Understanding Breast Cancer Survivors' Information-Seeking Behaviors and Overall Experiences: A Comparison of Themes Derived from Social Media Posts and Focus Groups

Journal:	*European Health Psychology Society*
Manuscript ID	GPSH-2019-0524.R2
Manuscript Type:	Psychology and Health
Keywords:	Breast Cancer, social support, focus groups, Information-seeking

Authors:

*Corresponding Author
Jennifer Currin-McCulloch, PhD, LSW
Colorado State University School of Social Work
jen.currin-mcculloch@colostate.edu

Amelia Stanton, PhD
Harvard Medical School Department of Psychiatry

Ryan Boyd, PhD
Lancaster University Psychology Department

Margaret Neaves, LMSW
Satellite Healthcare Department of Social Work

Barbara Jones, PhD, MSW, FNAP
University of Texas at Austin Steve Hicks School of Social Work

Title: Understanding Breast Cancer Survivors' Online Information-Seeking Behaviors and Overall Experiences: A Comparison of Themes Derived from Social Media Posts and Focus Groups

Abstract

Objective*:* Using two different analysis techniques, this study explored differences and similarities in information-seeking discourse and overall breast cancer experiences between posters to a Reddit board and breast cancer survivor focus groups.

Design*:* This study incorporates two qualitative methods for determining themes in breast cancer survivors' information-seeking behaviors and overall cancer experiences. First, posts from a breast cancer-specific Reddit community were extracted and analyzed using the meaning extraction method (MEM) to determine core themes. Then, investigators performed a thematic analysis of two focus groups of breast cancer survivors (N = 18). Finally, themes derived from each analysis method were compared.

Main Outcome Measures: Outcome measures include themes extracted from Reddit posts and themes generated from breast cancer survivor focus groups.

Results*:* Findings between qualitative methodologies represent similar yet nuanced themes in survivors' discourse. The MEM resulted in seven themes: diagnosis, treatment process, social support, existentialism, risk, information-seeking, and surgery. Focus groups revealed the same initial four MEM themes plus the following: disclosure, coping, and fears.

Conclusions*:* The MEM is a cost-effective research mechanism for informing common themes of experiences of cancer patients and survivors and may offer initial data to guide psychosocial oncology research design and recruitment.

Key Words: breast cancer; meaning extraction methods; social support; focus groups; information-seeking

Word Count: 4982

Introduction

When diagnosed with breast cancer, women receive an onslaught of information about medical treatments, potential changes to their body, and shifts in their familial, social, and occupational roles (Boyages et al., 2016; Brandão, Schulz, & Matos, 2017; Male, Fergus, & Cullen, 2016). They encounter numerous healthcare providers and well-meaning loved ones who share varying opinions on what types of treatments they should or should not receive (Berger et al., 2017; Laidsaar-Powell et al., 2017). Medical advances increase burdens of decision-making: a multitude of decisions involving surgery, reconstruction, radiation, or promising clinical research must be made. If a woman has a family history of cancer, genetic counseling and testing may also enter into her decision-making algorithm (Wevers et al., 2017). Women often weigh benefits between future health and cosmetic results with impacts of family and careers (Fasse et al., 2017; Swanberg, Nichols, Ko, Tracy, & Vanderpool, 2017; Wallner et al., 2017). Sorting through the possibilities and the potential future impacts of their treatment-related decisions often produces tremendous stress and anxiety (Drageset, Lindstrom, & Underlid, 2009). The complex decision-making process can become overwhelming and increase a woman's proclivity to negative coping (Dragesetet al., 2009; Williams & Jeanetta, 2015).

Breast cancer survivors frequently turn to online forums for decision-making and social support. The number of online breast cancer forums has increased exponentially since 2006, with significantly more posts on breast cancer forums appearing each month (Quinn et al., 2013). Online forums and support groups provide emotional benefits to cancer survivors, including information sharing, receiving support, and decreased isolation through social interaction (Kim et al., 2012).

Groups also promote hope and nurture a sense of altruism (Shaw, McTavish, Hawkins, Gustafson, & Pingree, 2000), and they may offer patients some control over their decisions and the ability to transition from a passive to active role in their care (Quinn et al., 2013). Though the abundance of information about breast cancer in these forums may help women choose among viable treatment options and formulate appropriate questions for their providers, easy access to facts, figures, and personal narratives may also lead to information burden.

In recent years, social scientists have used natural language data, collected through traditional experimental studies as well as from online support groups and social media sites, to study complex psychological phenomena, such as personality (Yarkoni, 2010), core values (Boyd et al., 2015), and sexual self-schemas (Stanton, Boyd, Pulverman, & Meston, 2015). Natural language data offers a number of unique advantages to researchers (see Boyd, 2017 for a thorough review of these advantages). The volume of text that is publicly available on these sites is unprecedented (Tausczik & Pennebaker, 2010), and researchers can easily extract the user-generated data at little to no cost. Other benefits of language data include reliability over time, internal consistency, inherent ecological validity, low cost, and considerable variety between people (Boyd & Pennebaker, 2017). Though there have been concerns about self-selection biases and data quality in online settings (Peek, Holmes, Sun, & 2014; Strasser et al., 2012), research suggests that social media sites can be used effectively as data sources, specifically for data on health-related outcomes (Alshaikh, Ramzan, Rawaf, Majeed, 2014). Text extracted from social media sites may be used as a precursor or even an alternative to laboratory measurement of certain complex phenomena that are difficult to measure (Stanton, Meston, & Boyd, 2017).

