

# Write for Life: Persisting in Online Health Communities with Expressive Writing and Social Support

HAIWEI MA, University of Minnesota, USA

C. ESTELLE SMITH, University of Minnesota, USA

LU HE, University of Minnesota, USA

SAUMIK NARAYANAN, University of Minnesota, USA

ROBERT A. GIAQUINTO, University of Minnesota, USA

RONI EVANS, University of Minnesota, USA

LINDA HANSON, University of Minnesota, USA

SVETLANA YAROSH, University of Minnesota, USA

Expressive writing has been shown to improve physical, mental, and social health outcomes for patients struggling with difficult diagnoses. In many online health communities, writing comprises a substantial portion of the user experience, yet little work has explored how writing itself affects user engagement. This paper explores user engagement on CaringBridge, a prominent online community for writing about personal health journeys. We build a survival analysis model, defining a new set of variables that operationalize expressive writing, and comparing these effects to those of social support, which are well-known to benefit user engagement. Furthermore, we use machine learning methods to estimate that approximately one third of community members who self-identify with a cancer condition cease engagement due to literal death. Finally, we provide quantitative evidence that: (1) receiving support, expressive writing, and giving support, in decreasing magnitude of relative impact, are associated with user engagement on CaringBridge, and (2) that considering deceased sites separately in our analysis significantly shifts our interpretations of user behavior.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**;

Additional Key Words and Phrases: Online health community; user engagement; death; social support; expressive writing; CaringBridge

## ACM Reference Format:

Haiwei Ma, C. Estelle Smith, Lu He, Saumik Narayanan, Robert A. Giaquinto, Roni Evans, Linda Hanson, and Svetlana Yarosh. 2017. Write for Life: Persisting in Online Health Communities with Expressive Writing and Social Support. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 73 (November 2017), 24 pages. <https://doi.org/10.1145/3134708>

This work is supported by CaringBridge and the National Science Foundation (grant 1464376). A student on this work is supported by the Graduate Assistance in Areas of National Need (GAANN) Fellowship.

Authors' addresses: Haiwei Ma, University of Minnesota, 200 Union Street SE, Minneapolis, MN, 55455, USA, maxxx979@umn.edu; C. Estelle Smith, University of Minnesota, 200 Union Street SE, Minneapolis, MN, 55455, USA, smit3694@umn.edu; Lu He, University of Minnesota, Minneapolis, USA; Saumik Narayanan, University of Minnesota, Minneapolis, USA; Robert A. Giaquinto, University of Minnesota, Minneapolis, USA; Roni Evans, University of Minnesota, Minneapolis, USA; Linda Hanson, University of Minnesota, Minneapolis, USA; Svetlana Yarosh, University of Minnesota, 200 Union Street SE, Minneapolis, MN, 55455, USA, lana@umn.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2017 Association for Computing Machinery.

2573-0142/2017/11-ART73

<https://doi.org/10.1145/3134708>

## 1 INTRODUCTION

A 2011 Pew Internet report indicates that 80% of American Internet users utilize the Internet to satisfy health-related needs [26]. Compared to traditional medical services, the Internet can provide easier access to information and 24-hour support than face-to-face or telephone interactions [101]. Furthermore, the wealth of resources available online can increase patients' self-awareness of disease [29, 92, 93].

In response to high demand for exchanging health information and support, online health communities have emerged as a means of helping patients connect with others. These Internet support groups cover various illnesses including cancer [58], substance use disorders [78], mood disorders [46], established bad habits [82], and dementia [30]. Researchers have discovered that participation in online support groups has beneficial effects on mental, physical, and social health by reducing stress, depression, and psychological trauma [5, 94], enhancing optimism about life [77], receiving social support [23], and coping more effectively with disease [50].

Besides offering information and support, online health communities can provide accessible opportunities for writing about personal experiences of traumatic events (also called expressive writing or disclosure writing) [34] and have been found beneficial to individual health outcomes (see [67] for a review). Whereas factors such as receiving and giving support are known to be beneficial [8, 43, 90, 91], it remains unknown how expressive writing affects *member engagement* in online groups. Therefore, our first research question (RQ1) is:

- *How does expressive writing affect member engagement in online health communities, relative to social support?*

RQ1 might, however, be impacted by a significant confound, i.e. survival bias. An important consideration in the study of online health communities is that many patients battle life-threatening or terminal diagnoses, such as many types of cancer. This difficult reality implies that some community members *literally* die as a result of their medical condition. Historically in HCI, the term "user death" has referred to the phenomenon of a user's site ending for any unspecified reason. In order to increase the clarity and precision of our language, we avoid using this term and introduce two replacement terms: (1) we define sites which community members may have intentionally or unintentionally left as "**relinquished sites**", and (2) we define sites where there is clear evidence that a patient has died as "**deceased sites**". We do not know of prior work that systematically identifies deceased sites or rigorously considers their statistical impact. Yet we believe that formally differentiating deceased sites in our analyses will help to mitigate survival bias and enable CSCW researchers to draw more valid conclusions in this space. Therefore, RQ2 is:

- *Are deceased sites automatically detectable? If so, how do they affect large-scale quantitative analyses of online health communities?*

To address our research questions, this paper employs a mixed-methods quantitative approach to analyze personal CaringBridge sites for cancer patients. First, we develop a machine learning model to classify relinquished from deceased sites. These two conditions are then treated appropriately in our second statistical method, Cox survival analysis, for which we operationalize variables for expressive writing and compare its effects against those of social support. Finally, we perform a validity check by running the exact same survival analysis model *without* considering the distinction between relinquished and deceased sites.

In the next section, we provide an overview of related literature. We go on to describe our research platform and design, expand upon the theoretical foundations of our research questions, explain our methods, and present results. Finally, we discuss these results, explaining the limitations of our approach and highlighting opportunities for future research and implications for the design of online health communities.

## 2 RELATED WORK

### 2.1 Online Health Communities and User Engagement

Because many people are not comfortable discussing health conditions on general-purpose social networking sites, niche online health communities often become an irreplaceable asset [61] for patients with cancer [58], substance use disorders [78], mood disorders [46], bad habits [82], and dementia [30]. However, several challenges also face online health communities, including protecting members' privacy [74], maintaining balance between people who seek and give support [73], and mediating divergent viewpoints and models [51].

Retaining participation is another challenge that has been extensively studied in contexts like Wikipedia [100], Yahoo! Answers [22, 96], MovieLens [39], and online gaming [7, 40]. Only a few studies have explored user engagement in online health communities [55, 91, 95]; these communities were primarily oriented towards users seeking information from people (usually strangers) who faced similar diagnoses. As we describe in the "Research Design" section, CaringBridge is oriented towards the *existing* social networks of patients rather than strangers, which provides an interesting complement to the few studies that do exist in this area. Importantly, because the design of CaringBridge emphasizes *storytelling* about health journeys over information-seeking, it also affords an ideal opportunity to introduce a new variable for consideration by the CSCW community when examining user engagement, i.e. expressive writing. In order to contextualize this work in the literature, we also test two factors previously shown to influence engagement, i.e. receiving and giving social support. The next section describes related work on these three factors.

### 2.2 Expressive Writing vs. Social Support

In the past three decades, many studies have found that writing about emotional personal events improves physical, mental, and social health [3, 48, 59, 63, 99]. The seminal experimental paradigm was to ask participants to write about emotional experiences for 15-20 minutes on 3-5 occasions, which resulted in better health outcomes than writing on neutral or superficial topics [68]. For example, the number of physician visits decreased [15], and immune function and long-term mood improved [32, 71]; these benefits were recently replicated in online settings [20]. Based on early findings, researchers developed a variety of expressive writing interventions, and systematic reviews suggest that most interventions have been beneficial to health [59, 63, 99]. Upon attempts to define boundary conditions, some results suggest that expressive writing is broadly beneficial for many different topics and across different ages, genders, education levels, languages, ethnicities, and cultures [21, 67, 68, 70, 76, 84].

While facilitated interventions may benefit participants, traveling to clinics or labs is not always feasible and may create additional burdens for patients [36]. Therefore, it is valuable to explore the *unfacilitated* expressive writing that occurs in online communities, which are far more easily accessible. In this paper, however, we do not examine any health benefits, per se; given prior literature, we assume expressive writing to be therapeutic. Rather, to identify implications for the design of platforms that offer writing as a prominent functionality, we investigate how expressive writing affects user engagement.

It is important to note that although many other works have investigated the social support exchanged online [6, 19, 23, 80, 90, 91], the expressive writing component remains understudied (see [2, 34] for exception). In general, previous findings show that receiving social support contributes to better health outcomes [4, 35], and longer participation in communities [13, 90, 91, 95]. Altruistically giving support to others by sharing advice, knowledge, and experience has also been found to be beneficial to health [8, 43, 77, 81, 97], though it has been understudied as a factor affecting user engagement. In the section 3.3 titled "Theoretical Foundations for Operationalization", we provide

detailed explanations for how we operationalize expressive writing, receiving and giving social support, and user engagement. Our statistical analysis aims to understand how these variables are associated with continued participation, relative to each other and to prior results. However, because many patients face life-threatening diagnoses, it is possible that the literal death of patients may be a confounding latent variable in any statistical analyses of these platforms; the next section provides a summary of work related to user death.

### 2.3 Death and Survival Bias in HCI

Deceased community members do not receive special attention in most communities because it is either time-consuming or difficult to detect this occurrence, and because it may not be of particular relevance to every community. However, in online health communities, especially those focused on life-threatening conditions, many patients *literally* do not survive. If models do not incorporate this fact, results are likely to be skewed by survival bias. For instance, Yang et al. recently analyzed an online cancer community and determined that death does not explain dropout because they found “similar results” when running their analysis separately on cancer types with different five-year survival rates [95]. Although results may be “similar”, they may nonetheless be significantly affected by death. Therefore it is interesting and also necessary to mitigate survival bias and ensure valid quantitative results.

