Published in final edited form as:

Health Commun. 2018 September ; 33(9): 1158–1165. doi:10.1080/10410236.2017.1339370.

Sentiment Analysis of an Online Breast Cancer Support Group: Communicating about Tamoxifen

Mark L. Cabling, MSc [Research Assistant],

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, 3300 Whitehaven St., NW, Suite 4100, Washington, DC 20007

Jeanine W. Turner, PhD [Associate Professor],

Communication, Culture, and Technology Program, Georgetown University, 3520 Prospect St., Suite 311, Washington, DC 20057, Office: (202)-687-6927, Fax: (202)-687-4031

Alejandra Hurtado-de-Mendoza, PhD [Research Instructor],

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, 3300 Whitehaven St., NW, Suite 4100, Washington, DC 20007, Office: (202)-687-8916, Fax: (202)-687-0305

Yihong Zhang, MSc [Data Analyst],

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, 3300 Whitehaven St., NW, Suite 4100, Washington, DC 20007, Office: (202)-687-8914, Fax: (202)-687-0305

Xinyang Jiang [Biostatistics Intern],

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, 3300 Whitehaven St., NW, Suite 4100, Washington, DC 20007, Office: (202)-687-8917, Fax: (202)-687-0305

Fabrizio Drago, MA [Research Assistant], and

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, 3300 Whitehaven St., NW, Suite 4100, Washington, DC 20007, Office: (202)-687-8915, Fax: (202)-687-0305

Vanessa B. Sheppard, PhD [Professor]

Virginia Commonwealth University, 830 E Main St., Richmond, VA 23219

Abstract

A better understanding of the breast cancer online narrative is important for a clearer conceptualizing of the role of online platforms in mediating health related support. Sentiment analysis was conducted on a breast cancer online support group regarding Tamoxifen to understand users' emotions and opinions. This analysis was then contextualized within online social support literature. Out of the 498 users, the *most active* users were 80% more positive than *least active* users while *least active* users were 48% more negative than *most active* ones; both differences were statistically significant. The higher the stage of cancer a user had, the less

likelihood that she would have posted, and if she were to post, the post would have focused on her side effects and the anxiety/sadness that tailgates those side effects. The lower the stage of cancer a user had, the more likelihood that she would have posted, and additionally remained active on the forum and encouraged more (online) social support. This finding suggests that the online support platform may provide a context that exacerbates support for like-minded users where stronger ties are created around a specific sentiment within the community with less connection from those with dissimilar sentiments to the dominant group.

Keywords

sentiment analysis; online support groups; social identity-deindividuation theory; breast cancer; tamoxifen

Online support groups (OSGs) provide a variety of opportunities for patients seeking information and support. Breast cancer survivors are among the most prolific users of online platforms to communicate and garner support (e.g. Portier et al., 2013). Tamoxifen is a common topic of discussion because many breast cancer survivors have taken or have to take it. Despite research suggesting that it effectively reduces breast cancer recurrence rates, Tamoxifen's regimen is often not adhered to. A better understanding of OSG narratives is important for a clearer conceptualization of the role of online platforms in mediating health related support. The goal of this paper is to analyze the conversations taking place in a breast cancer OSG in order to better characterize how it may mediate support in the context of Tamoxifen.

Online Social Support

In a review of 81 studies, Uchino and colleagues (1996) concluded that there were links between social support and beneficial outcomes in the cardiovascular, endocrine, and immune systems of patients. Smart phone technology, wireless accessibility, and a critical mass of users within most disease populations means that patients can talk to other patients about specific diseases anytime and anywhere, thus making online platforms attractive for seeking and providing support. Indeed, 61% of Americans use Internet to seek health information, which includes socializing on OSGs (Fox, 2011). Research in this sphere has shown that patient-to-patient interaction has led to the provision and receipt of informational, emotional, and instrumental support (Ancker et al., 2009), and empowerment (e.g. Robinson, Turner, Levine, & Tian 2011), regardless of whether patients were active participators or merely passive lurkers (van Uden-Kraan et al., 2008).

However, Internet use has also been linked to some problematic behaviors (e.g. Chung, 2014). For example, in a sample of 562 adults, 11% discontinued treatment based on Internet information and reported higher levels of OSG engagement (Weaver, Thompson, Weaver, & Hopkins, 2009). Calling for a better understanding of OSG users that voiced opinions against recommended treatments, Weaver and colleagues (2009) suggested the importance of exploring a potential for digital group thinking processes to understand their implications for patient adherence. The potential influence of OSGs can thus result in

incorrect information being spread. While OSG members can correct the information (Esquivel, Meric-Bernstam, & Bernstam, 2006), all messages are still available over time.