One family of methods commonly used to analyze natural language data is broadly known as topic modeling. The Meaning Extraction Method (MEM; Chung & Pennebaker, 2008), a specific type of topic modeling, identifies clusters of words that tend to co-occur across a corpus of text in an automated fashion. These word clusters represent latent components or “themes” that are present within the corpus. Using the MEM, themes in a given body of text can be quantified and analyzed over time and across groups (e.g., by demographic characteristics, between clinical populations and non-clinical populations, comparing individuals randomized to a treatment versus control group, etc.). One strength of the MEM rests in its ability to combine the wide sampling power that is typically associated with quantitative data with the detail and nuance that is characteristic of qualitative data (Stanton et al., 2015).

Though content analyses of online interactions related to breast cancer have been conducted, the MEM has not yet been used to isolate themes of women’s experiences with breast cancer. In a previous content analysis that used qualitative methods to assess breast cancer survivors’ coping mechanisms and physical adaptations, topics in online posts revolved around information seeking, with a focus on diagnosis and pathology as well as on suggestions for symptom management and healthy living (Rubenstein, 2015). Other posts were intended to provide encouragement and empathy (Rubenstein, 2015). Although the Rubenstein paper provides critical information about the ways in which social support and health information are exchanged online, the researchers’ postings and interactions with participants within the online forum may have influenced participants’ posts and study findings. Language analysis techniques like the MEM enable researchers to examine themes across massive corpora of natural language data, which is challenging to accomplish using traditional qualitative methodology. Given the quantity of data that is available online, the MEM may be helpful in extracting information that complements or expands upon the findings of previous qualitative studies.

The current study sought to build upon past work that has demonstrated significant overlap in the language content/themes found in controlled laboratory environments and social media (e.g., Stanton, Meston, & Boyd, 2017). Specifically, we sought to compare the themes of breast cancer survivorship that emerged from social media data with those that emerged via content analysis of focus group data. Interviews and focus groups are the most common methods of collecting data in qualitative healthcare research (Gill, Stewart, Treasure, & Chadwick, 2008). To our knowledge, themes extracted from natural language data via the MEM have yet to be compared to themes gleaned from focus groups, which are useful in generating a nuanced understanding of participants' experiences and beliefs about a certain topic. The current study offers insight on the association between themes of breast cancer survivorship that are expressed on social media and the self-reported experiences that breast cancer survivors share with researchers and/or medical providers.

Methods

Social Media Data

Archival data was collected from the Reddit website. An entertainment, social networking, and news website, Reddit is similar to an online bulletin board and other internet forums; registered users can submit various forms of content (e.g., hyperlinks, personal narratives) to topic-specific discussion boards, which are voted up or down by other users to determine the positions of the posts on each of the site's pages. We extracted data from the *r/breastcancer* subforum, which has over 3,000 members and is described as a "community of support and information for those affected by this disease through their personal struggle or that of a loved one." Each Reddit subforum is structured such that users can either leave 1) an original post (OP) to which other users can reply, or 2) a comment on those posts written by others. We included both OPs and comments in response to the OPs (excluding hyperlinks) to parallel the structure of a focus group, in which participants actively engage with each other in a group context.

The data from the *r/breastcancer* board was extracted using a custom script that pulled archival data from Reddit's application programing interface (API). This Reddit API is publicly accessible and allows researchers to acquire language data directly from the site without using the typical web interface. Unfortunately, Reddit does not collect thorough demographic data on the site's users, so we cannot describe the characteristics of the sample. The data posted through March 2016 were extracted chronologically. Ultimately, the sample included 1051 OPs and comments.

Social media data analysis. We used the MEM to extract the core themes of breast cancer survivorship. The details of the MEM's analytic technique have been described thoroughly elsewhere (Blackburn, Yilmaz, & Boyd, 2018; Boyd, 2017; Chung & Pennebaker, 2008; Pulverman, Boyd, Stanton, & Meston, 2017), but we offer a brief description here. First, all 1051 OPs and comments were analyzed to determine word prominence. The words in the corpus were "lemmatized," or converted to their basic inflections (e.g., "eats", "ate", and "eating" are converted to "eat"), using the Meaning Extraction Helper (Boyd, 2014). Then, common closed-class (i.e., "function") words and uncommon open class (i.e., "content") words were removed. Content words were considered uncommon if they appeared in less than 7.5% of all observations; this percentage falls within the recommended range proposed by Chung and Pennebaker (2008) when they developed the MEM. All remaining content words that are used by a set minimum of participants (5%) were then given a binary score (0 = absent, 1 = present) for each observation. Finally, these scores were submitted to a principle components analysis (PCA), the results of which revealed clusters of words (which we refer to as "themes") that commonly co-occurred across observations.

Focus groups

To compare themes generated by computerized data extraction methods with those described by breast cancer survivors through interpersonal group dialogues, we held two focus groups. We partnered with a local breast cancer support organization to recruit breast cancer survivors for these groups. Table 1 displays sociodemographic and cancer treatment variables from participants in the focus groups. To arrange the initial focus group, the organization emailed women in their database inviting them to participate in the study during the time normally allotted for a survivor support group. To set up the second focus group, the support organization invited their staff of breast cancer navigators to participate, thus enabling us to gather a broader perspective of survivors' experiences through thousands of client interactions. The staff of the support organization were all breast cancer survivors and professionally trained breast cancer navigators. This navigator focus group occurred at an offsite facility to foster a sense of safety for sharing experiences.

To ensure adherence to the research aims and questions, two of the authors (AS and JC) facilitated both group sessions. Prior to the start of the focus groups, each participant completed a one-page demographic survey that asked about their age, race, marital status, education, cancer diagnosis and treatment modalities, and online presence. The first focus group included eleven women and the second group included seven women. Each focus group lasted 90 minutes and drew from an interview guide that included four main topics: (1) brief description of their experiences at diagnosis and during breast cancer treatment; (2) engagement and/or disengagement of internal and external resources during diagnosis, treatment, and recovery; (3) experiences in accessing the internet for breast cancer information or support; (4) discussion of themes that emerged from computerized data extraction methods to discover whether themes matched individual experiences. After analyzing the Reddit data with the MEM and the focus group data with thematic analysis, the coding team compared the themes that were derived from the two methods. The objective of the comparison was not to find whether one method produced superior themes; but rather, to perform a narrative synthesis of the data emerging from each method, highlighting the ways in which themes converged or diverged. Comparison dialogues centered around the emergence of the themes within each data set, as well the as words within each theme (within the Reddit data) that had the highest loadings, noting the context of the dialogues across the two samples.