Prior work has studied how people use technology to respond to the loss of a loved one [9, 10], and how technology could be better designed to support bereavement [54, 57]. These qualitative studies suggest the possibility of detecting death based on death-related writing (i.e. mourning and bereavement) on the pages of deceased community members. Building on this work, this paper differentiates relinquished sites from deceased sites by extracting textual features from journals written by surviving community members.

## 3 RESEARCH DESIGN

### 3.1 CaringBridge Description and Research Definitions

Established in 1997, CaringBridge (CB) is a 501(c)(3) non-profit organization with the objective of using “compassionate technology” to enhance social connectedness. Specifically, CB brings people together in an online social network to help overcome the isolation often experienced with a medical journey. In the calendar year 2016, CB served over 215K people daily and over 27 million people total in over 230 countries and territories around the world.

**3.1.1 Platform Description.** The primary functionality of CB is similar to a blog. Users create dedicated online spaces for sharing, following, and supporting the health journeys of patients. Though this social network may be used for any health journey, the majority face serious, life-threatening illnesses. In order to disambiguate terms used in this paper and maintain the vernacular used by CB, we here provide formal definitions:

- **Patients:** Patients diagnosed with any illness who may or may not be CB users. (Users who are *not* patients may maintain CB sites *on behalf of* patients.)
- **Caregivers:** Social connections (e.g. spouses, parents, children, or close friends) who assume primary responsibility for meeting patient needs. *Note: Although not technically disallowed, this does not typically include professional health care providers.*
- **Site:** A public or private personal website dedicated to sharing a specific patient’s health journey. Each site’s home page has an (optional) patient photo, banner photo, and brief overview of their story, plus a variety of features we define below. *Note: CB has additional features we do not define because they are unrelated to our research questions.*

- **Journal:** Each site has one journal, which is CB's most prominent feature. Similar to a blog, a journal is a collection of multiple health "updates" by or about a patient. It is designed like a timeline, with the most recent update on top.
- **Updates:** Similar to blog posts, updates are comprised of text, and may additionally or only contain photos or links.
- **Well Wishes:** Each site has one Well Wishes area, which is a section at the bottom of the site's home page where visitors can leave messages that are not associated with any particular journal update.
- **Messages:** Messages left by visitors either in the Well Wishes area, or as journal comments on individual updates.
- **Authors:** CB users who post updates to a journal. Sites can have one or more contributing authors.
- **Visitors:** CB users who visit sites, but are not authors. Visitors are automatically subscribed to receive notifications when new updates are posted to a journal, and can opt out if they choose. Visitors cannot post updates, but they can leave messages.

In contrast to online health communities where users seek support from strangers, CB users generally have social ties to the patient, though not necessarily to each other. All users have the option to complete minimal profiles (with a photo and short description), but the design of CB emphasizes sites rather than user profiles. Both offline and online relationships between users, while common, are not formally specified on the platform. Figure 1 demonstrates how a site and journal update appear on CB.

Importantly, CB site ownership is not limited to one user. Patients may maintain sites themselves, or caregivers can maintain sites without patient involvement, the latter being common when patients are too young or ill to be authors. Frequently, patients and caregivers co-author sites, each posting updates to the journal at various points along the journey. We discuss implications of this situation in our discussion, however it is important to keep this group-authorship situation in mind while reading our paper. We do not explore *exclusively* single-authored site survival; rather, we explore single- and group-authored site survival together, hoping to find ways to support engagement for both patient- and caregiver-authors.

### 3.2 Ethics and Relationship with CaringBridge

This work proceeds from a research collaboration between CB and an interdisciplinary team from Center for Spirituality and Healing, College of Computer Science and Engineering, and School of Nursing at University of Minnesota. In accordance with CB's Privacy Policy & Terms of Use Agreement, the dataset includes de-identified information about 588,210 sites and 22,333,379 users between June 1, 2005 and June 3, 2016. Conducted with permission of the CB leadership, this study was reviewed and deemed exempt from further IRB review by the University of Minnesota Institutional Review Board. All text mining searches and computations performed in this study took place on the Minnesota Supercomputing Institute's (MSI) high performance computing clusters, which provided a secure location for the data and analysis.

As a collaborative endeavor, we find it important to introduce data use constraints to protect the privacy of individuals who use CB. CSCW and related fields have recently hosted a number of conversations to decide appropriate practices and norms around data use and dissemination. On one hand, it is critical to protect the rights of participants (above and beyond IRB review) [12] and no de-identification technique is without fail [60]. On the other hand, open data dissemination supports better science and forms of replication that could strengthen the field as a whole [37]. We respect both sides of this conversation. Given the sensitive nature of this data, it is not possible

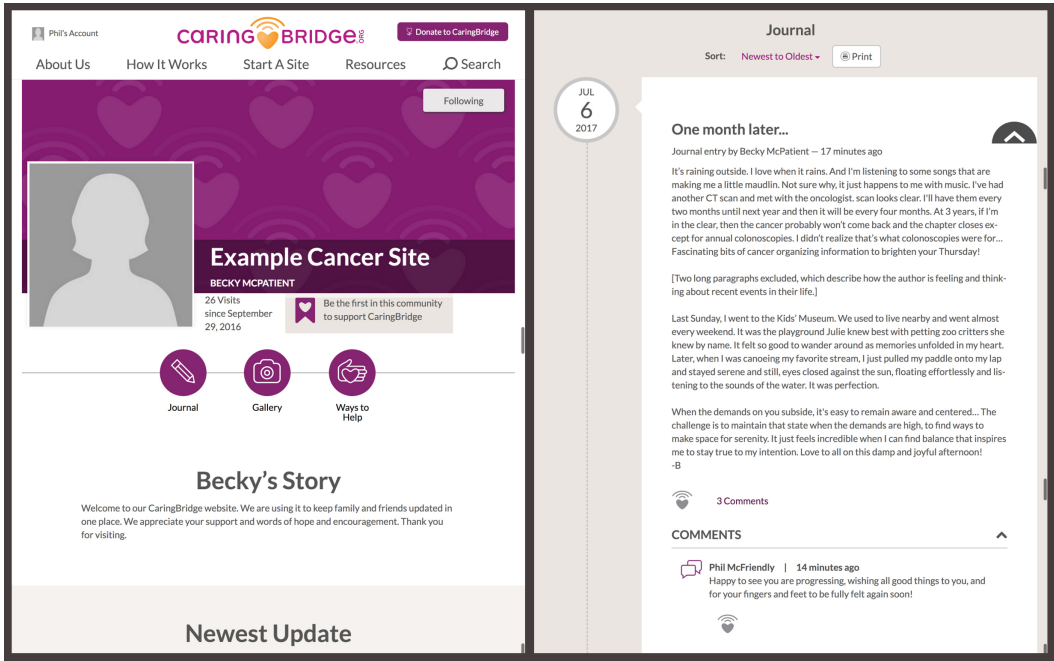


Fig. 1. An example site's homepage and journal. The update depicted in this journal is modified from real user data, in order to provide a representative text sample.

for us to publicly release the dataset used for analysis in this paper. However, we invite questions about the project and dataset by contacting CB directly or the investigators who conducted this study. We find this to be a reasonable compromise between the priorities of ethical protection and of replicable science.

### 3.3 Theoretical Foundations for Operationalization

**3.3.1 Expressive Writing: Introduction.** There is no universally acknowledged definition of expressive writing. In a seminal 1986 study, James W. Pennebaker instructed participants to write their deepest thoughts and feelings about a past trauma [68]. Since then, this paradigm has been repeatedly revised in attempts to understand how it works. Although exact mechanisms remain uncertain to this day, expressive writing is nonetheless effective when participants write about a broad range of topics, including disclosed or undisclosed traumas [31], imaginary traumas [32], positive aspects of traumatic experiences [86], illnesses [17], experiences of going to college [42], job loss [84], intensely positive experiences [14], and future goals [41, 52]. On one hand, this diversity suggests generalizable utility for expressive writing across many contexts. On the other, it makes expressive writing quite complex to precisely define and identify.

Previous studies of *facilitated* expressive writing provide slightly varying instructions, yet all share two implicit assumptions: (1) participants will follow the instructions, and (2) writing that results from following the instructions will be “expressive”. In the current study of *unfacilitated* expressive writing, we cannot make any assumptions about the conditions under which the writing was completed. However, in order to clarify the core concept, we here provide a definition adapted from these facilitated writing experiments for the context of CSCW research. **We define**



**online “expressive writing” to be writing completed using functionalities afforded by on-line communities to disclose users’ thoughts and feelings about personal experiences.** In online health communities in particular, expressive writing is completed by patient- or caregiver-authors who share experiences about personal health-related issues. By this definition, every journal update on CB could potentially be written expressively. That is to say, we do not assign a score of “expressiveness” to each update, or identify one update as “expressive” and another as “non-expressive”. Instead, we extract features from updates that *cumulatively* represent prominent aspects of expressive writing, namely emotional expression, cognitive adaptation, and dose of writing, all three of which have emerged as important qualities in prior research. The following sections describe each of these aspects in more detail.

**3.3.2 Expressive Writing: Emotional Expression.** Emotional expression involves confronting stressor-related feelings [47]. This type of expression helps people cope with illness or trauma possibly because it allows individuals to examine, understand, and regulate entrenched emotions and suffering [79, 85]. Despite variation in instructions, all previous studies emphasized writing about one’s deepest emotions in the experimental condition, and writing unemotionally or factually in the control condition [27, 28, 83]. Thus, many researchers use “emotional writing” or “emotional disclosure” interchangeably with expressive writing [24, 33, 75].

