Factors Contributing to Dropping-out in an Online Health Community: Static and Longitudinal Analyses

Shaodian Zhang, Noémie Elhadad, PhD
Biomedical Informatics, Columbia University, New York, NY

Abstract

Dropping-out, which refers to when an individual abandons an intervention, is common in Internet-based studies as well as in online health communities. Community facilitators and health researchers are interested in this phenomenon because it usually indicates dissatisfaction towards the community and/or its failure to deliver expected benefits. In this study, we propose a method to identify dropout members from a large public online breast cancer community. We then study quantitatively what longitudinal factors of participation are correlated with dropping-out. Our experimental results suggest that dropout members discuss diagnosis- and treatment-related topics more than other topics. Furthermore, in the time before withdrawing from the community, dropout members tend to initiate more discussions but do not receive adequate response from the other members. We also discuss implications of our results and challenges in dropout-member identification. This study contributes to further understanding community participation and opens up a number of future research questions.

Introduction

Internet and mobile technologies, especially newly emerging social networking applications, are revolutionizing how patients exchange information and social support with care providers, family members/friends, and other patients. Traditionally, patients with life-threatening conditions receive most of the information about their disease from their care providers. While providers tend to focus on the clinical impact of the disease and might ignore the impact of the disease on a patient's emotional wellbeing and daily life,[1] support groups and more recently online health communities (OHCs), act as a complementary source of support for patients.[2,3] Previous research suggested that peer patients are able to appreciate each other's conditions better than health providers, family members and friends, and to exchange necessary emotional support and practical advice of daily health management.[4–13] Recent years have witnessed increasing popularities of OHCs, with a wide range of formats ranging from discussion forums,[2] to Facebook groups,[9,14,15] to dedicated communities like Patientslikeme.[16–18]

Critical to studying OHCs' impact on their members is characterizing and understanding the patterns of participation in a community.[19] Researchers have studied whether users actively participate or lurk,[20] as well as when they decide to withdraw from the community permanently.[21] Lurking—the phenomenon of users browsing the content but not actively participating in discussions— has been shown to correlate with lower perceived social support and diminished emotional benefits when compared to active participation in a community.[20,22–24] Dropping-out— i.e., stopping participation or leaving an intervention, such as an online community, altogether— when examined across members indicates the level of activity in an OHC. For instance, Eysenbach and colleagues reported that the phenomenon of attrition (or dropout) is particularly common in online-based interventions,[21] with more than 90% of study subjects quitting throughout Internet-based studies. In the case of OHCs, understanding factors associated with dropping-out might help identify opportunities for more targeted support of members, and more generally identify for which members participation in an OHC is beneficial and for which it is not. Wang and colleagues examined how type of information received affect users' choices between staying and leaving, and suggested that informational support is positively correlated with dropping-out while emotional support is positively correlated with staying active in the community.[25] Zhang suggested that information and small group interactions, like emotions, also play a key role in retaining users.[26] Sadeque and colleagues proposed a supervised model to predict dropping-out, and found that factors like time since last activity were predictive.[27] To date, however, it is still unclear which other factors of individual members are moderating dropping-out from online health communities, such as topic of discussions, users' sentiment expressions, and interactions among users.

Previously we carried out longitudinal analysis of members' posting history in an online health community to explore how participation affects members' sentiment, and how different factors impact sentiment change through time.[11] We found that members show an increasingly positive sentiment at the early phases of participation, and that later changes may be correlated with different variables like age and cancer stage. In this study we carry out a series of static and longitudinal analyses, which take topic of discussions, sentiment, and user interactions as variables of interest. We explore if and how these factors correlate with members' decisions of dropping-out. Because there is no explicit marker for any participant to convey they dropped out of the community, we explore different approaches to determining that a member dropped out. To explore factors in context of dropping-out, we leverage established machine-learning-based methods for sentiment analysis and topic classification of a given member's posts. We hypothesize that dropout members discuss more disease-specific topics, express more negative sentiments, and interact with other members less actively than the members who stay active in the community. Furthermore, we hypothesize that characteristics of dropping-out can be detected by investigating patterns of changes of these factors.

Methods

The basic workflow of our analysis is as follows. First, we identify members that have dropped out from the community, i.e., members who had history of active participation in community, but have been inactive for a certain amount of time. We collect a set of variables for each members throughout their history of participation in the community. We compare distributions of each variables between dropout members and other members. The longitudinal analyses focus on dropout members to investigate if any patterns of changes of variables exist *before* they drop out the community, with respect to their sentiment expressions, topics of discussions, and interactions with other members.

