# Does Sustained Participation in an Online Health Community Affect Sentiment?

**Shaodian Zhang[1], Erin Bantum, PhD[2], Jason Owen, PhD, MPH[3], Noémie Elhadad, PhD[1]**
**[1]Columbia University, New York, NY; [2]University of Hawai'i Cancer Center, Honolulu, HI; [3]VA Palo Alto Health Care System, Menlo Park, CA**

**Abstract**

*A large number of patients rely on online health communities to exchange information and psychosocial support with their peers. Examining participation in a community and its impact on members' behaviors and attitudes is one of the key open research questions in the field of study of online health communities. In this paper, we focus on a large public breast cancer community and conduct sentiment analysis on all its posts. We investigate the impact of different factors on post sentiment, such as time since joining the community, posting activity, age of members, and cancer stage of members. We find that there is a significant increase in sentiment of posts through time, with different patterns of sentiment trends for initial posts in threads and reply posts. Factors each play a role; for instance stage-IV members form a particular sub-community with patterns of sentiment and usage distinct from others members.*

## Introduction

Online health communities, such as forums, blogs, and health-related Facebook or Yahoo groups, have become popular places for patients to exchange information with and seek psychosocial support from their peers[1–3]. Content analysis of online health communities has shown that there are primarily two types of support members provide to each other: informational and emotional support[4–7]. For patients with chronic or life-threatening diseases, there is evidence that psychological distress and anxiety related to medical decision making process as well as daily coping with the disease can be alleviated through the emotional support obtained in a community[8]. Like for other conditions, patients with breast cancer as well as caregivers of patients, rely on cancer-specific online health communities for both informational and emotional support[4,9–11]. Observational studies of breast cancer communities based on analysis of questionnaires and surveys of their members have indicated a positive association between a member's community participation and emotions such as empathy and satisfaction[10,12–15]. Meanwhile, content analysis of online patient-authored text has provided new perspectives on the health impact of online social networking[16,17], but such analysis usually requires manual annotations which can be costly when the contents are in large scale. As such, automated solutions that can be leveraged to study outcomes of online participations are needed.

More recently, automatic sentiment classification methods have been exploited to investigate sentiments of forum posts published by patients. For instance, studies found that thread originators change their sentiment in a positive direction through reviewing others' replies and self-replying[18], and such changes are largely resulting from postings of influential users[19]. The studies also find that within threads, sentiment changes are correlated with several factors such as number of self-replies, number of replies by others, and length of replies. In the general natural language processing community, sentiment analysis has been carried out on various genres of texts such as product reviews[20], news and blogs[21], and tweets[22].

In this study, we focus on a large online breast cancer community and seek to understand the effect of changes in post sentiment overall through sustained participation in a community. We leverage automated sentiment analysis to conduct large-scale analysis over all the posts in the community. But instead of examining sentiment changes within threads, we examine changes of sentiment from a longitudinal standpoint. We seek to answer the following two research questions: (1) does member participation in the community over different periods of time have an impact on the member posts' sentiment? And (2) do the following factors contribute to changes in posts' sentiment: age of members, cancer stage of members, duration of membership, and amount of posting affect?

## Methods

To explore changes in post sentiments in an online health community, we carried out the following steps. First, we collected all the posts in a large, public community. We trained and evaluated an automated sentiment analysis tool, specific to the community at hand. We applied the sentiment analysis to all posts and assessed the changes in sentiment through various factors of interest both in a static and longitudinal fashion. The study was reviewed and approved by the Columbia University Medical Center IRB.

***Dataset***

We crawled, collected, and analyzed data from the publicly available discussion board of breastcancer.org, one of the most popular online breast cancer communities. The discussion board is organized in several forums, each with threads and posts. At the time of collection, dataset consisted of 291,528 posts in 31,034 threads, published by 12,819 community members between May 2004 and September 2010[23]. Metadata including user profiles was also extracted, consisting of self-reported demographics, diagnosis histories, and treatment histories.

***Automated Sentiment Analysis – Annotation, Training, and Testing***

Since this study relies on automated sentiment analysis on all posts, we built our own sentiment analysis classifier, specific to our dataset, to ensure accuracy and robustness on this particular community.

**Sentiment Annotation.** A random sample of 1,000 posts from the dataset was manually annotated by two annotators according to the sentiment they conveyed overall[24]. To ensure annotators chose a polarity, we restrained the annotation to positive or negative only (no neutral), and provided guidelines and examples to the annotators. Overall, a post was considered positive if its author conveyed typical positive emotions, like joy, happiness, gratitude, as well as curiosity, independently of the topic discussed. Conversely, a post was considered negative if it conveyed negative emotions, such as anger, anxiety, sadness, and hopelessness. Disagreements between the two annotators were adjudicated, resulting in a dataset of 1,000 posts annotated as either positive or negative sentiment.

**Sentiment Classification.** The annotated 1,000 posts were used to train and test binary sentiment classifiers. We experimented with three established robust classifiers: Maximum Entropy[25], Adaboost[26], and Support Vector Machine (SVM)[27]. Among them, Adaboost outperformed other models in a similar sentiment classification task[18], but over a dataset different from ours and with different features. Each classifier was evaluated through 5-fold cross validation according to accuracy, AUC of the ROC curve, and F measures of positive and negative classes.