One experimental study found that writing either one’s “deepest thoughts and feelings” or “only positive thoughts and feelings” about their cancer experience had beneficial effects, whereas writing only about the facts of cancer and treatment did not [86]. In several quantitative studies, researchers analyzed participants’ essays using the Linguistic Inquiry and Word Count (LIWC) tool to measure emotional expression by calculating percentages of positive vs. negative emotion words, with mixed results about whether affective word use is associated with health outcomes [38, 45, 49, 59, 63, 72, 99]. Since CB provides a unique place to communicate emotions that are hard to present on other social media [61], and because of the LIWC precedent in prior work, **we also use LIWC to capture emotional expression in updates written during the first month.**

**3.3.3 Expressive Writing: Cognitive Adaptation.** Cognitive adaptation, or cognitive appraisal, is the process of discovering meaning through construction of narrative [18, 48]. It may be helpful in coping with illness or trauma because it allows people to interpret and make sense of the event and reconcile existing self- and world-views [16, 18, 48].

In one experimental study, researchers found that writers focusing on cognitions and emotions related to a trauma or stressor developed greater awareness of the positive benefits of the event than those focusing on emotions or facts alone [89], providing evidence for the importance of cognitive adaptation. Other evidence comes from content analysis. Interestingly, direct measures of LIWC cognitive words (i.e. insight and causal reasoning words), or of manually coded cognitive processing sentences, did not significantly influence health outcomes [18, 49]. However, an *increase over time* in the use of insight and causal words, as calculated by LIWC, is associated with better health outcomes [44, 72, 88]. Adopting this methodology, **we use LIWC to calculate the increase of cognitive word usage between the first and second half of writing during the first month.**

**3.3.4 Expressive Writing: Dose.** The “dose” or volume of expressive writing completed may also significantly influence outcomes. Some meta-analyses show that the number of disclosure sessions (fewer than three vs. at least three sessions) and length of disclosure (less than 15 vs. at least 15 min) were related to the overall effect size [27, 83]. In our current research, participants do not attend disclosure sessions; instead, CB authors freely choose when and how much to write “in the wild”. **Therefore, as a proxy for the number of writing sessions, we measure the frequency**

**of writing updates in the first month. Similarly, we measure the length (word count) of updates as a proxy for the length of expressive writing sessions.**

**3.3.5 Social Support.** People join online health communities to seek and offer social support [23, 77]. Previous well-known findings show that receiving social support contributes to better health outcomes [4, 35], and longer participation in the community [13, 90, 91, 95]. In those studies, researchers used the number of messages received as a proxy for the amount of support received. **In our study, we calculate the amount of support received as two variables, the number of messages received per visitor, and the number of visitors.** (Multiplying those two quantities is equivalent to the number of total messages received.) We make this distinction because the number of messages received per visitor shows the average intensity of supporting relationships between visitors and authors, while the number of visitors demonstrates the size of the support group.

Little work has directly examined how *providing* social support relates to user engagement, although two studies have found it to be beneficial to well-being [8, 43]. **We measure similar variables for giving social support: the number of messages sent per site visited and the number of sites visited by the authors of the focal site.** The number of messages sent approximates the average intensity of supporting relationships between authors and other sites. The number of sites visited approximates how many other sites are supported by the authors.

**3.3.6 User Engagement.** User engagement is a complex term describing peoples' experiences with technologies and is not defined consistently in past work [62]. There are various ways of measuring user engagement, including diversity of activities [39], number of activities [100], interaction with other users [95], attachment [95, 98], continued participation [55, 91, 95], etc. **Similarly to prior work in online health communities, we use "continued participation", i.e. the length of time users continue to participate in the community, to measure user engagement [55, 91, 95].**

Predicting entire duration of stay by behaviors during the first 30 days is a technique used in prior studies, e.g. [96]. This technique is also appropriate in our study of CB sites that self-select cancer as health condition, because the number of updates posted in the first month covered about a quarter of all updates and started to decline much more slowly after the first month. Utilizing this technique for our study, variables measuring expressive writing and social support during the first month are regarded as a "treatment" or as "early user experiences", which may differ from user to user and can affect how long they ultimately remain in the community. An alternative interpretation might be considering expressive writing and social support in the first month as "short-term engagement", and time length of continued participation in the community as "long-term engagement". Our study seeks to understand how different "short-term engagement" patterns possibly transform into "long-term engagement" through association.

### 3.4 Development of Inclusion Criteria

Unlike some forum-based communities, CB does not focus on a single topic. Members have diverse diagnoses ranging from chronic conditions like cancer to acute conditions such as childbirth. Upon creating a site, CB allows (but does not require) authors to self-select a health condition. Within the whole CB dataset, 35.5% of 588,210 sites specify health conditions. Of those self-specified conditions, 50.0% select a cancer condition. Thus, cancer constitutes the largest proportion of all self-identified health conditions, and represents 17.8% of *all* CB sites, regardless of whether health condition is specified or not. Because the literature indicates that the effects of expressive writing are long-term, chronic conditions like cancer are of particular relevance to our research questions. Therefore, we



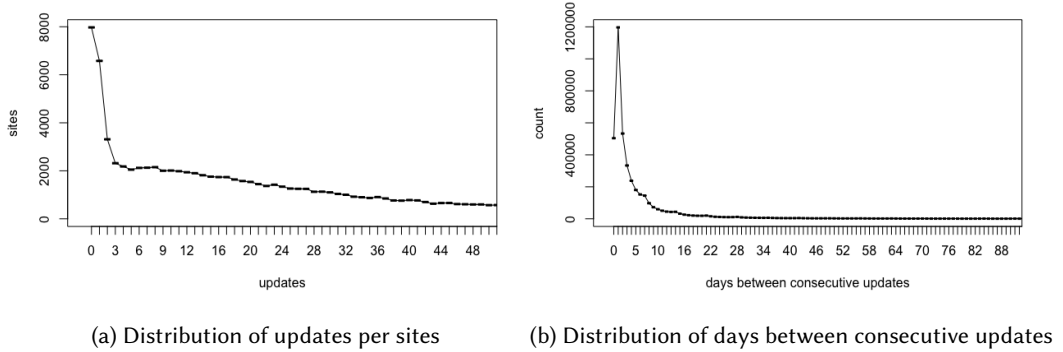


Fig. 2. (a) There is a huge drop between the number of sites with two updates and those with three. (b) 98% of updates are posted within 90 days since last update.

limit our focus to the approximately 104,500 sites that self-identify as cancer. Without additional specifying, we refer to these as “cancer sites” throughout our paper.

A minimum number of updates is necessary to study the effects of expressive writing. Based on three considerations, we include only cancer sites with at least three updates. First, writing only one or two updates might not have enough power, as the writing paradigm used in prior literature requires participants to write on at least three occasions. Second, as shown in Figure 2a, there is a huge drop between the number of sites with two updates and those with three, suggesting a possible difference between user groups before and after the three-update threshold. Third, as we will describe in our Methods, we extract features from the final three updates to classify sites as deceased or relinquished, and therefore require all sites in our dataset to have at least three updates. Among 104,500 sites, 86,637 have at least three updates. Furthermore, we look at the effects of expressive writing during the first month to predict long-term user engagement. Only 84,007 sites have updates in the first month. We refer to these as “eligible cancer sites”.

We define a site to have come to an end if it does not post a new update within 90 days of the latest post in the data, using the timestamp of site creation to signal the start of participation, and of the last update to signal the end of participation. Figure 2b shows the distribution of time intervals between consecutive updates. 98% of updates are posted within 90 days since last update, which indicates that a 90-day threshold is a reasonable choice. Some previous studies also define “leaving the group” as failing to post within 90 days of the last post [91, 98, 100]. Regarding the remaining 2% of updates which were not posted within 90 days, we noticed an interesting pattern that some updates are written years apart from each other. That is, some authors return to CB sites to write updates that memorialize a deceased friend or family member, or that revisit past events such as completion of treatment for a serious illness. A limitation of this study is that those sites which *will* post again, but have *not* yet posted within 90 days of the end of data collection are considered ended.

## 4 METHODS

To address our research questions, we utilize techniques from machine learning and statistics. First, we build a “death classifier” to distinguish between relinquished and deceased sites. This methodology is necessary because it provides a means of automatically differentiating sites into

two conditions, which are then treated accordingly in our second method. Incorporating predictions from our classifier, we next apply Cox survival analysis to predict user engagement with explanatory variables derived from components of expressive writing and social support. Finally, we perform a validity check by re-running the Cox survival model *without* differentiating between the relinquished and death conditions. The following sections provide in-depth explanations of these methods.

#### 4.1 Automatic Classification of Deceased Sites

Because of our decision to analyze cancer sites and because cancer can be a terminal condition, it is reasonable to believe that some patients' sites will end because of their literal deaths; yet we do not know of prior work that provides a means of detecting deceased vs. relinquished sites. Therefore, using techniques from machine learning, we build a "death classifier" to label sites as deceased ('1') or relinquished ('0'). Here, we describe our annotation and feature selection processes. We report our final model selection in Results.

**4.1.1 Annotation of Training Data.** To begin annotation, three investigators read 48 of the final three updates of randomly selected CB sites marked as ended, including cancer and non-cancer sites. We read the *last three* rather than the *final update alone* because death announcements are not always present in the very last update. Due to the quantitative nature of this paper, we will not elaborate on the qualitative contents of these updates; in our discussion, we present some insights about site endings to inform future work. In order to contextualize our feature selection process, however, we feel it necessary to provide high level observations that we noted and discussed while annotating.

