Figure 1: Interface for a fictional CaringBridge site. We predict retention (publishing journal updates (a)) from engagement (receiving likes (b) and comments (c)).

How much is a "like" worth? Engagement and Retention in an Online Health Community

Ruyuan Wan
University of Minnesota
Minneapolis, Minnesota, USA
wanxx199@umn.edu

Zachary Levonian
University of Minnesota
Minneapolis, Minnesota, USA
levon003@umn.edu

Svetlana Yarosh
University of Minnesota
Minneapolis, Minnesota, USA
lana@umn.edu

ABSTRACT

Online health communities are designed to help their users acquire social support, but developing self-sustaining communities capable of providing that support requires long-term user retention. Therefore, understanding the factors early in a user's experience that predict their long-term retention is important. In this study, we explore the impact of short-term visitor engagement on long-term user retention. We study users of CaringBridge.org, an online health community for communicating about health journeys, using survival analysis methods to quantify the impact of engagement on retention. First, we explicitly compare the impact of non-text "likes" to text comments, surprisingly finding that likes exceed comments in their impact on retention for some users. Second, we compare less active

CSCW '20 Companion, October 17–21, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8059-1/20/10.
https://doi.org/10.1145/3406865.3418320

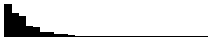
All time	First 30 days
Journal Updates	
% one: 40.6%	% one: 41.9%
M=13.0; SD=33.1	M=6.5; SD=12.3
	
Likes	
% none: 61.8%	% none: 63.5%
M=96.8; SD=572.5	M=39.1; SD=199.5
	
Comments	
% none: 58.7%	% none: 59.9%
M=80.6; SD=258.2	M=33.8; SD=93.4
	

Table 1: Summary stats for 320,063 sites. Histograms show the approximate distribution of each statistic to a max of 100, omitting zero counts (and one-counts for journal updates) for legibility. For modeling, we removed 262 outlier sites with more than 100 updates in the first 30 days, as visual inspection suggested that nearly all of these sites were unfiltered spam.

[1]We partnered with CaringBridge to help them predict retention from visitor engagement early on a site so that they might improve the new-author experience.

and more active users in their response to visitor engagement, finding that more active users are less affected by short-term visitor engagement. We discuss the implications of our findings for the design of OHCs and for future work on visitor engagement.

INTRODUCTION

The goal of this research is to understand the effect of short-term engagement—such as receiving comments and "likes"—on long-term retention of users in online health communities (OHCs). Long-term retention is important both to maintain a self-sustaining community [2] and because continued membership is necessary for giving and receiving support [11]. Social support is associated with decreased OHC user stress [10] and improved health outcomes [4]. Receiving supportive engagement from others in an OHC has been previously associated with user retention [7], but the specific relationship between engagement and retention is underexplored, both in terms of the *medium* of that engagement within the interface and potential differential effects on users with a varying activity level. By addressing this knowledge gap, we aim to facilitate the design of OHCs to encourage retention.

In this research, we study CaringBridge.org, a large multi-condition online health community that provides personal websites for users to share their health journeys. To understand visitor engagement and retention on CaringBridge, we use survival analysis methods to compare the retention of authors on CaringBridge based on the engagement they receive in their first 30 days on the site. We found that users respond differently to engagement depending on the medium of the message ("like" vs comment) and their activity level. These findings address two research questions:

RQ1: Are text comments more impactful on retention than non-text "likes"?
RQ2: Are less active authors affected by engagement differently than more active authors?

