writer profile:

age at diagnosis: 20-30 years

time since diagnosis: ---

children: ---

partner: ---

treatment received:
☐ breast conserving therapy
☐ mastectomy
☐ radiation therapy
☐ chemotherapy
☐ hormonal therapy
☐ immuno therapy (herceptin)
☐ breast reconstruction
☐ lymph node dissection

phase in the course of disease: ---

Figure 1 Screenshot of the search page of the combination group.

Table 1 Number of participants (n, %) who searched for the topic in at least one of their searches of those who could use the story topics search facility (n=79). When a participant searched for several subcategories of a main category, this participant is counted once in the main category.

Domain	Topic[a]	n (%)
I. Diagnosis	**1. Feelings about diagnosis**	**17 (22%)**
II. Treatment	**1. Decision-making about treatments**	**42 (53%)**
	a. having no choice	8 (10%)
	b. breast conserving surgery or mastectomy	12 (15%)
	c. adjuvant treatment or not	10 (13%)
	d. reconstruction or not	18 (23%)
	e. prophylactic mastectomy of other breast	10 (13%)
	f. ovary removal	15 (19%)
	g. other	10 (13%)
	2. Coping with treatment	**18 (23%)**
	3. Treatment side effects	**47 (60%)**
	a. arm lymphedema	17 (22%)
	b. hair loss	15 (19%)
	c. fatigue	27 (34%)
	d. nerve pain	18 (23%)
	e. hot flashes	24 (30%)
	f. skin problems	12 (15%)
	g. other	14 (18%)
III. Health care system	**1. Delay in/errors at diagnosis**	**9 (11%)**
	2. Waiting for test results	**8 (10%)**
	3. Contact with medical staff	**9 (11%)**
	a. support	9 (11%)
	b. lack of understanding	7 (9%)
	4. How treatment was performed	**8 (10%)**
	a. satisfied	8 (10%)
	b. dissatisfied	5 (6%)
	5. Second opinion	**3 (4%)**

Table 1 (Continued)

Domain	Topic	n (%)
IV. Living with it	**1. Work and insurances**	**12 (15%)**
	2. Family and friends	**24 (30%)**
	a. support	13 (17%)
	b. lack of understanding	16 (20%)
	c. talking with and worrying about	19 (24%)
	3. Body image and sexuality	**23 (29%)**
	a. (partly) missing a breast	20 (25%)
	b. partner's reaction	14 (18%)
	4. Pregnancy issues	**4 (5%)**
	a. pregnant at diagnosis	2 (3%)
	b. wanting to become pregnant after treatments	3 (4%)
	5. Coping with breast cancer	**36 (46%)**
	a. thinking (emotional-focused coping)	35 (44%)
	b. doing (problem-focused coping)	29 (37%)
	6. Practical advice	**19 (24%)**
	7. Concerns about heredity	**13 (17%)**
	8. Coping with metastasized breast cancer	**27 (34%)**

[a] Topics in bold are main categories; topics in non-bold subcategories.

To answer our first research question (patients' information need from stories), we analysed how many of the 79 (=40+39) participants who had the story topics search facility at their disposal searched for the different topics in at least one of their searches. When a participant searched for several subcategory topics of a main category topic, this participant is counted once in the main category topic. Furthermore, we analysed how many of the 82 (=43+39) participants who had the writer characteristics search facility at their disposal searched for the different writer characteristics in at least one of their searches. When a participant searched for several of the eight treatments, this participant is counted once in the overarching category 'treatment received'.

Table 2 Number of participants (n, %) who searched for the writer characteristic in at least one of their searches of those who could use the writer characteristics search facility (n=82). When a participant searched for several of the eight treatments, this participant is counted once in the overarching category 'treatment received'.

Writer characteristic	n (%)
Age at diagnosis	**59 (72%)**
Time since diagnosis	**57 (70%)**
Children	**54 (66%)**
Partner	**55 (67%)**
Treatment received	**70 (85%)**[a]
Breast conserving therapy	23 (28%)
Mastectomy	45 (55%)
Radiation therapy	38 (46%)
Chemotherapy	42 (51%)
Hormonal therapy	35 (43%)
Immunotherapy (herceptin)	12 (15%)
Breast reconstruction	26 (32%)
Lymph node dissection	30 (37%)
Phase in the course of disease	**55 (67%)**

[a] The overarching category 'treatment received' could not be searched for directly; the percentage is calculated based on the checked tick boxes of the eight underlying treatments.

To test whether the associations found in previous studies between younger patients and general information need [8,9] and recently diagnosed patients and general information need [10-12] also hold for information need from stories, we examined associations between participants' time since diagnosis and searching for the topics 'decision-making about treatments', 'treatment side effects' and 'coping with breast cancer', and associations between participants' age at study participation and searching for the topic 'body image and sexuality' and the writer characteristics 'breast conserving therapy' and 'metastasised cancer' (category of 'phase in the course of disease'). Due to the relatively small numbers and some skewness of the variables time since diagnosis and age at study participation, the associations were tested with Mann-Whitney U-tests. P-values lower than or equal to .05 were considered significant.

To answer the second research question (social comparison at the level of writer characteristics), we analysed for the searches containing a particular writer characteristic how often the category that was searched for corresponded with the category of the participant. For example, all searches with 'children: yes' and 'children: no' were examined. The own category was searched for when a participant had children herself and searched for 'children: yes' or when a participant had no children and searched for 'children: no'. Percentages were calculated on the basis of the number of searches: participants who had multiple searches on 'children' counted multiple times. In searches in which the own ‘time since diagnosis’ was not searched for, we examined which category was searched for instead in order to get insight in whether patients search for long-time survivors as encouraging examples, as resulted from our earlier interviews [7].

To answer the third research question (specificity of searches), we analysed for each search containing topics how many topics were combined. We used the definition that one could search for 36 topics per search: 9 main category topics without subcategories and 27 subcategory topics. Thus, for the 8 main category topics with subcategories we counted the number of subcategories that was checked. In the case the main category was directly checked by the participant all subcategories were automatically checked and included in our count. In a similar way, we analysed for each search containing writer characteristics how many writer characteristics were combined. One could search for 13 writer characteristics per search : 8 tick boxes (treatments) and 5 drop-down menus.

3 Results

3.1 Participant characteristics

Table 3 presents participants’ demographic and disease characteristics. The mean age of the participants was almost fifty years: the youngest participant was 25 years of age, the eldest 70 years of age. The mean time since diagnosis was over three years, but there was much variation in time since diagnosis. Over three quarters of the participants lived with partner and children. Two thirds of the participants had undergone chemotherapy and more than half had undergone mastectomy, lymph node dissection and radiation therapy. Most participants were cancer free at the time of the study.

The mean number of searches participants performed was 2.02 (SD=1.86). The number of searches per participant ranged from 1 to 11 searches. Of the 122 participants 75 participants performed one search, 22 participants two searches and 25 participants performed more than two searches.

Table 3 Participants' demographic and disease characteristics (n=122).

Characteristic		(mean (SD); min-max)
Age at study in years		48.5 (9.0); 25-70
Age at diagnosis in years [a]		45.1 (9.1); 24-68
Time since diagnosis in months [a]		39.1 (41.8); 0-229
Characteristic	Category	n (%)
Family situation	With partner	95 (78%)
	With children	95 (78%)
Treatment received	Breast conserving surgery	53 (43%)
	Mastectomy	71 (58%)
	Lymph node dissection	68 (56%)
	Radiation therapy	68 (56%)
	Chemotherapy	80 (66%)
	Hormonal therapy	57 (47%)
	Immunotherapy (herceptin)	12 (10%)
	Breast reconstruction	23 (19%)
Phase in the course of disease	Just before or in first treatment period	17 (14%)
	Just after first treatment period	12 (10%)
	Breast cancer for second time	11 (9%)
	Metastases to other parts of the body	15 (12%)
	Cancer free	67 (55%)

[a] Unknown n=2

3.2 Patients' information need from stories

Table 1 shows for the participants who used the story topics search facility (n=79) which topics they searched for in at least one of their searches. The two topics most participants searched for were from the domain 'Treatment': 'treatment side effects' (60%) and 'decision-making about treatments' (53%). The topic 'coping with breast cancer' from the domain 'Living with it' was searched for by almost half of the participants (46%) in their search session. Topics from the domain 'Health care system' were least often used by participants in their search session.

Table 2 shows for the participants who used the writer characteristics search facility (n=82) which writer characteristics they searched for in at least one of their searches. The majority of participants (n=70; 85%) searched in their search session for one or more of the eight treatments. After 'treatment received' the writer characteristic 'age at diagnosis' was searched for in most of the search sessions (n=59; 72%). Each writer characteristic was searched for by at least two thirds of the participants in their search session.

Participants who searched for the topic 'treatment side effects' were more recently diagnosed than participants who did not search for this topic (mean time since diagnosis 26.7 versus 54.5 months; P=0.012). Participants' time since diagnosis was not associated with whether or not searching for the topic 'decision-making about treatments' (mean time since diagnosis 30.6 versus 45.8 months; P=0.330) and whether or not searching for the topic 'coping with breast cancer' (mean time since diagnosis 30.0 versus 46.6 months; P=0.092). Participants' age at study participation was not associated with whether or not searching for the topic 'body image and sexuality' (mean age at study 49.1 versus 48.1 years; P=0.553), nor was it associated with whether or not searching for the writer characteristics 'breast conserving therapy' (mean age at study 45.4 versus 47.9; P=0.519) and 'metastasised cancer' (mean age at study 50.4 versus 46.6; P=0.168).

3.3 Social comparison at the level of writer characteristics

Table 4 shows for the searches in which a particular writer characteristic was searched for how often the category that was searched for corresponded with the category of the participant. For almost all writer characteristics was found that in the majority of the searches the participant's own category was searched for.

In 100% of the searches for 'breast conserving therapy' the participant herself had undergone breast conserving therapy. Also the other treatments scored high in searching for the own category: the percentages ranged from 71-96% (table 4). An exception was 'breast reconstruction': in 34% of the searches in which breast reconstruction was searched for the participant herself had undergone breast reconstruction.

Furthermore, in the majority of searches the own children status (97%), the own partner status (96%) and the own age at diagnosis (79%) was searched for. In a smaller percentage of searches the own phase in the course of disease (62%) and the own time since diagnosis (65%; 64 of 99 searches) was searched for. Of the 35 searches in which the own time since diagnosis was not searched for, 20 searches were for writers with a shorter time since diagnosis and 15 searches for writers with a longer time since diagnosis. In 22 of the 35 searches a category of 'time since diagnosis' that was adjacent to the own category was searched for.

Table 4 Number (%) of searches for writer characteristics that corresponded with participants' own category.

Writer characteristic		Number of searches (n)	Searches that correspond with own category of participant (n, %)
Age at diagnosis		108	85 (79%)
Time since diagnosis		99	64 (65%)
Children		95	92 (97%)
Partner		96	92 (96%)
Treatment received	Breast conserving therapy	42	42 (100%)
	Mastectomy	77	74 (96%)
	Radiation therapy	74	63 (85%)
	Chemotherapy	73	65 (89%)
	Hormonal therapy	58	50 (86%)
	Immunotherapy (herceptin)	21	15 (71%)
	Breast reconstruction	38	13 (34%)
	Lymph node dissection	53	46 (87%)
Phase in the course of disease		96	59 (62%)

3.4 Specificity of searches

A total of 161 searches was performed in which one or more topics were searched for: 92 searches by the 40 participants in the story topics search facility group and 69 searches by 36 participants in the combination search facility group (three of the 39 participants in the latter group did not search for topics, only for writer characteristics). Per search a mean number of 4.22 topics was searched (SD=4.84; range 1-34). Figure 2 shows the distribution.

A total of 134 searches was performed in which one or more writer characteristics were searched for: 78 searches by 43 participants in the writer characteristics search facility group and 56 searches by 29 participants in the combination search facility group (ten of the 39 participants in the combination search facility group did not search for writer characteristics, only for topics). Per search a mean number of 6.94 writer characteristics was searched (SD=2.94; range 1-12). Figure 3 shows the distribution.

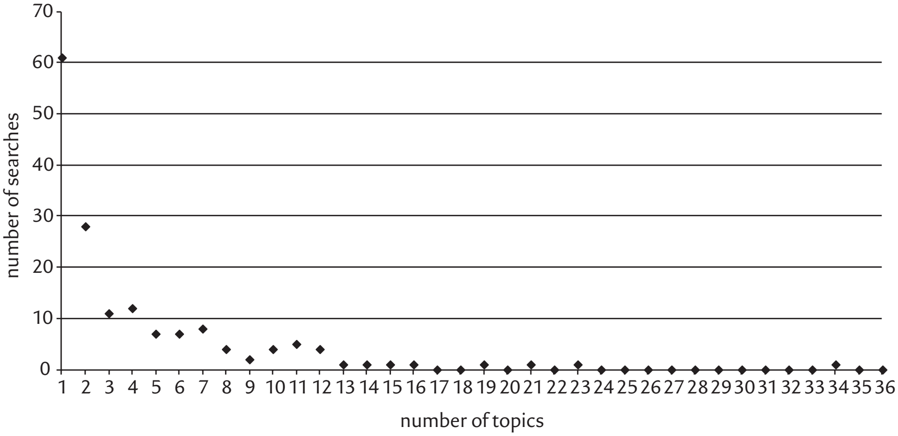


Figure 2 The number of topics that was searched for in searches with topics (n=161). A maximum number of 36 topics could be searched for (9 main category topics without subcategories + 27 subcategory topics).

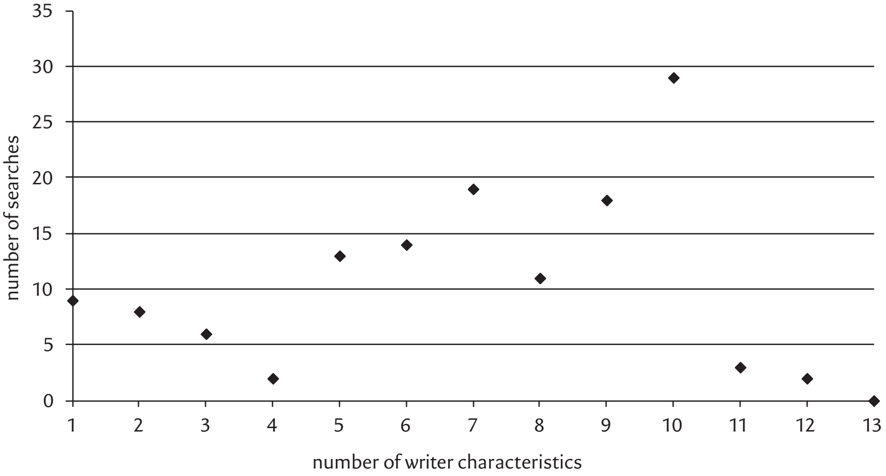


Figure 3 The number of writer characteristics that was searched for in searches with writer characteristics (n=134). A maximum number of 13 writer characteristics could be searched for (8 tick boxes (treatments) + 5 drop-down menus).

4 Discussion

Stories of other patients can provide breast cancer patients with support and information. In this study we examined what kind of stories breast cancer patients searched for when they used a search facility for story topics and/or writer characteristics.

A clear result is that participants who could use the writer characteristics search facility made full use of this search facility. Each writer characteristic was searched for by at least two thirds of the participants in their search session. Participants searched predominantly for writers with similar characteristics as they themselves have. The result that the majority of participants searched for writers who had undergone similar treatments as they themselves had supports the finding of Rozmovits & Ziebland [3] that patients appreciate it to have access to stories of others who had opted for similar treatment as they had opted. The finding from our earlier interviews [7] that patients want to search for writers who received the diagnosis longer ago than they did and therefore can serve as encouraging examples, was not supported by the results: only in a minority of searches writers with a longer time since diagnosis were searched for.