The interview guide included the same topics for each group; however, the facilitators encouraged the participants in the navigator group to respond from a broader perspective of trends they encounter in caring for breast cancer survivors; however, some of the navigators briefly shared personal experiences.

Focus group data analysis. Following each focus group, one investigator (MN) watched the video-recorded session and transcribed the focus group data into a word document. The data was then transferred into an Excel document for coding. Two of the investigators (AS and JC) performed data analysis utilizing thematic analysis methodology as described by Guest, MacQueen and Namey (2012). Analysis began with reading through each transcription to gather an overall impression of the data. Data was segmented based on interview questions to assist with the identification of descriptive themes related to diagnosis, treatment, social support, and coping processes. Two investigators (AS and JC) coded the first focus group individually then met to discuss coding methods and variance in codes. An analysis of variance between coders rendered (83.4%) agreement in codes from the first focus group and (84.6%) agreement in codes for the second focus group. Both coders discussed differences in codes until they achieved consensus. Another author was available to resolve differences if the lead authors failed to reach consensus.

Following discussion and consensus on codes from the first focus group, the investigators created a codebook. Data analysis for the second focus group followed that of the first focus group. They entered data from the second focus group into the existing codebook, noting new codes that arose within the data. Next, one of the investigators reviewed the master codebook, highlighting coding nuances and commonalities. Both investigators then reviewed codes to discover key themes within the data. While coding the data from the second interview, the researchers discovered that participants' experiences closely matched those of the women in the first focus group. Specifically, no new themes or meaning codes arose within the second focus group to further elaborate the first focus group's core seven themes. Therefore, the researchers determined that they had achieved thematic and meaning saturation and ended participant recruitment. Lastly, they compared final themes from the focus group data with themes from the computer extraction method.

Qualitative Comparison of Themes: After analyzing the Reddit data with the MEM and the focus group data with thematic analysis, the coding team compared the themes that were derived from the two methods. The objective of the comparison was not to find whether one method produced superior themes; but rather, to perform a narrative synthesis of the data emerging from each method, highlighting the ways in which themes converged or diverged. Comparison dialogues centered around the emergence of the themes within each data set, as well the as words within each theme (within the Reddit data) that had the highest loadings, noting the context of the dialogues across the two samples.

Ethics: The investigators received approval from their academic institutional review board prior to initiation of participant recruitment. Each participant reviewed the research protocol, signed the informed consent, and provided permission for video-taping. Participants received a $20 gift card to offset travel expenses. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Results

The presentation of results begins by portraying thematic commonalities between both qualitative methods and then shifts to the unique presentations of themes within each analytic method. Ten distinct themes emerged from the data sets: (1) screening/diagnosis, (2) treatment process, (3) social support, (4) existentialism, (5) disclosure, (6) coping, (7) fears, (8) risk, (9) information-seeking, and (10) surgery. Notably, four themes emerged in both samples and centered on physical and psychosocial experiences during screening, diagnosis, and treatment, as well as the ways in which women found social support and existential meaning throughout their illness.

Screening and diagnosis dialogues centered around discussions of cancer testing and emotionally disturbing diagnosis dialogues with healthcare providers. Women also revealed insurance and financial barriers, as well as challenges in finding medical providers to perform their mammograms and biopsies. The women encountered troubling times in deciding which treatment option(s) would best fit their personal and familial needs. Several women also expressed worries at the time of diagnosis about the potentially negative consequences their illness may have on their families, in particular, their children, regardless of the child’s age. Treatment-related dialogues focused on physical and psychosocial side-effects and the consequences on self and loved ones such as depression, anxiety, fatigue, chemo brain, body image changes, nausea, and hair loss.

Social support appeared similarly for women from Reddit and the focus groups with many turning to family, friends, faith communities, and other breast cancer survivors. Intimate and social relationships experienced some strain due to physical and emotional changes, as well as communication barriers. Social support appeared differently for survivors based on the communities in which they identified. For example, women in the Latina/Latinx communities had extensive social support systems, while older women struggled to find transportation and caregivers to provide physical care following surgeries and chemotherapy or radiation treatments. Discussions emerged similarly around role adaptations at work and other social settings. Resoundingly, social support arose as instrumental in fostering survivors' motivation, emotional well-being, and management of physical needs.

Although cancer presented numerous challenges for posters to the Reddit forum and women in the focus groups, many participants found existential benefits from their diagnosis. They described new-found personal strength and clarity in the people and activities in their life that brought joy. The "silver linings" of cancer appeared as acknowledgment of gratitude and a reprioritizing of relationships and wellness activities that might reduce stress and risks for cancer metastasis. Survivors also explored the deeper, existential sides of themselves including their spirituality and higher purpose, as well as creative outlets for self-expression through writing and art.

Social Media Results

The MEM analysis resulted in a model with seven distinct components. Seven theme labels were chosen by the experimenters to describe the co-occurring words that were associated with each of the seven distinct themes: (1) diagnosis, (2) social support, (3) risk, (4) existentialism, (5) treatment process, (6) information-seeking, and (7) surgery (see Table 2 for MEM themes and loading values). Each observation was given a score for each theme by summing the total percentage of theme-relevant words in each observation. Table 3 displays quoted text from observations with high scores for each of the seven themes extracted from all Reddit posts, as well as corresponding text from focus group participants.