Data set

The data set we relied on is from the publicly available discussion board of breastcancer.org, one of the most popular online breast cancer communities.[28] The discussion board is organized in several forums, each with threads and posts. At the time of data collection (January 2015), the dataset consisted of 3,282,008 posts in 121,474 threads, published by 57,424 community members. We extracted post content, public available meta-data like author ID and author name, as well as post signatures including user profiles of self-reported demographics, diagnosis and treatment histories.

Identifying dropout members

Identifying which members in a public community dropped out is not a trivial task. In practice, it is impossible to determine with absolute certainty a dropout member from a public community solely based on changes of posting activity, since an inactive user can always return to the community and resume participation. Moreover, in many communities, like our community of interest in this study, there is no publicly available information about members and their login patterns; and as such the only available information relates to their posting activity. Thus, a member could withdraw from posting content, but still act as a lurker.

To identify the cohort of dropout members for our study, we explored different heuristics. We defined a user in the breast cancer forum as a dropout member, if she has posted more than n times in the community (i.e. had some history of posting activity), but has been inactive for at least t years at the time of data collection. The first cut-off is to ensure that users we identify are users who participated in the community discussion meaningfully, instead of one-time information seekers or users who just chimed in a limited number of discussions without real information or support exchanges with other members. The second threshold is to exclude members that may return to the community in the near future, as we assume that users who have been inactive for longer time are less likely to return.

In this particular study, n and t were experimentally set as 10 posts and 3 years. As such, for the remaining part of this paper, dropout members refer to those who have posted more than 10 times in the community, and whose most recent post was before January 2012 (three years before January 2015).

Longitudinal analysis of dropout members

Three specific variables were studied to examine if they are correlated with dropout: topic of discussion, interaction with other members, and sentiment expression. These three variables are important building blocks of OHC content and member characteristics, and have been investigated in a wide range of previous studies.[11,12,29]

Our research hypotheses are as follows:

Table 1. Topics considered for analyzing breast cancer forum posts.

Topic	Abbreviation	Description
Alternative	ALTR	alternative, complementary, and integrative medicine
Daily	DAIL	daily cancer-related experiences
Diagnosis	DIAG	diagnoses, measurements, and results of tests
Finding	FIND	health findings, signs, symptoms, and side effects
Health Systems	HSYS	health systems patients interact with, e.g., nurses, doctors, practices, hospitals, and insurance companies
Miscellaneous	MISC	greetings, and uninformative text, or which does not fit under any other label
Nutrition	NUTR	nutrition
Personal	PERS	personal anecdotes, information
Resources	RSRC	links, pointers, and quotes towards informational resources
Test	TEST	testing procedures (but not results of tests)
Treatment	TREA	treatments, including procedures, medications and therapeutic devices

1. Dropout members are more likely to discuss certain topics such as cancer treatments and their side effects, and show certain patterns in topic transitions, before they drop out. These topics and topic transitions may indicate end of cancer treatment journeys, which are usually followed by participation withdrawal.

2. Dropout members receive inadequate social support from other members. They ask questions and seek support more often than other members, but receive less responses. These may indicate lower levels of social support reception leading to decresed sense of belonging, a phenomenon known to be vital to self-perceived effectiveness of community usage.[30]

3. Dropout members express more negative sentiment in general, or in their final stage of participation, which indicates a declining level of satisfaction towards community participation.

Dropout and topics. To investigate whether topics of discussions correlate with dropping-out, topics of posts must be identified. In this study, topics of posts were identified using a supervised machine-learning tool based on convolutional neural networks (CNN).[31] The tool was trained on an annotated data set consisting of 9,016 posts, which was sampled from the same data set used in this study. Eleven topics were considered, which are relevant to describing the information needs of the breast cancer community members. The topics range from disease diagnosis, treatment, to more personal issues like daily lives and nutrition; Table 1 lists all topics with their descriptions. Of note, each post can be annotated with several topics (and as such, the topic identification task is cast as a multi-label classification). Overall, the classifier can identify topics of discussions with around 65% F score across all 11 topics. Further methodological details and detailed system evaluation of the CNN classifier can be found in the cited paper.[31]

To characterize the impact of discussed topics, for each user (either dropout or non-dropout members), we aggregate numbers of topics of all posts authored by the user, and average the topic numbers by the total number of the user's posts. As such, an eleven-dimensional distribution of topics can be established for a member in the forum, representing frequencies of topics discussed by the user.