For 'breast reconstruction' we found that only in a minority of searches the own category was searched for: stories of writers who had undergone a breast reconstruction were relatively often searched for by participants who themselves did not have had a breast reconstruction. An explanation for this finding is that these participants may consider undergoing a breast reconstruction and want to read others' experiences with it. In this context it is important to notice that undergoing a breast reconstruction differs from undergoing, for example, a mastectomy. A patient can take time to decide in advance whether or not she wants to undergo a breast reconstruction, whereas in undergoing a mastectomy in most cases the patient has no choice: it happens to her and she has to cope with it.

A striking result for specificity of searches was that the number of writer characteristics that was combined in searches was higher than the number of topics that was combined. Participants searched on average for 7 writer characteristics per search (of the maximum number of 13 writer characteristics). This indicates that patients are looking for stories of writers with a clearly defined profile. It seemed less important for participants that multiple topics were simultaneously present in a story: in the majority of the searches with topics only one or two topics were searched for per search (of the maximum number of 36 topics). The type of search facility might be an explanation for the result: mainly drop down menus for writer characteristics and tick boxes for topics. Participants may have thought that they had to choose a category from each drop down menu. Another explanation for the relatively low number of topics per search might be that there is some kind of natural limit to the number of items participants check when they can choose from a lot of items.

Many participants who had the disposal of the story topics search facility searched in their session for the topics 'treatment side effects' and 'coping with breast cancer'. These two topics are broadly consistent with the top three topics resulting from our interviews with breast cancer patients about favourite topics in stories of others: coping with emotions, impact on daily life and physical discomforts [7]. Furthermore, we found that participants who searched for the topic 'treatment side effects' were significantly more recently diagnosed than participants who did not search for this topic. This result is in line with the findings of earlier studies on general information need in breast cancer patients [10-12]. However, we did not find any other associations between participants' time since diagnosis and searching for certain topics and between participants' age at study participation and searching for certain topics and writer characteristics. Yet, these associations may have been expected given the results from earlier studies on general information need [8-12]. An explanation for the result that these associations were not found in the present study on information from patient stories could be that patients prefer to receive information about certain topics from medical specialists and not from other patients. It may also be that other background variables that were not considered in the present study play a role.

4.1 Limitations

The present study was an analysis of the log data of a previous conducted experiment [16]. As a result, we were dependent on the research design of this earlier experiment.

In order to examine specificity of searches, searches from participants in the combination group were taken together with searches from participants in the story topics search group (to calculate number of topics per search) or with searches from participants in the writer characteristics search group (to calculate number of writer characteristics per search). Yet, in the combination group the number of topics searched for may be influenced by the fact that one could also search for writer characteristics, and vice versa.

The fact that eight main category topics were divided in subcategory topics may have led to some interpretation difficulties. Firstly, by checking the tick box of a main category topic with subcategory topics, all the subcategory topics were automatically included in the search and their number counted in the specificity of searches (while the participant checked only one tick box). Secondly, to determine information need participants who searched for at least one subcategory topic were counted in the main category topic. The latter was also true for the overarching category 'treatment received', with the difference that this overarching category could not be checked directly (no separate tick box).

Participants' reasons for participating in our study are unknown. In our call for participation we asked patients with interest in stories of others to participate in our study with the aim to improve the ways to search for stories on the Internet. Some patients may have participated not so much for their own information need, but rather to help us in testing the search facilities.

4.2 Conclusions

This study has shown with respect to information need from stories that there is no major discrepancy between what breast cancer patients want to read in stories of others and what they actually search for in stories when they have the disposal of a topics and/or a writer characteristics search facility. Patients use relatively more writer characteristics to search for stories, than topics. Herein they prefer to search for stories of patients with similar characteristics as they themselves have, probably because of their informational value. Given the variety of patients, stories provided on the Internet should reflect this diversity.

Acknowledgements

This paper is part of the Narrator project addressing the problem of storage and accurate retrieval of illness narratives. This project is conducted under the umbrella program ToKeN2000 and is funded by the Netherlands Organization for Scientific Research (NWO).

Conflicts of interest

The authors report no conflicts of interest.

References

1. BreastCancerStories.org (2008) Available at: http://www.BreastCancerStories.org (accessed 18 August 2008).
2. Healthtalkonline (formerly DIPEx) (2008) Available at: http://www.healthtalkonline.org (accessed 11 May 2008).
3. Rozmovits L and Ziebland S (2004) What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Education and Counseling* 53(1): 57-64.
4. Shaw BR, McTavish F, Hawkins R, et al. (2000) Experiences of women with breast cancer: exchanging social support over the CHESS computer network. *Journal of Health Communication* 5(2): 135-159.
5. Wise M, Han JY, Shaw B, et al. (2008) Effects of using online narrative and didactic information on healthcare participation for breast cancer patients. *Patient Education and Counseling* 70(3): 348-356.
6. Ziebland S and Herxheimer A (2008) How patients' experiences contribute to decision making: illustrations from DIPEx (personal experiences of health and illness). *Journal of Nursing Management* 16(4): 433-439.Bellizzi
7. Overberg RI, Alpay LL, Verhoef J, et al. (2007) Illness stories on the Internet: what do breast cancer patients want at the end of treatment? *Psycho-Oncology* 16(10): 937-944.
8. Degner LF, Kristjanson LJ, Bowman D et al. (1997) Information needs and decisional preferences in women with breast cancer. *Journal of the American Medical Association* 277(18): 1485-1492.
9. Molenaar S, Sprangers M, Oort F, et al. (2007) Exploring the black box of a decision aid: what information do patients select from an interactive Cd-Rom on treatment options in breast cancer? *Patient Education and Counseling* 65(1): 122-130.
10. Bilodeau BA and Degner LF (1996) Information needs, sources of information, and decisional roles in women with breast cancer. *Oncology Nursing Forum* 23(4): 691-696.
11. Luker KA, Beaver K, Leinster SJ, et al. (1995) The information needs of women newly diagnosed with breast cancer. *Journal of Advanced Nursing* 22(1): 134-141.
12. Rutten LJ, Arora NK, Bakos AD, et al. (2005) Information needs and sources of information among cancer patients: a systematic review of research (1980-2003) *Patient Education and Counseling* 57(3): 250-261.
13. Buunk AP and Gibbons FX (2006) Social comparison orientation: A new perspective on those who do and those who don't compare with others. In: Guimond S (ed) *Social Comparison and Social Psychology: Understanding Cognition, Intergroup Relations and Culture*. Cambridge: University Press, pp.15-32.
14. Bellizi KM, Blank TO and Oakes CE (2006) Social comparison processes in autobiographies of adult cancer survivors. *Journal of Health Psychology* 11(5): 777-786.
15. Bennenbroek F (2003) *Social comparison and coping with radiation therapy: The significance of different dimensions of comparison*. PhD Thesis, University of Groningen, The Netherlands.
16. Overberg R, Otten W, De Man A, et al. (2010) How breast cancer patients want to search for and retrieve information from stories of other patients on the Internet: an online randomized controlled experiment. *Journal of Medical Internet Research* 12(1): e7.
17. De Amazones, jonge vrouwen met borstkanker (young women with breast cancer) (2007) Available at: http://www.de-amazones.nl (accessed 2 January 2007).

CHAPTER 7

Searching for breast cancer stories on a website using verbose natural language queries: an exploratory study

Regina Overberg[1]
Wilma Otten[2]
Andries de Man[1]
Eduard Hoenkamp[3]
Bertie Zwetsloot-Schonk[1]

[1] Clinical Informatics Group, Leiden University Medical Center, Leiden
[2] Expertise Centre Life Style, TNO, Leiden
[3] Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

Abstract

Online patient stories provide other patients with support and information. Several search facilities have shown their usefulness in finding relevant stories, but also have limitations. Latent Semantic Indexing (LSI) may overcome these. This mathematical technique places stories and queries in a multidimensional space: distances between them provide information about similarity in content. No human reading is necessary to tag and categorize stories. Searching involves typing natural language queries, observing a list of retrieved stories (adapts while typing) and narrowing search results by refining queries (after accessing stories). We examined how twenty-four breast cancer patients use an LSI application to search for stories. Natural language queries ranged from complete sentences to one or two keywords. Sixteen participants refined their search queries. Most participants used the facilities that LSI offers and were quite satisfied with the search process and the dynamic list of retrieved stories. More research is needed to implement LSI in searching for patient stories.

1 Introduction

Stories of other patients can provide patients with emotional support, information, reassurance and practical advice [1]. Nowadays, the Internet has become an increasingly important source of patient stories [2-7]. Several websites offer patients an comprehensive set of stories of other patients (e.g.[8,9]). A crucial element in these sites is how people can retrieve the stories that fulfill their needs. This paper examines a search facility based on Latent Semantic Indexing (LSI), whose development does not require human reading to tag and categorize stories.

Regular search facilities include the following. The standard search facility of the browser (CTRL-F) can be used to search for specific words on a website. However, synonyms are not searched for and only the open page is searched through. Some major web search engines offer a facility that makes it possible to search within a specific website and allows users to type long queries in natural language. Yet, this facility also searches in other parts of the website than the part of interest and it is unclear which search algorithms are used and whether the whole query is used (trade secret).

When special search facilities are developed, they are usually based on concepts instead of text words. For example, articles in the PubMed database can be searched for using the MeSH terms that convey concepts [10, 11]. Developing such a search facility requires two steps: 1) establishing a list of concepts; and 2) tagging the content of a document to these concepts. Concept-based search facilities are also applied to patient stories, like on the healthtalkonline website [8]. Studies showed that patients who search for stories of other patients appreciate concept-based search facilities [1, 12]. Yet, to develop a concept-based search facility human effort is required (establishing concepts, tagging content) [13]. Other disadvantages are that users are restricted to pre-defined search categories [11] and that tagging content is subject to human interpretation.

Latent Semantic Indexing (LSI), a technique from the field of Information Retrieval (IR) [14], seems to overcome the previous mentioned shortcomings. A set of documents is placed in a multidimensional space: this space is a vector space that has the words in the documents as coordinate axes (the dimensions) and the documents as points in that space. In constructing this multidimensional space it is taken into account to what extent words and dimensions are distinctive. A word that occurs equally often in each document is not distinctive and will hardly count in constructing the space. Two or more dimensions that almost coincide will be reduced to one dimension [15]. The technique is purely mathematical: meanings of words and documents are not examined. The idea of LSI is that documents that are close together in the multidimensional space will be quite similar in content. Studies have shown that LSI divides a set of documents in the same clusters human subjects would assign when they are asked to group the documents by content [14, 16].

In the context of searching for illness stories LSI can be applied as follows [17]. A set of patient stories is placed in a multidimensional space. Subsequently, patients can search for stories by typing any text in a search box: one or more words, phrases or sentences. These natural language queries are considered mini-stories and are placed in the same multidimensional space as the set of stories. In this step, the whole natural language query is used. Stories with the shortest distance to a query are retrieved and presented to the user. While typing a search query, the location of the query in the multidimensional space continuously changes and thus also the list of stories retrieved changes. Figure 1 shows that the initial query is close to Story 1, but that the final query is close to Story 5. After accessing one or more stories, a search query can be refined without having to start a new search. Research has shown that the longer the search query, the more targeted the search results [18].

To our knowledge LSI has not been studied before in the context of searching for illness stories.

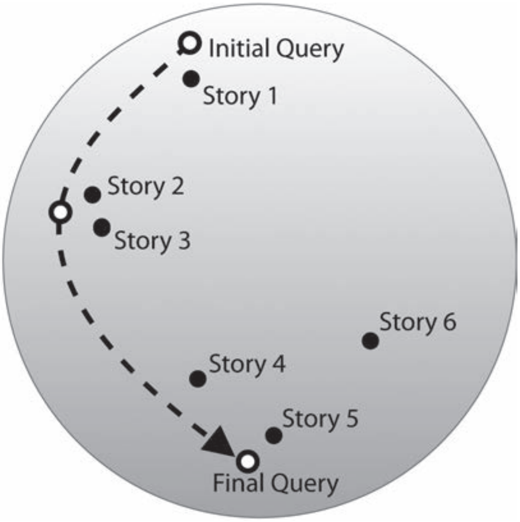


Figure 1 Example of a multidimensional space containing several stories. During the typing of the search query, it will travel through the document space [17].

In this paper we describe a study in which 24 breast cancer patients used a search facility based on LSI to search for stories of other patients. The aim of this exploratory study was to get an idea whether LSI would be suitable to disclose breast cancer stories. To this end we wanted to get insight into whether participants use the LSI search facility optimally: the search results of the LSI application are most targeted when a user types verbose natural language queries and refines search queries after accessing stories. Furthermore, we wanted examine whether satisfaction with an LSI search facility is in the same range as satisfaction with a story topics search facility or with a writer profile search facility. If breast cancer patients use the LSI search facility optimally and are quite satisfied with it, it will be more cost effective to disclose stories with LSI than to disclose stories with a story topics search facility and/or a writer profile search facility, since building an LSI search facility takes less time and less human effort than building a story topics and/or a writer profile search facility.

The research questions we addressed, were:

- How do breast cancer patients use the facilities LSI offers?
    a. Do participants refine their search queries after accessing stories in order to get more specific results?
    b. In what ways do participants use natural language in their search queries?
- How satisfied are breast cancer patients with the search process and the dynamic list of retrieved stories?

2 Methods

2.1 Design and procedure

Study design

We developed a search application based on LSI to search for breast cancer stories. The stories were downloaded from the website of *De Amazones* [19]. This website was founded by a group of young women with breast cancer and provides stories that are spontaneously submitted by patients for publication on the website. In April and May 2008 women with breast cancer were invited to the Leiden University Medical Centre to use our application in a computer room. They could not see each other's screen and one of the investigators was present for questions. The women were encouraged to use search queries consisting of natural language and to search for whatever they wanted. They could search for and read in the stories as long as they liked. Their search queries and the time involved were automatically saved in a database. At the end women completed a final questionnaire about their satisfaction with the search process and their background characteristics, also automatically saved in the database. For details see Appendix I and [20].

Recruitment process

Recruitment announcements were disseminated online via several breast cancer forums and offline via local newspapers, the magazine of the LUMC and several support groups. Women who subscribed were sent additional information, including an informed consent form. Twenty-four women returned the form. They received a book token for their participation and restitution of travel costs.

Ethical aspects

The board of *De Amazones* foundation gave us permission to conduct the study. The women who submitted their story to the website of *De Amazones*, i.e. the writers of the stories that could be searched for in the LSI application, were not identifiable since they used nicknames when submitting their story. Our research proposal was presented to the Ethical Committee of the Leiden University Medical Centre. The Committee concluded that our study involved no medical intervention and that we could proceed.

2.2 Development of the application

Set of stories

In January 2007 all 171 stories were downloaded from the website of *De Amazones* [19]. The stories were all written in the Dutch language and in the first person. All stories were 'completed' individual stories; they were not part of an interactive forum. Length, structure and content of the stories differed. The mean length of the stories was 759 words (SD=723 words) and ranged from 55 words to 5,112 words (median 568 words).

The document space

Latent Semantic Indexing [14] was used to place the stories of *De Amazones* in a multidimensional space. After removing stop words, the words in the set of stories were reduced to word stems. A word stem-by-story matrix was constructed, with columns for stories and rows for word stems. This matrix was filled with weighted values. For each cell in the matrix the following ratio was calculated: the frequency of the word stem in the story divided by the number of stories the word stem occurred in. The number of dimensions of the document space that was generated in this way, was lowered by applying a special technique purported to represent meaning underlying the stories rather than words [15]. For technical details see Appendix II.