The themes that were selected for inclusion had Eigenvalues above 4.25 and significantly contributed to the total variance of the model. Words that were retained within each of the seven themes had loadings of .25 or higher. Overall, the seven identified themes accounted for 16.13% of the total variance. This percentage may seem small in the context of traditional factor analysis, but it falls well within the range of variances that have been reported in other studies that applied principle components analyses to natural language data (Chung & Pennebaker, 2008; Stanton et al., 2015; Wolf, Chung, & Kordy, 2010).

Of the seven themes that emerged from the Reddit data, three (risk, information seeking, and surgery) were unique to the online context. The first unique theme, risk, captured descriptions that discussed genetic and other health risks associated with increased likelihood of a breast cancer diagnosis. Posts that had high scores on the risk theme highlighted the link between the BRCA gene, inherited risk, and family histories of cancer. Other posts discussed the importance of genetic counseling and the ways in which counseling can improve coping for those who carry high genetic risk. Posts that scored highly on information seeking, the second unique theme extracted from the Reddit data, addressed strategies for finding accurate information about breast cancer and associated treatments, sometimes providing specific resources for readers. In addition, some posts with high scores on the information seeking theme offered support on each stage of the diagnosis and treatment process, noting specific resources that apply to certain phases but not others. Finally, posts that scored highly on the last unique theme to emerge from the Reddit data, surgery, described the various surgical treatment options and offered personal anecdotes about these options, sometimes delving into the emotional tolls of each procedure. Other posts with high scores on the surgery theme described not only the initial surgeries (e.g., lumpectomies, mastectomies) but also the subsequent surgeries (e.g., reconstruction, surgeries to remove defective implants) and the ways in which theses surgeries were associated with hope for recovery and pain, both physical and emotional.

Focus Group Results

Seven main themes arose from the focus group data; the first four listed were the same as the MEM-derived themes and the final three represent themes that were unique to the focus groups. The themes from the groups included (1) screening/diagnosis, (2) treatment, (3) social support, (4) existentialism, (5) disclosure, (6) coping, and (7) fears. Table 4 displays the focus group themes and codes within these themes.

The first unique theme from the focus groups, disclosure, related to women's comfort in sharing her breast cancer diagnosis with family, friends, work colleagues, and current or future intimate partners. The women mostly based their disclosure decisions on potential impacts of others. Several of the women withheld their diagnosis from loved ones for extended periods of time to protect them from the emotional burden of their diagnosis. Specifically, those with young children or ailing parents strategized the best ways to disclose their diagnosis and sometimes hid the information by disguising their hair loss and body changes or by avoiding visits with parents. These women shared that social media offered the opportunity to disclose their diagnosis in a safe setting and also the chance to gather insight about ways to share their news with loved ones. The single women in the group shared an additional concern about when and how to tell people that they date about their breast cancer. The process of disclosure depleted women's emotional energy and appeared as one of the most challenging aspects of their cancer diagnosis.

The women expressed many fears related to changes in their body, independence, and relationships with others. Almost unanimously, the women talked about their fears of their cancer returning or dying from their illness. Due to the extensive surgeries and side effects from treatment, many of the women feared body image changes, not feeling themselves or embarrassing others with their looks, such as their young children. They worried about becoming dependent and requiring help from others. Many feared that illness progression would leave them debilitated like other women they witnessed in treatment. Younger women expressed worries about dating as well as infertility from their chemotherapy treatments. Lastly, the women shared concerns that changes to their physical abilities would leave them less valuable to their children, partners, or friends.

Finding ways to cope with the distress of waiting for a cancer diagnosis, managing treatment-related side effects, and balancing the uncertainties of survivorship transitions required the women to draw on their consistent support mechanisms, as well as develop new coping strategies. Most of the women turned to their existing support mechanisms to draw strength. They also benefited from becoming immersed in their roles within family or careers. Caregiving roles, especially as parents, enabled women to refocus their thoughts onto other's needs. They sought ways to maintain a sense of normalcy and attempted to hold onto their daily routines. The ability to remove negative forces in their lives such as relationships and distressing cancer information afforded them a clearer space to process their thoughts. A small group of the women engaged a hypervigilant stance and sought as much medical information as they could about their disease and treatment options. In telling stories of their coping processes, the women most commonly chose to explore positive aspects of their illness and participated in benefit-finding to decipher small and large existential outcomes from their illness experience.

Discussion

Our findings reveal many similarities and some notable differences between both breast cancer support formats as well as between both thematic analysis techniques. The seven themes arose from the MEM (diagnosis, social support, risk, existentialism, treatment process, information-seeking, and surgery) and seven themes were derived from focus groups (screening/diagnosis, treatment, support systems, existential concerns, disclosure, coping and fears). Women in the focus groups shared experiences with diagnosis, treatment and social support that were similar to the posters on the Reddit forum; however, the in-person dialogue presented a more nuanced and vulnerable description of the participant's coping processes, fears of dependency and dying, and experiences surrounding disclosure of their diagnosis. Marshall and colleagues (2016) conducted a comparison of online text extracted from a social media outlet with items endorsed on breast cancer symptom checklist. Utilizing the K-medoid clustering method, the study analyzed over 50,000 messages generated by users of the breast cancer forum on MedHelp.org. Overall, the online data significantly overlapped with the checklist data. Findings from this study mirror that of Marshall and colleagues (2016) and demonstrate the effectiveness of topic modeling in this domain.