**Table 1.** Features used for sentiment classification.

| General linguistic features | |
|---|---|
| *Words*: number of words<br>*PosWords*: number of emotionally positive words<br>*NegWords*: number of emotionally negative words<br>*AvgWdLen*: average word length<br>*Sen*: number of sentences<br>*Qmarks*: number of question marks | *Emarks*: number of exclamation marks<br>*NltkProb*: probability of being positive generated by the online NLTK based sentiment classifier<br>*NltkProbNtr*: probability of being neutral generated by the online NLTK based sentiment classifier |
| **Domain-specific features** | |
| *Symp*: number of domain-specific symptoms mentioned | *Meds*: number of domain-specific medication or treatment methods mentioned |
| **Genre-specific features** | |
| *PosEmo*: number of positive emoticons<br>*NegEmo*: number of negative emoticons | *Person*: number of person names |

We exploited several types of features to build the sentiment classifiers, from general linguistics to domain and genre specific features, as listed in Table 1. Features specific to the genre of online community and social media included emoticon lists for extracting *PosEmo* and *NegEmo* features (from http://en.wikipedia.org/wiki/Emoticon), as well as presence of people's names (personal names were tagged automatically using the Stanford Named Entity Recognizer [28] over the dataset as part of the pre-processing). General linguistic features included number of words in the post, and dictionary matching based features like number of emotionally positive/negative word stems. *PosWords* and *NegWords* were extracted by looking up two adjective lists: *glad, happy, relieved, grateful, excited, thrilled, thankful, great, lucky, pleased, blessed, fortunate, hopeful, inspiring, encouraging;* and *scared, sad, anxious, embarrassing, disappointing, confused, heartbreaking, frightened, frustrated, angry, upset, distress, stress, discouraging,* as well as their morphological variants (e.g. frustrated -> frustrating). Finally, to include other general linguistic features, we leveraged the output of a robust sentiment classifier which uses the NLTK package[29] and returns the probability of a post to be negative or otherwise, its probability of being positive.

For domain-specific features, we focused on mentions of medical terms in the posts, like treatments and side effects[30]. As such, recognizing these domain-dependent medical terms, which form a sublanguage of breast cancer communities, is a critical step in our analysis[31]. For example, in our dataset, since Tamoxifen is a widely used medication for breast cancer patients, there are a large amount of abbreviations and misspellings such as “tamox”, “tamo”, and “tamoxifan” referring to this medicine. In order to capture these variations without relying on dataset-specific knowledge, we used an unsupervised, domain independent, distributional semantics based method[32] to generate two lexicons for symptoms and medications, respectively (features *Symp* and *Meds*).

### *Impact of Different Factors on Post Sentiment*

The automated sentiment analysis output for each post a predicted probability of being positive, or sentiment score. The sentiment scores are useful, because they allow us to compare posts against each other. As such, the scores are not absolute representation of sentiment, but rather enable us to rank posts according to their sentiment polarity.

Armed with such sentiment score for each post in the dataset, we conducted the following analyses. The primary objective for our study was to assess if participation in the community has an impact on sentiment. We thus compared average sentiment scores of posts published in different periods of time with respect to user’s registration date, and tracked changes of sentiment. As such, each data point is the average sentiment of all posts in a given time slice (e.g., all posts published by their authors after 3 weeks of their joining the community). To visualize the changes in sentiment through time, we plotted in addition to the individual data points a fitted curve.

For our second research question, we considered three factors (age of members, cancer stage of members, and amount of posting) in both static and longitudinal analyses to examine their impact on post sentiment. In the static analysis, members were stratified by age/stage/amount of posting, and average post sentiments were calculated for each group. Statistical tests (ANOVA and TukeyHSD[33]) were carried out to detect differences across groups. In the longitudinal analysis, sentiment scores were compared across stratified groups and duration of participation in the community to identify the patterns of sentiment change across members from different groups through time. All p-value were adjusted for multiple comparisons with the Bonferroni correction.

For both research questions, we distinguished in our analyses the initial posts (i.e., first posts that initiate threads) and all other posts. Previous research[19] found that community members expressed significantly different polarities of emotions in the initial post of a thread compared with other posts. This could be explained by the fact that the post originators were more likely to express concerns and seek support, while responses to such posts tended to be more positive by conveying encouragement and empathy.

## Results

### *Data Annotation and Sentiment Classification*

The manual sentiment annotation of the 1,000 yielded good inter-annotator agreement (Cohen’s Kappa of 0.798)[34]. After adjudication and resolving disagreements, 728 out of 1,000 posts were annotated as positive, and 272 were annotated as negative. Examples of two positive and two negative posts are given in Table 2.