Despite these negative aspects, OSGs' potential to increase information/social support (Fogel, Albert, Schnabel, Ditkoff, & Neugut, 2002) makes them important tools for patients as they turn to these spaces to participate or read about alternatives available to them (e.g. Robinson et al., 2011). Furthermore, socially supportive messages within online monitoring systems have linked online support with satisfaction, information gathering, empowerment, and improved A1C levels in diabetics (Turner et al., 2013) – all without the logistical challenges of attending face-to-face support groups (Turner, Grube, & Meyers, 2001).

Tamoxifen

One out of eight US women is diagnosed with breast cancer per year (National Cancer Institute, 2012). Eighty percent of these cases are hormonal receptor positive (HR+), which is treated with adjuvant hormonal therapy (AHT), the most common being Tamoxifen. As breast tumors from this type of cancer need estrogen or progesterone to grow, AHT is used to deprive cancer cells of them (National Cancer Institute, 2012), reducing recurrence by 40% and mortality by 30% (Alkner et al., 2009). Despite the success of the medication, 50% of women do not fully adhere to the recommended 5-10 year AHT regimen (e.g. Hershman et al., 2011). Demographics, healthcare factors, patient-provider communication, and social support have been associated with AHT adherence (e.g. Livaudais et al., 2012; Murphy, Bartholomew, Carpentier, Bluethmann, & Vernon, 2012). As social support is expanded online, breast cancer survivors are more likely to seek support on OSGs (Davison, Pennebaker, & Dickerson, 2000). Despite the growing scope of OSGs and the significant public health concern of poor adherence to AHT, OSG conversations regarding AHT has not been studied.

Unique Nature of Online Support

The nature of computer mediated communication (CMC) creates the opportunity for users to selectively represent themselves, idealize other users, and foster personal exchanges that reinforce specific perspectives by creating an exaggerated sense of connection (Walther & Boyd, 2002). This sense of connection is built on social identity and deindividuation (SIDE) theory (e.g. Reicher, Spears, Postmes, 1995). SIDE theory predicts that in the absence of face-to-face (FtF) cues and prior personal knowledge, social context cues present in CMC may lead to over-attribution of similarity among users and an emphasis on group identity over the individual (Walther, 1996). Additionally, CMC interactions can prompt users to group other users according to social categories and norms rather than interpersonal cues (Tidwell & Walther, 2002). For example, a SIDE theory study concluded that in CMC users showed increased willingness to express normative attitudes against out-group interests as a function of support (Spears, Lea, Corneliussen, Postmes, & Haar, 2002). This is facilitated by the asynchronous nature of CMC, which allows participants to plan their responses more mindfully than in FtF interactions (e.g. Hiemstra, 1982; Walther & Boyd, 2002). A related approach found that OSGs targeting breast cancer patients are conducive to the development of hyper-personal relationships. Turner et al. (2001) found that breast cancer OSG

participants perceived the support received from their online community as more intimate than the support received from a significant other.

OSGs thus provide an opportunity for participants with similar health challenges to provide support for one another. While research has suggested that OSGs provide the opportunity for more intense connections and support than those relationships found in FtF contexts (Tidwell & Walther, 2002), an understanding of how users communicate and connect with each other within the OSG is less investigated. Although research has coded themes of interaction within these groups (e.g., Broom 2016), an understanding of the collective comments can be categorized with particular users is less explored. Therefore, there is limited examination of some facets of CMC, which may shed light into different types of OSG interactions. To explore these collective comments, sentiment analysis of OSG messages may be insightful. Examining users' sentiments can provide a more nuanced understanding of how specific interactions mediating support may lead to the development of in-group and out-group dynamics and hyper-personal communication within OSGs. This examination can reveal what users are talking about, how they are talking about specific patient protocols, and how those that participate actively may engage in different topics than those who do not participate actively.

Sentiment Analysis

Breast cancer patients are one of the most active participants in health related OSGs and they exchange emotional support most frequently (Rodgers & Chen, 2005). Sentiment analysis (SA) provides a unique opportunity for analyzing large amounts of information shared through OSGs as well as the structure and characteristics of the discourse. According to Liu (2012) SA is the study of attitudes, evaluations, and emotions around particular subjects. SA is based on the presumption that opinion affects human activity and is indicative of human behavior. Publicly available social media may reflect public opinion, for example, what people *say* they "feel" about something on Twitter is reflective of how they feel (Liu, 2012). Due to the breadth of available information (e.g., via Twitter) concerning the sentiments of subjects (e.g., movies), automated modes of SA are merited due to the limitations of collecting vast amounts of data manually (Liu, 2012).