Focus groups have historically served as the gold standard for psychosocial cancer research in developing, testing, and confirming soundness of research questions; however, focus groups require extensive resources and introduce methodological limitations, including experimenter bias. For example, focus groups require extensive time in planning, recruitment, and transcription. Validity of findings become dependent upon participant recruitment and the influence of normative discourse or group power dynamics (Smithson, 2000). Language analysis techniques like the MEM enable researchers to examine themes across massive corpuses of natural language data, which becomes challenging using traditional qualitative methodology. Data collected from focus groups and individual interviews may be affected by experimenter bias, in that the questions and associated probes that are used in these methodologies are generated by the researcher or the moderator. Indeed, by facilitating interaction among members, offering transitional summaries, and noting non-verbal responses, the moderator plays a large role in the focus group (McLafferty, 2004); language data extracted from online forums like Reddit are not shaped by a moderator or by the experimenter. The MEM can also be used to track discourse over time, which may highlight treatment-related changes, problematic thinking styles, and the emergence of adaptive coping strategies (Pulverman et al., 2016).

Findings from this study highlight the benefits of the MEM in advancing psychosocial cancer research. However, several limitations should be acknowledged. The focus group sample had limited racial/ethnic and educational diversity. Women in the focus groups may have met previously in their support group and have shared a familiar and trusting environment in which to share more vulnerable experiences. As well, focus group participants had familiarity with the researchers' backgrounds in counseling fields and breast cancer and this may have increased their likelihood to share sensitive topics. Focus group attendees were affiliated with a breast cancer support organization; thus, their responses may more likely reflect experiences of well-informed and socially connected survivors. Reddit, as an online social media platform, may cater to younger or more technologically savvy and educated individuals. Given our focus group samples' young age and high educational attainment, our findings between methodologies may be comparable. Importantly, the posts extracted from Reddit may have been written by a range of individuals—those who were recently diagnosed, are currently undergoing treatment, are long-term survivors, as well as family members, friends, and colleagues of individuals across the diagnosis, treatment, and survivorship phases.

This kind of diversity speaks to another potential confound that may also be considered a strength of this mixed-methods approach. Given the different populations assessed by the two thematic analysis methods, different themes are likely to emerge from each method. Of the 10 resultant themes from this study, four themes co-occurred in both samples, with slight variations potentially due to the context of dialogues among participants in each setting. Although there was variability in the themes that emerged from each method, this variability does not discredit the validity of either approach; rather, it highlights relevant nuances in experiences that may be driven by contextual factors, including familiarity among members, willingness to express vulnerability, and cultural norms in online versus in-person settings. For example, fears, one of the themes that emerged from the focus groups but was not found in the Reddit data, may have been easier to discuss in person among fellow survivors than to share over social media. Similarly, a recent linguistic analysis of posts on breastcancer.org revealed that content shied away from challenging topics, such as the fear of dying, and instead focused on protecting social relationships and personal well-being (Malloch & Taylor, 2019). The purpose of this comparison is therefore not to demonstrate the superiority of one method over the other, as we must expect some variation given the characteristics of each technique; rather, it is to build a rich and nuanced understanding of the breast cancer journey, acknowledging that both methods are ultimately useful in identifying the specific needs of this population and tracking improvements as well as persistent concerns.

Passive online sampling through language analysis techniques could offer oncology researchers additional mechanisms for exploring relationships that warrant more attention, relationships that could be challenging to assess in a laboratory setting, or relationships that may not be as strong as initially hypothesized. Researchers may consider engaging in more active sampling using this methodology; that is, posing an open-ended question to an online group and then analyzing written responses with the MEM. This approach has already been applied to written responses on sensitive topics, such as childhood sexual abuse and sexuality in general, to characterize experiences and to describe complex sensations (Handy, Stanton, & Meston, 2019;

Kilimnik, Boyd, Stanton, & Meston, 2018). Given the physical and psychosocial barriers that cancer poses for research and data collection, the MEM opens opportunities to locate large samples of participants that may historically be hard to find. The MEM, and social media data in general, may also enable the inclusion of voices of cancer survivors who have physical or social limitations that hinder in-person research participation.

Acknowledgements

The first author was supported by a Doctoral Training Grant in Oncology Social Work Renewal, DSWR-18-087-03, from the American Cancer Society and a PEO Scholar Award. The second author was supported by an NIMH T32 training grant (5T32MH116140-02). The authors would express their gratitude for the breast cancer survivors for sharing their experiences.

Disclosure statement

The authors report no financial and/or business interests.

Data availability

Supporting data for this study can be requested through the corresponding author.

References

Alshaikh, F., Ramzan, F., Rawaf, S., & Majeed, A. (2014). Social network sites as a mode to collect health data: A systematic review. *Journal of Medical Internet Research*, *16*(7), e171.

Berger, A. M., Buzalko, R. J., Kupzyk, K. A., Gardner, B. J., Djalilova, D. M., & Otte, J. L. (2017). Preferences and actual chemotherapy decision-making in the greater plains collaborative breast cancer study. *Acta Oncologica*, *56*(12), 1690-1697.

Blackburn, K. G., Yilmaz, G., & Boyd, R. L. (2018). Food for thought: Exploring how people think and talk about food online. *Appetite*, *123*. https://doi.org/10.1016/j.appet.2018.01.022

Boyages, J., Kalfa, S., Xu, Y., Koelmeyer, L., Mackie, H., Viveros, H., ... & Gollan, P. (2016). Worse and worse off: The impact of lymphedema on work and career after breast cancer. *Springerplus*, *5*(1), 657.

Boyd, R. L. (2014). MEH: Meaning Extraction Helper (Version 1.2.42) [Software].

Boyd, R. L. (2017). Psychological Text Analysis in the Digital Humanities. In S. Hai-Jew (Ed.), *Data Analytics in Digital Humanities* (pp. 161–189). https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, *18*, 63–68. https://doi.org/10.1016/j.cobeha.2017.07.017

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R. (2015, April). Values in words: Using language to evaluate and understand personal values. In *International Conference of Web and Social Media* (pp. 31-40).