Partnership & data description

CaringBridge.org is an OHC for patients and non-professional caregivers to create personal blogs called "sites" to communicate with their existing support networks [8] and to engage with a supportive community [6].[1] On sites, authors write textual *journal updates* to describe their health journey. Journal updates range from a few words to several paragraphs. Visitors to an author's site can leave "likes" (called *hearts* on CaringBridge) and text comments on individual journal updates, depending on the privacy settings of the site. Figure 1 shows the interface. We use likes and comments as proxies for engagement with a site and explore the impact of receiving likes and comments on site retention. In this study, we operationalize site retention as continuing to publish journal updates. CaringBridge shared anonymized user data consisting of 320,063 unique sites created between January 2014 and February 2019. These sites contain 4.2 million journal updates that collectively received 12.4 million likes and 10.8 million comments. Summary statistics are shown in Table 1.

Variable	HR	S.E.
ER=both	0.172^{***}	0.010
ER=comments only	0.294^{***}	0.015
ER=likes only	0.215^{***}	0.032
CC (log)	0.746^{***}	0.008
LC (log)	0.843^{***}	0.007
JC	0.863^{***}	0.001
JC : ER=both	1.159^{***}	0.001
JC : ER=comments	1.088^{***}	0.003
JC : ER=likes	1.105^{***}	0.006

Table 2: Hazard ratios and standard errors for the full model. ER is *engagement received*, CC is *comment count*, LC is *likes count*, JC is *journal count*. Note: $^{***}p < 0.001$

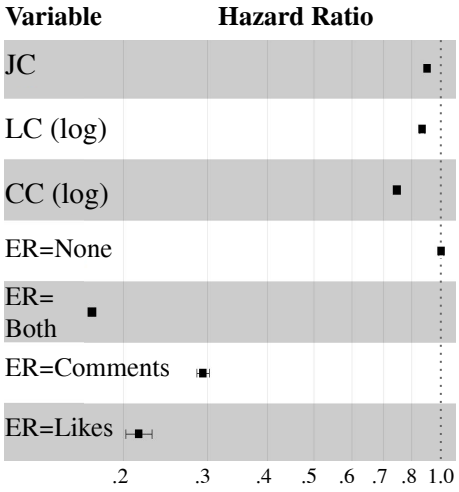


Figure 2: Hazard ratios for the main effects. Hazard ratios that are less than 1 increase the probability of a site surviving.

METHODS

To address our research questions, we used survival analysis to model the impact of short-term engagement on long-term user retention. Survival analysis is a set of statistical approaches used to investigate the time it takes for an event to occur. In our case, the "event" is an author leaving the the community permanently and publishing no additional journal updates on a site. Sites with a final update that occurs after the end of our study period are *right censored*; imitating prior work on CaringBridge, we assume that sites with updates within 90 days of the end of data collection are right censored [7]. This definition treats 7.9% of sites as right-censored.

We used Cox proportional hazards regression to quantify the effect of multiple variables on site survival. All variables are computed only from the first 30 days of site data; we chose the first 30 days because it is a short-enough period to reason about long-term effects and a long-enough period to allow time for multiple journal updates and possibilities for engagement. To address RQ1, we constructed a categorical variable—*engagement received*—with 4 levels describing if a site received one or more likes ($n = 1,973$), one or more comments ($n = 13,496$), both ($n = 114,909$), or neither ($n = 189,423$). Further, we include interaction features for the number of likes and comments received, log-transformed due to their wide range. To address RQ2, we included as a variable the number of journal updates published in the first 30 days. Furthermore, we fit four separate Cox models, stratifying on the journal update quantiles, to specifically compare sites by the frequency of their authors. We include an interaction variable between engagement received and journal update count.

RESULTS

We present the model's coefficients as hazard ratios in Table 2, visualizing their relative impact in Figure 2. The concordance of the model is 0.87, well above a baseline of 0.5. For the main effects, we first find that receiving short-term engagement is associated with increased probability of survival and longer retention. A site with the median two journal updates that receives a single like and comment has a 41.4% probability of surviving past one month, compared to a 2.2% probability for a site that receives no engagement. Second, a greater update frequency in a site's first 30 days is associated with an increased survival probability. Receiving no interactions, a site with only a single update is predicted to have only a 1.2% probability of surviving past one month, compared to a 30.9% probability for a site with 10 updates.