SA has been increasingly used in health, including in treatment decision-making (e.g., Bantum & Owen, 2009). For instance, Cobb and colleagues (2013) used SA in an OSG and found that exposure to positive messages about a new smoking cessation drug was associated with increased likelihood of switching to or continuing to use that drug. Most analyses of breast cancer OSGs have been conducted manually with small samples (e.g., Bender, Jimenez-Marroquin, & Jadad, 2011; Quinn et al., 2013). Few studies have examined breast cancer OSGs using SA techniques. For example, Vlahovic and colleagues (2014) assessed information/emotional support exchange among 90,000 users in a breast cancer OSG using a variety of methods including the Linguistic Inquiry Word Count (LIWC). Results suggested that certain message characteristics (e.g., questions) elicited different types of support. Emotional support (vs. information) was associated with greater participation rates (Vlahovic, Wang, Kraut, & Levine, 2014). Furthermore, a study by Portier and colleagues (2013) used sentiment analysis to examine the potential sentiment

change of forum thread initiators across breast and colorectal OSGs. Among the highest average sentiment changes found in their results were the OSGs specifically dedicated to breast cancer decision-making.

To our knowledge there are no such prior studies that look at Tamoxifen in breast cancer OSGs. With the growing diffusion of OSGs and hormone receptor positive breast cancer (Berry et al., 2006), research must explore how patients discuss Tamoxifen. The present study addresses the following questions. **RQ1:** What are the sentiments, emotions, and opinions of the OSG users? **RQ2:** Is there a dominant narrative surrounding Tamoxifen within the OSG? **RQ3:** Do most active users discuss topics differently than least active users?

Sample.

Breastcancer.org is a non-profit organization that provides information about breast cancer and encourages women to talk about their issues online. It has 151,025 members in 76 forums discussing 119,866 topics. In an average month, the website accrues 36,100 English posts from 2,306 unique authors. For this study, we used the OSG within Breastcancer.org that has the most number of threads: *Bottle o' Tamoxifen*, which is located within the forum called *Hormonal Therapy – Before, During, and After* (23,200 results), all of which are prompted by the OSG's initial post by *ashaby*: "Dear All...Got my [Tamoxifen] today and will start in a week. Where are my tamoxi pals? It says weight loss for one of the side effects. Now that's a good thing." This topic post is also the most viewed, with a total of 424,808 views. Since *ashaby*'s first post in June 2007, the threads have grown rapidly and have lasted until November 3, 2014, the same day as data collection. This thread, *Bottle o' Tamoxifen*, is substantiated by an active, enduring, and large thread that involves approximately 800 webpages and 498 unique authors discussing Tamoxifen. The topic is the AHT Tamoxifen, which is a selective estrogen receptor modulator and the most prescribed hormone therapy for survivors with HR+ breast cancer (Berry et al., 2006).

Method.

All posts from *Bottle o' Tamoxifen* were automatically downloaded by crawling through the URL provided. Each post was then parsed into relevant parts consisting of username, Dx medical information, and comment.¹ We eliminated common terms that do not provide substantive meaning. First, we removed stop words, or the most common words in the English language (e.g., prepositions) using existing lists (Liu, 2012). Secondly, we removed domain-specific common terms that are not relevant (e.g., "ER+" does not indicate sentiment).²

Preliminary sentiment analysis.—We ran a script³ to detect and extract the sentiment words in the lexica⁴ with their associated polarity value and frequency in the corpus. Then

¹We developed Python script to crawl the site and parse the downloaded data, using libraries such as 'BeautifulSoup.'

²We ran frequency analysis using Python script to extract most common domain-specific terms and through this, compiled our own domain specific stop word list.

³A script was written (Megardoomian, 2012) in order for the sentiment proportions to be found in an automated fashion with which the results of the proportion also came with a list of negative words and positive words with each of their frequencies.

we used the text concordance software, CasualConc, to study the most frequent sentiment terms within the context of their use. For example, through the “Cluster” function, we found the most frequented bigrams (two-word pairings) that the most frequent terms are a part of. This allowed the identification of terms that were misclassified, e.g., Liu’s dictionary tagged the term “hot” as the second most used positive term in the initial output but CasualConc showed that “hot” is used as part of the *most frequent* bigram “hot flashes” 99% of the time.⁵ Based on these findings, the sentiment lexica for the domain were modified, and a second SA was performed. The results were used to compute positivity/negativity scores.

LIWC.—LIWC is a computational linguistic software built by social psychologists to categorize any given text based on 72 dimensions. These categories are not merely positive or negative sentiments but specific sub-types of sentiment (e.g., sad for negative sentiment), social psychological categories of words (e.g., social) and degree of sentiment (i.e., psychological processes e.g., insight, tentativeness, etc.). For example, the linguistic dimension *Subject* also has the sub-dimensions I, we, you, she/he, and they, while the psychometric dimension *Social* encompasses swear words, social processes, family, friends, and humans.