**Table 2.** Example posts of positive and negative sentiments.

| *Positive posts* | *Negative posts* |
|---|---|
| The recovery from my lumpectomy was easy. Really. Nowhere near as difficult as I imagined. Very little pain at all. I never needed any pain meds after surgery. Good luck. | I had a mastectomy about three weeks ago and will be starting chemo at the end of the month (Dec. 27th). I wake up every morning anxious and scared. When does this go away? |
| I'm so happy you're feeling better!! Strange, but hey, that's our life these days. ! | Just had a 6month followup with my onc. My second round of scans came out clean. However in 3 months I will be doing bloodwork for tumor markers. She didn't discuss it with me and I don't know what it is about. I understand my cancer is aggressive, but what am I not understanding here? :( |

The classification performances of the three classifiers are given in Table 3. To demonstrate the effectiveness of machine learning models, performance of a baseline system is also given, which simply classified all posts as positive. The best performing system was Maximum Entropy (MaxEnt), followed by SVM and AdaBoost. Both MaxEnt and AdaBoost tended to classify posts as positive, caused by the uneven distribution of positive and negative samples in the training set. For MaxEnt, once the threshold of prediction was calibrated towards favoring negative (i.e., a post is classified as negative once the predicted probability was lower than 0.6 rather than 0.5), the F score of negative was dramatically improved. Fortunately, in our application to the entire dataset, we are more concerned with probabilities rather than discrete labels, since our modeling was based on the average likelihood of various groups of posts being positive or negative, rather than number of predicted positive and negative instances. In the remainder of the study, we relied on the MaxEnt classifier to output a sentiment score for each post.

**Table 3**. Performance of different sentiment classifiers according to Area under Curve of ROC, accuracy, and F scores for positive and negative sentiment polarity respectively. The baseline system classified all posts as positive.

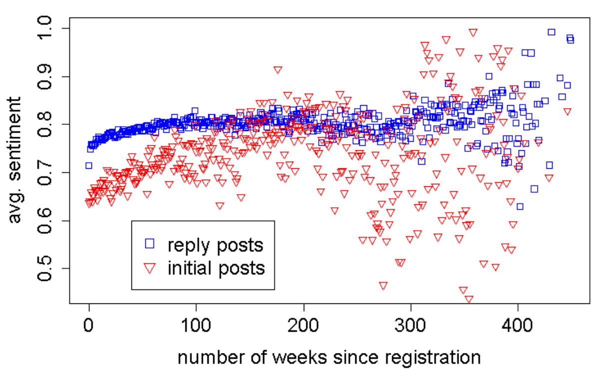
| | AUC of ROC (95% CI) | Accuracy (95% CI) | F (positive) (95% CI) | F (negative) (95% CI) |
|---|---|---|---|---|
| MaxEnt | 82.0% (2.7) | 79.4% (1.8) | 86.8% (1.9) | 53.7% (2.8) |
| AdaBoost | 76.0% (3.3) | 76.3% (1.5) | 84.6% (2.1) | 48.5% (3.9) |
| SVM | 78.1% (2.9) | 73.4% (2.9) | 68.4% (1.9) | 58.4% (1.4) |
| Baseline | 49.8% (0.8) | 72.9% (0.7) | 84.3% (1.0) | 0% (0) |

We analyzed the impact of individual features on the MaxEnt classifier, which assigns a weight to each feature after training, indicative of its discriminative power for the given task. Among all features, *NltkProb* (weight +2.7) had the strongest correlation with positive emotion, while *NegEmo* (weight -1.9) and *NegWords* (weight -1.2) were most correlated with negative emotion. On the contrary, *Words* (weight 0.003) and *Emarks* (weight 0.03) were borderline features, suggesting similar distributions of these features in positive and negative samples.

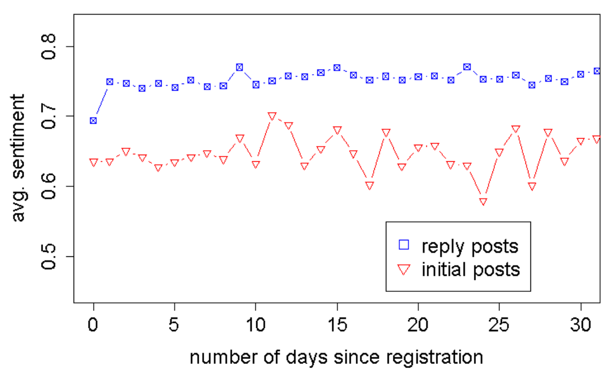
***Participation and Posts Sentiment – Static and Longitudinal Analyses***

The best performing classifier, the MaxEnt, was applied to the entire dataset based on the model trained with the 1,000 annotated samples. For each post in the dataset, a sentiment score (probability of post being positive) was calculated. The average sentiment score of the entire dataset was 0.785 (0.210 standard deviation). For the initial posts, the average sentiment score was 0.695 (0.263 standard deviation). In general, our research aligned with previous work on other online health communities that found initial posts to be less positive.

In order to examine the impact of participation through time in online discussion on sentiment overall, we plotted how sentiment scores changed through time, as computed since members' registration date. The registration dates of users were provided in the profile information of metadata. Figure 1 shows the average sentiment scores of posts that were published after membership creation at both weekly (a) and daily (b) intervals. For example, the left-most blue data point in Figure 1(a) represented the average sentiment score of all reply (i.e., non-initial) posts published by all users respectively within one week of their joining the community.




(a)




(b)

**Figure 1**. Sentiment changes by length of membership at the time of posting, by number of weeks in (a) and number of days in (b). A colored point at (x, y) in the graph represents that the average sentiment score of all posts published by all users in the xth week (a) or day (b) after their registration is y.