Brandão, T., Schulz, M. S., & Matos, P. M. (2017). Psychological adjustment after breast cancer: A systematic review of longitudinal studies. *Psycho‐oncology*, *26*(7), 917-926.

Browall, M., Kenne Sarenmalm, E., Persson, L.-O., Wendstrom, Y., & Gaston-Johansson, F. (2015). Patient-reported stressful events and coping strategies in post-menopausal women with breast cancer. *European Journal of Cancer, 25,* 324-333.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality*, *42*(1), 96-132.

Drageset, S., Lindstrom, T. C., & Underlid, K. (2009). Coping with breast cancer: Between diagnosis and surgery. *Journal of Advanced Nursing, 66*(1), 149-158.

Fasse, L., Flahault, C., Vioulac, C., Lamore, K., Van Wersch, A., Quintard, B., & Untas, A. (2017). The decision-making process for breast reconstruction after cancer surgery: Representations of heterosexual couples in long-standing relationships. *British Journal of Health Psychology*, *22*(2), 254-269.

Fiszer, C., Dolbeault, S., Sultan, S., & Brédart, A. (2014). Prevalence, intensity, and predictors of the supportive care needs of women diagnosed with breast cancer: A systematic review. *Psycho‐Oncology*, *23*(4), 361-374.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. *British Dental Journal*, *204*(6), 291-295.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Los Angeles: Sage Publications.

Hamama-Raz, Y., Pat-Horenczyk, R., Roziner, I., Perry, S., & Stemmer, S. M. (2019). Can posttraumatic growth after breast cancer promote positive coping?—A cross-lagged study. *Psycho‐oncology*, *28*(4), 767-774.

Handy, A. B., Stanton, A. M., & Meston, C. M. (2019). What does sexual arousal mean to you? Women with and without sexual arousal concerns describe their experiences. *The Journal of Sex Research*, *56*(3), 345-355.

Kilimnik, C. D., Boyd, R. L., Stanton, A. M., & Meston, C. M. (2018). Identification of nonconsensual sexual experiences and the sexual self-schemas of women: Implications for sexual functioning. *Archives of Sexual Behavior*, *47*(6), 1633-1647.

Kim, E., Han, J. Y., Shaw, B., Shah, D., McTavish, F. & Gustufson, D. (2012). The process and effect of supportive message expression and reception in online breast cancer support groups. *Psycho-oncology, 21, 531-540.*

Laidsaar-Powell, R., Butow, P., Charles, C., Gafni, A., Entwistle, V., Epstein, R., & Juraskova, I. (2017). The TRIO Framework: Conceptual insights into family caregiver involvement and influence throughout cancer treatment decision-making. *Patient Education and Counseling*, *100*(11), 2035-2046.

Maheu, C., Hébert, M., Louli, J., Yao, T. R., Lambert, S., Cooke, A., ... & Kyriacou, J. (2019). Revision of the fear of cancer recurrence cognitive and emotional model by Lee-Jones et al with women with breast cancer. *Cancer Reports*, e1172.

Male, D. A., Fergus, K. D., & Cullen, K. (2016). Sexual identity after breast cancer: sexuality, body image, and relationship repercussions. *Current Opinion in Supportive and Palliative Care*, *10*(1), 66-74.

Malloch, Y. Z., & Taylor, L. D. (2019). Emotional self-disclosure in online breast cancer support groups: Examining theme, reciprocity, and linguistic style matching. *Health Communication*, *34*(7), 764-773.

Marshall, S. A., Yang, C. C., Ping, Q., Zhao, M., Avis, N. E., & Ip, E. H. (2016). Symptom clusters in women with breast cancer: An analysis of data from social media and a research study. *Quality of Life Research, 25,* 547-557.

McLafferty, I. (2004). Focus group interviews as a data collecting strategy. *Journal of Advanced Nursing*, *48*(2), 187-194.

Peek, N., Holmes, J. H., & Sun, J. (2014). Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics. *Yearbook of Medical Informatics*, *23*(01), 42-47.

Pulverman, C. S., Boyd, R. L., Stanton, A. M., & Meston, C. M. (2017). Changes in the sexual self-schema of women with a history of childhood sexual abuse following expressive writing treatment. *Psychological Trauma: Theory, Research, Practice, and Policy*, *9*(2), 181–188. https://doi.org/10.1037/tra0000163

Quinn, E. M., Corrigan, M. A., McHugh, S. M., Murphy, D., Mullane, A. D., & Redmond, H. P. (2013). Who's talking about breast cancer? Analysis of daily breast cancer posts on the internet. *The Breast, 22,* 24-27.

Rubenstein, E. L. (2015). "They are always there for me": The convergence of social support and information in an online breast cancer community. *Journal of the Association for Information Science and Technology, 66*(7), 1418-1430.

Shaw, B., McTavish, F., Hawkins, R. P., Gustafson, D., & Pingree, S. (2000). Experiences of women with breast cancer: Exchanging social support over the CHESS computer network. *Journal of Health Communication, 5,* 135-149.

Smithson, J. (2000). Using and analysing focus groups: limitations and possibilities. *International Journal of Social Research Methodology*, *3*(2), 103-119.

Stanton, A. M., Boyd, R. L., Pulverman, C. S., & Meston, C. M. (2015). Determining women's sexual self-schemas through advanced computerized text analysis. *Child Abuse & Neglect*, *46*, 78-88.

Stanton, A. M., Meston, C. M., & Boyd, R. L. (2017). Sexual Self-Schemas in the Real World: Investigating the Ecological Validity of Language-Based Markers of Childhood Sexual Abuse. *Cyberpsychology, Behavior, and Social Networking*, *20*(6). https://doi.org/10.1089/cyber.2016.0657

Strasser, S. M., Smith, M., Pendrick-Denney, D., Boos-Beddington, S., Chen, K., & McCarty, F. (2012). Feasibility study of social media to reduce intimate partner violence among gay men in metro Atlanta, Georgia. *Western Journal of Emergency Medicine*, *13*(3), 298-304.