Results.

Research question one explored what sentiments, emotions, and opinions were expressed by OSG users. Thirty-six percent of *Bottle o’ Tamoxifen*’s aggregate sentiment is made up of negative terms while 64% of it is made up of positive ones. Out of the total number of unique authors in the OSG, 155 (31.1%) posted only once; we defined this group as the least active users. The most active users were the top 25 (5%) authors by number of posts, which was calculated to be any user that posted 82 times or more.⁶ The goal of comparing these two groups was to explore potential reasons why users decided to post and how their sentiment may have played a role in seeking or providing online support as compared to those who did not post.

Liu’s dictionary.—Research question two explored whether there was a dominant narrative around Tamoxifen within the OSG. Most active users were 80% more positive than least active ones while least active users were 48% more negative than most active ones; both differences were statistically significant (Table 1). Barring the top descriptive technical bigrams (e.g., “latissimus dorsi”), the most common bigram was “hot flashes.”

Finally, research question three examined whether most active users differ from least active users. Data from Liu’s dictionary python scripts show that most and least active users share the 10 most frequent positive negative words in 50% and 60% of the cases respectively. For example, the words “pain”, “bad”, “good”, and “hope” were most used negative and positive

⁴Liu’s (2012) open source positive and negative lexicon dictionaries were used in order to find the proportion of each corpus’s negativity (negative word count divided by total words) and positivity (positive word count divided by total words).

⁵If there were words that each author thought were not indicative of sentiment or were indicative of the wrong sentiment (e.g., if CasualConc showed that the reason why a word has high frequency is due to it being a signature of a top user, or if CasualConc showed that most context indicated the use of negation, sarcasm, or irony), then the authors would decide via consensus whether to keep the word tagged as a sentiment.

⁶We tried to transform the data to a normal distribution but the range of the data is too large and the distribution of the data highly right skewed. Therefore, least active versus most active was defined based on cumulative frequency of authors/users.

words by both most and least active users. However, terms such as “cancer” and “fine” were most in least active users but not in most active users. This study’s results also show that when same terms were most frequent in both groups, they were used with similar and different meanings. For example, the term “hot” was used by both groups as part of the bigram “hot flashes”. However, while least active users most frequently used the term “problem/s” to signify problems are caused by medication, e.g., “[Tamoxifen] has a history of causing tooth problems”, most active users used it with a negation, e.g. “not a problem”.⁷

Clinical characteristics.—Of the users that provided their clinical information in their profiles, approximately 11% had stage 0 cancer, 45% stage I, 31% stage II, 10% stage III, and 3% stage IV. Almost all these users (99%) were estrogen receptor positive (ER+) and/or progesterone receptor positive (PR+) and/or HER positive (HER+), all of which had options of using hormone therapy (Cancer Research UK, 2014). The most active users were characterized by the following cancer stages: 16% stage 0, 48% stage I, and 20% stage II (16% of most active users did not provide stage information). The least active users’ cancer stages are: 8% stage 0, 35% stage I, 42% stage II, 10% stage III, and 5% stage IV (Figure 1).

CasualConc and bigrams.—“Hot flashes” was the most frequent bigram used by most active users, which corroborated with the bigram findings in the corpus as a whole. But while the top users tended to have top bigrams that consisted of clinical and technical language such as “latissimus dorsi” or “uterine issues,” the least active users had the top bigram (besides “hot flashes”) “side effects,” which did not appear to be a top bigram in the most active corpus. The least active users used less technical terms and focused on side effects in general.

LIWC and difference ratios.—*Bottle o’ Tamoxifen* corpus’s differences (by having a difference ratio of at least 100%) by most active and least active users were found in 18 (out of the 72) LIWC dimensions. Least active users used “I” 16.5 times more than “we” (while most active users used “I” 12.7 times more times than “we”), and used “I” half as much as the most active users (Table 2). Most active users used “we,” “swear words,” “friends,” “family,” and “sexual,” twice more than least active users. Most active users also used “money” four times more than least active users. Least active users used “negative emotion,” “sadness,” “causation,” “tentativeness,” “certainty,” “exclusion,” “achievement,” and “question” twice more than most active users. Least active users also used “anxiety,” “feel,” “religion,” and “death” three times more than most active users (Table 2).

Discussion

According to our data, LIWC, and Liu’s sentiment dictionaries reinforced that the sentiment of this corpus was far more positive than negative. However, it was only by comparing the most active users with the least active users that we uncovered the nuances. The least active users differed in that their rates of negative sentiment words were double than the most active users’ ones and more than one and a half times when compared to the negative rates of the OSG as a whole, differences that were statistically significant (Table 1). It is also

⁷A table with complete frequencies descriptions is available from the corresponding author.

telling that the negative term “cancer” in an OSG *about* cancer is part of the least active user’s top terms but not the most active users.