RQ1: Are text comments more impactful on retention than non-text "likes"?

The hazard ratio for receiving *both* likes and comments is lower than receiving likes or comments alone. Compare likes and comments specifically, receiving likes is associated with an increased probability of survival relative to receiving comments. However, the comment count and like count variables indicate that receiving 1 additional comment has a greater impact on survival probability relative to

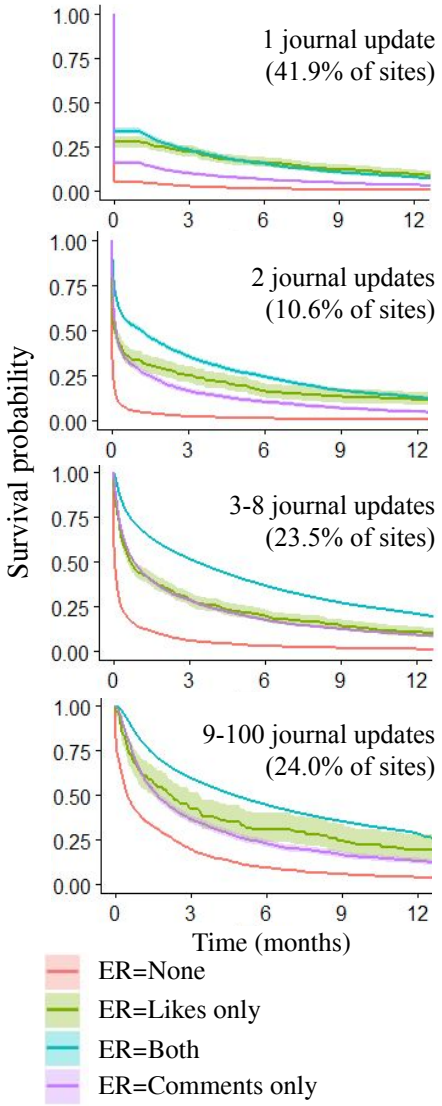


Figure 3: Survival curves for Cox models stratified by journal count

receiving 1 additional like. Thus, *many comments* is more beneficial for survival than *many likes*, but in small amounts receiving likes is more important than receiving comments.

RQ2: Are less active authors affected by engagement differently than more active authors?

The three hazard ratios for the interaction variables between engagement received and update count are all greater than 1, indicating that sites that update more frequently are less affected by having received engagement. By stratifying sites by their journal count, we can specifically examine the differential impact of receiving likes vs comments. Figure 3 shows the survival curve for Cox models trained only on sites with those number of updates. These survival curves show two clear trends: more journal updates is associated with a higher survival probability and receiving both types of engagement is more associated with survival than receiving only one type or receiving neither. Further, we can see that differences in survival due to receiving *only* likes vs *only* comments is primarily driven by sites that update only once in the first 30 days and receive little engagement: in the full model, a single-update site that receives a single like has a 34.9% probability to survive the first month, while a single comment is associated with a significantly lower 24.3% probability (p <0.05).

DISCUSSION & LIMITATIONS

CaringBridge and similar OHCs involve disclosing highly personal health information [3], so visitors to CaringBridge sites may feel they don't know what to say [5] or may be unsure whether it's appropriate for them to respond [1]. Our results suggest that writing a text-based comment may not be necessary; just leaving a "like" can greatly increase the probability of an author continuing to write, particularly if it is the first engagement on a site. OHCs should consider emphasizing likes in their visitor interfaces, encouraging likes in the same way CaringBridge currently encourages comments.

We found that sites with higher journal update frequency are less influenced by engagement, which suggests differences in users' motivations for use of CaringBridge. An author using CaringBridge for reflective writing about their health condition [3, 7], as a personal therapeutic practice [9], or as a platform for disseminating health information may place less importance on visitor engagement. Future research should explore motivations for use of CaringBridge in more detail and consider interface changes to more explicitly support these activities.

