# Understanding and facilitating peer communication in online health communities

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Zachary Levonian

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Loren Terveen, Svetlana Yarosh

July, 2022

# Acknowledgements

I would like to thank my advisors: Lana Yarosh and Loren Terveen. I hope I'm only the first of many to benefit from your complementary advising styles—and senses of humor. I am further indebted to my committee members: I thank Martin Michalowski for his valuable disciplinary perspective and Stevie Chancellor for always cutting through to the real issues.

I am immeasurably lucky to have benefited from high-quality mentorship during this PhD. My thanks to Max Harper, Vanessa Murdock, Daniel Kluver, Isaac Johnson, and Aaron Halfaker for advice and guidance. To my GroupLens colleagues past and present, thank you. A special thanks to Adel Al-Dawood, Jasmine Jones, Haiwei Ma, Fernando Maestre, Sarah McRoberts, Hannah Miller Hillberg, Sabirat Rubya, Irene Ye Yuan, Matthew Zent, and Qian Zhao. And an *extra*-special thanks to my collaborators and friends: C. Estelle Smith and Hao-Fei Cheng. GroupLens would not have been the same without you.

It gives me great delight to thank the talented undergraduate and Master's students I worked with over the last five years—too many to list here! To all of you: thank you, I learned so much from you. My work would not have been possible without the support of the CaringBridge organization. I'd like to acknowledge Brigid Bonner, Pat McMorrow, Dennis Still, Dale Durham, Jon Linn, and my other collaborators at CaringBridge for their support and enthusiasm.

I want to thank my friends who made living in Minneapolis a delight, including Alex Aspell, Phoebe Chastain, Miles Douglas, Margo Fritz, Joel Jaeger, Greg Michel, David Murro, Rose Prullage, and Tony Tran. Particular credit to Greg Michel for making certain introductions and to David Murro for encouraging various unsafe sporting activities.

Finally, I would like to thank my family. I thank my grandparents, to whom this dissertation is dedicated. I thank my parents, Peter Levonian and Betsy Morgan, for teaching me curiosity and for their unflagging support and understanding. I thank my brother, Adam Levonian, for a lifetime of friendship. And I thank my partner, Sunita Chepuri, for her love and support.

# Dedication

For my grandparents.

## Abstract

When a person has a health crisis, the availability of social support affects both their physical and mental health. Online communities can make support available by providing a place to connect with peers who have had similar experiences.However, finding relevant peers to talk to and learn from is challenging. Algorithmic systems for peer matching could help people find relevant peers, but designing such systems requires an understanding of how people use online communities for support—when, how, and to whom they connect. I collaborated with a large existing online community—CaringBridge.org—to understand how patients experiencing a health crisis and their non-professional caregivers use CaringBridge to seek and receive support. Based on this understanding, I created a recommendation system to facilitate peer connections on CaringBridge. CaringBridge users of my system received email recommendations for peer users they may wish to connect with. By collecting survey and usage feedback, I advance an understanding of when support seekers and providers connect with potentially-supportive peers. Taken together, my work describes quantitatively and qualitatively the use of health-related online communities for receiving and providing social support. My work has implications for the deployment of peer-matching systems that facilitate supportive communication.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Social support is associated with improved physical and mental health. In a health crisis, social support may be particularly important, especially from people who have had similar experiences. However, a person's existing offline support community may be insufficient or lack individuals with valuable similar experiences. Online communities offer a promise of scale—a place to find supporters not available in their existing offline support communities. While support is exchanged online on diverse social media platforms, online health communities (OHCs) are those specifically intended for health-related discussion and support. OHCs include open forums, organizing spaces, listservs, information dumps, support groups, and more. The social support provided and received by OHC participants includes access to useful information and advice, a sense of community and belonging, and direct emotional support and coping assistance. To realize the potential of always-available, personalized support, OHCs are designed to facilitate finding information and supporters from among thousands or millions of posts and users.

Peer discovery is the problem of identifying supportive people to connect with online. *Peers* are people with similar health experiences who can leverage that similarity to provide effective support, and OHCs facilitate peer discovery by affording reach and visibility to users. OHC interfaces for peer discovery might require a user to explicitly search or filter for people of interest. For example, users of the mental health community 7Cups.com can specify "topic tags" to identify an appropriate conversational partner [6]. However, in health-related contexts, support seekers and providers may find it challenging to articulate what they are looking for in a way the system can understand—as a user's support goals are diverse and the characteristics that make a potential support provider effective are correspondingly diverse [7]. Algorithmic systems attempt to bridge the gap between the intent of a support seeker and potential providers. These systems match or rank users to facilitate the exchange of support and the identification of peers.