LIWC takes this further by categorizing the negative words into specific clusters (Table 2). For example, we knew that the least active users expressed more negativity towards the medication Tamoxifen and its side effects but it was through LIWC’s categorization that we noticed that this negativity (relative to the whole corpus and the corpus of the most active users) was expressed through albeit similar anger levels, more sadness (by 1.4 times more than that of the whole corpus and twice as much as that of the most active), and far more anxiety (by almost twice more than that of the whole corpus and almost triple that of the most active).

The findings of the SA suggest that the OSG may create specific in-group and out-group dynamics which influence users’ interactions and behaviors. The users that expressed and supported the dominant sentiment of the OSG were likely to stay and engage actively in the community. Those who did not express and support the dominant sentiment of the OSG tended not to participate or to leave the community.

The development of a dominant opinion within OSGs has been identified in past research. Askay (2015) found that the dominant opinion expressed through positive bias in review websites created a fear of isolation that reduced the likelihood of users to post neutral or negative reviews. The sentiment within our study’s OSG provides a similar context in which the dominant opinion expressed by the most active users may have reduced the willingness of the least active users to voice their opinion and to actively participate in the OSG. Future research should explore the potential silencing behavior that might exist within OSGs, which can alienate those with alternative opinions and perspectives.

The sentiment expressed by the OSG users is also linked to their cancer stage. The most active users were made up of breast cancer survivors who ranged from having stages 0—IIa with none having more than stage IIa. Indeed, 77% of most active users had stage 0 or stage I, which are highly treatable. There is a 98% survival rate at these stages (American Cancer Society, 2016). Contrastingly, least active users range from having stages 0—IV with 60% of these users having stage II (90% survival rate), stage III (72% survival rate), and stage IV (22% survival rate) (Figure 1).

This suggests that the higher the stage of cancer a person had, the less likely that person would post in this OSG, and if they were to post, the post would focus on the individual’s side effects and the anxiety that goes along with them. Significantly, if a user had stage III cancer that user had a 100% chance of posting once and with negative sentiment. On the other hand, the lower the stage a person had, the more likely that user will remain active on the OSG and encourage more (online) social support. Consequent posts may focus on not just sympathetic and personal language but hyperbolically positive vocabulary that elicits more engagement by other users. Between the most active and least active users, 90% of those with stage II cancer were most active users. Past studies have suggested that active participation in an OSG was negatively associated with depression and positively associated with emotional well-being (Batenburg & Das, 2014), that patients experienced difficulty

being confronted with negative sides of their disease (Barak, Boniel-Nissim, & Suler, 2008), and that users withdrew from OSGs when discussing details about cancer became too painful to bear (Sandaunet, 2008).

When we contextualized the negative and positive words using CasualConc, we saw how the dimensions of anxiety, sadness, negative emotion, tentativeness, and exclusion were made more salient by the least active users. Although they had a similar percentage of positive terms compared to the most active users, their positive terms differed significantly. Both groups included in their top ten positive terms “glad,” “happy,” “good,” “luck,” and “hope.” While the most active users frequently used the terms “better,” “fun,” “great,” “love,” and “welcome,” the least active users used the most frequent words of “smile,” “positive,” “normal,” “fine,” and “able.” The former’s positive terms were *more* positive e.g., things are *fun* instead of just merely *fine*, *better* instead of merely *normal*, *great* instead of solely *fine*. More telling was the use of such a romantic word as *love* and a communal word such as *welcome* in the most active users’ text. This sentimentality and social vocabulary frequented by the most active users revealed that the least active users may be less community-oriented in the OSGs (e.g., exclusive) and that the most active users used language that was part of the LIWC dimensions more conducive to social support (e.g., family).

This finding reinforces past cancer patient communication research. If cancer patients overly exhibit negative sentiments (e.g., sadness and anxiety), those sentiments consume most of their energy, leaving little room for engagement with other OSG users (Kim, Shah, Namkoong, McTavish, & Gustafson, 2013). The more severe stage of cancer experienced by the least active users translates into the increased feeling of anxiety in their posts. If those posts are not welcomed into the OSG, least active users may feel less motivated to continue to participate. This supports Han and colleagues’ (2008) finding that using a higher frequency of positive words is associated with less anxiety while using a higher frequency of negative words is associated with more worry regarding breast cancer. Future research should further explore the extent to which participants are welcomed (or not) based on the sentiment of their posts.