Armed with distributions of topics for all users in the community users, we first did a multivariate t-test to examine the difference of topic distributions between posts of dropout members and posts of other members. For each topic, we then carried out a univariate t-test, adjusted by Bonferroni correction due to multiple comparisons, between the dropout members and other members in the community to test if a significant difference exists. These two static analyses identify the distributional differences between topics of discussions between dropout members and other members.

Finally, we examined how the averaged frequencies of topics change through time for dropout members before they actually quit the community from a longitudinal standpoint, to investigate whether certain patterns of changes could be detected.

Dropout and interaction. Member interaction is the primary medium of exchanging social support, which can be complex in online health communities.[2,3] In this study, we considered two basic aspects of user interactions: number of initial posts versus number of reply posts, and average number of responses received from other members in the community. Initial posts are those posts initializing threads of discussions, which are usually question asking or help seeking which represent needs of support requesting. Previous research has reported that initial posts are vital part of interactions amongst members, and are usually more negative emotionally.[11] Reply posts, usually representing support providing, are those posts responding to the initial posts, which can exert positive influence on the discussion originator (i.e., author of the initial post).[32] As such, the ratio of number of initial posts to the number of reply posts can be seen as how often the user seek support from others rather than actively provide support to others. Average number of responses received when initializing discussions, on the other hand, represent how much social support in average members receive from other ones. Previous studies have suggested that support providing and receiving may have different effects on perceived benefits.[33,34]

For each member, we counted the number of their initial posts, the number of their reply posts to other member's threads, and the number of responses received from other people when initiating a thread. We then calculated the two measures described above, and examined how these numbers differ between dropout members and other members. We relied on a Chi-square test (for initial vs. reply) and t test (for number of replies). Like for the topics, we also examined how these numbers change longitudinally before members' dropping-out.

Dropout and sentiment. Sentiment expression reveals how positive the author's emotion is when posting. We rely on a supervised classifier for sentiment analysis, which is described in.[11] For each of the post, a sentiment score was calculated, representing the degree of positiveness of the overall sentiment expression. Based on the sentiment scores, we first identified if a significant difference exists between the averaged sentiment scores of posts published by dropout members and posts published by other members, by doing a t-test. Second, we illustrated how sentiment of posts changed through time as dropout members approached the time point when they withdrawn from the community, to see if a decline of sentiment actually happened as suggested by our hypothesis.

Results

Identifying dropout members

6,338 dropout members were identified using our definition, corresponding roughly to 11% of all users that have posting history in the breast cancer forum. When accounting for all users who have posted more than 10 times (i.e., "meaningfully active") in the community, the dropout members amounted to 42% of these 15,199 users. The identified dropout members posted 570,932 posts in total in the breast cancer forum, with each one posting 90.1 posts in average. The average posting number is roughly the same as the average across all users posted more than 10 times (91.8). 195 out of these 6,338 dropout members have been highly active in the forum, with each of them posted more than 500 times. These "super-users", although relatively small in number, contributed to roughly 45% of posts identified.

Longitudinal analysis of dropout members

Dropout and topics. The multivariate t-test between the topic distributions of posts contributed by dropout members and other members respectively yielded a result which supports a difference with p-value less than 10^{-16}. Average prevalence of each topic for the two types of members is given in table 2 with corresponding p-values based on the univariate t-tests. We did not include MISC in the table because it is a default topic category only given to those posts which are not assigned any topics otherwise. We used 0.001 as the threshold of p-value for significance. Five topics amongst all ten show significant differences in average numbers between dropout members and other members. Specifically, dropout members posted more relevant to diagnosis and treatment, but less about nutrition and daily matters. The hypothesis that dropout members discuss more about treatment and diagnosis than other members is thus supported.

Figure 1 shows how topic frequencies change through time as dropout members approach the time point of withdrawing. The way we illustrate the changes is as follows. For each topic category, we plotted change of its average frequencies in all posts that were published a certain length of time before their authors' respective dropout time. We used week, days, and post orders as three different measures to show both long term and short term effects. For exam-

Table 2. Average prevalence of topics (per post) in posts of dropout members and other members. P-values are calculated by a t tests adjusted by Bonferroni correction. We use 0.001 as the threshold of p-value for significance.

	ALTR	DAIL	DIAG	FIND	HSYS	NUTR	PERS	RSRC	TEST	TREA
Dropout	0.002	0.059	0.099	0.063	0.081	0.034	0.274	0.013	0.009	0.053
Others	0.002	0.074	0.093	0.063	0.078	0.039	0.279	0.017	0.010	0.046
p-value	0.226	<0.001*	<0.001*	0.953	0.030	<0.001*	0.162	<0.001*	0.247	<0.001*

ple, a point (1, 0.3) in Figure 2(a) or 2(d) represents that the average frequency of the corresponding topic of all posts that are published in the final week of their authors' participation is 0.3. Except for an trend for a higher frequencies of DIAG and HSYS posts in the final weeks, no significant changes of topic frequencies were identified before members' dropping-out.