Swanberg, J. E., Nichols, H. M., Ko, J., Tracy, J. K., & Vanderpool, R. C. (2017). Managing cancer and employment: Decisions and strategies used by breast cancer survivors employed in low-wage jobs. *Journal of Psychosocial Oncology*, *35*(2), 180-201.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24-54.

Tighe, M., Molassiotis, A., Morris, J., & Richardson, J. (2011). Coping, meaning and symptom experience: A narrative approach to the overwhelming impacts of breast cancer in the first year following diagnosis. *European Journal of Oncology Nursing, 15*, 226-232.

Wallner, L. P., Li, Y., McLeod, M. C., Hamilton, A. S., Ward, K. C., Veenstra, C. M., ... & Hawley, S. T. (2017). Decision-support networks of women newly diagnosed with breast cancer. *Cancer*, *123*(20), 3895-3903.

Wevers, M. R., Aaronson, N. K., Bleiker, E. M., Hahn, D. E., Brouwer, T., van Dalen, T., ... & Vrouenraets, B. C. (2017). Rapid genetic counseling and testing in newly diagnosed breast cancer: patients' and health professionals' attitudes, experiences, and evaluation of effects on treatment decision making. *Journal of Surgical Oncology*, *116*(8), 1029-1039.

Williams, F., & Jeanetta, S. C. (2015). Lived experiences of breast cancer survivors after diagnosis, treatment and beyond: Qualitative study. *Health Expectations*, *19,* 631-642.

Wolf, M., Chung, C. K., & Kordy, H. (2010). Inpatient treatment to online aftercare: E-mailing themes as a function of therapeutic outcomes. *Psychotherapy Research: Journal of the Society for Psychotherapy Research, 20*(1), 71–85.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363-373.

Table 1

Focus Group Participant Demographics and Cancer Treatment Characteristics (n=18)

Category	N(%)	M(SD)	Range
Current age		53 (9.05)	39-72
Age at diagnosis		48 (9.53)	36-71
Time since diagnosis (in years)		4.5 (4.99)	1-17
Stage of cancer			
0	3 (16.7)		
I	3 (16.7)		
II	8 (44.4)		
III	3 (16.7)		
IV	0 (0.0)		
Other	1 (5.6)		
Surgery			
No	1 (5.6)		
Yes	17 (94.4)		
Radiation			
No	7 (38.9)		
Yes	11(61.1)		
Chemotherapy			
No	6 (33.3)		
Yes	12 (66.7)		
Hormones			
No	9 (50.0)		
Yes	9 (50.0)		
Genetic Testing			
No	5 (27.8)		
Yes	13 (72.2)		
Education			
Some high school or less	1 (5.6)		
High school graduate/GED	0 (0.0)		
Some college	4 (22.2)		
4 years of college	11(61.1)		
Advanced degree	2 (11.1)		
Relationship status			
Single, not dating	3 (16.7)		
In a committed relationship	2 (11.1)		
Married	12 (66.7)		
Separated/divorced	1 (5.6)		
Widowed	0 (0.0)		
Ethnicity			
Hispanic/Latina	5 (27.8)		
Not Hispanic/Latina	13 (72.2)		

Race	
American Indian/Alaska native	0 (0.0)
African American/Black	2 (11.1)
White	10 (55.6)
Asian American	1 (5.6)
Pacific Islander	0 (0.0)
Other	4 (22.2)
Internet usage	
I have never used the internet	0 (0.0)
I use the internet a couple of times a month	1 (5.6)
I use the internet on a weekly basis	17 (94.4)
I use the internet daily	0 (0.0)
Which internet sites do you use?	
Cancer information sites	
No	4 (22.2)
Yes	14 (77.8)
Breast cancer support sites	
No	6 (33.3)
Yes	12 (66.7)
Breast cancer blogs	
No	12 (66.7)
Yes	6 (33.3)
Medical center websites	
No	5 (27.8)
Yes	13 (72.2)
Other sites	
No	13 (72.2)
Yes	5 (27.8)

Table 2

Breast Cancer Themes, including~~s~~ Most Commonly Co-Occurring Words and Loading Values, from the Reddit Sample ~~Extraction~~

Theme 1: Diagnosis		Theme 2: Social Support		Theme 3: Risk		Theme 4: Existentialism		Theme 5: Treatment Process		Theme 6: Information Seeking		Theme 7: Surgery	
~~Code~~	Loading Value	~~Code~~	Loading Value	~~Code~~	Loading Value	~~Code~~	Loading Value	~~Code~~	Loading Value	~~Code~~	Loading Value	~~Code~~	Loading Value
Lump	0.67	Husband	0.50	BRCA	0.76	Live	0.41	Lymph	0.73	Diagnosis	0.39	Double	0.55
Ultrasound	0.62	Recovery	0.39	Test	0.71	Die	0.41	Node	0.72	Question	0.34	Mastectomy	0.53
Mammogram	0.53	Friend	0.38	Genetic	0.66	Issue	0.46	Radiation	0.51	Answer	0.33	Negative	0.52
Biopsy	0.51	Expect	0.35	Mutation	0.65	Life	0.39	Chemo	0.43	Story	0.33	Stage	0.49
Tissue	0.50	Time	0.33	Gene	0.63	Cancer	0.36	Spread	0.42	Information	0.30	Reconstruction	0.36
Told	0.45	Give	0.32	Risk	0.53	Effect	0.37	Lumpectomy	0.36	Group	0.29	Surgery	0.32
Said	0.42	Work	0.32	History	0.50	People	0.31	Treatment	0.33	Read	0.26	Week	0.31
Size	0.40	Stress	0.31	Insurance	0.44	Side	0.31	Mom	0.30	Learn	0.26	PM	0.31
Breast	0.38	Kid	0.30	Positive	0.43	Year	0.29					Wife	0.30
Doctor	0.36	Care	0.30	Family	0.41	Change	0.29					Month	0.29
Worry	0.34	Sound	0.29	High	0.35	Woman	0.28					Diagnose	0.29
Small	0.34	Hard	0.29	Mother	0.33	Important	0.28						
Tumor	0.34	Stay	0.29	Cover	0.32	Treat	0.27						
Left	0.33	Deal	0.28	Age	0.28	Mean	0.26						
Right	0.33	Share	0.28	Chance	0.27								
Large	0.33	Hospital	0.28										