Our SA shows how discourse used by the most active users may reflect a bias towards emotional well-being via the positivity of its sentiment and aversion to negative sentiment. Hyper-personal theory suggests that minimal cues environments (e.g. OSGs) can create this intense and interpersonal environment in which users tend to connect to others with similar views. This environment may also exaggerate the influence of the sentiment expressed, making it even more prevalent and reinforcing a specific perspective (Walther, 1996). The data supports the argument that a community can create norms associated with a sentiment (e.g. positive) and encourage that sentiment in a way that discourages users that express a different sentiment (e.g. neutral, negative) to participate in the community. In this study, we found that users have the potential to create an intense environment for interaction around the topic of Tamoxifen and that they may quickly categorize posts as positive or negative. In doing so, users may associate the norms of positive users differently than negative users, reinforcing a norm of a specific sentiment of interaction. While we do not know whether users of negative posts were explicitly told to leave or not to continue posting, we do know

that there was considerable difference between the sentiment norms of active users and least active users.

SIDE theory provides a lens for understanding this in-group and out-group behavior within CMC. While past research has explored how CMC can be used to foster support for in-group behavior in opposition to an out-group opinion (Spears et al, 2002), the present study provides an opportunity to explore in-group and out-group opinions within the same OSG and to examine how norms influence most active and least active user interactions. This study shows that within this OSG there is a clear demarcation between most active users and least active users based on how they discuss breast cancer. Future research should further examine OSG interactions as they relate to how users chronologically reveal their cancer stage and sentiment. Examining these interactions would help to indicate how users affiliate with an in- or out-group.

The importance of in-group and out-group attributions within OSGs lies not only in understanding posting behavior but also in providing insights into treatment adherence. Within this OSG, the support for adhering to medication regimens was strong for in-group members. Since AHT has low adherence rates, it may be important for healthcare organizations to foster OSG participation around support for medication adherence. It would be insightful to survey members of OSGs who support adherence to explore whether the positive discussions around AHT influenced their ongoing decision to take medication. Future research could continue to explore the links between adherence to in-group behavior within OSGs.

This analysis further reinforced other findings that showed how breast cancer patients used hopeful and positive stories from online peers as a tool of inspiration (Helgesson & Taylor, 1993). The positive sentiment of the most active users may indeed be comparably far more positive than the least active users precisely because they initiated optimistically, engaged with others who are also enthusiastic, and actively ignored (and therefore prevented further engagement from) those who initiated discussion in the OSG via primarily negative sentiment.

Conclusion

We acknowledge that the sentiment of *Bottle o' Tamoxifen* may represent a marginal voice because one user was responsible for 9% of posts, and only 10 users were responsible for 30% of posts. However, this also reinforces other studies on OSGs that show how most online forums, despite being open to the public and containing a plurality of voices, tend to be dominated by a marginal group of active users (e.g. Dahlberg 2007). Notwithstanding these big data limitations, this paper has shown how studies can use SA to analyze the diffusion of positive and negative messages about Tamoxifen. Through the synthesis of big data SA and research exploring the development of group cohesion within CMC, a richer narrative is revealed about what might be happening in OSGs and provides reflection for moderators to be alert to sentiment dynamics.

For further inquiries, variables such as time of post can be included in data collection 1) in order to take into account the change and evolution of sentiment over time as a registered user (or group of users) continues to be active on the OSG; 2) to refine the definition of “most active” and “least active” users as the latter may have *just* posted and therefore have lacked the time to post more; and 3) to distinguish how language used about Tamoxifen develops as a registered user first posts to when the user stops. Future research can also aim to measure the similarities and differences between members of the same activity group (e.g., for the most active user that posted 2,063 times – how does this user compare to the next most active user, who posted approximately half that amount, and does this most active user significantly affect the sentiment due to mere presence?). The present research suggests that the OSG can highlight or emphasize a specific sentiment and therefore encourage connections between those patients that also support or post that perspective. It would be interesting to further explore the nature of the messages within the OSG to better understand how alienation or welcoming behavior might be understood. Future research into OSGs can provide insight into the differences and similarities between OSGs and in-person support groups. Applying the methodology we used on other forums (e.g., regarding “Alternative Therapy”) may assess previously unseen connections between sentiment and support needs. This understanding could indicate when moderators may best intervene. The benefit of utilizing SA can especially be implicated in the context of patient centeredness in the current era of technologically mediated health support.

Acknowledgements

We would like to acknowledge the contribution of Dr. Karine Megerdooian, who we consulted for our use of automated tools. We would also like to acknowledge Breastcancer.org for their support. This work was partly funded by a Georgetown University internal grant (‘Reflective Engagement for the Public Interest’, Sheppard and Hurtado-de-Mendoza: MPI) and a R01 grant from NIH (Sheppard: PI, grant number: CA154848).