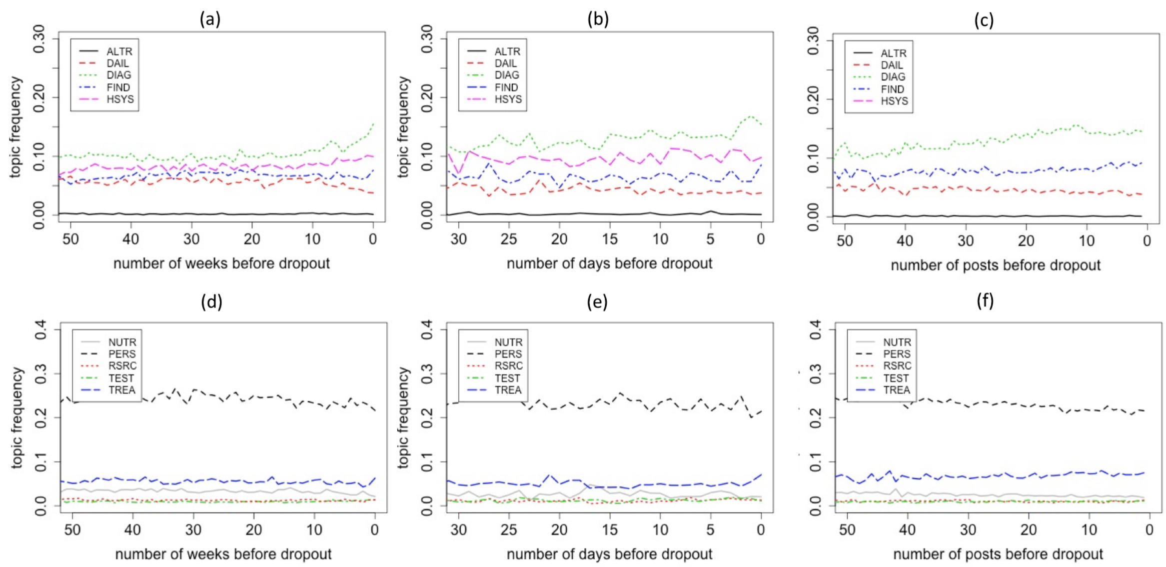

Figure 1. How topic frequencies change through time before members' dropping-out. X axes, which are in reserve order, represent the time point before members' dropping-out. Y axis is the average topic frequency of all posts that are published in the corresponding time. Units of x axes in (a)(d), (b)(e), and (c)(f) are weeks, days, and post orders, respectively.

Dropout and user interaction. 121,193(3.9%) of all posts in the forum are initial posts of threads. Among them, 31,277 were posted by dropout members, which are 5.5% of all dropout member publications. However, the Chi-squared test indicates no significant difference between dropout members and other members in terms of ratio of initial to reply posts, with a p-value over 0.9. Across the entire forum, an initial post can receive 24.4 replies in average. Dropout members, in particular, can receive an average number of 23.7 replies throughout their community engagement when initializing discussions. A t-test between the numbers of dropout members and other members indicate no significance with p value 0.69. As such, the hypotheses that dropout members receive less reply from other people and that post initial posts more often in the community are both rejected.

In contrast, the ratio of initial posts increases towards dropout time (Figure 3). It is particularly significant from a longer term standpoint, where the ratio of initial posts dramatically increase from around 5% to over 10% in the last 10 weeks of participation before dropping-out. We carried out a supplementary t-test, in which we compare all posts in final 10 weeks and posts before 10 weeks in terms of the initial/reply ratio, and indeed found a significant difference between the two with p value less than 0.001. Short term changes can also be observed, particularly in the final 5 days. Meanwhile, in term of number of replies received, a landslide can be observed in the week view, which roughly accompanies temporally the ratio increase of initial posts.