The search queries of the participants went in-real-time through the same process as the stories and were also placed in the document space. Stories with the shortest distances to the search query were retrieved and presented to the user. A search query of one word that did not appear in the set of stories could not be positioned in the document space.

Study website

The search page of the search facility consisted of an explanation, a search box, a Reset button and a dynamic list of stories retrieved (Figure 2). The search application was explained as follows:

> *On this page you can search for stories on the website of De Amazones. You can search by typing text in the box below. This can be your own story, for example, but it can also be some text about a topic you want to know more about. While you are typing, stories that are similar to the text you type will be retrieved. The more text you type, the more accurately stories can be retrieved.*

There was no Search button, since the application started searching automatically when a participant completed or deleted a word or stopped with typing for 3 seconds.

For each story retrieved the writer's nickname was presented as well as zero to five pink ribbons (Figure 2). The number of pink ribbons indicated the degree a story matched the query. The more pink ribbons, the higher the similarity. Retrieved stories were presented in descending order of the number of pink ribbons. Ten stories were presented to the user and all stories that had the same similarity degree as the tenth story.

Retrieval of stories continued non-stop during the typing of a search query: stories could disappear from the list, other stories could appear and the number of pink ribbons could change. While a participant was typing a search query, she could see which changes her typing caused in this dynamic list.

A user could at anytime access one or more stories in the dynamic list by clicking on the nickname of a writer. A user who accessed stories could return to the search page at anytime. After accessing stories a user could choose to refine the search query already typed by adding or removing text or to completely remove the search query already typed and start typing a new query. With the Reset button text in the search box could be removed.

To finish the application participants could click on the Questionnaire button. After clicking this button they could not return to the application.

On this page you can search for stories on the website of De Amazones.
You can search by typing text in the box below. This can be your own story, for example, but it can also be some text about a topic you want to know more about.
While you are typing, stories that are similar to the text you type will be retrieved. The more text you type, the more accurately stories can be retrieved.

I would like to know more about how De Amazones have coped with the diagnosis and treatments and how they went on with their lives

List of stories retrieved:
The number of pink ribbons (0-5) indicates to what extent a story matches your search.

- leyre
- mitchy
- marjo2
- rina_ruth
- keeke
- january
- maisa
- ing38
- conny
- nika
- Chica

Figure 2 Screenshot of the search page showing the LSI search facility.

2.3 Data analysis

Data from the database were imported in the statistical software package SPSS 17.0. Descriptive statistics were performed. Time spent using the application was defined as the total time elapsed between starting with typing the first search query and clicking on the Questionnaire button. Time spent reading stories was defined as the time spent using the application subtracted by time spent searching. The mean reading time per story per participant was calculated by dividing the time reading stories by the number of stories that was accessed. In order to interpret the degree of satisfaction, the results of the present study were compared to the results of an earlier conducted online randomized controlled experiment in which the same questions about search satisfaction were asked [20]. The answers to the last question of the final questionnaire asking whether participants had any remarks about the study, were examined in order to get insight into participants' remarks regarding satisfaction with the search facility.

Participants' search queries (sequence, verbatim texts) and the number of stories accessed thereupon were imported from the database into MS Excel to count the number of searches a participant performed. We defined a *new search* as typing a query in an empty search box: i.e. at the start or after complete removal of a query already typed. Furthermore, for each participant the verbatim text of the query that contained the most words was word-counted. Also, the number of stories accessed was counted. A story that was accessed multiple times in the same search session, was counted multiple times.

Moreover, refinement fractions were calculated. A *search refinement* was defined as altering the query already typed after accessing stories (i.e., adding text at the end, the beginning or in the middle of a query; or, removing and/or replacing part of the query text). A search could contain more than one refinement. A participant's *refinement fraction* was defined as the number of searches with refinements divided by the total number of searches. The refinement fraction lies between 0 and 1. The closer to 0, the smaller the percentage of searches refined by a participant. The closer to 1, the greater the percentage of searches refined by a participant.

The verbatim texts of the queries provided insight in the use of natural language in the queries. The grammatical structures of the queries (range: complete sentences – separate keywords) and the descriptions of information needs (personal experience, in what person, in question form, etcetera) were examined. The search queries that are quoted in this article as illustrations were translated from Dutch into English.

3 Results

3.1 Participant statistics

Table 1 shows the demographic and disease characteristics of the participants and their use of the Internet and *De Amazones* website before study participation. One participant was excluded from the analyses due to not having accessed any stories.

Table 1 Baseline characteristics of the participants (n=23).

Baseline characteristics	N (%)[a]
Demographic characteristics	
Age in years (mean, SD)	51.9 (9.4)
Married or living together	16 (70)
Children	19 (83)
Religious	8 (35)
Higher professional education or university degree	11 (48)
Employed	14 (61)
Disease characteristics	
Time since diagnosis in months (mean, SD)	61.6 (62.4)
Diagnosed with one tumour	15 (65)
Size of tumour ≥2 cm	16 (76)
Cancer in axillary lymph nodes at diagnosis	12 (55)
Metastases to other parts of the body	0 (0)
Breast conserving surgery	9 (39)
Mastectomy	17 (74)
Radiation therapy	13 (57)
Chemotherapy	16 (70)
Hormonal therapy	13 (57)
Cancer free	21 (91)
Use of the Internet and De Amazones website	
Daily Internet use	19 (83)
Familiar with searching online for specific information	23 (100)
Familiar with accessing fellow patients' stories on the Internet	21 (91)
Visited De Amazones website at least once before participation	15 (65)
'Rather well' or 'well' familiar with De Amazones website	10 (67[b])
Read half or more of De Amazones stories before	7 (47[b])

[a] n(%) is shown unless noted otherwise

[b] Percentages based on the n=15 participants who visited *De Amazones* website before.

3.2 Use of the LSI application

Search behaviour

Time spent using the application varied from 8.9 to 76.2 minutes, with a mean of 48.6 minutes (SD=20.0). Participants performed on average 7 searches (SD=5; range 1-21). Their longest search query contained 2 to 130 words, with a mean of 22 words (SD=30). Participants accessed on average 26 stories (SD=14; range 6-52) and their mean reading time per story was on average 2.0 minutes (SD=1.8; range 0.6-7.0).

Figure 3 and 4 provide illustrations of the search sessions of two participants with different search strategies. Figure 3 shows a participant who performed one search in which she refined her search query twice after accessing stories. She started with typing a search query (SQ1) and accessed 13 stories of the list of stories retrieved. Then, she added text to the query already typed (SQ2), accessed 6 stories of the changed list, and again added text (SQ3) and accessed 6 stories. In her whole session, she accessed 25 stories, four of which she accessed twice (these are double counted). Her longest search query consisted of 13 words (=SQ3, word count in Dutch). Her refinement fraction is 1.0 (=1 search with refinements/1 search).

Figure 4 shows a participant who performed 12 searches. In her first search she refined her search query once after accessing stories. In her whole session she accessed 45 stories, six of which she accessed twice (these are double counted) and one of which she accessed three times (this one is triple counted). In three of her searches she did not access any stories. Her longest search query consisted of 6 words (=SQ2, word count in Dutch). Her refinement fraction is 0.08 (=1 search with refinement/12 searches).

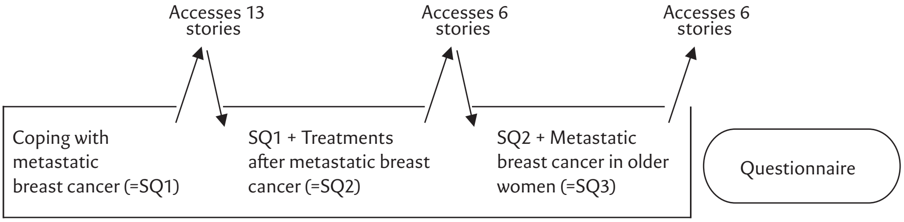


Figure 3 Search session of a 61-year-old participant who spent 76.2 min using the application with a mean reading time of 2.6 min per story. SQ=search query. SQ2 is formed by expanding SQ1, SQ3 by expanding SQ2.

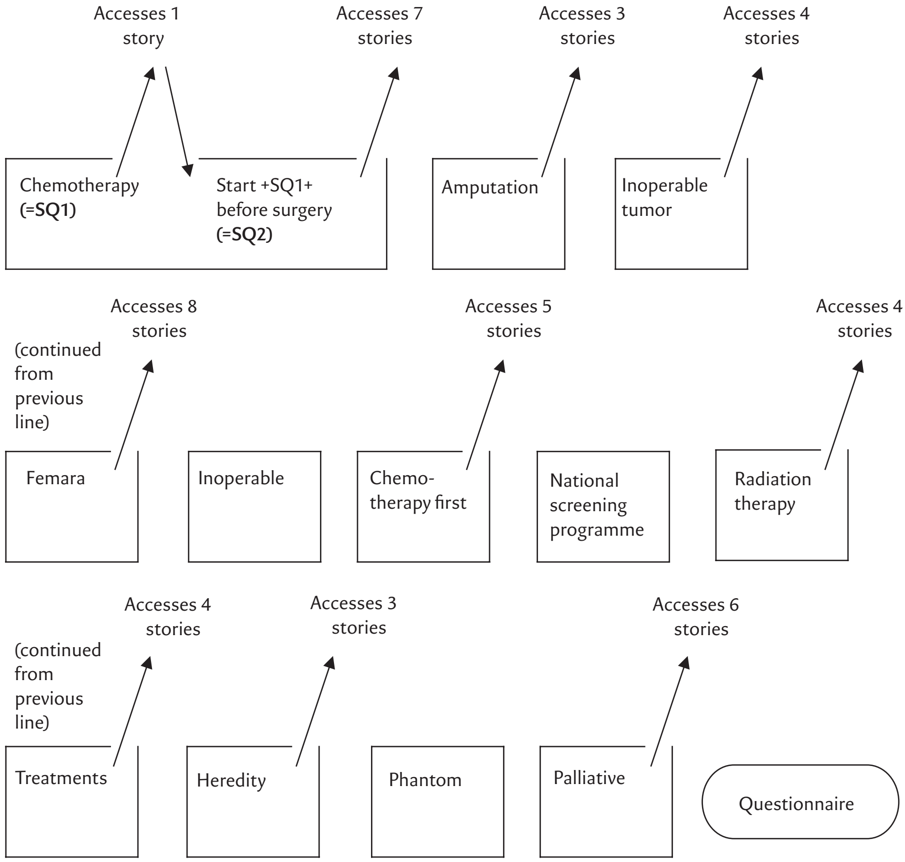


Figure 4 Search session of a 59-year-old participant who spent 58.1 min using the application with a mean reading time of 1.1 min per story. SQ=search query. SQ2 is formed by expanding SQ1.

Refinement of search queries

Figure 5 shows the refinement fractions of the participants. The mean refinement fraction of the participants was 0.25 (SD=0.30) and the median was 0.17. Seven participants (30%) refined none of their searches and therefore had a refinement fraction of 0. Two participants (9%) refined all of their searches and therefore had a refinement fraction of 1.

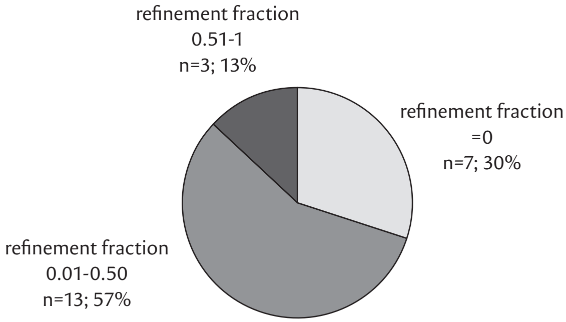


Figure 5 Refinement fractions of the participants (n=23).

Natural language use in search queries

The analysis revealed that natural language was used in three different ways (note that a participant who performed more than one search and/or refinement can use several categories in her session):

1. *Complete sentences*. Twelve participants used complete sentences in their queries. These sentences can be divided in three subcategories:
   a. *Own experience in the first person*. Four participants typed their own experience in the first person. For example:
   'After chemo and radiation therapy I have developed many problems with my condition and joints. I am therefore very limited in doing things.' (54-year-old participant)
   b. *Description of information need in the first person*. Three participants decribed what they were searching for in the first person. For example:
   'I want to know more about possible metastases.' (58-year-old participant)
   c. *Questions in general or literally directed to other patients*. Seven participants formulated questions. One participant addressed her questions directly to her fellow patients:
   'The weight gain since taking Tamoxifen, how did you lose those extra pounds? This was hard for me, and also the menopausal symptoms were disappointing.' (59-year-old participant)
   Others asked questions more in general, for example:
   'Breast reconstruction and nipple reconstruction, is that painful?' (35-year-old participant)
2. *Topics described in a short grammatical structure*. Fourteen participants formulated search queries in which they used a short grammatical structure to describe the topic they were looking for. Examples are:
   - 'experiences with breast reconstruction and lymphedema' (58-year-old participant)
   - 'coping of children after they heard that their mother has breast cancer' (38-year-old participant)

3. *Topics described in one or two keywords*. Fourteen participants used search queries consisting of one or two keywords (word count in Dutch). Examples are:
   - 'heredity' (53-year-old participant)
   - 'hormonal therapy' (38-year-old participant)

   Two of these fourteen participants used no other type of search queries: they used only search queries consisting of one or two keywords.

3.3 Satisfaction with the search process

Table 2 shows participants' satisfaction with the search process. All outcomes measured on a scale from 1 to 5 scored 3 or higher. The outcome measure 'opinion about the list of stories displayed after a search' scored highest with a mean of 4.4 (SD=0.6). The mean of the outcome measure 'overall satisfaction with the search facility' was 6.7 (SD=1.7) (scale 1-10).

Table 2 Participants' satisfaction with the search process (n=23). Higher means indicate better outcomes.

Outcome measure	Mean (SD)
Opinion about the search facility (range 1-5)	3.9 (0.7)
Opinion about the list of stories displayed after a search (range 1-5)	4.4 (0.6)
The extent to which search options enable finding information one was looking for (range 1-5)	2.9 (1.0)
Recommendation to others and future own use (range 1-5)	3.7 (0.9)
Overall satisfaction with the search facility (range 1-10)	6.7 (1.7)

The results for satisfaction with the LSI application lay in the same range as the results for satisfaction with a story topics search facility or with a writer profile search facility [20]. The story topics search facility group scored on 'overall satisfaction with the search facility' a mean of 7.3 (SD=1.4) and the writer profile search facility group scored a mean of 7.1 (SD=1.6). The outcome measure on which LSI seems to score somewhat higher is 'opinion about the list of stories displayed after a search'. On this outcome measure the story topics search facility group scored a mean of 3.6 (SD=0.9) and the writer profile search facility group a mean of 3.8 (SD=0.9).

The answers to the last question of the questionnaire asking whether participants had any comments about the study, yielded some useful remarks regarding satisfaction with the LSI search facility. Seven participants commented on the content of *De Amazones* stories. In their opinion the stories were predominantly written by young women and by women who just received the diagnosis and contained quite personal experiences. These participants were more interested in stories of older women who received the diagnosis longer ago and in stories containing experiences that were supported by scientific research. One participant commented on the search facility. According to her there was sometimes no link between her search query and the stories retrieved and when using certain keywords as search query no stories were retrieved at all. This participant used predominantly one or two keywords as search queries.