Previous qualitative studies have shown that death-related mourning and bereavement on a user's page can signal literal death [9–11, 53, 54, 56, 57]. Our study strongly confirms this finding. In CB updates about death, caregiver-authors write obituary or funeral announcements, bereavement messages, and highly emotional passages about grief, all of which provide clear linguistic evidence that the site's patient has passed away. Thus, our assumption that some users pass away is valid, but *how many* sites end due to the death of a person?

We arbitrarily denote updates in which we find evidence that the patient has died as '1' (indicating a deceased site), and those with no such evidence as '0' (indicating a relinquished site). All three coders began by independently annotating the same 48 updates as '1' or '0'. We initially achieved consensus for 46 of 48 updates, with a Fleiss' Kappa inter-rater reliability score of 0.939, i.e. near perfect agreement. We resolved the only two cases of disagreement by developing the following rule: *Even if the text of a update indicates that a patient might die soon, the update should still be marked '0'.* That is, we agreed to annotate conservatively, marking updates as '1' only in cases of absolute certainty. After the initial 48, all remaining updates were independently annotated by one coder each.

We consider a site to have ended in death if one (or more) of the final three updates to its journal was marked '1'. In total, we manually annotated 2070 unique updates, corresponding to 690 sites, approximately half of which were randomly selected from *all* CB sites with at least three updates, the other half of which were randomly selected from those with at least three updates *and* a self-identified cancer condition. (*Note: We included non-cancer sites so that our model remains generalizable for future analyses.*) Of these 690 sites, 194 were marked '1'. In our final training set, we included all 194 sites marked '1' and randomly selected an additional 194 sites from the remaining 496, for a total training set of 388 sites.

As we read these final updates, we continually recorded and discussed death-related words or phrases that appeared in updates marked '1', along with any other textual characteristics that might

be useful as features for a machine learning classifier. These discussions and records helped us to identify and refine appropriate features.

**4.1.2 Features: LIWC Death-Related Unigrams.** The Linguistic Inquiry and Word Count (LIWC) tool provides research-validated dictionaries containing sets of words related to linguistic categories [69] and has already been applied for text analysis in online health communities with promising results [1, 64, 65, 91]. In standard practice, LIWC is used to count words, assign them to categories, and produce percentages of categories present in a given text.

LIWC offers a validated set of 29 words in a “death” category. Whereas some LIWC death-related words like “funeral\*”, “died”, and “grief” appeared regularly in ‘1’ updates, we did not observe others, such as “murder\*”, “suicid\*”, and “decay\*” in the training data. Furthermore, we noticed that usage of *some* LIWC death words was more often idiomatic than literal. For instance, “kill” and “dead” are LIWC death words, but we observed usages like “*let’s go kill some cancer cells*” and “*I missed the dead line*” which do not indicate the death of a person. Nevertheless, the presence of many LIWC death words correctly corresponds with literal death. Therefore, adapting a technique used by [87], we consider LIWC death-related words as unigrams and calculate the term frequency for each, dividing the number of occurrences of each unigram by the total number of words in the last three updates.

**4.1.3 Features: Data-Derived Death-Related N-grams.** We noticed that when writing about death in CB updates, authors do not often use literal words like “terminated” or “deceased”. Rather, they frequently use gentler, euphemistic expressions like “passed away” or “completed her journey.” Many updates utilize religious language like “went to heaven” or “with our Lord.” Others detail pragmatic considerations, such as “in lieu of flowers” or “memorial service.” Because LIWC’s death category does not include these words and phrases, we compiled a second set of 26 non-LIWC death-related n-grams (in units of one, two, three, four, or five words) that appeared in the training data. Identically to LIWC death-words, we utilized a frequency variable for each data-derived n-gram, dividing the number of occurrences of each N-gram by the total number of words in the last three updates.

**4.1.4 Features: Summary.** Combining 29 LIWC and 26 non-LIWC death-related term frequencies calculated from the last three updates, we extracted 55 features for each site. These features provide linguistic evidence for whether or not the site’s patient passed away. A deceased site will have higher values for these features, compared to a relinquished site. We extracted these 55 features for all sites in our dataset, fed them into six different machine learning methods, and evaluated all models through 10-fold cross validation on our 388-site training dataset. Finally, we also evaluated their performance on a separate 202-site test dataset which had not been seen during annotation, in order to ensure that the features selected from training data remained effective on unseen data. We present evaluation metrics and our final model selection in Results.

## 4.2 Statistical Analysis of User Engagement

**4.2.1 Survival Analysis: Overview.** Survival analysis is a branch of statistics that focuses on analyzing the expected duration(s) of time until one or more events happen. In our case, we use information gathered from the *beginning* (i.e. first month) of a CB site in order to predict how long it will be before the site *ends*. Site ending is therefore our “event of interest”. We use the Cox Proportional Hazards Model to examine the effects of different factors on length of CB participation. This model is not dependent on any assumption of the distribution of survival times. All independent variables are standardized with a mean of zero and standard deviation of one before being entered into the survival regression model. Recall that CB sites can be authored by patients, caregivers, or

some combination of both; our data does not afford us a means of differentiating sites based on number of authors, so this analysis cumulatively describes sites that are single- or group-authored. In the following sections, we define the terminologies and parameters of our model.

**4.2.2 Survival Analysis: Hazard Ratio.** In order to understand survival analysis, it is important to understand the notion of “Hazard Ratio” (HR). In this method, different variables are used to predict expected lengths of time. If a particular variable contributes to the “event of interest” occurring sooner, then this variable has a higher (greater than one) HR. Conversely, if a variable contributes to the event of interest occurring later, it has a lower (less than one) HR. In this context, a HR greater than one indicates that a given variable contributes to a site ending sooner, whereas a HR less than one contributes to a site enduring for a longer period of time. In addition, because all explanatory variables have been standardized, HR should be interpreted as the predicted change in the probability of dropout from the group for increasing one standard deviation of the corresponding variable, when all other variables are at their mean levels. We define the baseline hazard to be the probability of dropout when all variables are at their mean levels. We similarly define baseline survival.

**4.2.3 Survival Analysis: Right Censoring.** Compared to standard regression techniques, survival analysis offers the advantage of being able to “censor” certain types of sites. That is, censored sites are understood within the model to have *incomplete* data. In particular “right censoring” indicates that the true value of the data is likely to be higher, although there is no way to know by how much. **Right censoring is what enables us to statistically differentiate relinquished and deceased sites.** That is, in the present research, we right-censor sites for two reasons: 1) the site is lost to follow-up because of the literal death of its primary subject, or 2) the site does not come to an end before the end of data collection (i.e. the latest update was posted within 90 days of end of data collection, and might therefore still be active). We now proceed to describe model parameters.

	Mean	Median	Std
posemo	0.052	0.051	0.016
negemo	0.013	0.012	0.006
causalIncrease	0.000	0.000	0.008
insightIncrease	-0.000	-0.000	0.011
length	207.1	168.8	155.6
frequency	0.416	0.333	0.337
visitors	83.6	51	128.3
msgPerVisitor	1.259	1.122	0.845
visits	18.7	1	921.5
msgPerVisit	1.725	1	3.978

(a) Descriptive statistics

Variable Pair	Corr.
(posemo, negemo)	-0.13
(posemo, length)	-0.24
(negemo, length)	0.14
(causalIncrease, insightIncrease)	0.11
(length, visitors)	0.13
(frequency, visitors)	0.29
(frequency, msgPerVisitor)	0.44
(msgPerVisitor, msgPerVisit)	0.14

(b) Correlation

Table 1. (a) Descriptive statistics for variables included in the survival analysis model. Numbers are rounded to tenths for length, visitors, and visits. Numbers are rounded to thousandths for all other variables. The mean and median of causalIncrease are small positive numbers, rounded to 0.000. The mean and median of insightIncrease are small negative numbers rounded to 0.000. (b) Correlation between variables larger than 0.1 or smaller than -0.1 calculated from covariance.

**4.2.4 Survival Analysis: Variables.** Because of the size of our dataset, it would be possible to measure the effects of any number of arbitrarily selected variables and simply cherry-pick those that

produced significant effects. However, in the section 3.3 “Theoretical Foundations for Operationalization”, we provide detailed explanations about how our selection of variables is grounded in prior literature. **That is, we test and present only variables for which there could be theoretical significance.** For the sake of concision, we here provide a bulleted summary of the dependent and independent variables computed for each site in our survival model:

*Dependent variable*

- **Time-to-event:** Duration of time between site creation and the latest update. The event of leaving the community is observed if the site has been inactive for at least 90 days following the last post, and was not classified as a deceased site.

*Independent variables*

- **Expressive Writing: Emotional Expression**
  - **Positive emotion:** Average percentage of positive LIWC emotion words in updates written during the first month after site creation.
  - **Negative emotion:** Average percentage of negative LIWC emotion words in updates written during the first month after site creation.
- **Expressive Writing: Cognitive Adaptation**
  - **Causal increase:** Increase in percentage of causal LIWC words between the second and first half of updates written during the first month after site creation.
  - **Insight increase:** Increase in percentage of insight LIWC words between the second and first half of updates written during the first month after site creation.
- **Expressive Writing: Dose**
  - **Length:** Average number of words per update during the first month after site creation.
  - **Frequency:** Average number of updates per day during the first month after site creation.
- **Receiving Support**
  - **Visitors:** Number of visitors to a site during the first month after site creation.
  - **Messages per visitor:** Average number of messages received from each visitor to a site during the first month after site creation. If there are no visitors, set to 0.
- **Giving Support**
  - **Visits:** Sum of the number of other sites visited by authors of the focal site during the first month after site creation.
  - **Messages per visit:** Average number of messages sent to other sites visited by authors of the focal site during the first month after site creation. If there are no visited sites, set to 0.