References

- Alkner S , Bendahl PO , Fernö M , Nordenskjöld B , & Rydén L (2009). Tamoxifen reduces the risk of contralateral breast cancer in premenopausal women: Results from a controlled randomised trial. *European Journal of Cancer*, 45, 2496–2502.19535242
- American Cancer Society. (2016, August 16). Breast cancer survival rates. Retrieved from <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>
- Askay DA (2015). Silence in the crowd: The spiral of silence contributing to the positive bias of opinions in an online review system. *New Media & Society*, 17, 1811–1829.
- Bantum EOC , & Owen JE (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological Assessment*, 21, 79.19290768
- Barak A , Boniel-Nissim M , & Suler J (2008). Fostering empowerment in online support groups. *Computers in Human Behavior*, 24, 1867–1883.
- Batenburg A , & Das E (2014). Emotional coping differences among breast cancer patients from an online support group: A cross-sectional study. *Journal of Medical Internet Research*, 16(2), 1–13
- Bender JL , Jimenez-Marroquin MC , & Jadad AR (2011). Seeking support on facebook: A content analysis of breast cancer groups. *Journal of Medical Internet Research*, 13(1), 1–12.
- Berry DA , Cirincione C , Henderson IC , Citron ML , Budman DR , Goldstein LJ , & Hudis C (2006). Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *Journal of the American Medical Association*, 295, 1658–1667.16609087

- Broom A (2005). The eMale: Prostate cancer, masculinity and online support as a challenge to medical expertise. *Journal of Sociology*, 41, 87–104.
- Cancer Research UK. (2014, August 4). Trastuzumab (Herceptin). Retrieved from <http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/cancer-drugs/trastuzumab>
- Chung JE (2014). Social networking in online support groups for health: How online social networking benefits patients. *Journal of Health Communication*, 19, 639–659.23557148
- Cobb NK , Mays D , & Graham AL (2013). Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. *Journal of the National Cancer Institute Monographs*, 47, 224–230.
- Davison KP , Pennebaker JW , & Dickerson SS (2000). Who talks? The social psychology of illness support groups. *American Psychologist*, 55, 205–217.10717968
- Dahlberg L (2007). Rethinking the fragmentation of the cyberpublic: From consensus to contestation. *New Media & Society*, 9, 827–847.
- Esquivel A , Meric-Bernstam F , & Bernstam EV (2006). Accuracy and self correction of information received from an internet breast cancer list: Content analysis. *British Medical Journal*, 332, 939–942.16513686
- Fogel J , Albert SM , Schnabel F , Ditkoff BA , & Neugut AI (2002). Internet use and social support in women with breast cancer. *Health Psychology*, 21, 398–404.12090683
- Fox S (2011). *The social life of health information*. Washington, DC: Pew Internet & American Life Project.
- Helgeson VS , & Taylor SE (1993). Social comparisons and adjustment among cardiac patients1. *Journal of Applied Social Psychology*, 23, 1171–1195.
- Hershman DL , Shao T , Kushi LH , Buono D , Tsai WY , Fehrenbacher L , ... & Neugut AI (2011). Early discontinuation and non-adherence to adjuvant hormonal therapy are associated with increased mortality in women with breast cancer. *Breast Cancer Research and Treatment*, 126, 529–537.20803066
- Hiemstra G (1982). Teleconferencing, concern for face, and organizational culture *Annals of the International Communication Association*, 6, 874–904.
- Kim SC , Shah DV , Namkoong K , McTavish FM , & Gustafson DH (2013). Predictors of online health information seeking among women with breast cancer: The role of social support perception and emotional well-being. *Journal of Computer-Mediated Communication*, 18, 98–118.24634575
- Liu B (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- Livaudais JC , Li C , John EM , Terry MB , Daly M , Buys SS , & Coronado GD (2012). Racial and ethnic differences in adjuvant hormonal therapy use. *Journal of Women's Health*, 21, 950–958.
- Murphy CC , Bartholomew LK , Carpentier MY , Bluethmann SM , & Vernon SW (2012). Adherence to adjuvant hormonal therapy among breast cancer survivors in clinical practice: A systematic review. *Breast Cancer Research and Treatment*, 134, 459–478.22689091
- National Cancer Institute. (2017, February 14). Hormone therapy for breast cancer. Retrieved from <https://www.cancer.gov/types/breast/breast-hormone-therapy-fact-sheet>
- Noelle-Neumann E (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, 24, 43–51.
- Portier K , Greer GE , Rokach L , Ofek N , Wang Y , Biyani P , & Yen J (2013). Understanding topics and sentiment in an online cancer survivor community. *Journal of the National Cancer Institute Monographs*, 47, 195–8.
- Quinn EM , Corrigan MA , McHugh SM , Murphy D , O'Mullane J , Hill AD , & Redmond HP (2013). Who's talking about breast cancer? Analysis of daily breast cancer posts on the internet. *The Breast*, 22, 24–27.22683246
- Reicher SD , Spears R , & Postmes T (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, 6, 161–198.
- Robinson JD , Turner JW , Levine B , & Tian Y (2011). Expanding the walls of the health care encounter: Support and outcomes for patients online. *Health Communication*, 26, 125–134.21294020