*Recommender systems* for peer matching suggest OHC users or content without explicit guidance from the seeker. Such systems typically incorporate the user's past behaviors as an implicit signal to identify and recommend potential matches. However, designing recommendation systems and their associated interfaces requires an understanding of how people use OHCs for support. Understanding when, how, and who is involved in supportive online interaction enables interventions to increase supportive communication on OHCs. In this dissertation, I conducted an intensive study of a large existing OHC to understand the use of this community and how peer recommendation systems can be designed in this sensitive health context. After formative work aiming to understand how OHC users communicate, I designed, implemented, and evaluated a peer recommendation system during a 12-week field study. Ultimately, this dissertation contributes an understanding of peer interaction on an OHC and an intervention to facilitate more peer interaction in the future.

## 1.1 Collaboration with CaringBridge.org

This dissertation arises from a 5-year collaboration with CaringBridge.org[1], a large online health community based near the University of Minnesota. CaringBridge is designed for patients experiencing some kind of health crisis to communicate with their existing support network. CaringBridge hosts blogs called *sites* on which *authors* publish blog posts called *Journal updates*. As we will see in Chapter 4, most authors are non-professional caregivers writing on behalf of the patient—such as a spouse or parent—rather than the patient themselves. By providing the ability for site visitors to comment or react to individual Journal updates, CaringBridge sites serve both to inform a support community about medical updates and to centralize supportive messages from that community. Chapters 3 and 4 focus on understanding authors usage of CaringBridge—including a surprising amount of between-site author interaction. Chapter 5 builds on CaringBridge to provide a peer recommendation system to authors. Further discussion of the features and interface of CaringBridge will be introduced as needed in

[1] `https://www.caringbridge.org/`

subsequent chapters.

## 1.2 Understanding support seekers and providers

To understand how patients experiencing a health crisis use CaringBridge to seek and receive support, I conducted a mixed-methods study of cancer patient authors. **When and how do cancer patients use a health blogging community?**

Chapter 3 describes a study of cancer patients and their use of CaringBridge over the course of a health journey. Using quantitative classification methods based on prior qualitative work, I identified the "phases" of cancer during which authors join, use, and leave CaringBridge. For example, I found that while most cancer patients join CaringBridge before they start primary treatment, many have already begun treatment, a finding that has implications for onboarding new authors and providing timely support—e.g. by recommending peers who have shared recent treatment experiences. To learn about about support seekers' needs, I further explored the responsibilities that authors discuss in their Journal updates. But understanding users' support needs is insufficient to design for supportive communication; we first need to understand how authors use CaringBridge for supportive peer communication, the subject of Chapter 4.

In Chapter 4, I aim to understand peer communication and how that communication leads to supportive peer relationships. CaringBridge lacks existing "peer finding" features, which enables us to learn about authors' online relationship preferences when these features are absent. I found a surprising amount of peer interaction among CaringBridge authors, and I model these interactions to identify factors associated with the initiation of new connections and the reciprocation and growth of peer relationships. A particular focus of this chapter is the importance of distinguishing between patients and their caregivers; patients and caregivers interact differently, revealing both preferences for and barriers to additional supportive communication with peers. To address some of those barriers, I developed a peer recommendation system with the goal of facilitating supportive peer communication on CaringBridge.

## 1.3 Facilitating peer communication

**How can recommendation systems facilitate peer discovery in a health blogging community?** I designed an algorithmic recommendation system to facilitate peer connections on CaringBridge. I conceptualize peer recommendation as an intervention designed to increase two specific behaviors: reading about peer experiences and interacting with peers. Chapter 5 describes the system I designed and an evaluation of its feasibility during a 12-week field study. During the field study, 79 active CaringBridge authors received a weekly email with personalized peer site recommendations. Based on survey feedback and log data analysis, I determined that peer recommendation is feasible on CaringBridge. I observed evidence of demand for peer recommendation, identified key implementation requirements and trade-offs, evaluated the acceptability of recommendations to participants, and estimated the efficacy of the system at increasing interactions and encouraging the formation of supportive peer relationships. Chapter 6 reflects on these findings, identifying promising areas for future study of peer recommendation—as made accessible through the findings in this dissertation.

## 1.4 Contributions

I contributed to four refereed conference papers related to my collaboration with CaringBridge. The first, led by my colleague C. Estelle Smith, identified a mismatch between support needs as expressed by CaringBridge authors and the visitors who read those author's writings [8]. The remaining three constitute Chapters 3-5 of this dissertation. Chapter 3 was published at ICWSM [9]. Chapter 4 was published at CSCW [10]. Chapter 5 is in submission as of this dissertation. Throughout this collaboration, I have worked with many talented MS and undergraduate students, reflected on the author lists for those submissions. Two students published posters at CSCW as a result. Li et al. developed machine learning models to predict the health conditions discussed in Journal updates [11]. Wan et al. developed survival analysis models to predict the impact that reactions and comments have on the retention of site authors [12].