Figure 1(a) indicates that, for both responding and initial posts, sentiment gets more and more positive through at least 100 weeks (2 years) of participation, with such changes most significant right after joining the community. Members, in their first days joining the community, publish posts, which are significantly more negative than later on. This is particularly true for initial posts, suggesting that newcomers to the community (likely newly diagnosed patients) express more anxiety and concerns than later in their questions to the community. Figure 1(b) provides a more granular view over the sentiment changes in the first 30 days of participation in the community, confirming that reply posts are significantly more positive than initial posts, and the increase of sentiment of initial posts does not happen until later on, at least 1 month into participation in the community. We do note a drastic increase in sentiment from posts published on the first day of joining the community to the later days, when looking at all posts (replies and initial posts combined).

In our dataset, the average length of membership of all users was 2 years 5 months (around 120 weeks); therefore, most of posts published after 200 weeks were written by a small portion of long-time users. We found that most of them were stage IV patients and showed a slight sentiment decline between 200 and 300 weeks. Topics of these posts were primarily about chemotherapy or metastasis/recurrence. While this set of posts is indeed homogeneous in sentiment and topic, it is difficult to assess the value of the analysis on such a small sample for the posts written by members who have been more than four years active in the community.

In order to obtain a more concrete understanding of how sentiment changed through sustained participation in the community, we grouped posts into nine groups, considering both short-term and long-term periods of participation. The nine groups were posts published within one day of registration, 1-3 days, 3 days to 1 week, 1 to 2 weeks, 2 weeks to 1 month, 1 to 3 months, 3 months to 1 year, 1 to 2 years, and more than 2 years since registration. An ANOVA test was carried out for the groups, for all posts and initial posts respectively, followed by a TukeyHSD test to illustrate the significances of differences between all possible group pairs. ANOVA test showed significant difference among groups in both cases (p values << 0.001). Post distribution, average sentiment scores, and p values compared with previous category given by TukeyHSD test are listed in Table 4. In this table as well as following tables, "all posts" represent initial posts and reply posts. Results showed same pattern as Figure 1, and demonstrated that the dramatic sentiment change after the first day was statistical significant in the case of all posts, while we could only see long term (3 months and then 1 year) significant changes for initial posts.

**Table 4.** Post distribution, average sentiment scores, and p values compared with previous category returned by TukeyHSD test, for all posts and initial posts respectively. The first p value for <1d is not available since there is no previous category to compare sentiment to. P values are adjusted for multiple comparisons with the Bonferroni correction.

| | | <1 d | 1-3 d | 3d – 1w | 1-2w | 2w-1m | 1-3m | 3m-1y | 1-2 y | >2 y |
|---|---|---|---|---|---|---|---|---|---|---|
| All posts | Sentiment | .693 | .748 | .745 | .753 | .756 | .766 | .782 | .800 | .804 |
| | # posts | 8,369 | 4,203 | 4,361 | 6,235 | 9,906 | 32,302 | 89,304 | 60,944 | 75,781 |
| | p value | N/A | <<0.001 | 1.000 | 1.000 | 1.000 | 0.025 | <<0.001 | <<0.001 | 0.577 |
| Initial posts | initial | .636 | .642 | .637 | .656 | .644 | .664 | .685 | .728 | .760 |
| | # posts | 3,304 | 732 | 734 | 1,064 | 1,487 | 3,842 | 8,085 | 5,134 | 6,641 |
| | p value | N/A | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.032 | <<0.001 | <<0.001 |

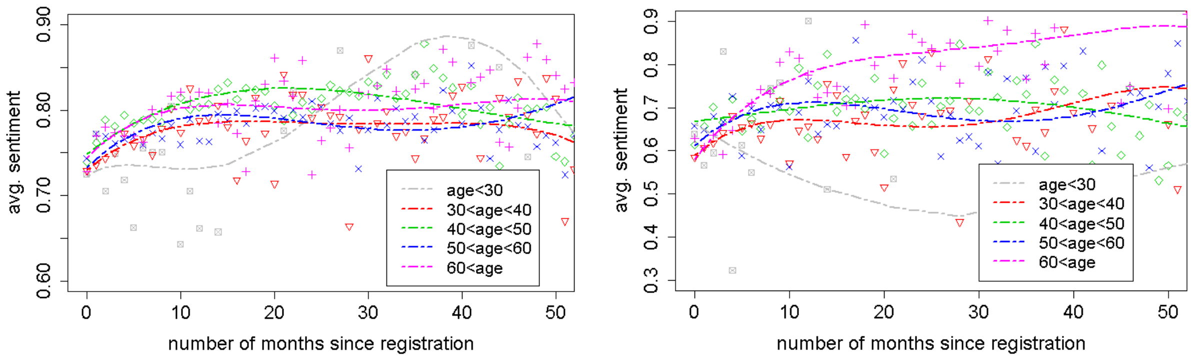
### *Impact of Members' Age on Sentiment*

The posts in the dataset were published by 12,819 users, while a total of 14,919 user profiles were filled at least partially in the online breast cancer community and there were about 60,000 members overall. This meant that a very large majority of members were so called lurkers [35,36], who never published anything but were likely to browse some of the posts. Behavior of lurkers was beyond the scope of this study. Rather, we focused on members who had posted content. Among all non-lurkers, 1,211 provided date of birth in their profiles. Members born between 1960 and 1970 were the most dominant at the time of data collection, and the average age of all users were 47.5 (standard deviation 9.6 years), an older mean than in some other online health communities, such as weight loss forums[37].