4 Discussion

The purpose of this study was to get insight into how breast cancer patients use the facilities that LSI offers in searching for stories of other patients and how satisfied they are with an LSI search facility. This LSI search application was built without the need for human tagging and categorizing of the stories and offers the user the benefits of searching in natural language queries, a dynamic list of retrieved stories which adapts while typing a query, and -after accessing stories- refining queries to narrow search results.

In building the LSI search facility we experienced that the fine-tuning of the multidimensional space took quite some time and energy. An explanation for this might be that breast cancer stories are quite similar: each of the stories discusses treatments and coping with illness, although with a slightly different focus. LSI can correctly cluster a set of texts about entirely different topics [14, 16], but may have more difficulty with texts that are more similar.

The result that participants differed in search behaviour is consistent with previous studies. Analyses of large numbers of user queries on Web search engines revealed that users differ in the number of queries and results pages viewed per query, and in the mean number of terms per query [21, 22]. Kim [23] found that a person's search behaviour –such as time spent in retrieving information- was associated with cognitive style, information search experience and information search task. So, in order to interpret the patients' search behaviour for stories, patients' information need and reasons for searching should be known. For a patient with a specific information need conducting one search might indicate a good performance, while a patient with a wider, not yet fully defined information need might need to conduct more searches for a good performance. In this exploratory study such relationships were not examined.

More than two thirds of the participants refined a search query after accessing stories, nearly one third did not. By refining a query search results may be more targeted. Explanations for not refining any of their search queries might be that participants are not used to search in this way or that the stories they had retrieved so far already satisfied their information need. In previous studies it was found that most users searched one query only and did not follow with successive queries [21, 22]. In future research it would be informative to find out the reasons why some participants did not refine their search queries.

Participants used natural language queries in three ways: 1) complete sentences, 2) short grammatical structures, and 3) keywords. Short grammatical structures and keywords were most used as queries. Our results were in line with the results of Zeng et al. [24] who found that consumers who searched for health information on the MEDLINEplus website tended to use short and general text queries (rarely more than one or two words). Spink et al. [21] showed in their study of consumer use of a large Web search engine that the mean number of terms in queries was 2.4. An explanation might be that people are most accustomed to using short queries. Their experience with other search engines might be that general, short queries retrieve results that give sufficient satisfaction. In addition, it takes less time and energy to type a short query than to type a whole sentence. Yet, with LSI it is expected that the longer the search query, the more targeted the stories retrieved.

Seven participants addressed in complete sentences a question directly to other patients. We might say that this is in line with the intentional stance theory of Daniel Denett: they saw the LSI application as if it were a 'live' interaction partner and assigned an intention to it [25].

Participants were as satisfied with the LSI application as with two other search facilities used in a previous study, that is, a story topics search facility and/or a writer profiles search facility [20]. However, the two studies are not entirely comparable since the participants of our earlier study were younger, participated online without direct contact with the investigator, and the groups were larger [20].

A striking result is that the outcome measure 'Opinion about the list of stories displayed after a search' is relatively positive in the LSI group compared to the other two search facilities. The LSI-users might have a high satisfaction on this outcome measure because they could observe the list of retrieved stories while typing a search query and could immediately see any changes in this list. With qualitative research, for example a think-aloud study, it could be examined why participants have a high satisfaction on this outcome measure.

The Internet is highly dynamic. The focus is shifting from searching for information using search engines into sharing information via social media [26]. Members of the online patient community PatientsLikeMe, for example, send their experiences to other members who -according to their profiles- may most likely benefit from them [27]. Yet, we think that search facilities remain important for patients who do not wish to create a personal profile on social media.

4.1 Limitations

The present study was an exploratory study. The number of participants was relatively low. Also, the mean age of the participants was somewhat higher than the mean age of the writers of the stories the participants could search for (mean age at diagnosis 35.8 years (SD=6.2)) [28]. The difference in age between writers and participants may have affected participants' search satisfaction. Furthermore, the questionnaire did not go into detail on participants' reasons for their search behaviour and satisfaction.

4.2 Conclusion

Our study showed that participants did use the facilities that LSI offers in searching for other patients' stories. The majority of the participants refined their search queries and used short grammatical structures or complete sentences in their search queries. Furthermore, satisfaction with LSI was in the same range as satisfaction with a story topics search facility or a writer profile search facility. Participants appreciated the dynamic list of retrieved stories during the typing and formulating of their query. Given these promising results of the use of the LSI search facility by patients to search for stories of other patients, further research would be useful. With an extensive quantitative research the LSI search facility can be statistically compared to other search facilities. A qualitative research can provide insight into the motivations of patients.

Acknowledgements

The authors wish to thank the participants for their time and interest. We also wish to thank the organizations that disseminated the call for participation.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO). The paper is part of the Narrator project addressing the problem of storage and accurate retrieval of illness narratives, conducted under the umbrella program ToKeN2000.

References

1. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Educ Couns* 2004; 53: 57-64.
2. Ziebland S, Chapple A, Dumelow C, et al. How the internet affects patients' experience of cancer: a qualitative study. *BMJ* 2004; 328: 564-569.
3. McTavish FM, Gustafson DH, Owens BH, et al. CHESS (Comprehensive Health Enhancement Support System): an interactive computer system for women with breast cancer piloted with an underserved population. *J Ambul Care Manage* 1995; 18: 35-41.
4. Herxheimer A, McPherson A, Miller R, et al. Database of patients' experiences (DIPEx): a multi-media approach to sharing experiences and information. *Lancet* 2000; 355: 1540-1543.
5. Pitts V. Illness and Internet empowerment: writing and reading breast cancer in cyberspace. *Health (London)* 2004; 8: 33-59.
6. Hardey M. 'The story of my illness': personal accounts of illness on the Internet. *Health (London)* 2002; 6: 31-46.
7. Chou WYS, Hunt Y, Folkers A, et al. Cancer Survivorship in the Age of YouTube and Social Media: A Narrative Analysis. *J Med Internet Res* 2011; 13(1): e7.
8. Health Experience Research Group. Database of personal and patient experiences (formerly Dipex), www.healthtalkonline.org (accessed 11 May 2008).
9. BreastCancerStories.org, www.breastcancerstories.org (2006, accessed 18 December 2011).
10. PubMed. Combining MeSH Terms Using the MeSH Database, http://www.nlm.nih.gov/bsd/viewlet/mesh/combining/mesh2.html (2011, accessed 18 December 2011).
11. Richter RR, Austin TM. Using MeSH (Medical Subject Headings) to Enhance PubMed Search Strategies for Evidence-Based Practice in Physical Therapy. *Phys Ther* 2012; 92: 124-132.
12. Overberg RI, Alpay LL, Verhoef J, et al. Illness stories on the internet: what do breast cancer patients want at the end of treatment? *Psychooncology* 2007; 16: 937- 944.
13. Ziebland S, Herxheimer A. How patients' experiences contribute to decision making: illustrations from DIPEx (personal experiences of health and illness). J *Nurs Manag* 2008; 16: 433-439.
14. Deerwester S, Dumais ST, Furnas GW, et al. Indexing by Latent Semantic Analysis. *J Am Soc Inf Sci* 1990; 41: 391-407.
15. Hoenkamp E. Unitary Operators on the Document Space. *J Am Soc Inf Sci* 2003; 54: 314-320.
16. Landauer TK, Dumais ST. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychol Rev* 1997; 104: 211-240.
17. Hoenkamp E, Overberg R. Computing Latent Taxonomies from Patients' Spontaneous Self-Disclosure to Form Compatible Support Groups. *Stud Health Technol Inform* 2006; 124: 969-974.
18. Hoenkamp E, Bruza P, Song D, et al. An Effective Approach to Verbose Queries Using a Limited Dependencies Language Model. In: *Advances in Information Retrieval Theory: Second International Conference on the Theory of Information Retrieval, ICTIR 2009* (eds L Azzopardi, G Kazai, S Robertson, et al.), Cambridge, UK, 10 September-12 September 2009, pp. 116-127. Springer.
19. De Amazones, young women with breast cancer (In Dutch: De Amazones: jonge vrouwen met borstkanker), www.de-amazones.nl (accessed 2 January 2007).
20. Overberg R, Otten W, De Man A, et al. How breast cancer patients want to search for and retrieve information from stories of other patients on the internet: an online randomized controlled experiment. *J Med Internet Res* 2010; 12(1): e7.

21. Spink A, Wolfram D, Jansen BJ, et al. Searching the web: the public and their queries. *J Am Soc Inf Sci* 2001; 53: 226-234.
22. Jansen BJ, Spink A, Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the web. *Inf Process Manag* 2000; 36: 207-227.
23. Kim KS. Information seeking on the Web: Effects of user and task variables. *Libr Inf Sci Res* 2001; 23: 233-255.
24. Zeng QT, Kogan S, Plovnick RM, et al. Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval. *Int J Med Inform* 2004; 73: 45-55.
25. Dennett DC. True believers: The Intentional Strategy and Why It Works. In: Dennett DC (eds) *The Intentional Stance*. Cambridge: MIT Press Mass, 1987, pp. 13-36.
26. Lober WB, Flowers JL. Consumer Empowerment in Health Care amid the Internet and Social Media. *Semin Oncol Nurs* 2011; 27: 169-182.
27. Frost JH, Massagli MP. Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data. *J Med Internet Res* 2008; 10: e15.
28. Overberg R, De Man A, Wolterbeek R, et al. Spontaneously published illness stories on a website for young women with breast cancer: do writers and themes reflect the wider population? Submitted.
29. ORACLE. Dutch (nl) Default Stoplist, http://web.archive.org/web/20020111230441/downloadwest. oracle.com/otndoc/oracle9i/901_doc/text.901/a90121/astopsu4.htm (2001, accessed 29 November 2007).
30. Kraaij W, Pohlmann R. Porter's stemming algorithm for Dutch. In: *Informatiewetenschap 1994: Wetenschappelijke bijdragen aan de derde STINFON Conferentie* (eds LGM Noordman and WAM De Vroomen), Tilburg, The Netherlands, pp. 167-180. Tilburg: STINFON.

CHAPTER 8

General discussion

Spontaneously published breast cancer stories on the Internet are a unique source of support and information for other breast cancer patients for three reasons: 1) these stories are usually complete stories; 2) the writers are not selected and their stories are not edited; and 3) the number of these stories is large and continuously increasing. According to the narrative approach [1], social comparison theory [2], and social cognitive theory [3], reading stories written by others in a similar situation may have the most positive impact on how the reader copes with her own situation. However, such stories can be difficult to find. Therefore, the primary aim of this thesis was to gain insight into which search facilities help breast cancer patients find spontaneously published stories written by other patients in a similar situation to themselves.

This was examined following a user-centred design: users of search facilities (i.e., the patients) were involved in each phase of the research. This ensured that the search facilities were tailored to the users' needs and preferences, rather than forcing users to adapt their search behaviour to search facilities developed without their involvement. Concretely, we examined which breast cancer patients spontaneously publish their story online and which themes they write about in their stories (*Chapter 3*), and how these stories are presented and disclosed on websites (*Chapter 2*). We then examined which topics and writer characteristics patients prefer to search for (*Chapter 4*) and translated these preferences into search facilities. Finally, we examined which search facilities had the most positive effects on patient satisfaction and search success, and how patients actually searched for stories written by others (*Chapters 5-7*).

In this chapter, we will reflect on certain aspects of spontaneously published stories on the Internet, using the findings from *Chapters 2 and 3*. Secondly, on the basis of the results presented in *Chapters 4-7*, we will discuss aspects of search facilities for patient stories. Subsequently, we will describe methodological issues and limitations and practical implications. Finally, we will conclude the chapter with an overall conclusion and reflection.

Spontaneously published breast cancer stories

In this thesis, spontaneously published breast cancer stories are defined as first-person texts written by breast cancer patients in which the patients describe their experiences with the illness trajectory, their feelings about having breast cancer, and how they cope with the disease, and which they published at a given moment on the Internet as a 'completed' story (i.e., without later additions and not as part of an interactive forum). The stories can be published and read anonymously.

Stories play an important role in how people learn about the world, and we are used to telling and hearing them from childhood onward (e.g. [4]). More specifically, patient stories contain valuable patient expertise on strategies for coping with day-to-day personal health issues gained through trial and error of the lived experience [5]. The Internet allows all breast cancer patients who want to share their stories with others to do so by publishing their story on websites that offer this option. Other patients can read these spontaneously published stories for information and support. The studies in *Chapters 2 and 3* highlight a number of important aspects of spontaneously published stories. Here we discuss the aspects ‘representativeness’, ‘authenticity’ and ‘reliability’. In this context, we also discuss the role of physicians and new developments on the Internet.

Representativeness

To gain an idea of the writers and their stories, we conducted a content analysis of all 167 breast cancer stories available on the Dutch-language website “De Amazones” [6] at the time of our study (*Chapter 3*). We found that women were more likely to be a writer if they had been diagnosed at a younger age, had undergone a mastectomy, or were in the first treatment period. The analysis of the themes in the stories showed that some themes were addressed less often than would be expected based on their prevalence in quantitative studies. For example, the theme ‘body image and sexuality’ was discussed in 20-25% of the stories, while prevalence studies showed that approximately half the patients experience problems regarding this theme.

Our findings point to a spontaneous selection of both the patients who publish their story on the Internet and the themes they write about. One explanation for the finding that writers were relatively often in the first treatment period might be that patients find themselves in the middle of the coping process during this period and therefore feel a strong need to write [7-12]. Regarding the topics, patients mainly write about topics relating to their demographic and disease characteristics, topics that are hard to handle, or topics with which they still have to cope (see also *Chapter 4*). Furthermore, for some patients the Internet may provide the privacy and anonymity to write about more personal topics than they would do in other settings (e.g. [13]).

In conclusion, our content analysis of the stories on the website of De Amazones (*Chapter 3*) indicates that spontaneous selection occurs with regard to both the patients who publish their story and the topics about which they write. The question arises whether this spontaneous selection poses a problem for readers who want to use the stories for support and information. With regard to patient decision aids that include patient stories there is a clear consensus among an international team of patients, practitioners, policy makers, and researchers that a range of patient perspectives and experiences should be included [14]. This principle has been applied on the website Healthtalkonline (formerly DIPEx) [15], where the developers have used maximum variation sampling when interviewing patients for their website to ensure that a wide range of patient perspectives and experiences are represented [16]. They thus offer a set of stories that is qualitatively representative: for each topic that appears to be of interest in coping with the disease, an approximately equal number of positive and negative perspectives/experiences is presented [16]. However, even such a ‘balanced’ presentation of views can potentially give false impressions [17]: the number of positive and negative experiences related to each topic seems to be equally balanced, while in reality perhaps 90% of patients had negative experiences regarding a particular topic. Therefore, it might be argued that readers are best served by a set of stories that meets quantitative representativeness. Thus, if, for example, a quarter of patients are struggling with topic x, this topic should be covered in about a quarter of the stories. Nonetheless, even if a set of stories meets criteria for qualitative or quantitative representativeness, it remains unclear whether this representativeness will reach the reader since they may select only certain stories to read and thus may not read the whole set of stories. This is illustrated by our interviews (*Chapter 4*) in which patients mentioned that they did not read stories about patients with metastatic cancer to avoid being confronted with this subject. Moreover, a reader can select stories from other sources of information besides a single website with a representative set of stories and, consequently, the representativeness of the original stories may be affected by reading further stories on other websites. On the other hand, an advantage of websites with a set of stories that does meet qualitative or quantitative representativeness is that it enables the reader to interpret the representativeness of the set of stories.