Table 3

Quotes Representing Each Theme

Theme	Meaning Extraction Method	Focus Groups
Diagnosis	Tuesday, I go to see my doctor. He gives me a clinical breast exam. He feels the lump but he also told me something completely unexpected. He started to feel on the outer side of my left breast and asked me if I had noticed how thick and firm the tissue was on that side, going into my underarm. I simply thought I would go in there and he would tell me not to worry.	You get the diagnosis you have breast cancer, but actually that is only one part of the story... and there's this really insane rollercoaster that happens, where you get one call about one test result and another call about another and you don't know half the time what it means when they tell you… There's just like that terrorizing few weeks in the beginning that are just universally just so hard on everybody who gets a diagnosis.
Social support	It's awesome that your husband is supportive, and support him right back. It's not easy for anyone - but supporting each other will help.	... [I]t's the immediate family, the Spanish community is very close, and you never see them come alone to a meeting, they always have somebody that they want to bring, even a cousin a neighbor. The Hispanic community help each other, and you know, when they don't have financials they actually, the family and friends and neighbors will get together to pitch in and everybody give a dollar.
Risk	I was just diagnosed this year and have learned a lot in a short amount of time…While you can inherit the risk for breast cancer from your mother, the most telling would be if she's had genetic testing done for BRCA 1/2 and a few other genes that carry a high risk. If she's positive for those genes, you might want to look into being tested to see if you've inherited them.	——
Existentialism	At least personally, I feel like the main benefit of living a long and healthy life is that I would have extra time to spend with my partner.	And I've seen the spiritual go both ways. It can be an existential crisis for a lot of people. I've seen people both embrace the spiritual and the religious life and others move away from it completely.
Treatment process	A cancer diagnosis is always scary, but doctors have a lot more tools for fighting it…They are waiting on the biopsy because the info they get from that will determine what treatment plan they suggest.	[T]here are just a lot of issues that happen ongoing and it does not end when your active treatment ends and I think there's also this misperception, I think, you know, that doctors are not good at explaining to patients that are going through treatment. You know, they're trying to keep them positive to get them through treatment,

		but then when treatment ends, they just think they’re cancer free and walk away. And it’s all done. It’s over. And it’s not over.
Information-seeking	Radiation and chemotherapy are the main treatments, these have also changed over the years, so don't let people's horror stories get to you. Even if someone went through cancer 5 years ago, things have changed. Good luck to your mom in her fight. If she needs it, there is probably a support group in your area. For now, here is an internet hug for you and her. If you want to read some info these are good sites: [breastcancer.org](http://www.breastcancer.org/) [CDC site](http://www.cdc.gov/cancer/breast/basic_info/) [cancer.net](http://www.cancer.net/cancer-types/breast-cancer)”	——
Surgery	She opted for a nipple/skin sparing double mastectomy and her cosmetic results look fantastic (there are some scary mastectomy pics out there). Most of those are older procedures, look for a surgeon who knows how to do his procedure (if your cancer locale permits it). My wife's breasts have some scars but by far the best looking boobs I've seen post mastectomy.	——
Disclosure	——	Of the young women I think two things, their fear is that they don’t want to scare their children, they don’t want to look different for their children, they want it to be as normal as possible. And then fear that, if they’re single, they are not going to be able to date, right? What are they gonna, how are they gonna tell them, you know, their breasts look different? And when do they tell them.
Coping	——	I was diagnosed October 2nd of ’15 and I took half days off to have chemo. I took three weeks to have surgery and then I just had to keep going, you know? The energizer bunny. But I did have a huge support group at my [job]. And I have a huge family so they’re a support group too.
Fears	——	But I’m scared for recurrence. I’m more scared. I was not scared at chemo. I was not scared with radiation. I’m scared what will happen in future. How we can prevent a recurrence.

Table 4

Themes Derived from Breast Cancer Focus Groups

Screening/Diagnosis	Treatment	Social Support	Disclosure	Existentialism	Coping	Fears
Delays	Surgery	Friends	Not comfortable sharing	Prioritizing	Normalcy	Asking for help
Mammogram	Chemotherapy	Family	Family	Gratitude	Reframing	Dying
Young Age	Radiation	Significant other	Kids	Wellness	Accepting help	Recurrence
Stage	Reconstruction	Church	Dating	Self-discovery	Strength from others	Not being needed
Tests/labs	Motivated by fear of death	Neighbors	Engaging with social media	Personal strength	Roles	Body changes
Decision-making	Physical and emotional healing	Community members	Timing	Self-efficacy	Work	Not being self
Fear	Side effects	Internalized stigma	Consider family's needs	Higher purpose	Caregiving	Dating
Avoidance of support	Chemo brain	Support groups		Role model	Removing negative energy	Infertility
Withdrawal	Complications			Comforting others	Hypervigilance	Embarrassing others
Medical distrust					Humor	Comparing to others
Shame					Pain meds	
Feeling a burden					Seeking information	
Protecting others						