**Table 5.** Average sentiment scores and number of posts published by different age groups, for all posts and initial posts respectively. This analysis is restricted to posters who provided date of birth in their profile only, 1,211 members overall.

| Age group (# users) | | <30 (38) | 30-40 (198) | 40-50 (485) | 50-60 (358) | 60+ (132) |
|---|---|---|---|---|---|---|
| All posts | Sentiment | 0.742 | 0.768 | 0.793 | 0.778 | 0.791 |
| | # posts | 278 | 6,417 | 22,180 | 14,479 | 4,217 |
| Initial posts | Sentiment | 0.614 | 0.643 | 0.681 | 0.681 | 0.744 |
| | # posts | 54 | 841 | 1,873 | 1,323 | 339 |

To study whether age affected sentiment, we considered members who disclosed their date of birth, and grouped them into 5 groups: below 30 years old, between 30 and 40, between 40 and 50, between 50 and 60, and above 60 years old. There were 47,571 posts in the dataset published by members with date of birth information. We calculated averaged post sentiment scores, and carried out statistical tests for the groups. Table 5 shows numbers of posts published by each age group and average sentiment score of posts of each group. The ANOVA test showed significant differences among groups for both all posts and initial posts. For all posts, TukeyHSD test found that difference between all pairs of groups were significant, except between <30 and 30-40, <30 and 50-60, and between 40-50 and 60+. For initial posts, differences between <30 and all other groups were not significant. We suspect that this is caused by the very low number of members in the age group <30, as expected in a community for a disease that affects older women predominantly. Members older than 60 showed markedly more positive sentiment than younger members, especially while publishing initial posts to start new threads. These facts might be explained by previous psychological finding of effects of older age on lower levels of psychological distress[38–40].




(a) all posts (b) initial posts

**Figure 2.** Sentiment changes by length of membership at the time of posting for different age groups, for (a) all posts and (b) initial posts. A colored point at (x, y) in the graph represents that the average sentiment score of all posts (a) or initial posts (b) published by users in corresponding age group in the xth month after their registration is y. Polynomial curves fitting each group were drawn for the sake of visualization.

To illustrate age's impact on longitudinal sentiment, sentiment changes over time after registration for different age groups were plotted, along with polynomial curves fitting each set of points to visualize the tendencies (Figure 2). Keeping in mind the very low sample size for members <30 years old, we do not attempt to interpret their longitudinal sentiment changes. For all other groups, however, the general trend observed earlier holds true independently of age: the longer the members participate in the community, the more positive their posts are on average. The observation that older members (>60 years old) post more positive posts, especially initial posts is visible as well on the plots.

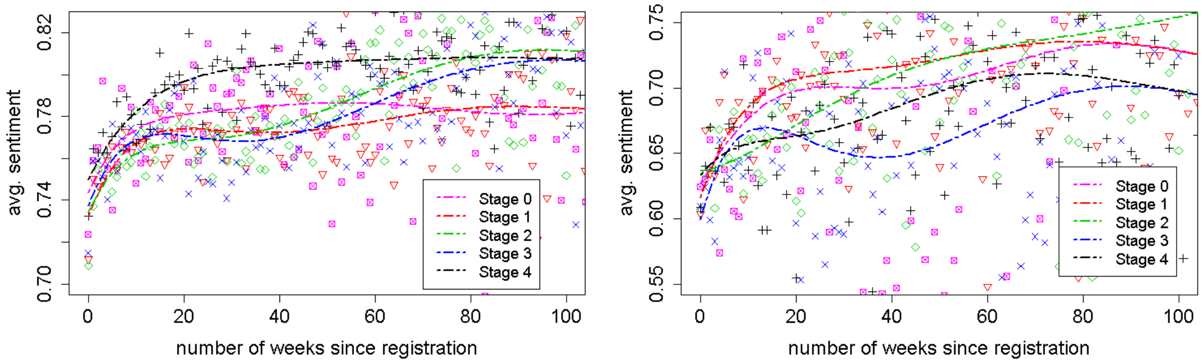
***Impact of Member's Cancer Stage on Sentiment***

In our dataset, 4,602 users (who published 172,566 posts) had self-reported cancer stage information. Among them, 442 members were stage 0 patients, 1,407 were stage I, 1,544 were stage II, 650 were stage III, and 559 members were stage IV. Table 6 provides numbers and average sentiment scores of posts published by members in different stages. Although there were significantly fewer stage IV patients than stage I and II patients, they published many more posts and formed the most active cancer stage group in breast cancer forum[23]. Moreover, stage IV patients were the most positives posters in term of the emotion expressed through the reply posts they wrote, but not initial posts. For all posts, comparisons between stage 0, stage I, and stage II, returns non-significant results according to adjusted p values.. For initial posts, only the differences between stage I and stage III and between stage II and stage III were significant.

**Table 6.** Average sentiment scores and number of posts published by patients in different stages, for all posts and initial posts respectively.

| Cancer stage (# users) | | Stage 0 (442) | Stage I (1,407) | Stage II (1,544) | Stage III (650) | Stage IV (559) |
|---|---|---|---|---|---|---|
| All posts | Sentiment | 0.775 | 0.771 | 0.776 | 0.782 | 0.796 |
| | # posts | 9,229 | 36,422 | 39,398 | 27,806 | 59,711 |
| Initial posts | Sentiment | 0.675 | 0.690 | 0.687 | 0.661 | 0.675 |
| | # posts | 820 | 3,344 | 4,218 | 2,534 | 4,829 |

Figure 3 illustrates longitudinal sentiment of different cancer stage groups. Not only were the stage IV users the most positive, but they also showed the fastest change towards positive after registering in the breast cancer forum. However, these findings were specific to reply posts. These findings indicate that stage IV users seek support through starting threads with negative posts, but are very active in providing emotional support to their peers, through posting positive replies.