For a set of spontaneously published stories, representativeness will always remain an issue due to the above-mentioned self-selection of writers and topics and due to the dynamic nature of such a story set where stories are constantly being added and removed. For example, the 2013 addendum to Chapter 2 shows that, of the 12 sites described in 2006, three no longer exist and four no longer contain patient stories. Thus, five of the twelve sites still contain breast cancer stories. On one of these sites, the number of stories increased from 49 to 790 [6]. In addition, in 2013 twelve new websites with patient stories were found. In conclusion, the addendum illustrates the dynamic nature of patient stories online.

Authenticity

In our study on how Dutch-language breast cancer stories are disclosed on websites (*Chapter 2*), we defined authentic stories as those written or told by real patients. The addendum to the study, written in 2013, revealed that the stories on three of the sixteen websites examined could definitely be ascribed to real patients since these patients had been filmed during hospital visits (*Chapter 2*). However, the literature shows that there are different opinions about authenticity. Some feel that even the authenticity of stories written by real patients can be questioned since, in their stories, patients may tend: a) to apply a little self-deception in order to feel better about their situation; b) to portray themselves more favourably than they really are; and/or c) to comply with the dominant narrative conventions of their culture and society, such as the restitution narrative (patient gets sick, patient receives medical intervention, patient recovers and returns to pre-illness life) or the quest narrative (disease is required for a positive change) [12;18;19]. On the other hand, others feel that regardless of what drives the story, it remains the story the patient wanted to tell and, as such, it is authentic [18]. Moreover, stories that are not written by real patients might also offer support. Research on social comparison shows that people are able to compare themselves with non-existent persons because they are able to cognitively construct others [2]. The aspects discussed above raise the question as to which criteria make a story authentic and how important this authenticity is for the readers.

Reliability

Our study of sixteen websites with Dutch-language breast cancer stories (*addendum to Chapter 2*) revealed that the stories on six websites had been reviewed by an editor prior to publication. The question is whether review by an editor results in reliable stories. As discussed above, one view is that stories written by patients are always ‘true’, because stories describe the patients' experiences [18;19]. However, besides patient experiences, stories also contain medical information. The medical information described by patients in their stories may differ from the conventional medical standards. To prevent the stories containing unconventional medical information, an editor/health care professional may review the stories prior to publication to ensure that the medical information is reliable. However, in practice this might be difficult for stories that are published spontaneously, because it is not known when new stories will be added. Moreover, those who publish their own story online may feel hurt if their story is criticized or simply corrected for facts [4], since a story is a very personal form of expression.

It is not known how many spontaneously published stories contain medically incorrect information. Studies on the accuracy of medical information in online support groups have shown that only few postings are false or misleading [20;21] and that most false or misleading statements are rapidly corrected by participants in subsequent postings [20]. However, for spontaneously published stories, correction by others is not possible due to the absence of an interactive component. Another type of medical information covered by stories, and possibly of more concern than medical inaccuracies, is the writer's treatment plan. A reader can mistakenly think that she needs the same treatments as the writer, because the writer's situation seems very similar to that of the reader. However, the writer may have omitted to report important background information that was decisive in determining her treatment plan. This issue will not be remedied by editing a story, because the bottleneck lies in the interpretation by the reader.

Role of health care professionals

We saw that readers of spontaneously published stories on the Internet may be affected by the stories' representativeness, authenticity and reliability. It is important to discuss the role of health care professionals in how these stories are interpreted by patients. Studies show that many cancer patients discuss information that they find on the Internet with their health care professionals [22-24]. Some physicians react positively: patients are better informed about their disease and treatment options and it strengthens the doctor-patient relationship because it becomes a partnership rather than a process of one-way information provision. Other physicians feel threatened in their expertise, have difficulty discussing the information, and think that it makes patients anxious and confused [22-24]. Therefore, it is important for physicians to reflect on how to deal with patients who have read other patients' stories on the Internet and who have questions about these stories or who believe they should receive the same treatments as the writer.

Another issue is the role of physicians in referring to health-related websites, such as websites with patient stories. Research [22] shows that physicians do not refer to health-related websites often and that they seldom or never refer to online support groups. One reason for this is that many physicians find it difficult to stay up to date with reliable Internet sites for patients [22]. Given the above-mentioned aspects of spontaneously published stories, it can indeed be difficult for physicians to determine which websites with patient stories they can 'safely' refer their patients to. One solution may be to develop a quality label for websites with patient stories, providing readers with information about the representativeness, authenticity and reliability of the set of stories. Such quality labels already exist for websites with medical information, for example the HONcode developed by the Health On the Net Foundation [25] and 'ZegelGezond' developed by a Dutch organization [26]. However, for a site with spontaneously published patient stories it may be difficult to keep a quality label up to date, since the compilation of the set of stories changes constantly. Therefore, another solution may be to develop a checklist with a set of criteria that enables readers to focus on those aspects that are important for a set of patient stories. For example, readers should check whether the stories contain both positive and negative experiences and whether information is provided about why patients decided to share their story. Such checklists have already been successfully developed for patient decision aids. For example, the International Patient Decision Aid Standards (IPDAS) Collaboration has developed a checklist with a set of criteria that enables users of a decision aid to assess its quality [14]. The outcome of this research on patient decision aids may also be helpful in the development of checklists for spontaneously-published patient story sets.

New developments in accessing patient experiences on the Internet

In this thesis, we focused on websites with 'completed' spontaneously published breast cancer stories. However, there are various other forms in which patients can read experiences on the Internet.

The Internet provides patients with a broad range of online support groups, which have been widely investigated (e.g. [27-30]). The experiences published in such online support groups differ from the stories focused on in our research since online support groups allow patients to pose questions and/or reply to each other. Also, the excerpts in online support groups are usually shorter than complete stories.

In the last few years, there have been several new developments regarding patient experiences on the Internet. One example is the website PatientsLikeMe [31], which is an online community built to support information exchange between patients. The site provides customized disease-specific outcome and visualization tools to help patients understand and share information about their condition [31]. Another development is the opportunity for patients to publish their personal story on the Internet through videos on YouTube [32].

More recent developments are social media, such as Twitter and Facebook, which serve many functions including health-related purposes, such as sharing illness experiences [33-37]. Social media enable users to post messages for everyone to read (public) or only for certain friends or followers (customized) and allow users to follow individuals and organizations in order to receive new posts from these people or groups. In addition, users can forward possibly relevant posts to friends or followers. The use of social media for sharing illness experiences raises new questions: for example, how can patients find each other, how can they retrieve relevant experiences posted by others, and how should they deal with privacy issues when using real name accounts.

Moreover, the Internet is no longer a medium that can only be accessed from behind a computer screen. With the introduction of the Internet on smartphones, everyone has the Internet at their fingertips, wherever they are and whatever they are doing. Smartphones also offer the ability to support patients through apps. For example, the

Trimbos Institute is developing an app for women newly diagnosed with breast cancer that provides mental support during diagnosis and treatment [38].

The forms described above are characterized by consisting of (relatively) short fragments, having interactive components, being moderated, and having a continuous flow of communication (threads). The 'completed' stories, focused on in this thesis, are characterized by the absence of an interactive component and by being non-moderated. In terms of these 'completed' stories, we found that if these stories were disclosed using a 'story topic' search facility, the stories retrieved had a positive effect on patients' self-reported coping with cancer (*Chapter 5*). Future research could investigate whether this finding can be extrapolated to the various other forms of online patient experience-sharing. Moreover, it could be investigated which particular features of these other forms of online patient experience-sharing have positive effects on writers and readers and, more specifically, how they impact a reader's coping with cancer.

Search facilities

Our study from 2013 on how stories are presented on websites (*addendum to Chapter 2*) revealed that two of the sixteen websites offer a search facility for writer characteristics [6;39], while none offer a search facility using content keywords. Nine websites offer a full-text search facility (e.g. [40]). Full-text search engines only retrieve documents in which the word that is searched for occurs literally (for example, a search for 'bestraling' [*radiation*] will retrieve no stories in which this is called 'radiotherapie' [*radiotherapy*]).

The question is how a lack of search facilities affects the reader. Our interviews (*Chapter 4*) showed that breast cancer patients prefer story fragments about specific topics from patients with particular personal and/or disease characteristics. Without search facilities, such stories might be difficult to retrieve. Moreover, our online randomized controlled experiment (*Chapter 5*) showed that search facilities for story topics (based on manual indexing) and writer characteristics contributed positively to readers' satisfaction with the search process (outcome measure 1) and the stories retrieved (outcome measure 2). The 'story topics' search facility also contributed positively to patients' self-reported coping with breast cancer (outcome measure 3). However, this positive effect was not found for the 'writer characteristics' search facility. Overall, the 'story topics' search facility had markedly more positive effects per outcome measure than the 'writer characteristics' search facility, and combining the two search facilities barely increased the positive effect. It might therefore be concluded that, for stories to provide readers with support and information, they should, as a minimum, be disclosed according to story-specific topics.

On the other hand, it is striking to note that analysis of the log data from the online experiment revealed that, per search, patients were more likely to search for writer characteristics than for topics (*Chapter 6*), even though the online experiment had shown that searching for topics had far more positive effects on satisfaction and search success than searching for writer characteristics (*Chapter 5*). So, given these positive effects, why do patients not search for more topics per search? In the interviews (*Chapter 4*), one patient explained that she wanted to be able to search for writer characteristics (treatment undergone) to gain information about a topic (decision-making about treatment). Thus, one explanation might be that writer characteristics and story topics cannot be clearly distinguished from one another when searching for stories. Another explanation might be that the nature of a topic is more important than the number of topics, while for writer characteristics the number might be more important than the nature. In other words, the higher the number of writer characteristics searched for per search, the more specific the search results will be in terms of writer characteristics. This is important in social comparison, since social comparison theory suggests that comparison with a writer on a specific dimension (regardless of the direction of the comparison) will have the most impact when other dimensions of the writer, such as their personal situation or characteristics, are similar to those of the reader [2;41;42]. To ensure cost-effective implementation of search facilities, it is important to examine which writer characteristics and which topics should, as a minimum, be included in search facilities to maintain the positive effects on satisfaction and search success.

Furthermore, the *addendum to Chapter 2* showed that there is an increase in the number of patient videos in which patients tell their story. The question arises how these patient videos could be made accessible to other patients.

Selection of writer search characteristics

The analysis of the log data (*Chapter 6*) showed that patients predominantly search for stories by writers with similar demographic and disease characteristics to their own. This might enable patients to identify with these writers, whose stories they may therefore find most informative and supportive. The question arises whether there are other story characteristics that may be relevant for a reader to identify with the writer, such as the writer's writing style, personality and/or coping style. Based on social comparison theory [2] and social cognitive learning/modelling [3], which suggest that people can learn from observing the behaviour of others (vicarious learning), stories will be more appealing to the reader if the reader observes similarities between him/herself and the writer. An interesting theme for future research would therefore be whether search facilities that include aspects such as writing style, personality and/ or coping style increase the usefulness and impact of stories retrieved. Such research would face the challenging task of clearly defining these aspects and tagging stories according to these definitions.

Feasibility and responsibility

Two challenges in implementing search facilities for spontaneously published stories are feasibility and responsibility. Regarding feasibility, it is important to assess the most cost-effective balance between, on the one hand, the number and type of search facilities (topics, writer characteristics and/or other characteristics) and, on the other hand, the potential positive impact that stories retrieved by these search facilities may have on readers. Another cost-effectiveness issue is tagging stories efficiently to build the search facilities (tagging by human reading and/or semi-automatic tagging).

Regarding responsibility, the question arises as to who will implement the search facilities and who will keep them up to date, as this costs both time and money. Two cost-effective solutions for building search facilities seem to be available. Firstly, writers may tag their own stories, for example by using the coding scheme that was developed for our content analysis (*Chapter 3*). Nowadays, content tagging by users is common on the Internet; the distinction between website builders and website users is less clear. On Twitter and Facebook, for example, users can tag their own posts and those of others. There are also social bookmarking websites that collect information about the content tags assigned to websites by users and which may serve as an alternative to search engines [43]. Another solution might be a technique like Latent Semantic Indexing (LSI). This mathematical technique places stories and search queries in a multidimensional space: distances between them provide information about similarity in content. No human reading is necessary to tag and categorize stories and searching involves typing natural language queries. Our exploratory study showed that patients used the features of the LSI search facility and that they were quite satisfied with the search process (*Chapter 7*). However, the LSI was also associated with some difficulties. For example, fine-tuning the multidimensional space was time-consuming and one participant indicated that she saw no relationship between her search query and the stories retrieved by LSI. Thus, more research is needed before LSI can be used as a search facility to find relevant stories.

Methodological issues and limitations

The research described in this thesis is a combination of qualitative and quantitative studies. Following a user-centred design, we involved the user in the process of developing the search facilities from the outset. The results of one study served as input for the design of the next study: the results of the qualitative interviews with breast cancer patients (*Chapter 4*) served as input for the design of the online randomized controlled experiment (*Chapter 5*), and the results of the online randomized controlled experiment triggered the idea for exploring the use of Latent Semantic Indexing as a search facility (*Chapter 7*). We consider this combination of qualitative and quantitative studies and the user-centred design to be strong methodological features of our research.

However, we wish to critically address two methodological issues. The first issue relates to researching the Internet. Firstly, it is impossible to examine the entire Internet: to keep the research manageable, one needs to focus on a defined part of the Internet. We focused on several Dutch websites with breast cancer stories. Secondly, websites are constantly changing, while researching a website is conducted at a particular moment in time. The set of stories on the De Amazones website, for example, consisted of 49 stories in 2005 (*Chapter 2*) and grew to 167 stories by 2007 (*Chapter 3*) and 790 stories in 2013 (*addendum Chapter 2*). Besides the number of stories, the way in which stories are presented and disclosed might also change. The *addendum to Chapter 2* shows that, in 2013, two websites offered a search facility for writer characteristics, while in 2006 none of them offered such a search facility. However, as in 2006, in 2013 no websites offer a search facility for story topics (based on content keywords). Finally, it is important to realize that our online experiment (*Chapter 5*), although coming close to searching a real website, remains an experimental setting that could have affected the participants' behaviour and opinions.

A second methodological issue is that the qualitative study and the exploratory study involved relatively small numbers of participants. We interviewed 26 breast cancer patients about their preferences for topics and writer characteristics (*Chapter 4*). However, it is important to mention that the results of the interviews are not isolated, but were included in the online experiment and resulted in positive effects on search satisfaction and search success (*Chapter 5*). In the LSI study, 23 breast cancer patients participated (*Chapter 7*). This study was exploratory in nature and more research is required before LSI can be used as a search facility for stories.

Practical implications

The research in this thesis has led to three practical implications:

- If health care professionals decide to refer their patients to websites with spontaneously published breast cancer stories, it is important that they discuss these websites' opportunities and risks with their patients. As mentioned earlier, opportunities and risks lie in the representativeness, authenticity, and reliability of stories and in the availability of search facilities.
- Organizations and individuals who offer spontaneously published breast cancer stories on their websites are advised to provide search facilities to enhance the accessibility of the stories for readers. Since, in patient decision aids, patient stories are being used increasingly frequently [44;45], developers of these aids may also consider including search facilities for stories. Appropriate search facilities increase the likelihood that a reader will find relevant stories that provide support. Currently, the most practical and cost-effective way to implement search facilities might be to ask the patients themselves (writers and readers) to tag the stories according to, for example, writer characteristics and/or topics.
- This thesis examined online breast cancer stories, but it can be assumed that the two above-mentioned practical implications can also be applied to online stories written by patients with other diseases. Stories on the Internet might be especially helpful in rare diseases or for men with breast cancer [46], where the probability of meeting a fellow patient in real life is small [47].