- Rodgers S , & Chen Q (2005). Internet community group participation: Psychosocial benefits for women with breast cancer. *Journal of Computer Mediated Communication*, 10(4), 00–00.
- Sandaunet AG (2008). The challenge of fitting in: Non participation and withdrawal from an online self help group for breast cancer patients. *Sociology of Health & Illness*, 30, 131–144.18254837
- Spears R , Lea M , Corneliussen RA , Postmes T , & Haar WT (2002). Computer-mediated communication as a channel for social resistance: The strategic side of SIDE. *Small Group Research*, 33, 555–574.
- Tidwell LC , & Walther JB (2002). Computer mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, 28, 317–348.
- Turner JW , Grube JA , & Meyers J (2001). Developing an optimal match within online communities: An exploration of CMC support communities and traditional support. *Journal of Communication*, 51, 231–251.
- Turner JW , Robinson JD , Tian Y , Neustadt A , Angelus P , Russell M , & Levine B (2013). Can messages make a difference? The association between e-mail messages and health outcomes in diabetes patients. *Human Communication Research*, 39, 252–268.
- Uchino BN , Cacioppo JT , & Kiecolt-Glaser JK (1996). The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin*, 119, 488–531.8668748
- van Uden-Kraan CF , Drossaert CHC , Taal E , Lebrun CEI , Drossaers-Bakker KW , Smit WM , & Van de Laar MAFJ (2008). Coping with somatic illnesses in online support groups: Do the feared disadvantages actually occur?. *Computers in Human Behavior*, 24, 309–324.
- Vlahovic TA , Wang Y , Kraut RE , & Levine JM (2014, 4 26). Support matching and satisfaction in an online breast cancer support community In Jones M , Palanque P , Schmidt A , & Grossman T (Eds.). *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paper presented at The ACM CHI Conference on Human Factors in Computing Systems, Toronto, CA (pp. 1625–1634). Toronto, CA: Special Interest Group on Human-Computer Interaction.
- Walther JB , & Boyd S (2002). Attraction to computer-mediated social support In Lin CA & Atkin D (Eds.), *Communication technology and society: Audience adoption and uses* (pp. 153–188). Cresskill, NJ: Hampton Press.
- Walther JB (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23, 3–43.
- Weaver JB , Thompson NJ , Weaver SS , & Hopkins GL (2009). Healthcare non-adherence decisions and internet health information. *Computers in Human Behavior*, 25, 1373–1380.

Figure 1. Users' cancer stage: Percentage of least active vs. most active*

*16% of most active users did not include a stage of cancer in their profiles.

Table 1.

Results of Liu dictionary and python scripts

Corpus characteristics	All users N (%)	Least active N (%)	Most active N (%)	P*
Total users	498 (100)	155 (31.12)	25 (5.02)	
Word count	1,110,262 (100)	6,669 (0.60)	890,313 (80.19)	
Sentiment terms	123,832 (11.15)	1,066 (15.98)	93,406 (10.49)	
Negative terms	43,616 (3.93)	495 (7.42)	31,896 (3.58)	< 0.05
Positive terms	80,216 (7.23)	571 (8.56)	61,510 (6.91)	< 0.05

Table 2.

LIWC output for biggest differences between least active and most active

Dimension characteristics	Least active %	Most active %	Difference (ratio)
we	0.19	0.43	2.263157895
swear words	0.08	0.17	2.125
family	0.14	0.4	2.857142857
friends	0.08	0.16	2
negative emotion	3.27	1.67	1.958083832
anxiety	0.89	0.32	2.78125
sadness	0.69	0.33	2.090909091
causation	1.85	0.87	2.126436782
tentativeness	3.59	1.83	1.961748634
certainty	1.75	0.99	1.767676768
exclusion	2.54	1.48	1.716216216
feel	2.01	0.72	2.791666667
sexual	0.36	0.85	2.361111111
achievement	2.27	1.21	1.876033058
money	0.17	0.72	4.235294118
religion	0.3	0.11	2.727272727
death	0.08	0.03	2.666666667
question	1.06	0.48	2.208333333