(a) all posts (b) initial posts

**Figure 3.** Sentiment changes by length of membership at the time of posting for different cancer stage groups, for (a) all posts and (b) initial posts. A colored point at (x, y) in the graph represents that the average sentiment score of all posts (a) or initial posts (b) published by users in corresponding cancer stage in the xth month after their registration is y. Polynomial curves fitting each group were drawn for the sake of visualization.

### *Impact of Member's Posting Activity on Sentiment*

The last factor we considered was the amount of posting by each individual. Table 7 groups members into 5 groups by number of posts, listing the distributions and average sentiment of each group. There were 8247, 3527, 757, 255, and 24 profiles in the 5 groups respectively. Although members who published less than 50 times wrote only 20% of all posts, approximately half of the initial posts were authored by these members. This suggests that new members tend to seek information and support while long-time members provided information and support more than they requested it. All differences of sentiment scores between groups, including both all posts and initial posts, were significant, except between group of < 5 and 5-50 for initial posts.

**Table 7.** Average sentiment scores, number of posts published by patients, and number of posts published per user in different stages, for all posts and initial posts respectively.

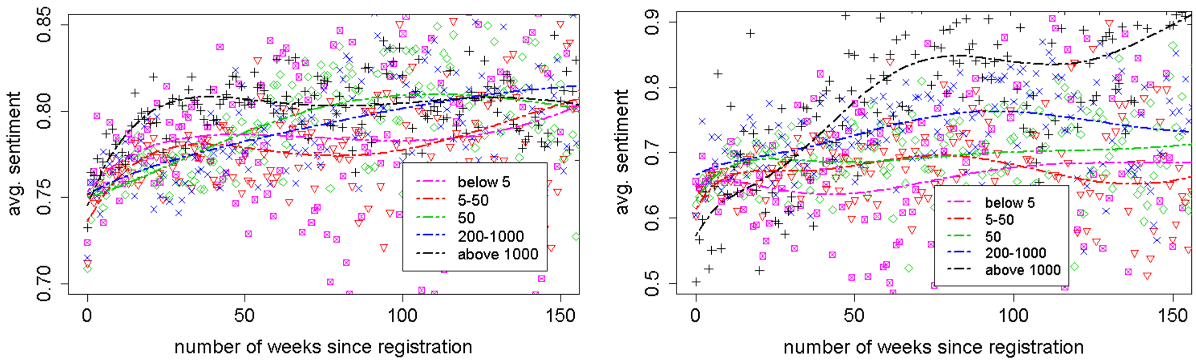
| User post number (#users) | | < 5 (8,247) | 5-50 (3,527) | 50-200 (757) | 200-1000 (255) | 1000+ (24) |
|---|---|---|---|---|---|---|
| All posts | Sentiment | 0.727 | 0.754 | 0.779 | 0.806 | 0.817 |
| | # posts | 16,725 | 36,422 | 73,951 | 102,466 | 39,944 |
| | avg # post | 2.0 | 10.3 | 97.7 | 401.8 | 1664.3 |
| Initial posts | Sentiment | 0.657 | 0.658 | 0.683 | 0.730 | 0.828 |
| | # posts | 4,565 | 9,445 | 7,399 | 6,635 | 2,990 |
| | avg # post | 0.6 | 2.7 | 9.8 | 26.0 | 124.6 |

Figure 4 illustrates how sentiment changed over time for different groups of members with different posting activity count. In general, active members (i.e., with more posts authored) were likely to gain sentiment improvement faster and more significantly. It is particularly interesting to note that although members posting more than 1,000 times throughout their time in the community, and who were long-time users, had a significantly higher sentiment score in average, their sentiments were as negative as other members when they just joined the forum, especially for their initial posts. The pattern seen in Table 7 and Figure 4 seems to suggest that long-time users, who suffered from cancer but benefited from hearing from their peers online at early stages of participation, changed their roles in the forum later and acted as information and support providers more than requesters. Such role change should be another important outcome of online discussion participation.

## Discussion

### *Principal Findings*

Our study results suggest that members benefit from sustained participation in a breast cancer community with respect to the sentiment they convey through their posts. At the early stages of participation, sentiment of users usually increased significantly, and the rate of improvement dropped after several weeks, followed by a slower positive sentiment increase which could last for as long as several years. Our study also showed that compared with




(a) all posts (b) initial posts

**Figure 4.** Sentiment changes by length of membership at the time of posting for different groups of posting amount, for (a) all posts and (b) initial posts. A colored point at (x, y) in the graph represents that the average sentiment score of all posts (a) or initial posts (b) published by users grouped by their number of posts in the xth month after their registration is y. Polynomial curves fitting each group were drawn for the sake of visualization.