Overall conclusion and reflection

The Internet has a role in exchanging 'complete' illness stories. There are patients who spontaneously publish their story on the Internet and there are other patients who read these stories. Although the writers and the readers may have different intentions, this does not present a problem as both writing and reading serve certain functions In my opinion, it is important that patients make their own decision about whether they want to publish their story on the Internet or whether they want to read stories written by other patients. Important others, such as health care professionals and family members, may recommend patients to write and read stories, but patients should not feel 'pressured' to do this.

Spontaneously published stories on the Internet provide both opportunities and risks. Opportunities lie in the possibility that stories may serve a functional purpose, for both writers and readers, such as offering emotional support or providing models by which to solve problems. Therefore, patients who feel the need to publish their own story or to read stories of others should continue to do so, and publishing and reading stories should not be regulated (an impossible task in itself). The research in this thesis has shown that the opportunities provided by online stories can be increased by implementing relevant search facilities to retrieve stories: participants indicated that stories retrieved using a ‘story topics’ search facility had a positive impact on coping with the disease. In view of this finding, I believe that there should be greater focus on developing search facilities for stories, since there is a very large number of freely-available online patient stories that may offer support and information to other patients. Not making use of these stories is a missed opportunity. Professionals should therefore facilitate patients in finding relevant stories written by others by implementing search facilities that fulfil patient preferences (user-centred design). Stories that are not adequately disclosed, for example due to a lack of search facilities, will have no impact on a reader.

Possible risks regarding spontaneously published stories on the Internet relate to the representativeness, authenticity and reliability of stories. There are two approaches to limit these risks. The first approach is to regulate which patients publish their stories online. However, this is neither desirable nor possible and therefore there will always be patients who publish ‘deviant’ and ‘unconventional’ stories. The second approach is to empower the readers. Since Internet use is part of our everyday life and since we are used to hear stories from childhood onwards, my expectation is that readers already have quite some skills to assess online stories and that there is common sense that not all stories reflect reality. Nonetheless, to further enhance readers’ skills in assessing patient stories, quality labels might be assigned to websites with patient stories or a checklist with criteria to assess websites with patient stories could be developed. Since quality labels may be difficult to keep up to date because stories are spontaneously published at unknown time points, a checklist may be preferable.

We have observed many new developments on the Internet through which patients can share their illness experiences. However, my expectation is that spontaneously published ‘completed’ stories will continue to coexist alongside these new developments, since there will always be patients who want to publish their whole story at once, without interacting with others and without their story being moderated. Future research could focus on the effects that the features of these different forms of online patient experience-sharing (i.e., completed story vs. short fragment, interactive vs. non-interactive, moderated vs. non-moderated, etc.) have on their writers and readers.

As mentioned above, it is neither desirable nor possible to regulate spontaneously published stories on the Internet. However, I do believe that it is important to maintain an open discussion about the opportunities and the risks presented by spontaneously published illness stories on the Internet. All four parties involved should participate in this discussion: patients (both writers and readers), patients’ family and friends, organizations and individuals who offer stories on their websites, and health care professionals. In addition, new developments in sharing illness experiences, such as Twitter and Facebook, should also be discussed. I think it is particularly important to discuss each group’s role and responsibilities in the context of patient experiences on the Internet. Such discussions may start spontaneously, as illustrated by the case of a family member who raised the ethical issue of posts about an ill relative on a real name account on Facebook [48]. Moreover, patient organizations may also play an important role in initiating discussions about this theme. A powerful discussion may help to increase the opportunities and reduce the risks presented by online patient experiences. In this discussion, the question of how stories can be made accessible to patients so that patients can quickly and easily find stories written by similar others should play a central role.

References

1. Polkinghorne DE. Narrative knowing and the human sciences. Albany: State University of New York Press; 1988.
2. Wood JV. What is Social Comparison and How Should We Study it? Personality and Social Psychology Bulletin 1996;22(5):520-37.
3. Bandura A. Social foundations of thought and action: A social cognitive theory. New York: 1986.
4. Ziebland S, Wyke S. Health and Illness in a Connected World: How Might Sharing Experiences on the Internet Affect People's Health? The Milbank Quarterly 2012;90(2):219-49.
5. Hartzler A, Pratt W. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians. Journal of Medical Internet Research 2011;13(3):e62.
6. De Amazones, jonge vrouwen met borstkanker [The Amazones, young women with breast cancer]. De Amazones 2007 [cited 2007 Jan 2];Available from: URL: http://www.de-amazones.nl
7. Pennebaker JW, Seagal JD. Forming a story: the health benefits of narrative. J Clin Psychol 1999 Oct;55(10):1243-54.
8. Smyth JM, Nazarian D, Arigo D. Expressive Writing in the Clinical Context. In: Vingerhoets AJJM, Nyklícek I, Denollet J, editors. Emotion Regulation. Conceptual and Clinical Issues. New York: Springer Science + Business Media; 2008. p. 215-33.
9. Sakalys JA. Restoring the patient's voice. The therapeutics of illness narratives. J Holist Nurs 2003 Sep;21(3):228-41.
10. Mathieson CM, Stam HJ. Renegotiating identity: cancer narratives. Sociology of Health & Illness 1995;17(3):283-306.
11. Anderson JO, Martin PG. Narratives and healing: exploring one family's stories of cancer survivorship. Health Commun 2003;15(2): 133-43.
12. Thomas-MacLean R. Understanding breast cancer stories via Frank's narrative types. Soc Sci Med 2004 May;58(9):1647-57.
13. Joinson AN. Disinhibition and the Internet. In: Gackenbach J, editor. Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications. 2 ed. Burlington/San Diego/London: Academic Press/Elsevier; 2007. p. 75-92.
14. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417-22.
15. Healthtalkonline. DIPEx. Healthtalkonline . 2008. 11-5-2008.
16. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Med Educ 2006 May;40(5):405-14.
17. Bekker HL, Winterbottom A, Butow P, Dillard A, Feldman-Stewart D, Fowler J, et al. Chapter E: Using personal stories. In: Volk R, Llewellyn-Thomas H, editors. 2012 Update of the International Patient Decision Aids Standards (IPDAS) Collaboration's Background Document. 2012.
18. Shapiro J. Illness narratives: reliability, authenticity and the empathic witness. Med Humanit 2011;37(2).
19. Frank AW. The Wounded Storyteller. Body, Illness, and Ethics. London: The University of Chicago Press; 1995.
20. Esquivel A, Meric-Bernstam F, Bernstam EV. Accuracy and self correction of information received from an internet breast cancer list: content analysis. BMJ 2006 Apr 22;332(7547):939-42.
21. van Uden-Kraan CF, Drossaert CHC, Taal E, Lebrun CEI, Drossaers-Bakker KW, Smit WM, et al. Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur? Computers in Human Behavior 2007.

22. van Uden-Kraan CF, Drossaert CHC, Taal E, Smit WM, Seydel ER, van de Laar MAFJ. Experiences and attitudes of Dutch rheumatologists and oncologists with regard to their patients' health-related Internet use. Clin Rheumatol 2010;29:1229-36.
23. Helft PR, Hlubocky F, Daugherty CK. American oncologists' views of internet use by cancer patients: a mail survey of American Society of Clinical Oncology members. J Clin Oncol 2003 Mar 1;21(5):942-7.
24. Broom A. Medical specialists' accounts of the impact of the Internet on the doctor/patient relationship. Health (London) 2005 Jul;9(3):319-38.
25. Wilson P. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet. BMJ 2002 Mar 9;324(7337):598-602.
26. ZegelGezond. Waardering gezondheidsinformatie. 2012.
27. Im EO, Chee W, Lim HJ, Liu Y, Guevara E, Kim KS. Patients' attitudes toward internet cancer support groups. Oncol Nurs Forum 2007 May;34(3):705-12.
28. van Uden-Kraan CF, Drossaert CH, Taal E, Shaw BR, Seydel ER, van de Laar MA. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-17.
29. Hoybye MT, Johansen C, Tjornhoj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. Psychooncology 2005 Mar;14(3):211-20.
30. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003 May;21(3):136-42.
31. Frost JH, Massagli MP. Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data. J Med Internet Res 2008;10(3):e15.
32. Chou WYS, Hunt Y, Folkers A, Augustson E. Cancer Survivorship in the Age of YouTube and Social Media: A Narrative Analysis. Journal of Medical Internet Research 2011;13(1):e7.
33. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Research Notes 2012;5:699.
34. Farmer AD, Bruckner-Holt CEM, Cook MJ, Hearing SD. Social networking sites: a novel portal for communication. Postgrad Med J 2009;85:455-9.
35. Chou WYS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social Media Use in the United States: Implications for Health Communication. J Med Internet Res 2009;11(4):e48.
36. De la Torre-Diez I, Diaz-Pernas FJ, Anton-Rodriguez M. A Content Analysis of Chronic Diseases Social Groups on Facebook and Twitter. Telemedicine and e-Health 2012;18(6):404-8.
37. Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups. J Med Internet Res 2011;13(1):e16.
38. http://www.trimbos.nl/nieuws/trimbos-nieuws/eerste-app-mentale-ondersteuning-van-vrouwen-met-borstkanker?pagnr=2. 2012.
39. www.borstborst.nl. 2013. 1-3-2013.
40. KWF Kankerbestrijding [Dutch Cancer Society]. <http://www.kankerbestrijding.nl> . 2008. 4-6-2008.
41. Bennenbroek F. Social comparison and coping with radiation therapy: The significance of different dimensions of comparison. University of Groningen; 2003.
42. Bellizzi KM, Blank TO, Oakes CE. Social comparison processes in autobiographies of adult cancer survivors. J Health Psychol 2006 Sep;11(5):777-86.
43. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health

care education. Health Information & Libraries Journal 2007;24(1):2-23.

44. Molenaar S, Sprangers MA, Rutgers EJ, Luiten EJ, Mulder J, Bossuyt PM, et al. Decision support for patients with early-stage breast cancer: effects of an interactive breast cancer CDROM on treatment decision, satisfaction, and quality of life. J Clin Oncol 2001 Mar 15;19(6):1676-87.
45. Khangura S, Bennett C, Stacey D, O'Connor AM. Personal stories in publicly available patient decision aids. Patient Educ Couns 2008;73(3):456-64.
46. Giordano SH, Buzdar AU, Hortobagyi GN. Breast cancer in men. Ann Intern Med 2002;137(8):678-87.
47. Lasker JN, Sogolow ED, Sharim RR. The role of an online community for people with a rare disease: content analysis of messages posted on a primary biliary cirrhosis mailinglist. J Med Internet Res 2005 Mar 31;7(1):e10.
48. http://www.guardian.co.uk/lifeandstyle/2012/jul/01/facebook-online-cancer-privacy-family?fb=native&CMP=FBCNETTXT9038. 2012.

SUMMARY

The primary aim of this thesis is to gain insight into which search facilities for spontaneously published stories facilitate breast cancer patients in finding stories by other patients in a similar situation. Spontaneously published stories are defined as first-person texts written by breast cancer patients in which these patients describe their experiences with the illness trajectory, their feelings about having breast cancer, and how they cope with the disease, and which are published at a given moment on the Internet as a ‘completed’ story (i.e. without later additions and not as part of an interactive forum). The research was conducted from a clinical informatics perspective, using knowledge from the fields of information retrieval, human-computer interaction and psychology.

The rationale behind the study is that spontaneously published breast cancer stories on the Internet are a unique source of support and information for other breast cancer patients, since these stories are mostly complete stories, the writers are not selected, the stories are not edited, and the number of these stories is large and continuously increasing. According to the narrative approach, social comparison theory, and social cognitive theory, reading stories about similar others may have the most positive impact. Breast cancer was chosen because of its relatively high prevalence and its impact on physical, psychological, and social levels.

The research followed a user-centred design: users of search facilities (i.e., patients who want to read stories written by others) were involved in every phase of the research so that the search facilities could be tailored to these users’ needs and preferences instead of forcing users to change their search behaviour to conform to search facilities that were developed without their involvement. Concretely, we examined which breast cancer patients spontaneously publish their story online and which themes they write about in their stories (*Chapter 3*), and how these stories are presented and disclosed on websites (*Chapter 2* and its *addendum* containing an update conducted in 2013). We then examined which topics and writer characteristics patients prefer to search for (*Chapter 4*) and translated these preferences into search facilities. Finally, we examined which search facilities show the most positive effects on patient satisfaction and search success, and how patients actually search for stories written by others (*Chapters 5-7*).

Spontaneously published breast cancer stories on the Internet

The study in the addendum to *Chapter 2* shows that on six out of the sixteen websites the stories had been edited by an editor, and that only for three websites could it be confirmed that part of the stories had actually been written or recounted by real patients. Whether stories are reviewed by an editor or written/told by real patients might affect the reliability and authenticity of the stories. However, story reliability and authenticity are difficult to define and the question is how these aspects affect readers of stories. Furthermore, the addendum to Chapter 2 shows that only two websites offer a search facility based on writer characteristics (for example: age, treatment undergone, etc.) and none of the examined websites offer a search facility for story topics based on content keywords.

The results of the study in *Chapter 3* point to a spontaneous selection of both breast cancer patients who publish their story on the Internet and the themes about which they write: the patients and the themes do not exactly reflect the larger population of breast cancer patients and the prevalence of problems they are facing. This result raises questions about the representativeness of a set of spontaneously published stories and the importance of representativeness for the reader.

When the results of *Chapters 2 and 3* are considered together, the question that arises is how a reader can be best helped in easily finding stories by other patients in a similar situation and which may therefore offer the most support and information. To answer this question, *Chapters 4-7* examine several aspects of search facilities for breast cancer stories.

Search facilities for stories

The interviews with breast cancer patients in *Chapter 4* showed that these patients preferred to read about the following themes in other patients' stories: coping with emotions, the impact of cancer on everyday life, and physical discomforts. Patients wanted to be able to select stories according to disease and demographic writer characteristics. The three most preferred writer characteristics were treatment undergone, age and presence of metastases. Most patients preferred part of a story about a specific theme to be presented in text format. Some of them wanted to be able to click on the complete story or on sound and/or video clips.

Chapter 5 describes an online randomised controlled experiment in which we implemented two of the user requirements found in Chapter 4 were implemented: a search facility for story topics and a search facility for writer characteristics. The experiment followed a 2x2 factorial design. The two factors were: searching for story topics 'yes/no' and searching for writer characteristics 'yes/no'. This resulted in four groups: 1) control group (no search facility); 2) search facility for story topics; 3) search facility for writer characteristics; and 4) combination group (a search facility for story topics as well as a search facility for writer characteristics). The results showed that, compared to those patients who were not able to search for story topics, breast cancer patients who were able to search for topics in the stories scored more positively in terms of satisfaction with the search process, satisfaction with the stories retrieved and self-reported impact of the stories on coping with breast cancer. Searching for writer characteristics resulted in fewer positive effects and only on two of the three outcome measures (satisfaction with the search process and satisfaction with the stories retrieved). Moreover, combining the two search facilities barely amplified the effect. Based on the results of Chapter 5, we can conclude that a search facility for story topics is the most favourable way to disclose breast cancer stories.

The analysis of the log data from this experiment (*Chapter 6*) yielded a contradictory result: the search facility for writer characteristics was used more intensively than the search facility for story topics. At least two thirds of the participants searched for each of the writer characteristics at least once in their search session. When participants searched for writer characteristics, they tended to search for characteristics that matched their own (social comparison with similar others). For example, in all the searches for 'breast conserving therapy' and in 96% of the searches for 'mastectomy', the participants themselves had undergone these treatments, and in 79% of the searches for 'age at diagnosis' participants searched for their own age at diagnosis. Per search, a mean number of 7 writer characteristics were searched for (out of a maximum number of 13 writer characteristics). With respect to topics, participants searched for an average of 4 topics per search (out of a maximum number of 36 topics). However, for most of the searches carried out, one or two topics were searched for.