Table 1a reports the descriptive statistics of these variables. The correlations between variables are mostly between -0.1 and 0.1, with a few exceptions noted in Table 1b. Note that when an author posts an update, all of the visitors to that site can receive a notification after logging in to CB (unless they opted out). These notifications might explain the relatively high positive correlation between frequency and both number of visitors and number of messages per visitor.

## 5 RESULTS

### 5.1 Building the Death Classifier

After experimenting with potential features, we found that n-gram frequencies yielded high classification accuracy. Table 2 displays the top 10 n-grams from each category, listed in order of their importance to our final model, Random Forest. Other non-LIWC terms that do not appear in the top 10, including “viewing”, “sorrow”, “complete [their] journey”, “get [their] angel wings”, “obituary”, “would have been [XX] years”, and “newest angel”, actually have larger Gini Importance than the top 10 LIWC word “mourn\*<sup>†</sup>”. These non-LIWC features are indeed good complements to

LIWC unigrams		non-LIWC n-grams	
Feature	Gini Importance	Feature	Gini Importance
funeral*	0.142	passed away	0.116
died	0.056	[memorial] service	0.101
death*	0.05	celebrate [their] life	0.066
die	0.029	in heaven	0.061
griev*	0.029	with [god]	0.051
cemet**	0.015	lieu of flowers	0.044
grave*	0.010	honor of	0.019
kill*	0.009	visitation	0.018
mourn*	0.008	[went to] heaven	0.015
ashes	0.008	lost [their] battle	0.014

Table 2. Feature importance (calculated by Gini Importance, the sum over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits) of top 10 n-grams from each category. Asterisks (\*) indicate word stems matched using regular expressions. Brackets([]) indicate multiple possible matching words; words displayed in brackets are representative examples.

the LIWC features. The complete list of features, in addition to the code used to extract and use them, is available upon request.

We built classification code using the scikit-learn Python module and selected a variety of learning models to test, including a baseline model that always predicts the most frequent label, L2 regularized binomial logistic regression, L2 regularized linear support vector machine, naive Bayes, decision tree, and random forest with 50 trees [66]. We built these six models using 55 features extracted from each site's journal (i.e. 29 LIWC and 26 non-LIWC n-gram frequencies per final three updates). We tested them using 10-fold cross validation on a balanced training set of 388 sites which were manually labeled during annotation. Table 3 presents the results of cross-validation. Figure 3a presents the ROC curve.

Because our non-LIWC features were developed specifically from the 388-site training set, evaluation results on the same data set might be biased. To test performance on unseen data, we developed a separate test set by manually annotating an additional 202 randomly selected eligible cancer sites. This test set is imbalanced with 70 sites labeled "1" and 132 labeled "0"; therefore, with 95% confidence, the true proportion of eligible deceased cancer sites should be between 27.7% and 41.7% (i.e. 34.7% (+/- 7%)). With all models trained on the entire 388-site training set, we test the performance of each classifier on this separate test set. The results of test set evaluation are also presented in Table 3. Figure 3b presents the ROC curve.

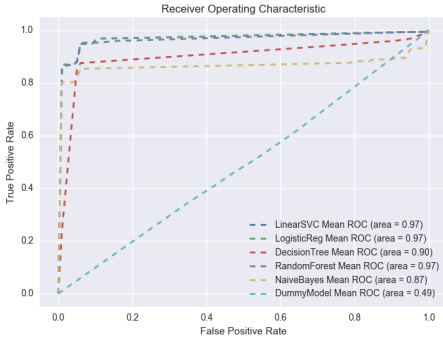
Excluding DummyClassifier, all other models achieved classification accuracies between 75%-94% on our training data, and between 86%-93% on our test data. This result partially addresses our second research question: **automatic classification of relinquished vs. deceased sites based on features extracted from the last three updates is indeed feasible, with high accuracy.** Yet how many sites end this way?

For our final model, we selected Random Forest because it has comparable performance to LinearSVC and Naive Bayes on the test set. However, Random Forest offers two advantages: (1) it is simpler to train than LinearSVC since it does not require parameter tuning, and (2) it provides a direct and more interpretable measure of feature importance than Naive Bayes. By Random Forest predictions, 36.9% of 84,007 eligible cancer sites are deceased sites. This estimate appears to be reasonable since it lies within our 95% confidence interval. Additionally, according to the United

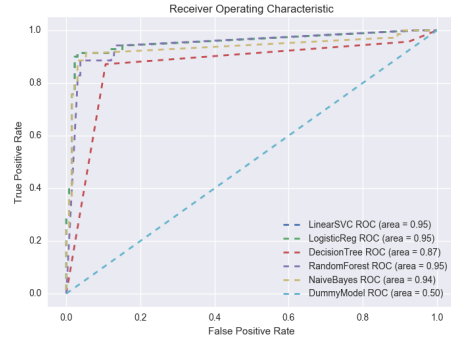


	cross validation					separate test set				
Classifier	ACC	PREC	REC	F1	AUC	ACC	PREC	REC	F1	AUC
DummyModel	0.50	0.50	1.00	0.67	0.49	0.65	0.00	0.00	0.00	0.50
LogisticReg	0.75	1.00	0.50	0.64	0.97	0.86	0.96	0.61	0.75	0.95
LinearSVC	0.85	1.00	0.71	0.82	0.97	0.93	0.95	0.84	0.89	0.95
NaiveBayes	0.90	0.94	0.85	0.89	0.87	0.92	0.85	0.91	0.88	0.94
DecisionTree	0.92	0.96	0.89	0.92	0.90	0.89	0.82	0.87	0.85	0.87
RandomForest	0.94	0.98	0.90	0.94	0.97	0.92	0.85	0.91	0.88	0.95

Table 3. Model Performance as evaluated by 10-fold cross validation on a balanced training set of 388 sites and a separate test set of 202 sites. Models include: baseline (DummyModel) that always predicts most frequent label, L2 regularized binomial logistic regression (LogisticReg), L2 regularized linear support vector machine (linearSVC), naive Bayes (NaiveBayes), decision tree (DecisionTree), and random forest with 50 trees (RandomForest). Column headers are average accuracy (ACC), precision (PREC), recall (REC), f1 score (F1), and area under the ROC curve (AUC).



(a) 10-fold cross validation



(b) separate test set

Fig. 3. ROC curves as evaluated by (a) 10-fold cross validation on a balanced training set of 388 sites and (b) a separate test set of 202 sites.

States CDC National Program of Cancer Registries, the relative five-year survival of all cancer sites from 2001-2012 (combining all cancers, ages, races, and genders, which aptly matches our dataset) is 65.9%; thus, 34.1% of cancer patients do not survive five years from diagnosis [25]. The average site duration in our dataset is 425 days (minimum, 0 days, and maximum, 3967 days or 11 years). Although this CDC statistic is not directly comparable because the time spans differ, the proximity of our estimate to this statistic suggests high validity.

## 5.2 Survival Analysis Results

We constructed a Cox Proportional Hazards survival model in order to explore the relationships between user engagement and expressive writing and receiving and giving social support. In this model, deceased sites are treated as right-censored data (see Table 4). For a validity check, we fitted a second survival model with the same set of explanatory variables, but did not incorporate classification results, such that deceased sites were treated equivalently to relinquished sites (see Table 5). Effects are reported in terms of Hazard Ratio (HR). (See the section 4.2.2 “Survival Analysis: Hazard Ratio” for a brief discussion of how to interpret HRs.) Note that although variables differ in

Survival Model with Death Classification Results				
Construct	Indep. Var.	HR	p value	[95% Conf. Int.]
EW: Emotional Expression	posemo	0.998	0.715	[0.990, 1.007]
	negemo	1.041***	<0.001	[1.032, 1.049]
EW: Cognitive Adaptation	causalIncrease	1.003	0.574	[0.994, 1.012]
	insightIncrease	0.977***	<0.001	[0.969, 0.986]
EW: Dose	length	0.891***	<0.001	[0.882, 0.900]
	frequency	0.845***	<0.001	[0.835, 0.855]
Receiving Support	visitors	0.817***	<0.001	[0.805, 0.828]
	msgPerVisitor	0.910***	<0.001	[0.900, 0.921]
Giving Support	visits	0.992	0.070	[0.983, 1.001]
	msgPerVisit	0.982***	<0.001	[0.973, 0.992]

Table 4. This table presents the survival model testing effects of expressive writing (EW), receiving support, and giving support on continued participation with predictions of death classifiers incorporated. \*\*\*:p<0.001, \*\*:p<0.01, \*:p<0.05.

Survival Model without Death Classification Results				
Construct	Indep. Var.	HR	p value	[95% Conf. Int.]
EW: Emotional Expression	posemo	1.020***	<0.001	[1.014, 1.026]
	negemo	1.044***	<0.001	[1.038, 1.051]
EW: Cognitive Adaptation	causalIncrease	0.983***	<0.001	[0.976, 0.990]
	insightIncrease	0.991*	0.015	[0.984, 0.998]
EW: Dose	length	0.887***	<0.001	[0.880, 0.894]
	frequency	0.873***	<0.001	[0.865, 0.881]
Receiving Support	visitors	0.884***	<0.001	[0.876, 0.893]
	msgPerVisitor	0.924***	<0.001	[0.915, 0.932]
Giving Support	visits	0.991**	0.009	[0.984, 0.998]
	msgPerVisit	0.992*	0.019	[0.985, 0.999]

Table 5. This table presents the survival model testing effects of expressive writing (EW), receiving support, and giving support on continued participation without incorporating predictions of death classifiers. \*\*\*:p<0.001, \*\*:p<0.01, \*:p<0.05.

units, each row of Table 4 and Table 5 analyzes a change in one standard deviation of that unit. The concordance of the model is 0.641 (with standard error 0.001), larger than 0.5. This implies the model has a relatively good prediction ability. The overall log-rank score is statistically significant ( $p < 0.001$ ), suggesting the model is a significantly better model than null model.