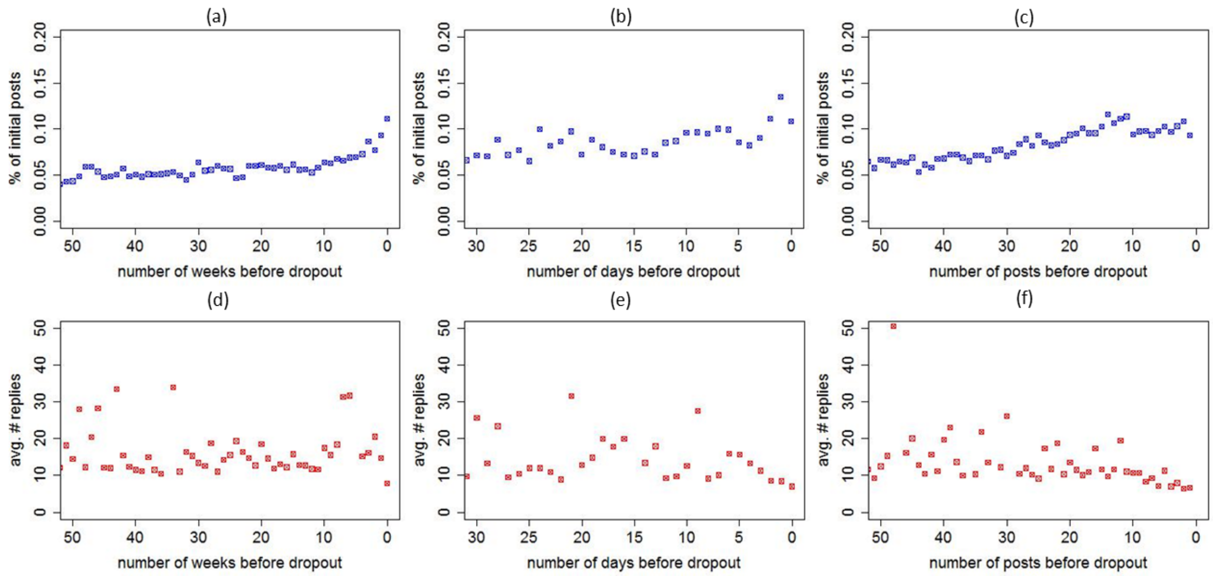

Figure 2. How percentage of initial posts and number of replies change through time before members' dropping-out. X axes, which are in reserve order, represent the time point before members' dropping-out. Units of x axes in (a)(d), (b)(e), and (c)(f) are weeks, days, and post orders, respectively.

Dropout and sentiment. The average sentiment score (probability of being positive) for all posts in the community is 0.786, while the average sentiment score of dropout member authoring is 0.788, with no significant difference according to a statistical t-test. Longitudinally, an insignificant decline of sentiment can be observed from the week view, but no other patterns can be found. Although we found a tendency of posting more initial posts in the final stage of participation in the previous analysis, no patterns of sentiment change is visible when initial posts and reply posts are considered separately. In contrast to our expectation, dropout members not necessarily express more negative emotion in discussion, and no significant changes of sentiment can be detected before they drop out.

Discussion

Principle findings

Our first hypothesis that dropout members are more likely to discuss certain topics is supported by our experimental results. We find that dropout members tend to discuss more about disease diagnosis and treatment, but less about daily issues and nutrition. Topics of treatment and diagnosis are common in posts that tell stories of one's cancer journey, or that describe cancer treatment experience. On the contrary, more daily-matter issues like exercises and nutrition are less focused by these users. Not many significant patterns of topic changes are identified longitudinally, except increased frequencies of health system and diagnosis in the final weeks before dropping out. This seems to suggest that although dropout members are more interested in certain topics in general, they do not necessarily shift their focus drastically throughout their participation. The increasing frequency of DIAG is interesting, however. One possible explanation may be that many dropout members were patients who were diagnosed with cancer recurrences or metastasis, which may be followed by the deterioration of the disease.

Our second hypothesis, with respect to user interactions with other members, is partially supported by the results. We originally expected that dropout members receive less replies from other members, which represents a lower level of social support received from other users, and that dropout members post initial posts more often, which represents that they are more likely to be information seekers rather than social support providers. Previous research in online social support groups suggested that emotional support providing is an important motivation of participation[35] and is beneficial to the providers themselves socially,[36] which is a factor that are expected to be negatively correlated with