Thus, it would seem that both searching for topics and searching for writer characteristics are valuable to the reader (*Chapters 5 and 6*). One relatively time- and cost-effective search facility that would enable simultaneous searches for both writer characteristics mentioned in stories and topics covered by the stories is Latent Semantic Indexing (LSI). This mathematical technique positions a set of stories in a multidimensional space. The position of a story in the space is determined by how often certain words appear in the story and in how many stories out of a set of stories these words appear. The user types a search query in natural language that is also positioned in the multidimensional space: the distance between the search query and a story provides information about the degree of similarity between them. *Chapter 7* reports on an exploratory study in which 24 breast cancer patients searched for stories using the LSI search facility. These patients reported being quite satisfied with the features offered by the LSI facility (typing natural language queries, refining queries after accessing stories). However, the facility was associated with certain difficulties: fine-tuning the multidimensional space took more time and effort than anticipated, and one participant indicated that she saw no relationship between her search query and the stories retrieved by LSI. In conclusion, more research is needed before LSI can be used as a search facility for finding relevant stories.

Conclusion and discussion

The Internet has a role in sharing illness stories: there are patients who spontaneously publish their story on the Internet and there are other patients who read these stories. Such spontaneously published stories on the Internet provide both opportunities and risks.

Opportunities lie in the possibility that stories may serve functional purpose, for both writers and readers, such as offering emotional support or models by which to solve problems. As a result, both writing and reading stories on the Internet should not be regulated: patients who feel the need should be able to publish their story or read stories written by others. The opportunities that online stories offer can be increased by implementing relevant search facilities to search for stories. For example, participants indicated that stories that were retrieved with a 'story topics' search facility had a positive impact on coping with the disease. Stories that are insufficiently disclosed, for example due to a lack of search facilities, will have no impact on a reader.

Possible risks regarding spontaneously published stories on the Internet relate to the representativeness, authenticity and reliability of stories. These concepts are difficult to define with respect to spontaneously published illness stories. However, they are important since they may affect readers. Therefore, readers should be empowered in evaluating spontaneously published stories. This could be done by assigning quality labels to websites containing patient stories or by offering a checklist with criteria that patients can use to assess the quality of such websites.

Besides the 'completed' stories examined in this thesis, there are many new developments on the Internet through which patients can share their illness experiences, such as Twitter and Facebook. Since there will always be patients who want to publish their whole story at once, without interaction with others and without their story being moderated, our expectation is that 'completed' stories will coexist alongside these new developments. However, future research could focus on how features of these new forms (for example, degree of interactivity) affect writers and readers.

It is important that the parties involved (such as patients, organisations that provide websites with stories, health care professionals and patient organisations) discuss the opportunities and risks of spontaneously published stories with each other and that they discuss each others' roles and responsibilities in this context. An effective discussion may help to increase the opportunities and reduce the risks of online patient stories and experiences.

SAMENVATTING

Het doel van dit proefschrift is inzicht te krijgen in welke zoekfaciliteiten naar spontaan geplaatste borstkankerverhalen op het internet borstkankerpatiënten ondersteunen in het vinden van verhalen van andere patienten die zich in een gelijke situatie als zijzelf bevinden. Spontaan geplaatste verhalen zijn gedefinieerd als teksten geschreven door borstkankerpatiënten in de eerste persoon waarin deze patiënten de eigen ervaringen, gevoelens en de manier waarop zij omgaan met het hebben van borstkanker beschrijven en die zij op een bepaald moment als 'compleet' verhaal op het internet hebben geplaatst (d.w.z. zonder toevoegingen op latere tijdstippen en niet als onderdeel van een interactief forum). Het onderzoek is uitgevoerd vanuit een klinisch informatiekundig perspectief met gebruikmaking van kennis uit de vakgebieden information retrieval, mens-machine interactie en psychologie.

De rationale achter het onderzoek is dat spontaan geplaatste borstkankerverhalen op het internet een unieke bron van steun en informatie zijn voor andere borstkankerpatiënten, omdat deze verhalen meestal complete verhalen zijn, de schrijfsters en inhoud van de verhalen niet gestuurd of geredigeerd zijn en dat het om een grote hoeveelheid verhalen gaat die continu in aantal toeneemt. Volgens de narratieve benadering, sociale vergelijkingstheorie en de sociaal cognitieve theorie zal het lezen van verhalen geschreven door gelijke anderen de meest positieve invloed hebben. Borstkanker is gekozen, omdat borstkanker relatief veel voorkomt en een grote invloed heeft op fysiek, psychologisch en sociaal vlak.

Het onderzoek is uitgevoerd volgens een *user-centred design*. Dit houdt in dat gebruikers van zoekfaciliteiten (dit zijn patiënten die verhalen van anderen willen lezen) zijn betrokken in elke fase van het onderzoek, zodat de uiteindelijke zoekfaciliteiten aan zullen sluiten bij hun wensen en behoeften in plaats van dat gebruikers gedwongen worden hun zoekgedrag aan te passen aan zoekfaciliteiten die ontwikkeld zijn zonder hun betrokkenheid. Concreet hebben we onderzocht welke borstkankerpatiënten spontaan hun verhaal online plaatsen en over welke onderwerpen zij schrijven in hun verhalen (*hoofdstuk 3*) en hoe deze verhalen gepresenteerd en ontsloten worden op websites (*hoofdstuk 2* en het bijbehorende *addendum* dat een update bevat van de situatie in 2013). Vervolgens is onderzocht op welke onderwerpen en schrijfsterkenmerken patiënten willen zoeken (*hoofdstuk 4*) en zijn deze patiëntenvoorkeuren vertaald in zoekfaciliteiten. Tot slot hebben we borstkankerpatiënten deze zoekfaciliteiten laten gebruiken en onderzocht welke zoekfaciliteiten de meest positieve effecten hadden op tevredenheid met het zoekproces, tevredenheid met de gevonden verhalen en zelfgerapporteerde impact van de verhalen op het omgaan met borstkanker (*hoofdstuk 5*). Ook is onderzocht hoe patiënten daadwerkelijk zoeken naar verhalen van andere patiënten (*hoofdstuk 6 en 7*).

Spontaan geplaatste borstkankerverhalen op het internet

Het onderzoek in het addendum van *hoofdstuk 2* laat zien dat de verhalen op 6 van de 16 websites (38%) geredigeerd worden door een eindredacteur en dat slechts van drie websites met zekerheid gesteld kan worden dat de verhalen geschreven zijn door echte patiënten. Redactie door een eindredacteur en echtheid van patiënten zegt mogelijk iets over betrouwbaarheid en authenticiteit van de verhalen. Echter, deze aspecten zijn moeilijk te definiëren en de vraag rijst hoe belangrijk deze aspecten zijn voor de lezer van verhalen. Daarnaast laat het addendum van *hoofdstuk 2* zien dat slechts twee websites een zoekfaciliteit op schrijfsterkenmerken (leeftijd, ondergane behandeling, etc.) bieden en dat geen van de websites een zoekfaciliteit op onderwerpen (gebaseerd op betekenis) in verhalen biedt.

In *hoofdstuk 3* bleek dat een set spontaan geplaatste verhalen een selectie kent zowel wat betreft de patiënten die hun verhaal plaatsen als wat betreft de onderwerpen waarover ze schrijven. Deze vormen geen precieze afspiegeling van de grotere populatie van borstkankerpatiënten en de prevalentie van problemen waar zij mee te maken hebben. Dit resultaat roept vragen op over de representativiteit van een set spontaan geplaatste verhalen en het belang van deze representativiteit voor de lezer.

De vraag die *hoofdstukken 2* en *3* samen oproepen, is hoe een lezer ondersteund kan worden in het relatief snel en makkelijk vinden van verhalen van andere patiënten die zich in een gelijke situatie bevinden als de lezer en vanwege die reden de meeste steun en informatie zouden kunnen bieden. Om deze vraag te beantwoorden zijn in *hoofdstukken 4-7* van dit proefschrift diverse aspecten van zoekfaciliteiten onderzocht.

Zoekfaciliteiten naar verhalen

Uit de interviews met borstkankerpatiënten in *hoofdstuk 4* kwam naar voren dat zij in verhalen van anderen bij voorkeur willen lezen over de volgende onderwerpen: omgaan met emoties, de invloed van kanker op het dagelijks leven en lichamelijke ongemakken. Patiënten wilden de verhalen kunnen selecteren op ziekte- en persoonkenmerken van schrijfsters. De drie belangrijkste kenmerken waarop patiënten verhalen wilden zoeken, waren: ondergane behandeling van de schrijfster, leeftijd van de schrijfster en of de schrijfster al dan niet uitzaaiingen had. Patiënten zagen een gevonden onderwerp in een verhaal het liefst weergegeven als een tekstfragment. Sommigen wilden de mogelijkheid door te klikken naar het complete verhaal of naar geluid- en/of videofragmenten. *Hoofdstuk 5* beschrijft een online gerandomiseerd gecontroleerd experiment waarin twee van de *user requirements* uit hoofdstuk 4 geïmplementeerd zijn, te weten: zoekfaciliteit op onderwerp en zoekfaciliteit op kenmerken van de schrijfsters. Het experiment bestond uit een 2x2 factoriële onderzoeksopzet. De twee factoren waren: wel/niet zoeken op onderwerpen in verhalen en wel/niet zoeken op schrijfsterkenmerken. Dit resulteerde in vier groepen: 1) controlegroep (geen zoekmogelijkheid), 2) zoekfaciliteit op onderwerpen in verhalen, 3) zoekfaciliteit op schrijfsterkenmerken en 4) combinatiegroep (zoekfaciliteit op onderwerpen in verhalen én schrijfsterkenmerken). De resultaten laten zien dat, vergeleken met patiënten die niet konden zoeken op onderwerpen, patiënten die konden zoeken op onderwerpen in de verhalen duidelijk gunstiger scoorden op tevredenheid met het zoekproces, op tevredenheid met (de informatie in) de gevonden verhalen en op zelfgerapporteerde impact van de verhalen op het omgaan met kanker. Zoeken op schrijfsterkenmerken bleek minder effecten te hebben en op slechts twee van de drie uitkomstmaten (tevredenheid met het zoekproces en tevredenheid met (de informatie in) de gevonden verhalen). De twee zoekfaciliteiten gecombineerd hadden geen versterkend effect op elkaar. Een zoekfaciliteit op onderwerpen in verhalen lijkt dus de meest aangewezen manier om borstkankerverhalen te ontsluiten voor patiënten.

De analyse van de log-files van het experiment (*hoofdstuk 6*) liet een tegengesteld resultaat zien: de zoekfaciliteit naar schrijfsterkenmerken werd intensiever gebruikt dan de mogelijkheid om te zoeken op onderwerp. Elk van de schrijfsterkenmerken waarop gezocht kon worden, is door minstens twee derde van de deelnemers minimaal één keer in hun zoeksessie gezocht. Als deelnemers op schrijfsterkenmerken zochten, zochten ze meestal naar kenmerken die ze zelf ook hadden (sociale vergelijking met gelijke anderen). Zo hadden bijvoorbeeld bij alle zoekacties naar borstsparende operatie en bij 96% van de zoekacties naar borstamputatie deelnemers deze behandeling zelf ook ondergaan en werd in 79% van de zoekacties op leeftijd tijdens diagnose door de deelnemers gezocht op hun eigen leeftijd tijdens de diagnose. Gemiddeld werd per zoekactie op 7 schrijfsterkenmerken gezocht (van de maximaal 13 schrijfsterkenmerken waarop gezocht kon worden). Bij de onderwerpen werd gemiddeld op 4 onderwerpen gezocht in een zoekactie (van de maximaal 36 onderwerpen waarop gezocht kon worden). Echter, in de meeste zoekacties werd gezocht op één of twee onderwerpen.

Zowel het kunnen zoeken op onderwerpen als het kunnen zoeken op schrijfsterkenmerken lijkt dus waardevol voor de lezer (*hoofdstuk 5* en *6*). Een zoekfaciliteit waarmee zowel naar kenmerken van de schrijfster die in het verhaal genoemd worden als naar onderwerpen die in het verhaal aan bod komen gezocht kan worden en die relatief weinig geld en tijd kost om te bouwen omdat geen handmatige codering nodig is, betreft *Latente Semantische Indexering* (LSI). Dit is een wiskundige techniek, waarbij een set van verhalen in een multidimensionale ruimte geplaatst wordt. De plaats van een verhaal in de ruimte wordt bepaald door hoe vaak bepaalde woorden in een verhaal voorkomen en in hoeveel verhalen van de set deze woorden voorkomen. De gebruiker kan in natuurlijke taal een zoekvraag typen, die ook in de multidimensionale ruimte geplaatst wordt: de afstand in deze ruimte tussen de zoekvraag en een verhaal geeft een indicatie voor de mate waarin zoekvraag en verhaal op elkaar lijken. *Hoofdstuk 7* beschrijft een exploratieve studie waarin 24 borstkankerpatiënten naar verhalen van andere patienten zoeken met de LSI-zoekfaciliteit. De resultaten laten zien dat zij de mogelijkheden die de LSI-faciliteit bood, gebruikten (doortypen aan een zoekvraag na openen verhalen, typen van de zoekvraag in natuurlijke taal) en redelijk tevreden waren. Echter, het opbouwen van de multidimensionale ruimte kostte meer tijd en energie dan voorzien en één deelnemer gaf aan geen verband te zien tussen haar zoekvraag en de gevonden verhalen. Om LSI te kunnen gebruiken bij het zoeken naar verhalen is dus meer onderzoek nodig.

Conclusie en discussie

Het internet heeft een rol in het uitwisselen van verhalen: sommige patiënten plaatsen spontaan hun verhaal op het internet en andere patiënten lezen deze verhalen. Aan spontaan geplaatste verhalen op het internet zijn kansen en risico’s verbonden.

De kansen zijn dat zowel het schrijven/plaatsen als het lezen van verhalen bepaalde functies dient, zoals het bieden van emotionele steun en van voorbeelden hoe men problemen op kan lossen. Vanwege deze functies is het van belang dat zowel het schrijven als het lezen van verhalen op het internet niet gereguleerd wordt: elke patiënt die daar behoefte aan heeft moet het eigen verhaal kunnen plaatsen of dat van anderen kunnen lezen. De kansen die verhalen geven, kunnen vergroot worden door relevante zoekmogelijkheden naar verhalen aan te bieden. Van verhalen die gevonden zijn met een zoekfaciliteit op onderwerpen geven patiënten aan dat deze positieve invloed hebben op hoe zij omgaan met kanker. Verhalen die onvoldoende ontsloten worden, bijvoorbeeld door gebrek aan zoekfaciliteiten, kunnen geen effect hebben op de lezer.

Mogelijke risico’s bij spontaan geplaatste verhalen houden verband met representativiteit, authenticiteit en betrouwbaarheid van verhalen. Deze begrippen zijn moeilijk definieerbaar als het om verhalen gaat, maar wel van belang vanwege de invloed die de verhalen op de lezer kunnen hebben. Lezers van spontaan geplaatste verhalen dienen daarom ondersteund te worden in het beoordelen van spontaan geplaatste verhalen. Dit kan bijvoorbeeld door keurmerken op websites met verhalen of door lijsten met criteria waarmee lezers zelf de kwaliteit van een website met verhalen kunnen inschatten.