reply posts, initial posts of threads were more emotionally negative, especially at the beginning of participation. Sentiment increases of initial posts were more dramatic but long term. A qualitative analysis over the forum data showed that newcomers of the forum were more likely to be newly diagnosed or post-treatment patients. For most of them, going online was the choice when some of their needs, either informational or emotional, could not be met in other settings such as family and hospitals. As a result, we found a large amount of posts with strong negative sentiments, especially initial posts, published by newcomers asking various questions about cancer symptoms, medication use and side effects, and choices of therapeutic method, which were the issues usually brought up by individuals with little cancer or treatment experiences. In contrast, long-time members were more likely to be cancer survivors or patients who were recovering or being treated as a routine part of their lives. It is likely they were more experienced, empowered, and acted more as informational and emotional support providers rather than requesters, and were expressing more encouragement and empathy in the threads in which they participated. The different patterns of reply posts and initial posts also suggested that people immersed themselves quickly into the discussion by learning to encourage others and provide information through replying, but were still concerned about their own issues.

Our study examined three factors' impacts on sentiment and sentimental changes: age, cancer stage, and amount of posting. We showed that all three factors had an impact on the members' sentiment on average. Statistically significant differences were found for every stratified group. For age, we found that users older than 60 years old showed the most positive sentiment, especially while publishing initial posts. There were no significant differences between longitudinal aspects of different age groups. With respect to cancer stage, although there were significantly fewer stage IV patients than any other stage, they published many more posts and formed the most active cancer stage group in the breast cancer forum. They showed the fastest change towards positive sentiment after registering in the breast cancer forum. They also were the most positive in their replies, while the most negative in their initial posts. The last factor, amount of posting, also made a difference. Members who published less than 50 posts, mostly newcomers and lurkers, were responsible for only 20% of all posts, but around half of the initial posts were authored by these users, which indicated that new users and lurkers tended to seek information and support while long-time members provided information and support more than requested it. Long-time members, who suffered from cancer but benefited from hearing from their peers online at early stages, later changed their roles in the forum later and acted more as information and support providers.

**Limitations**

Our study was exploratory and has several limitations. First, the analyses rely on the output of an automatic sentiment classifier, which while providing state-of-the-art accuracy is not 100% accurate. Further feature engineering has the potential to improve the classification accuracy. Second, the classification was defined as a two-category problem: positive and negative, and documents with neutral sentiment were simply regarded as ones whose sentiment scores lie near the boundary of the binary classification. Since the sentiment scores are a mean to comparing posts in aggregate, it might be also useful to leverage a more granular classification, or at least one that considers neutral as a category on its own[41]. Third, profile information, especially cancer stage, was extracted at the time of data collection, and such information might have been edited by members through time, as their disease evolved. Finally, this study was conducted on a single online health community. It will be interesting to see the impact of these factors in different communities specific to breast cancer as well as to or other chronic conditions.

### Future Work

This study brings up several research questions we would like to explore in the future. While one interpretation of our findings is that sustained participation in an online health community overall increases the sentiment of members' posts, we must acknowledge that there is a strong uncertainty to this interpretation due to right censoring issues, common to longitudinal observational studies. As the number of members who stay in the community decrease with time since registration, one must think of potential reasons for the right censoring of data: is it that members with adverse health outcomes were too sick to continue participating in the community, or even that individuals who did not receive appropriate support from their peers stopped participation. In other words, it is possible that only the people for which the community is beneficial emotionally are the ones that stay in the community through time, while others simply stop posting. Under this assumption, there is no causal link between participation in the community and positive sentiment on average.

We did not examine the impact of lurking (and of its duration before posting for the first time) on participation and sentiment in particular. Because the community we studied does not keep track of members' reading activity, this is a difficult question to study quantitatively, but an important one to consider in the future.

Another area of research we plan to explore further is that the sentiment of the posts alone is a rough representation of the sentiment or emotions of community members. As we refine our understanding of the different topics conveyed in an online health community, it will be critical to understand the relationship between sentiment and different topics. For instance, a member's anger at an insurance company refusing to reimburse her treatment and a member's anxiety faced with a dire test result represent very different aspects of sentiment with respect to cancer in general. The longitudinal evolution of topics and their associated sentiments for community members is an area in much need for further analysis.

### Conclusion

This paper carried out an exploratory study over a popular public online community for breast cancer and used automated sentiment analysis to investigate correlations between sentiment changes of users and different participation-relevant factors. Finding suggests that as participation is sustained, posts' sentiment increases towards positive. Further, members convey more positive posts when replying to their peers than when initiating a thread. In addition, we discovered that users in different ages, cancer stages, and stages of participation showed different sentiment patterns. Most significantly, members over 60 years old and stage IV members were expressing more positive sentiment than any other groups of people, while newcomers to the community tend to post more negative initial posts than long-time members. This study contributes to further the understanding of community participation on members' attitudes and opens up to a number of research questions to explore further.

### Acknowledgements

This work is supported by NSF (National Science Foundation) award #1027886 (NE) and NCI (National Cancer Institute) award R21 CA143632 (EB, JO, NE). We are grateful to Rimma Pivovarov for the helpful discussions.