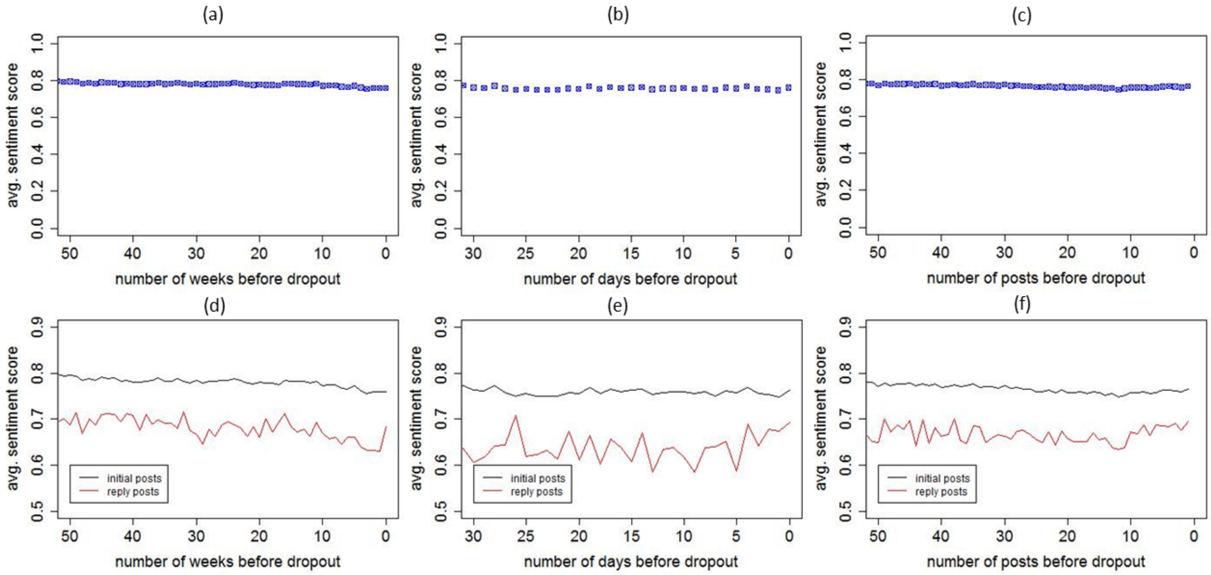

Figure 3. How average sentiment score changes through time before members' dropping-out. X axes, which are in reserve order, represent the time point before members' dropping-out. The first three figures show the average score of posts including both initial and reply, and the last three figures distinguish the two. Units of x axes in (a)(d), (b)(e), and (c)(f) are weeks, days, and post orders, respectively.

attrition.

However, in our static analyses, no significant differences are identified in the static analysis between dropout members and other members with respect to number of replies received, or ratio of initial posts to number of reply posts. The result may have two possible explanations. The first is that neither of the two measures can truly represent the degree of social support exchange in online health communities, and the other is that OHC users, particularly BC forum users, are different from online social support group members studied in previous research in how they perceive and understand benefits.

Although static comparison finds no difference, longitudinally we indeed find a rather significant increased ratio of initial posts at the end of user participation, as well as an insignificant drop of numbers of received replies, which is consistent with findings in the previous research that number of replies is important predictor of dropout.[27] The change is particularly dramatic in the final few weeks from the week view, and in the final 5 days from the day view.

This result, along with results from the static analysis of interactions as well as from previous analyses of topics, possibly shows a more complete picture of dropping-out: dropout members, in terms of support seeking and support providing, are identical to other community members in most of the times throughout their participations; however, certain events, which may be from the real lives of the users such as recurrence of cancer, trigger online behavioral changes and make the users seek much more support than before. At this moment, if these members don't receive adequate support, dropout may eventually happen.

Our final hypothesis that users express increasingly negative emotions in posts are not supported by our analysis. No significant difference is found between dropout members and other members, and no clear patterns can be identified longitudinally. The results contradict findings in previous research that usages of emotional keywords are associated with dropping out,[27] possibly because keywords of emotions cannot truly represent sentiment. Synthesizing the sentiment and interaction results seems to suggest that changes at the end of participation are mostly peaceful in sentiment, with no evident clue emotionally.

Table 3. Number of dropout members identified as the cut-off t changes.

t cut-off	1	2	3	4	5	6	7	8	9	10
# of dropout members	13,997	9677	6,338	3,864	2,311	925	210	76	32	11

Validity of our dropout member identification method

In this study, we rely on a straightforward method to identify more than 50 thousand dropout users, assuming that these users having an active history but having been inactive for at least 3 years. For a public online health community like breast cancer forum in which the community has no control over its members, identifying dropout members with absolute accuracy is infeasible. Because of the unavailability of gold-standard, evaluation of such identification method is challenging.