**5.2.1 Expressive Writing and User Engagement.** Results in Table 4 show the relationship between expressive writing and user engagement. Specifically, sites are more likely to survive when the percentage of insight words increases more over time, updates are longer, and updates are more frequently posted. Compared to baseline survival:

- Sites with an additional 1.1% increase in insight word usage are 2.3% more likely to survive.
- Sites with an additional 156 words per update are 10.9% more likely to survive.
- Sites with an additional 0.34 updates posted per day are 15.5% more likely to survive.

In this dataset, it appears that the effects of emotion words are mixed. Specifically, using slightly more negative emotion words decreases site engagement. Compared to baseline survival:

- Sites with an additional 0.6% negative emotion words are 3.9% less likely to survive.

However, positive emotion word usage and increasing usage of causal words are not associated with site survival. In summary, emotional words are not positively associated with sustained engagement, whereas increasing insight words, length, and frequency of updates are positively associated with sustained engagement.

**5.2.2 Social Support and User Engagement.** Results in Table 4 show the relationship between social support and user engagement. Specifically, the more visitors, the more messages received per visitor, and the more messages sent to each site visited by authors, the more likely sites are to survive. Compared to baseline survival:

- Sites with an additional 128 visitors are 18.3% more likely to survive.
- Sites with an additional 0.85 messages received from each visitor are 9.0% more likely to survive.
- Sites with an additional 3.98 messages sent to each site visited by authors are 1.8% more likely to survive.

However, visiting other sites is not associated with site survival.

**5.2.3 Comparing Expressive Writing and Social Support.** In order to address our first research question, we can directly compare Hazard Ratios to determine the relative importance of explanatory variables for expressive writing and social support, since all variables have been standardized. **Table 4 demonstrates that getting more visitors, posting updates more frequently, writing longer per update, and getting more messages per visitor are the top four contributing factors positively associated with sustained engagement.** Though giving support (sending more messages to each site visited by the authors) is also positively associated with user engagement, it has the least importance.

**5.2.4 Statistical Effects of Literal Death.** Comparing effect sizes when we incorporate death predictions in our model (Table 4) versus when we do not (Table 5), using the same set of explanatory variables helps to fully answer our second research question. That is, if we do not incorporate literal death in our model, three of the top four influential factors positively associated with site survival are underestimated:

- Adding 128 visitors to a site increases site survival by 18.1% if we incorporate death; this value drops to 11.6% when we do not incorporate death.
- The benefit of adding 0.34 updates per day to site survival drops from 15.5% to 12.7%.
- The benefit of adding an additional 0.85 messages/visitor drops from 9.0% to 7.6%.

The Hazard Ratios of these three variables estimated by the model without death predictions are beyond the 95% confidence interval of those estimated by the model with death predictions, suggesting their statistical difference at the 5% significance level. **Without incorporating death predictions, though the beneficial nature of these variables to site survival do not change, their magnitude is decreased.** The Concordance of the model without death classification is 0.611 (standard error 0.001), which is lower than the corresponding Concordance of the model with death classification, which is 0.641. The Akaike information criterion (AIC) computed for the model without death classification (AIC = 1707655) is larger than that for the model with death classification (AIC = 1070679). Both of these statistics provide evidence that the model incorporating literal death is a better fit to the data.

## 6 DISCUSSION

In this paper, we have shown that during the first month on CaringBridge (CB) cancer sites, early behaviors affect the expected duration of stay within the online health community, echoing similar phenomena in other online contexts, such as peer production. Our results indicate that receiving support, expressive writing, and giving support early on correlate with longer survival. Here, we reflect on these results, sharing implications for the design of online health communities. We also reflect on our methodology of automatically differentiating deceased vs. relinquished sites, in order to provide the CSCW community with insights to guide future research in this space. Finally, we present limitations of this approach and opportunities for future work.

### 6.1 Reflecting on Results

Our most interesting and novel result is that some variables that capture components of expressive writing have comparable (although smaller) effect sizes as receiving social support. This is, we confirmed previous well-supported findings that receiving social support benefits user engagement [13, 90, 91, 95]. We also showed that giving social support is weakly associated with longer survival. However, for the first time, we compared the relative importance of these factors against those of expressive writing in determining how long users will stay. We find it surprising that the dose of expressive writing (measured by *length* and *frequency* of updates) is comparable to receiving support (measured by *number of visitors* and *number of messages received per visitor*). It is possible that writing longer and more frequently in the early stage may increase users' initial commitment to the community, contributing to longer participation overall. It is also possible that users come to CB not only for support, but also because writing about health journeys may be intrinsically beneficial to health, as suggested by [27, 83]. Thus, designers should consider balancing social interaction activities with individual activities like expressive writing, which can have real health benefits for users, in addition to supporting their continued engagement in online health communities.

Prior literature also suggests that cognitive adaption to stressful events is beneficial to health [44, 72, 88]. Our results echo this finding in the context of user engagement, since we find that cognitive adaptation is associated with longer survival. Cognitive adaptation (measured by *increase of insight word usage*) reflects discovery of meaning as more constructive narratives are built, possibly contributing motivation to continue writing. In online health communities that provide space for expressive writing, designers should consider mechanisms that encourage users to engage cognitive reflection processes, and to thoroughly express themselves with some frequency. This could potentially help users construct more meaningful narratives and get more benefits out of the online community.

Finally, prior studies that have examined affective word use (negative vs. positive emotions) have had mixed results regarding benefits of emotional expression. Though a previous experimental study found that writing emotional experiences about cancer (instead of facts) had beneficial effects on health outcomes [86], we found that expressing negative emotions (measured by *negative emotion word usage*) was associated with shorter survival in the community, while expressing positive emotions (measured by *positive emotion word usage*) did not have significant effects. Prior work reported participants' short-term distress after writing about traumatic events [27, 28, 83]. In these *facilitated* expressive writing studies, however, participants must continue writing sessions, which eventually becomes beneficial. In *unfacilitated* expressive writing, there is no such external force to keep users writing. Our results suggest that expressing too much negative emotion at an early stage possibly correlates with decreased motivation to continue engagement. Therefore, encouraging users not to reflect too much on negative emotions in the very beginning may help to sustain newcomer participation.

## 6.2 Reflecting on Methods

For the first time, we examined how deceased sites affect survival analysis in online health communities. We find that three of the four most influential variables are more weakly associated with site survival if we do not incorporate death predictions in our analysis. Corresponding with five-year cancer survival statistics, we predict that for approximately 36.9% of cancer sites, the focal patient dies. Without differentiating these sites, they are treated as normal dropouts, thus biasing results in online health community analyses. Studying death and bereavement is an interesting research topic in and of itself. Yet in this paper, we demonstrate its prevalence in online health communities, and show how it can affect other lines of analysis in this space. Thus, we contribute to the CSCW community a fast way to automatically detect deceased sites based on simple textual features.

## 6.3 Limitations and Future Work

As with many large-scale log analyses, perhaps the primary limitation of this work is that our results are correlational rather than causal. In the future, experimental studies with random assignment of users to different writing conditions in online health communities could demonstrate a causal relationship between expressive writing and user engagement. This could also provide insights about potential differences between facilitated and unfacilitated expressive writing.

Cancer comprises the largest proportion among self-identified conditions on CB. However, by looking only at cancer sites, we lose generalizability to all health conditions. Future work could address this limitation by comparing the effects of expressive writing across different illness domains. Additionally, it is possible that CB attracts a certain type of user who enjoys expressive writing more than the online population at large, though the psychology literature indicates that the benefits of expressive writing do not depend on one's interest in writing, and improve health outcomes across diverse backgrounds [21, 67, 76, 84].

On CB, sites may be authored by patients, their caregivers, or some combination of these two groups. For this study, data was not available on the number of authors per site, and we did not have sufficient information per update to differentiate patient-authors from caregiver-authors. However, it is possible that behavior may differ significantly between these two groups. Thus, our results cumulatively show how individuals *and/or* small groups engage with their online community. We are not aware of any prior work that differentiates patient-authors from caregiver-authors in the analysis of user engagement. Future work could potentially address this issue through automatic classification of update authorship, possibly by examining relative usage of first- and third-person pronouns and other textual indicators. However, this group-authorship situation is itself interesting, and may provide insight into other such contexts.

The CB dataset affords a wealth of opportunities for future HCI research in online health communities. Building on the present work, we hope to look at the evolution of CB sites over time. This line of research could help to identify critical moments when computer-mediated interventions might help patients get the online support they need. Though not discussed in this paper, our reading of the data also revealed nuanced accounts of life and death; in-depth qualitative work could illuminate our understanding of why and how patients use their sites nearing end of life, and how their communities respond to death. For instance, we noticed that some sites simply end with a funeral announcement, whereas others endure, with some authors returning years later to write memorials about important medical events or deceased patients. Interestingly, some authors specifically write that they will take a CB site down because it is no longer needed, or has fulfilled its purpose. Future investigations can reveal how people turn to technology for help during challenging emotional events, and when their needs are perceived as being fully met.

## 7 CONCLUSION

This paper offers two major contributions to the CSCW community. First, we contribute a machine learning model for detecting literal death in online health communities. The resultant automatic predictions provide us with the ability to refine our quantitative analyses and draw more valid conclusions about user behavior. Second, we conduct a large-scale survival analysis, finding that the activity of expressive writing is associated with user engagement in online health communities. Furthermore, we show that receiving support, expressive writing, and giving support have a positive relationship with user engagement, in decreasing relative importance. These contributions provide CSCW with implications for the design of online health communities, as well as implications for future research in this domain.