## References

1. Medicine 2.0: Peer-to-peer healthcare. Available at: http://www.pewinternet.org/Reports/2011/Medicine-20.aspx.
2. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ*. 2004;328(7449):1166.
3. Ziebland S, Chapple A, Dumelow C, Evans J. How the internet affects patients' experience of cancer: a qualitative study. *BMJ*. 2004;328(7439):564.
4. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. *J Med Internet Res*. 2007;9(2):e12.
5. Wang Y, Kraut R, Levine J. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. *Proc ACM 2012 Conf Comput Support Coop Work*. 2012:833–842.
6. Chung JE. Social Networking in Online Support Groups for Health: How Online Social Networking Benefits Patients. *J Health Commun*. 2013.
7. Vlahovic TA, Wang Y, Kraut RE, Levine JM. Support Matching and Satisfaction in an Online Breast Cancer Support Community. *Proc 32nd Annu ACM Conf Hum factors Comput Syst*. 2014:1625–1634.

8. Kim E, Han JY, Moon TJ, et al. The process and effect of supportive message expression and reception in online breast cancer support groups. *Psychooncology*. 2012;21(5):531–40.
9. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Educ Couns*. 2004;53(1):57–64.
10. Overberg R, Otten W, de Man A, Toussaint P, Westenbrink J, Zwetsloot-Schonk B. How breast cancer patients want to search for and retrieve information from stories of other patients on the internet: an online randomized controlled experiment. *J Med Internet Res*. 2010;12(1):e7.
11. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. *Am Psychol*. 2000;55(2):205.
12. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. *J Am Med Informatics Assoc*. 2011;18(3):298–304.
13. Beaudoin C, Tao C. Modeling the impact of online cancer resources on supporters of cancer patients. *New Media Soc*. 2008;10(2):321–344.
14. Sharf BF. Communicating breast cancer on-line: support and empowerment on the Internet. *Women Health*. 1997;26(1):65–84.
15. Han JY, Shah D V, Kim E, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. *Health Commun*. 2011;26(2):185–97.
16. Bender JL, Jimenez-Marroquin M-C, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. *J Med Internet Res*. 2011;13(1):e16.
17. Ma X, Chen G, Xiao J. Analysis of an online health social network. *Proc 1st ACM Int Heal informatics Symp*. 2010:297–306.
18. Qiu B, Zhao K, Mitra P, et al. Get Online Support, Feel Better -- Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. *2011 IEEE Third Int'l Conf Privacy, Secur Risk Trust*. 2011:274–281.
19. Zhao K, Greer G, Qiu B, Mitra P. Finding influential users of an online health community: a new metric based on sentiment influence. *arXiv Prepr arXiv12116086*. 2012.
20. Cui H, Mittal V, Datar M. Comparative experiments on sentiment classification for online product reviews. *AAAI*. 2006.
21. Godbole N, Srinivasaiah M, Skiena S. Large-Scale Sentiment Analysis for News and Blogs. *ICWSM*. 2007.
22. Pak A, Paroubek P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREC*. 2010:1320–1326.
23. Jha M, Elhadad N. Cancer Stage Prediction Based on Patient Online Discourse. In: *Proc ACL BioNLP (Association for Computational Linguistics Bio Natural Language Processing) Workshop*.; 2010:64–71.
24. Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis. *Found Trends® Inf Retr*. 2006;1(2):91–231.
25. Pietra S Della. Inducing features of random fields. *Pattern Anal Mach Intell*. 1997;19(4):380–393.
26. Freund Y, Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55(1):119–139.
27. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. 1999;9(3):293–300.
28. Finkel JR, Grenager T, Manning C. Incorporating non-local information into information extraction systems by gibbs sampling. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*.; 2005:363–370.
29. Sentiment Analysis with Python NLTK Text Classification. Available at: http://text-processing.com/demo/sentiment/.
30. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. *J Med Internet Res*. 2007;9(2):e12.
31. Friedman C, Kra P, Rzhetsky A. Two biomedical sublanguages: a description based on the theories of Zellig Harris. *J Biomed Inform*. 2002;35(4):222–235.
32. Elhadad N, Zhang S, Driscoll P, Brody S. Characterizing the Sublanguage of Online Breast Cancer Forums for Medications, Symptoms, and Emotions. In: *Proc AMIA Annual Fall Symposium*.; 2014.
33. Winer BJ. *Statistical principles in experimental design*. McGraw-Hill Book Company; 1962.
34. Cohen J, others. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46.
35. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. *J Med Internet Res*. 2011;13(4):e122. doi:10.2196/jmir.1696.
36. Van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar M a FJ. Self-reported differences in empowerment between lurkers and posters in online patient support groups. *J Med Internet Res*. 2008;10(2):e18.
37. Hwang K, Ottenbacher A. Social support in an Internet weight loss community. *Int J Med Inform*. 2010;79(1):5–13.
38. Singer J, Rexhaj B, Baddeley J. Older, wiser, and happier? Comparing older adults' and college students' self-defining memories. *Memory*. 2007;15(8):886–98.
39. Hoffman K, McCarthy EP, Recklitis CJ, Ng AK. Psychological distress in long-term survivors of adult-onset cancer: results from a national survey. *Arch Intern Med*. 2009;169(14):1274–1281.
40. Kaiser NC, Hartoonian N, Owen JE. Toward a cancer-specific model of psychological distress: population data from the 2003--2005 National Health Interview Surveys. *J Cancer Surviv*. 2010;4(4):291–302.
41. Schler J. The importance of neutral examples for learning sentiment. In: *In Workshop on the Analysis of Informal and Formal Information Exchange during Negotiations*.; 2005.