Naast de ‘complete’ verhalen zijn er op internet veel nieuwe vormen in opkomst voor het uitwisselen van ervaringen tussen patiënten, zoals Twitter en Facebook. De verwachting is dat ‘complete’ verhalen altijd zullen blijven bestaan, omdat er patiënten zijn die er voorkeur aan geven hun hele verhaal in één keer te plaatsen, zonder interactie met anderen en zonder dat het verhaal gemodereerd wordt. Toekomstig onderzoek zou kunnen nagaan wat de eigenschappen van de verschillende vormen om ervaringen uit te wisselen (bijvoorbeeld mate van interactiviteit) voor effect hebben op schrijfsters en lezers.

Het is belangrijk dat betrokken partijen (zoals patiënten, organisaties die websites met verhalen bieden, artsen en patiëntenorganisaties) een open discussie met elkaar voeren over kansen en risico’s van spontaan geplaatste patiëntenervaringen en elkaars rollen en verantwoordelijkheid hierbij. Een vruchtbare discussie kan bijdragen aan het vergroten van de kansen en het verkleinen van de risico’s van online patiëntenervaringen.

APPENDICES

Appendices Chapter 5

Please view Multimedia Appendix 1 and Multimedia Appendix 2 on the Internet. You can find these at:
http://www.jmir.org/article/downloadSuppFile/1215/1405
http://www.jmir.org/article/downloadSuppFile/1215/885

Multimedia Appendix 3 (http://www.jmir.org/article/downloadSuppFile/1215/1406):
An overview of the constructs and items belonging to the three main outcome measures.

Constructs: 1. 'Satisfaction with the search process'
1.a. Opinion about the search facility (Cronbach's alpha = .88):
Five 5-point semantic differentials:

- poor … good
- difficult … easy
- not at all helpful … very helpful
- unpleasant … pleasant
- not convenient … convenient

1.b. Opinion about the number of search options (Cronbach's alpha = NA):
Response categories: 'too few', 'quite few', 'not few, not many', 'quite many', 'too many'

- What is your opinion on the number of search options?

(For analysis, this item was recoded into 3 points 1=negative opinion ('too few' or 'too many'), 2=slightly negative opinion ('quite few' or 'quite many'), and 3= positive opinion ('not few, not many')

1.c. The extent to which the search options enable finding the information one was looking for (Cronbach's alpha = .75):
Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'.
The ways in which I could search…

- …did *not* enable me to find what I wanted to know.
- …did enable me to find information that is important to me.
- …did *not* enable me to find stories that were helpful to me.
- …did enable me to find what I was searching for.

1.d. Recommendation to others' and future own use (Cronbach's alpha = .82):
Response categories: 'yes', 'probably yes', 'maybe yes, maybe not', 'probably not', 'no'.

- Would you recommend others to search for stories in this way?
- Would you yourself search more often for stories in this way?

1.e. Overall satisfaction with the search facility (Cronbach's alpha = NA):
One 10-point semantic differential:

- very poor ... excellent

Constructs: 2. 'Satisfaction with the stories retrieved'
2.a. Opinion about the stories retrieved (Cronbach's alpha = .71):
Six 5-point semantic differentials:

- poor ... good
- difficult ... easy
- not at all helpful ... very helpful
- not informative ... informative
- frightening ... not frightening
- I already knew everything ... everything was knew to me

2.b. Opinion about the number of stories retrieved (Cronbach's alpha = NA):
Response categories: 'too few', 'quite few', 'not few, not many', 'quite many', 'too many'.

- What is your opinion on the number of stories retrieved?

(For analysis, this item was recoded into 3 points 1=negative opinion ('too few' or 'too many'), 2=slightly negative opinion ('quite few' or 'quite many'), and 3= positive opinion ('not few, not many')

2.c. Opinion about the list of stories displayed after a search (Cronbach's alpha = .76):
Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'.

- It was clear to me which story from the list I had to click onto to find the information I was looking for.
- The list of retrieved stories displayed after a search, was *not* helpful to me.
- It was difficult for me to decide which story I should read first.
- The way in which the list of retrieved stories was displayed, was clear to me.

2.d. The extent to which the stories retrieved covered one's information need (Cronbach's alpha = .82):
Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'.

- I have found the information I was looking for.
- I am *not* satisfied with the information I have found.
- The information satisfied my expectations.
- I did *not* find the information I wanted to know.

2.e. Recommendation to others' and future own reading (Cronbach's alpha = .77):
Response categories: 'yes', 'probably yes', 'maybe yes, maybe not', 'probably not', 'no'.

- Would you recommend others to read these stories?
- Would you yourself read more of these kinds of stories?

2.f. Overall satisfaction with the stories retrieved (Cronbach's alpha = NA):
One 10-point semantic differential:

- very poor … excellent

Constructs: 3. 'The stories' impact on coping with breast cancer'

3.a. The stories' impact on coping with breast cancer (Cronbach's alpha = .85):
Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'.

Having read the stories:

- …I have learnt things.
- …I view things differently.
- …I am able to understand my feelings better.
- …I see that certain emotions accompany learning to live with breast cancer.
- …I know what to do.
- …I see that others have experienced the same things.

Appendices Chapter 7

Appendix I

Composition of the questions of which the results are presented in this article. The questions come from a questionnaire about search satisfaction that was used in a previous study [1].

- Opinion about the search facility (Cronbach's alpha = .88).
  Five 5-point semantic differentials:
  - poor … good
  - difficult … easy
  - not at all helpful … very helpful
  - unpleasant … pleasant
  - not convenient … convenient

- Opinion about the list of stories displayed after a search (Cronbach's alpha = .76).
  Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'.
  - It was clear to me which story from the list I had to click onto to find the information I was looking for.
  - The list of retrieved stories displayed after a search, was not helpful to me.
  - It was difficult for me to decide which story I should read first.
  - The way in which the list of retrieved stories was displayed, was clear to me.

- The extent to which the search options enable finding the information one was looking for (Cronbach's alpha = .75). Response categories: 'disagree', 'disagree a bit', 'neither disagree nor agree', 'agree a bit', 'agree'. The ways in which I could search…
    - o …did not enable me to find what I wanted to know.
    - o …did enable me to find information that is important to me.
    - o …did not enable me to find stories that were helpful to me.
    - o …did enable me to find what I was searching for.

- Recommendation to others' and future own use (Cronbach's alpha = .82). Response categories: 'yes', 'probably yes', 'maybe yes, maybe not', 'probably not', 'no'.
    - o Would you recommend others to search for stories in this way?
    - o Would you yourself search more often for stories in this way?

- Overall satisfaction with the search facility (Cronbach's alpha = NA). One 10-point semantic differential:
    - o very poor … excellent

Appendix II

Latent Semantic Indexing [2] was performed on the set of 171 Amazones stories. The ORACLE Dutch stop words list was used to remove stop words [3]. An exception was made for the Dutch word 'haar' meaning 'her', because this word also means 'hair' which is an important word in the context of cancer. The stemming algorithm "Dutch Porter for the UPLIFT project" was used to stem the words [4]. Word stems shorter than three characters were removed. Diacritics were reduced to non-diacritics. TF/IDF was calculated on the basis of the 6241 terms that remained in the 171 stories. The weighted word-by-story matrix consisted of 34343 non-zero values (cutoff=0.0002) and was thus filled for 3% (=34343 /(171 * 6241)). With Singular Value Decomposition (SVD) the number of dimensions was reduced to 12 (eigenvalues greater than 0.32). A quadratic vector normalisation and cosine similarity between stories were used.

References

1. Overberg R, Otten W, de MA, Toussaint P, Westenbrink J, Zwetsloot-Schonk B. How breast cancer patients want to search for and retrieve information from stories of other patients on the internet: an online randomized controlled experiment. J Med Internet Res 2010;12(1):e7.
2. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 1990;41(6):391-407.
3. ORACLE Dutch stop words list. ORACLE Dutch stop words list 2011.
4. Kraaij W, Pohlmann RC. Dutch Porter for the UPLIFT project. Research Institute for Language and Speech, Utrecht University; 1994.

CURRICULUM VITAE

Regina Overberg werd geboren op 24 augustus 1978 te Stadskanaal. In 1997 behaalde zij het eindexamen gymnasium B aan het Alexander Hegius Lyceum te Deventer. In datzelfde jaar begon zij haar studie Voeding en Gezondheid aan de Wageningen Universiteit. Deze studie heeft zij afgerond met twee afstudeeronderzoeken: één op het gebied van Humane epidemiologie bij de vakgroep Humane Voeding en Epidemiologie van de Wageningen Universiteit, en één op het gebied van Leefwijze en volksgezondheid bij het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) te Bilthoven. Na haar afstuderen in 2002 heeft Regina een jaar bij het Centrum voor Volksgezondheid Toekomstverkenningen (cVTV) van het RIVM gewerkt. Het promotieonderzoek dat in dit proefschrift beschreven is heeft zij in de periode 2003-2008 bij de groep Klinische Informatiekunde van het Leids Universitair Medisch Centrum (LUMC) uitgevoerd. Sinds 2008 voert zij epidemiologisch onderzoek uit binnen het team Advies & Onderzoek Publieke Gezondheid van GGD Kennemerland te Hoofddorp. Meer informatie is te vinden op: http://nl.linkedin.com/in/reginaoverberg

Regina Overberg was born on the 24th of August 1978 in Stadskanaal, the Netherlands. In 1997 she completed her secondary education at Alexander Hegius Lyceum in Deventer. In the same year she started to study Human Nutrition and Health at the Wageningen University. She graduated in 2002 after having completed two studies: one in the field of human epidemiology at the Department of Human Nutrition and Epidemiology of the Wageningen University, and one in the field of lifestyle and public health at the National Institute for Public Health and the Environment (RIVM) in Bilthoven. After her graduation, Regina worked a year at the Centre for Public Health Forecasting (cVTV) of the RIVM. She performed her PhD research that is described in this dissertation in the period 2003-2008 at the Clinical Informatics group of the Leiden University Medical Centre (LUMC). From 2008, she carries out epidemiological research within the team Advice & Research Public Health at the Municipal Health Service Kennemerland in Hoofddorp. More information is available at: http://nl.linkedin.com/in/reginaoverberg

LIST OF PUBLICATIONS AND PRESENTATIONS*

(in chronological order)

Overberg R, Toussaint P, Zwetsloot-Schonk B. Illness stories on the Internet: Features of websites disclosing breast cancer patients' illness stories in the Dutch language. *Patient Education and Counseling* 2006;61:435-442.

Hoenkamp E, **Overberg R**. Computing latent taxonomies from patients' spontaneous self-disclosure to form compatible support groups. *Studies in Health Technology and Informatics* 2006;124:969-974.

Wolf L, **Overberg R**, Toussaint P, Hoenkamp E, Reckman H. Design of the Narrator system – processing, storing and retrieving medical narrative data. *Journal of Integrated Design & Process Science* 2006;10(4):13-33.

Overberg RI, Campagne AE, Zwetsloot-Schonk JHM. Empowering patients – Use and appreciation of a web application that enables patients to keep their own medical data and to share illness experiences with fellow patients. Poster presentation at the *International Conference on Communication in Healthcare 2006 (EACH)*, 5-8 September 2006, Basle, Switzerland.

Overberg RI, Alpay LL, Verhoef J, Zwetsloot-Schonk JHM. Illness stories on the Internet: What do breast cancer patients want at the end of treatment? *Psycho-Oncology* 2007;16:937-944.

---

* This is a selection related to the content of this dissertation

Alpay LL, **Overberg RI**, Zwetsloot-Schonk JHM. Empowering citizens in assessing health related websites: a driving factor for healthcare governance. *International Journal of Healthcare Technology and Management* 2007;8:141-160.

Overberg R. Search facilities driven by patients' preferences for breast cancer stories on the Internet. Oral presentation at the *International Conference on Communication in Healthcare 2008* (EACH), 2-5 September 2008, Oslo, Norway. As part of the symposium Web-Based Interventions and Information Technology.

Overberg R, Otten W, De Man A, Toussaint P, Westenbrink J, Zwetsloot-Schonk B. How breast cancer patients want to search for and retrieve information from stories of other patients on the Internet: an online randomized controlled experiment. *Journal of Medical Internet Research* 2010;12(1):e7.

Overberg R, De Man A, Wolterbeek R, Otten W, Zwetsloot-Schonk B. Spontaneously published illness stories on a website for young women with breast cancer: do writers and themes reflect the wider population? *Informatics for Health and Social Care* 2013; 38(1): 54-66.

Overberg R, Otten W, De Man A, Zwetsloot-Schonk B. Searching for breast cancer stories online: What topics and/or writer characteristics do patients search for? *Submitted*

DANKWOORD

Allereerst wil ik graag alle patiënten bedanken die aan de onderzoeken in dit proefschrift hebben meegedaan voor hun interesse, tijd en energie. Zonder hen had ik de onderzoeken niet kunnen uitvoeren. De deelnemers die ik heb geïnterviewd wil ik in het bijzonder bedanken voor de persoonlijke verhalen die ze met mij hebben gedeeld. Daarnaast wil ik de Borstkankervereniging Nederland en Stichting Amazones bedanken voor hun betrokkenheid bij mijn onderzoek. Veel dank gaat uit naar Judith Westenbrink en Diana Gosens die toestemming gaven om de website van Stichting Amazones te gebruiken voor het online experiment naar zoekmogelijkheden naar verhalen. Verder wil ik Tineke van Dijk bedanken voor het feit dat ik haar gouache voor de voorzijde van mijn proefschift mocht gebruiken.

Voor de inspirerende gesprekken en overleggen in het kader van mijn onderzoek wil ik graag mijn promotor prof. dr. Bertie Schonk en mijn co-promotor dr. Wilma Otten bedanken. Ook alle mede-auteurs en oud-collega's van Klinische Informatiekunde wil ik bedanken voor hun ideeën en suggesties: Amy Ahluwalia, Laurence Alpay, Nicole Ezendam, Watze Hoekstra, Eduard Hoenkamp, Lieke Lim, Herman Lodder, Pieter Toussaint, Jan Vanderschoot, John Verhoef, Leo Wolf en Ron Wolterbeek. In het bijzonder wil ik Andries de Man bedanken voor het bouwen van de website voor het online experiment en de LSI-zoekfaciliteit en voor het in goede banen leiden van de enorme databestanden die deze onderzoeken opleverden. Ik heb daarnaast veel gehad aan de frisse blik van Henriët Werij die ik begeleid heb bij haar afstudeeronderzoek en van Kristel Schaap die de verhalen heeft gecodeerd. Verder wil ik Henk Herman Nap en Hilke Reckman bedanken voor de samenwerking binnen het Narrator-project van NWO.

Ook buiten het LUMC heb ik met veel mensen ideeën kunnen uitwisselen over de inhoud en voortgang van mijn onderzoek, die ik daar graag voor wil bedanken: de huidige collega's van GGD Kennemerland, in het bijzonder Jolanda ten Brinke, Resi Cluitmans, Saskia Mérelle, Claire Morssink en Arie Oosterlee, en, vrienden en oud-studiegenoten, in het bijzonder, Marije Brouwer, Daphne Cupido, Carolien van Loo-Bouwman, Anouk Pijpe en Bart van Praag. Speciale dank gaat in dit kader uit naar Joost Klaver.

Tot slot wil ik mijn familie bedanken voor hun vertrouwen, hulp en steun: mama en papa, Gisela en Gerhardus, zus, zwager en lief klein nichtje, Ingrid, Frank en Charli. Bedankt dat jullie er altijd voor me zijn.

ISBN 978-90-9027748-6