## ACKNOWLEDGMENTS

We would like to thank the CaringBridge management team, especially Liwanag Ojala, Brigid Bonner, Sharon Berry, Ann Sandgren, and Deb Gailloux for their support and facilitation of this research, along with our colleagues at the Center for Spirituality and Healing who provided feedback, especially MaryJo Kreitzer who facilitated the formation of our collaboration, and our colleagues at the Department of Computer Science and Engineering who provided substantial guidance of our methods, especially Arindam Banerjee, Raghav Pavan Karumur, and Jacob Thebault-Spieker. We would also like to extend thanks to our reviewers for their thoughtful comments, which have helped us to improve the clarity of this work. This work is supported by CaringBridge and the National Science Foundation (grant 1464376). A student on this work is supported by the Graduate Assistance in Areas of National Need (GAANN) Fellowship.

## REFERENCES

- [1] Georg W Alpers, Andrew J Winzelberg, Catherine Classen, Heidi Roberts, Parvati Dev, Cheryl Koopman, and C Barr Taylor. 2005. Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior* 21, 2 (2005), 361–376.
- [2] Isolde K Anderson. 2011. The uses and gratifications of online care pages: a study of CaringBridge. *Health communication* 26, 6 (2011), 546–559.
- [3] Karen A Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in psychiatric treatment* 11, 5 (2005), 338–346.
- [4] Antonina Bambina. 2007. *Online social support: the interplay of social networks and computer-mediated communication*. Cambria press.
- [5] Christopher E Beaudoin and Chen-Chao Tao. 2008. Modeling the impact of online cancer resources on supporters of cancer patients. *New Media & Society* 10, 2 (2008), 321–344.
- [6] Prakhar Biyani, Cornelia Caragea, Prasenjit Mitra, and John Yen. 2014. Identifying Emotional and Informational Support in Online Health Communities.. In *COLING*. 827–836.
- [7] Zoheb H Borbora and Jaideep Srivastava. 2012. User behavior modelling approach for churn prediction in online games. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. IEEE, 51–60.
- [8] Stephanie L Brown, Randolph M Nesse, Amiram D Vinokur, and Dylan M Smith. 2003. Providing social support may be more beneficial than receiving it results from a prospective study of mortality. *Psychological Science* 14, 4 (2003), 320–327.
- [9] Jed R Brubaker and Gillian R Hayes. 2011. We will never forget you [online]: An empirical investigation of post-mortem MySpace comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 123–132.
- [10] Jed R Brubaker, Gillian R Hayes, and Paul Dourish. 2013. Beyond the grave: Facebook as a site for the expansion of death and mourning. *The Information Society* 29, 3 (2013), 152–163.
- [11] Jed R Brubaker and Janet Vertesi. 2010. Death and the social network. In *Proc. CHI Workshop on Death and the Digital*.
- [12] Amy S Bruckman, Casey Fiesler, Jeff Hancock, and Cosmin Munteanu. 2017. CSCW Research Ethics Town Hall: Working Towards Community Norms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 113–115.



- [13] Moira Burke, Cameron Marlow, and Thomas Lento. 2009. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 945–954.
- [14] Chad M Burton and Laura A King. 2004. The health benefits of writing about intensely positive experiences. *Journal of research in personality* 38, 2 (2004), 150–163.
- [15] Linda D Cameron and Gregory Nicholls. 1998. Expression of stressful experiences through writing: Effects of a self-regulation manipulation for pessimists and optimists. *Health Psychology* 17, 1 (1998), 84.
- [16] Leslie F Clark. 1993. Stress and the cognitive-conversational benefits of social interaction. *Journal of social and clinical psychology* 12, 1 (1993), 25–55.
- [17] Melissa A Craft, Gail C Davis, and René M Paulson. 2013. Expressive writing in early breast cancer survivors. *Journal of Advanced Nursing* 69, 2 (2013), 305–315.
- [18] J David Creswell, Suman Lam, Annette L Stanton, Shelley E Taylor, Julianne E Bower, and David K Sherman. 2007. Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin* 33, 2 (2007), 238–250.
- [19] Carolyn E Cutrona and Julie A Suhr. 1994. Social support communication in the context of marriage: an analysis of couples' supportive interactions. (1994).
- [20] Jeremy Dean, Henry WW Potts, and Chris Barker. 2016. Direction to an internet support group compared with online expressive writing for people with depression and anxiety: a randomized trial. *JMIR mental health* 3, 2 (2016).
- [21] Benjamín Domínguez, Pablo Valderrama, María de los Angeles Meza, Sara Lidia Pérez, Amparo Silva, Gloria Martínez, Victor Manuel Méndez, and Yolanda Olvera. 1995. The roles of disclosure and emotional reversal in clinical practice. (1995).
- [22] Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor. 2012. Churn prediction in new users of Yahoo! answers. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 829–834.
- [23] Christine Dunkel-Schetter. 1984. Social support and cancer: Findings based on patient interviews and their implications. *Journal of Social Issues* 40, 4 (1984), 77–98.
- [24] Brian A Esterling, Michael H Antoni, Mary Ann Fletcher, Scott Margulies, and Neil Schneiderman. 1994. Emotional disclosure through writing or speaking modulates latent Epstein-Barr virus antibody titers. *Journal of consulting and clinical psychology* 62, 1 (1994), 130.
- [25] Centers for Disease Control and Prevention. 2015. Five-Year Relative Survival. [https://nccd.cdc.gov/uscs/Survival/Relative\\_Survival\\_Tables.pdf](https://nccd.cdc.gov/uscs/Survival/Relative_Survival_Tables.pdf). (2015). Online; Accessed: 2017-07-01.
- [26] Susannah Fox. 2011. *The social life of health information*, 2011. Pew Internet & American Life Project Washington, DC.
- [27] Joanne Frattaroli. 2006. Experimental disclosure and its moderators: a meta-analysis. *Psychological bulletin* 132, 6 (2006), 823.
- [28] Pasquale G Frisina, Joan C Borod, and Stephen J Lepore. 2004. A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of nervous and mental disease* 192, 9 (2004), 629–634.
- [29] David C Giles and Julie Newbold. 2011. Self-and other-diagnosis in user-led mental health online communities. *Qualitative Health Research* 21, 3 (2011), 419–428.
- [30] Robert L Glueckauf, Timothy U Ketterson, Jeffrey S Loomis, and Pat Dages. 2004. Online support and education for dementia caregivers: overview, utilization, and initial program evaluation. *Telemedicine Journal & E-Health* 10, 2 (2004), 223–232.
- [31] Melanie A Greenberg and Arthur A Stone. 1992. Emotional disclosure about traumas and its relation to health: effects of previous disclosure and trauma severity. *Journal of personality and social psychology* 63, 1 (1992), 75.
- [32] Melanie A Greenberg, Camille B Wortman, and Arthur A Stone. 1996. Emotional expression and physical health: Revising traumatic memories or fostering self-regulation? *Journal of personality and social psychology* 71, 3 (1996), 588.
- [33] Adam J Guastella and Mark R Dadds. 2006. Cognitive-behavioral models of emotional writing: A validation study. *Cognitive Therapy and Research* 30, 3 (2006), 397–414.
- [34] Michael Hardey. 2002. "The story of my illness": Personal Accounts of Illness on the Internet. *Health*: 6, 1 (2002), 31–46.
- [35] Vicki S Helgeson, Sheldon Cohen, Richard Schulz, and Joyce Yasko. 2001. Long-term effects of educational and peer discussion group interventions on adjustment to breast cancer. *Health Psychology* 20, 5 (2001), 387.
- [36] Erika A Henry, Rebecca J Schlegel, Amelia E Talley, Lisa A Molix, and B Bettencourt. 2010. The feasibility and effectiveness of expressive writing for rural and urban breast cancer survivors.. In *oncology nursing forum*, Vol. 37.
- [37] Kasper Hornbæk, Søren S Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough?: on the extent and content of replications in human-computer interaction. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3523–3532.
- [38] Jeffrey H Kahn, Renee M Tobin, Audra E Massey, and Jennifer A Anderson. 2007. Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American journal of psychology* (2007), 263–286.