Our results suggest that survival probability is greatly increased when even a single like or comment is provided on a site. If a platform were to intervene to support early engagement, such as by directing visitors to that site or even fabricating engagement, would we see the same retention benefits? Experimental work would be necessary to understand how changes in a user's early experience on a platform translate to long-term usage, but conducting such interventions ethically may first require a qualitative understanding of how OHC authors conceptualize visitor engagement and privacy. A key limitation of this work is that we examined only one OHC; a cross-platform study could verify our results in contexts with additional engagement mediums and pave the way for experimental work.

ACKNOWLEDGMENTS

We would like to thank Sunny Parawala, the Minnesota Supercomputing Institute (MSI) at the University of Minnesota, and our partners at CaringBridge.

REFERENCES

[1] Nazanin Andalibi and Andrea Forte. 2018. Responding to Sensitive Disclosures on Social Media: A Decision-Making Framework. *ACM Trans. Comput.-Hum. Interact.* 25, 6 (Dec. 2018), 31:1–31:29. https://doi.org/10.1145/3241044

[2] Antonina Bambina. 2007. *Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication*. Cambria Press. Google-Books-ID: DIW0YKEBcPsC.

[3] Jofen Han and Jo Wiley. 2013. Digital illness narratives: a new form of health communication. (2013). http://scholarworks.wmich.edu/ichita_transactions/18/?utm_content=bufferfb481&utm_source=buffer&utm_medium=twitter&utm_campaign=Buffer

[4] Julianne Holt-Lunstad, Timothy B. Smith, and J. Bradley Layton. 2010. Social Relationships and Mortality Risk: A Meta-analytic Review. *PLOS Medicine* 7, 7 (July 2010), e1000316. https://doi.org/10.1371/journal.pmed.1000316

[5] Maia Jacobs, Jeremy Johnson, and Elizabeth D. Mynatt. 2018. MyPath: Investigating Breast Cancer Patients' Use of Personalized Health Information. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 78:1–78:21. https://doi.org/10.1145/3274347

[6] Zachary Levonian, Marco Dow, Drew Erikson, Sourojit Ghosh, Hannah Miller Hillberg, Saumik Narayanan, Loren Terveen, and Svetlana Yarosh. 2020. Patterns of Patient and Caregiver Mutual Support Connections in an Online Health Community. *Proceedings of the 23rd ACM Conference on Computer-Supported Cooperative Work and Social Computing–CSCW'20* (Oct. 2020), 46. https://arxiv.org/abs/2007.16172

[7] 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 Through Expressive Writing and Social Support. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 73:1–73:24. https://doi.org/10.1145/3134708

[8] Michael Massimi, Jill P. Dimond, and Christopher A. Le Dantec. 2012. Finding a New Normal: The Role of Technology in Life Disruptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 719–728. https://doi.org/10.1145/2145204.2145314

[9] Anthony McCosker and Raya Darcy. 2013. Living with Cancer. *Information, Communication & Society* 16, 8 (Oct. 2013), 1266–1285. https://doi.org/10.1080/1369118X.2012.758303

[10] Bert N. Uchino, Kimberly Bowen, Robert Kent de Grey, Jude Mikel, and Edwin B. Fisher. 2018. Social Support and Physical Health: Models, Mechanisms, and Opportunities. In *Principles and Concepts of Behavioral Medicine: A Global Handbook*, Edwin B. Fisher, Linda D. Cameron, Alan J. Christensen, Ulrike Ehlert, Yan Guo, Brian Oldenburg, and Frank J. Snoek (Eds.). Springer New York, New York, NY, 341–372. https://doi.org/10.1007/978-0-387-93826-4_12

[11] 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 (CSCW '12)*. ACM, New York, NY, USA, 833–842. https://doi.org/10.1145/2145204.2145329