In our method, the most tricky part is to choose the *t* cut-off, which represents the minimal length of inactiveness for a member to be considered as dropped out. A larger *t* would definitely bring a set of dropout members with higher precision, but may excludes eligible dropouts incorrectly. Given the fact that most members joined the community in recent years and the forum was getting increasingly popular, a large t would lead to a small sample size. As such, the problem becomes a precision-recall trade-off, and our task is to finds the best value that balance the two properly. The oldest posts of our data set date back to Sep 2004, which is roughly 10 years before the data collection. To see how the cut-off impacts the sample size, we show in table 3 number of dropout members identified by setting t from 1 to 10. It can be seen that sizes of samples shrink rapidly when larger *t* is used.

The major false-positives of our method are the users that return to the community after long time inactiveness. To quantify the prevalence of these comebacks, we designed a sanity check experiment in which we calculate the percentage of users who have been inactive for more than 3 years in the community, anytime in the history, but return to the community after the long break, over the total number of users who have been active for more than 3 years. The number we get is 1.2%, which suggests a relatively good precision of our identification method.

What we learned from our topic analysis that dropout members focus more on diagnosis and treatment related themes also reminds us that users may drop out of the community because of death. Their escalated interest in diagnosis and treatment related issues may just be a signal of cancer metastasis, or unsuccessful treatments which may be followed by deterioration of the disease. These members leave the community not because of dissatisfaction towards community usage, and should usually be excluded in the attrition analysis. Similar to the issue of returning of inactive users, in public online communities there is no way to accurately identify dead members. To investigate how much this confounder impacts our results, we extract cancer stage information from user signatures, exclude cancer stage IV users, and replicate all analyses. The rationale is that stage IV users are the ones most likely to leave the community because of death, while stage 0 to stage III breast cancer are believed to have quite high 5-year survival rate. These supplementary analyses show identical findings as we demonstrated previously, and the exclusion of stage IV users does not impact the results. It is noteworthy, however, that the result does not indicate the nonexistence of impact of dead members on our study since not all users have accurate profile information in signatures.

Limitations and Future Work

There are several limitations of this work, which will be important parts of our future work. First, methodologically, our methods identifying dropout members, topic of discussion, and sentiment of posts are all based on automated algorithms that are not 100% accurate. Manual analysis over a small sample set might be a good way to complement the automated analyses. Efforts will also be made in creating better computational tools for topic and sentiment analysis in future work. Second, in this paper we only consider three variables of participation, while many other important ones are not covered such as members' age, social status, whether there are community debates or conflicts, etc. These variables of interest will be the focus of our future work to study how they impact users' decision makings. Third, our analyses, especially static ones, present only statistical differences, and such differences do not guarantee real clinical or psychological differences. In this paper, we do not try to identify causations and provide explanations; rather, we focus on detecting interesting correlational patterns that worth exploring by future research with rigorous experimental study designs. Our analysis in this study has captured some signals of what factors may be contributing to dropout, and we have proposed that it is likely that real life events may trigger the change the way members exchange social support and their foci of discussions. We believe that our study is a showcase of how quantitative methods can be used to analyze OHC content at scale for hypothesis discovery. Further analysis, both qualitative and experimental ones, can be carried out to examine these hypotheses. Forth, a better distinguish between dropout member and lurking members should be made more systematically. Identification of dropout members in public online health communities is a challenging task, and should worth more exploring in the future. For example, one possible solution might be doing expert annotations, followed by supervised learning models. 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 other conditions.

Conclusion

This paper presents a quantitative exploratory study over a popular and active public online breast cancer community to identify the characteristics of members that quit participation. We investigate correlations between community dropping-out and different factors: topic of discussion, sentiment of post, and user interactions in the community. When conducting such a quantitative study, one important methodological question pertains to identifying at scale the users who drop out. We explore strategies to identify such members, as well as static and longitudinal analyses of members' post history. Dropout members did not show any significantly different patterns in sentiment change when compared to other members in the community. Our findings suggest however that dropout members (1) tend to focus more on diagnosis- and treatment-related topics; and (2) exhibit increased needs of social support at the ending phase of participation, which are less and less fulfilled by other members.

Acknowledgments

This work is supported by a National Science Foundation award (#0941339) and a National Institute of General Medical Sciences award (R01 GM114355).