- [39] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. 2016. Early Activity Diversity: Assessing Newcomer Retention from First-Session Activity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 595–608.
- [40] Jaya Kawale, Aditya Pal, and Jaideep Srivastava. 2009. Churn prediction in MMORPGs: A social influence based approach. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, Vol. 4. IEEE, 423–428.
- [41] Laura A King. 2001. The health benefits of writing about life goals. *Personality and Social Psychology Bulletin* 27, 7 (2001), 798–807.
- [42] Kitty Klein and Adriel Boals. 2001. Expressive writing can increase working memory capacity. *Journal of experimental psychology: General* 130, 3 (2001), 520.
- [43] Neal Krause, A Regula Herzog, and Elizabeth Baker. 1992. Providing support to others and well-being in later life. *Journal of gerontology* 47, 5 (1992), P300–P311.
- [44] Ethan Kross and Ozlem Ayduk. 2008. Facilitating adaptive emotional analysis: Distinguishing distanced-analysis of depressive experiences from immersed-analysis and distraction. *Personality and Social Psychology Bulletin* 34, 7 (2008), 924–938.
- [45] Margaret Laccetti. 2007. Expressive writing in women with advanced breast cancer.. In *Oncology Nursing Forum*, Vol. 34.
- [46] Lynne Lamberg. 2003. Online empathy for mood disorders. *JAMA* 289, 23 (2003), 3073–3077.
- [47] Stephen J Lepore. 1997. Expressive writing moderates the relation between intrusive thoughts and depressive symptoms. *Journal of personality and social psychology* 73, 5 (1997), 1030.
- [48] Stephen J Lepore and Melanie A Greenberg. 2002. Mending broken hearts: Effects of expressive writing on mood, cognitive processing, social adjustment and health following a relationship breakup. *Psychology and Health* 17, 5 (2002), 547–560.
- [49] Carissa A Low, Annette L Stanton, and Sharon Danoff-Burg. 2006. Expressive disclosure and benefit finding among breast cancer patients: mechanisms for positive health effects. *Health Psychology* 25, 2 (2006), 181.
- [50] Diane Maloney-Krichmar and Jenny Preece. 2005. A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 201–232.
- [51] Jennifer Mankoff, Kateryna Kuksenok, Sara Kiesler, Jennifer A Rode, and Kelly Waldman. 2011. Competing online viewpoints and models of chronic illness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 589–598.
- [52] Traci Mann. 2001. Effects of future writing and optimism on health behaviors in HIV-infected women. *Annals of Behavioral Medicine* 23, 1 (2001), 26–33.
- [53] Michael Massimi and Ronald M Baecker. 2010. A death in the family: opportunities for designing technologies for the bereaved. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 1821–1830.
- [54] Michael Massimi and Ronald M Baecker. 2011. Dealing with death in design: developing systems for the bereaved. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1001–1010.
- [55] Michael Massimi, Jackie L Bender, Holly O Witteman, and Osman H Ahmed. 2014. Life transitions and online health communities: reflecting on adoption, use, and disengagement. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1491–1501.
- [56] Michael Massimi and Andrea Charise. 2009. Dying, death, and mortality: towards thanatosensitivity in HCI. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2459–2468.
- [57] Michael Massimi, William Odom, Richard Banks, and David Kirk. 2011. Matters of life and death: locating the end of life in lifespan-oriented HCI research. In *Proceedings of the sigchi conference on human factors in computing systems*. ACM, 987–996.
- [58] Andrea Meier, Elizabeth J Lyons, Gilles Frydman, Michael Forlenza, and Barbara K Rimer. 2007. How cancer survivors provide support on cancer-related Internet mailing lists. *J Med Internet Res* 9, 2 (2007), e12.
- [59] Erin L Merz, Rina S Fox, and Vanessa L Malcarne. 2014. Expressive writing interventions in cancer patients: a systematic review. *Health psychology review* 8, 3 (2014), 339–361.
- [60] Arvind Narayanan and Vitaly Shmatikov. 2010. Myths and fallacies of personally identifiable information. *Commun. ACM* 53, 6 (2010), 24–26.
- [61] Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. 2011. It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 341–350.
- [62] Heather L O'Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the Association for Information Science and Technology* 59, 6 (2008), 938–955.
- [63] Pok-Ja Oh and Soo Hyun Kim. 2016. The Effects of Expressive Writing Interventions for Patients With Cancer: A Meta-Analysis.. In *Oncology nursing forum*, Vol. 43. 468.

- [64] Jason E Owen, Janine Giese-Davis, Matt Cordova, Carol Kronenwetter, Mitch Golant, and David Spiegel. 2006. Self-report and linguistic indicators of emotional expression in narratives as predictors of adjustment to cancer. *Journal of behavioral medicine* 29, 4 (2006), 335–345.
- [65] Jason E Owen, Joshua C Klapow, David L Roth, and Diane C Tucker. 2004. Use of the internet for information and support: disclosure among persons with breast and prostate cancer. *Journal of behavioral medicine* 27, 5 (2004), 491–505.
- [66] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.
- [67] James W Pennebaker. 1997. Writing about emotional experiences as a therapeutic process. *Psychological science* 8, 3 (1997), 162–166.
- [68] James W Pennebaker and Sandra K Beall. 1986. Confronting a traumatic event: toward an understanding of inhibition and disease. *Journal of abnormal psychology* 95, 3 (1986), 274.
- [69] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.
- [70] James W Pennebaker, Michelle Colder, and Lisa K Sharp. 1990. Accelerating the coping process. *Journal of personality and social psychology* 58, 3 (1990), 528.
- [71] James W Pennebaker, Janice K Kiecolt-Glaser, and Ronald Glaser. 1988. Disclosure of traumas and immune function: health implications for psychotherapy. *Journal of consulting and clinical psychology* 56, 2 (1988), 239.
- [72] James W Pennebaker, Tracy J Mayne, and Martha E Francis. 1997. Linguistic predictors of adaptive bereavement. *Journal of personality and social psychology* 72, 4 (1997), 863.
- [73] Jenny Preece, Blair Nonnecke, and Dorine Andrews. 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior* 20, 2 (2004), 201–223.
- [74] Farzana Rahman, Ivor D Addo, and Sheikh I Ahamed. 2014. PriSN: a privacy protection framework for healthcare social networking sites. In *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems*. ACM, 66–71.
- [75] Martina Reynolds, Chris R Brewin, and Matthew Saxton. 2000. Emotional disclosure in school children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 41, 2 (2000), 151–159.
- [76] Jane M Richards, Wanda E Beal, Janel D Seagal, and James W Pennebaker. 2000. Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates. *Journal of Abnormal Psychology* 109, 1 (2000), 156.
- [77] Shelly Rodgers and Qimei Chen. 2005. Internet community group participation: Psychosocial benefits for women with breast cancer. *Journal of Computer-Mediated Communication* 10, 4 (2005), 00–00.
- [78] Sabirat Rubya and Svetlana Yarosh. 2017. Video-Mediated Peer Support in an Online Community for Recovery from Substance Use Disorders. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1454–1469.
- [79] P Servaes, AJJM Vingerhoets, G Vreugdenhil, JJ Keuning, and AM Broekhuijsen. 1999. Inhibition of emotional expression in breast cancer patients. *Behavioral Medicine* 25, 1 (1999), 23–27.
- [80] Bret R Shaw, Fiona McTavish, Robert Hawkins, David H Gustafson, and Suzanne Pingree. 2000. Experiences of women with breast cancer: exchanging social support over the CHESS computer network. *Journal of health communication* 5, 2 (2000), 135–159.
- [81] Elizabeth Sillence. 2013. Giving and receiving peer advice in an online breast cancer support group. *Cyberpsychology, Behavior, and Social Networking* 16, 6 (2013), 480–485.
- [82] Åsa Smedberg. 2008. *Online communities and learning for health: The use of online health communities and online expertise for people with established bad habits*. Ph.D. Dissertation. Institutionen för data-och systemvetenskap (tills m KTH).
- [83] Joshua M Smyth. 1998. Written emotional expression: effect sizes, outcome types, and moderating variables. (1998).
- [84] Stefanie P Spera, Eric D Buhrfeind, and James W Pennebaker. 1994. Expressive writing and coping with job loss. *Academy of management journal* 37, 3 (1994), 722–733.
- [85] Annette L Stanton and Sharon Danoff-Burg. 2002. Emotional expression, expressive writing, and cancer. (2002).
- [86] Annette L Stanton, Sharon Danoff-Burg, Lisa A Sworowski, Charlotte A Collins, Ann D Branstetter, Alicia Rodriguez-Hanley, Sarah B Kirk, and Jennifer L Austenfeld. 2002. Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology* 20, 20 (2002), 4160–4168.
- [87] Acar Tamersoy, Munmun De Choudhury, and Duen Horng Chau. 2015. Characterizing smoking and drinking abstinence from social media. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 139–148.
- [88] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [89] Philip M Ullrich and Susan K Lutgendorf. 2002. Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine* 24, 3 (2002), 244–250.

- [90] Xi Wang, Kang Zhao, and Nick Street. 2014. Social support and user engagement in online health communities. In *International Conference on Smart Health*. Springer, 97–110.
- [91] Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 833–842.
- [92] Marsha White and Steve M Dorman. 2001. Receiving social support online: implications for health education. *Health education research* 16, 6 (2001), 693–707.
- [93] Erin Willis. 2014. The making of expert patients: the role of online health communities in arthritis self-management. *Journal of health psychology* 19, 12 (2014), 1613–1625.
- [94] Andrew J Winzelberg, Catherine Classen, Georg W Alpers, Heidi Roberts, Cheryl Koopman, Robert E Adams, Heidemarie Ernst, Parvati Dev, and C Barr Taylor. 2003. Evaluation of an internet support group for women with primary breast cancer. *Cancer* 97, 5 (2003), 1164–1173.
- [95] Diyi Yang, Robert Kraut, and John M Levine. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6363–6375.
- [96] Jiang Yang, Xiao Wei, Mark S Ackerman, and Lada A Adamic. 2010. Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities. *ICWSM* 10 (2010), 186–193.
- [97] Woohyun Yoo, Kang Namkoong, Mina Choi, Dhavan V Shah, Stephanie Tsang, Yangsun Hong, Michael Aguilar, and David H Gustafson. 2014. Giving and receiving emotional support online: Communication competence as a moderator of psychosocial benefits for women with breast cancer. *Computers in human behavior* 30 (2014), 13–22.
- [98] Bowen Yu, Yuqing Ren, Lorene Terveen, and Haiyi Zhu. 2017. Predicting Member Productivity and Withdrawal from Pre-Joining Attachments in Online Production Groups. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1775–1784.
- [99] Chunlan Zhou, Yanni Wu, Shengli An, and Xiaojin Li. 2015. Effect of Expressive Writing Intervention on Health Outcomes in Breast Cancer Patients: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PloS one* 10, 7 (2015), e0131802.
- [100] Haiyi Zhu, Robert E Kraut, and Aniket Kittur. 2014. The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 281–290.
- [101] Sue Ziebland, Alison Chapple, Carol Dumelow, Julie Evans, Suman Prinjha, and Linda Rozmovits. 2004. How the internet affects patients' experience of cancer: a qualitative study. *Bmj* 328, 7439 (2004), 564.

Received April 2017; revised June 2017; accepted November 2017