References

[1] Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res. 2011;13(3):e62.
[2] Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol. 2000;55(2):205–217.
[3] Biyani P, Caragea C, Mitra P, Yen J. Identifying emotional and informational support in online health communities. In: Proc. COLING; 2014. p. 827–836.
[4] Cohen S, Underwood LG, Gottlieb B, editors. Social support measurement and intervention: A guide for health and social scientists. Oxford University Press; 2000.
[5] Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. Patient Educ Couns. 2013;93(3):472–479.
[6] Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ. 2004;328(7439):564.
[7] Medicine 2.0: Peer-to-peer healthcare;. http://www.pewinternet.org/Reports/2011/Medicine-20.aspx.
[8] 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(May):1–6.
[9] Das A, Faxvaag A. What influences patient participation in an online forum for weight loss surgery? A qualitative case study. Interact J Med Res. 2014;3(1):e4.
[10] Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res. 2014;16(1):e12.
[11] Zhang S, Bantum E, Owen J, Elhadad N. Does sustained participation in an online health community affect sentiment? In: AMIA Annu Symp Proc; 2014. p. 1970–1979.

[12] Wang YC, Kraut RE, Levine JM. Eliciting and Receiving Online Support: Using Computer-Aided Content Analysis to Examine the Dynamics of Online Social Support. J Med Internet Res. 2015;17(4).

[13] Park A, Hartzler AL, Huh J, McDonald DW, Pratt W. Automatically Detecting Failures in Natural Language Processing Tools for Online Community Text. J Med Internet Res. 2015;17(8):e212.

[14] Bender JL, Jimenez-Marroquin MC, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res. 2011 Jan;13(1):e16.

[15] Valle CG, Tate DF, Mayer DK, Allicock M, Cai J. A randomized trial of a Facebook-based physical activity intervention for young adult cancer survivors. J Cancer Surv. 2013 Sep;7(3):355–68.

[16] Frost J, Massagli M. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res. 2008;10(3):e15.

[17] Wicks P, Keininger DL, Massagli MP, de la Loge C, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. Epilepsy Behav. 2012;23(1):16–23.

[18] Hixson JD, Barnes D, Parko K, Durgin T, Van Bebber S, Graham A, et al. Patients optimizing epilepsy management via an online community: The POEM Study. Neurology. 2015;85(2):129–136.

[19] Zhang S, Bantum E, Owen J, Bakken S, Elhadad N. Online cancer communities as informatics intervention for social support: Conceptualization, Characterization, and Impact. J Am Med Inform Assoc. 2016;In Press.

[20] 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.

[21] Eysenbach G. The law of attrition. J Med Internet Res. 2005;7(1):e11.

[22] van Uden-Kraan C. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res. 2008 Jan;10(2):e18.

[23] Mo PKH, Coulson NS. Empowering processes in online support groups among people living with HIV/AIDS: A comparative analysis of lurkers and posters. Computers in Human Behavior. 2010;26(5):1183–1193.

[24] Han JY, Hou J, Kim E, Gustafson DH. Lurking as an Active Participation Process: A Longitudinal Investigation of Engagement with an Online Cancer Support Group. Health Commun. 2014;p. 37–41.

[25] Wang Y, Kraut R, Levine J. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In: Proc CSCW; 2012. p. 833–842.

[26] Zhang Y. Understanding the sustained use of online health communities from a self-determination perspective. J Assoc Inform Science Technology. 2015;.

[27] Sadeque F, Solorio T, Pedersen T, Shrestha P, Bethard S. Predicting Continued Participation in Online Health Forums. In: Proc LOUHI Workshop; 2015. p. 12.

[28] Jha M, Elhadad N. Cancer Stage Prediction Based on Patient Online Discourse. In: Proc ACL BioNLP Workshop; 2010. p. 64–71.

[29] Civan A, Pratt W. Threading together patient expertise. In: AMIA Annu Symp Proc. American Medical Informatics Association; 2007. p. 140–144.

[30] Høybye MT, Johansen C, Tjørnhøj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. Psycho-Oncology. 2005;14(3):211–220.

[31] Zhang S, Grave E, Sklar E, Elhadad N. Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. In: arXiv preprint; 2016. p. 1603.08458.

[32] Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. J Am Med Inform Assoc. 2014;21(e2):e212–e218.

[33] Namkoong K, Shah DV, Han JY, Kim SC, Yoo W, Fan D, et al. Expression and reception of treatment information in breast cancer support groups: How health self-efficacy moderates effects on emotional well-being. Patient Educ Couns. 2010;81:S41–S47.

[34] Han JY, Shah DV, Kim E, Namkoong K, Lee SY, Moon TJ, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. Health Commun. 2011;26(2):185–97.

[35] Chung JE. Social Networking in Online Support Groups for Health: How Online Social Networking Benefits Patients. J Health Commun. 2014;19(6):639–659.

[36] Rodgers S, Chen Q. Internet community group participation: Psychosocial benefits for women with breast cancer. J Computer Mediated Commun. 2005;10:1–27.