Taken together, the work described in this dissertation characterizes the use of online health communities for receiving and providing social support. The core contribution of this thesis is an investigation of peer recommendation as an intervention to increase peer communication on CaringBridge. Based on evidence that peer recommendation is acceptable to OHC users and effective at increasing beneficial user behaviors, I ultimately argue for the further study of peer recommendation in OHCs.

# Chapter 2

# Related Literature

For the reader interested in the topic of this dissertation and seeking a high-level introduction, I highly recommend the following books: Stephen Rains' "Coping with Illness Digitally" [13] and Antonina Bambina's "Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication" [14]. The rest of this chapter will take a deeper dive into online health communities, communication during health, and intervening to influence behavior and increase available support. Most of these topics are discussed again in Chapters 3-5, which each contain a more condensed Related Work summary as needed to understand that chapter in isolation.

## 2.1 Methodological underpinnings

Diverse methods are used in HCI [15]. This dissertation draws from both qualitative and quantitative methodological approaches, generally united under a positivist framework. Quantitative approaches use rigid categories and counts—a necessarily reductive summary of reality—to enable comparison and estimation of effect sizes. Qualitative approaches enable contextual description of categories and support the external validity of abstract constructs. Throughout this dissertation, I create a number of constructs, based on qualitative approaches or on prior literature, that I use to quantitatively describe phenomena on CaringBridge—for example, I assert rigid taxonomies for cancer phase, for structural health role, and for update content.

The creation of constructs and taxonomies is common in HCI. Social media data capture human activity that, if structured, can describe aspects of human behavior [16, 17]. The unstructured text of social media data constitutes a trace of human behavior, and those texts can inform us about humans' behaviors and beliefs [18]. Social media text has been used to infer ideology [19], personality [17], nutrition [20], and other aspects of human experience. Behavioral analysis via social media is often used to explore human behavior during periods of change like the birth of a child or a health crisis [21, 3]. To create useful taxonomies, we need to combine prior literature with qualitative insight in the process of *operationalization* [22].

### 2.1.1 Operationalization of Taxonomies

Researchers define taxonomic categories of behavior from three non-exclusive sources: unsupervised machine learning, experts, and qualitative inquiry. Unsupervised machine learning defines categories and the boundaries between them directly from patterns in the data, but it can be hard to validate automatically-inferred patterns or to determine their relevance to the research question at hand [23]. But, questions can be asked and answered using the resulting taxonomies without strictly adhering to prior expectations [24].

Expert-derived taxonomies are built from close collaboration with domain experts [25, 26], a manual reading of existing literature in the target domain [3, 27], or from codebooks of keywords uncovered from "expert" Internet sources [28]. While these taxonomies gain validity from their basis in expert knowledge, this top-down approach may limit the ability to detect novel categories in the data and in many cases the relevant domain expert may not exist.

An alternative is to operationalize a taxonomy from qualitative work. Zhang, Culbertson, and Paritosh aimed to develop a taxonomy from prior work, but found that existing work was too narrow, instead iteratively developing their own taxonomy with experts [29]. Singer et al. used hand-coded survey responses to construct a taxonomy and validated it with an additional survey [30]. While it is ideal for quantitative researchers to collaborate closely with qualitative ones on the same research questions, requiring that qualitative and quantitative experts work together synchronously limits the community's ability to learn from the existing body of qualitative work [31]. In Chapter 3, I articulate a process of taxonomy operationalization from qualitative themes.

### 2.1.2 Validity

Throughout this dissertation, I include analyses intended to support the validity of the core research questions investigated. To do so, I rely on validation methods adapted from psychometrics [32], ablation methods adapted from the machine learning literature [33], and converging lines of evidence from the use of mixed methods [22]. In general, the external validity of my findings for the CaringBridge platform is high, due to lack of sampling error introduced by access to the full CaringBridge database. But how much do my findings generalize to other online communities and support contexts? I argue that my findings are most relevant to communities with similar affordances to CaringBridge. Rains argues that communication technologies for social connection have four primary affordances: visibility, availability, control, and reach [13]. Control, for example, is the potential to manage interactions, so other text-based asynchronous communities provide similar potential for control. Availability is the potential to overcome time and space constraints in connecting with others; communication technologies that restrict access e.g. to people in a particular hospital [34] offer availability in a very different way than CaringBridge. Visibility is the potential to make one's self known to others or to observe others' behavior; on CaringBridge, visibility is linked to specific blogs and the communities that form around those blogs. Reach is the potential to contact specific individuals, groups, or communities. CaringBridge provides reach only to specific individuals known to the user by name; much of our investigation relies on this difference between CaringBridge and other OHCs, first to investigate peer connection when traditional social discovery features are absent and then to facilitate reach via recommendation.

## 2.2 Health journeys

Life disruptions lead to transitions [35]. Transitions are precipitated by change points such as a diagnosis and involve adopting a new role [36]. Meleis' Transitions theory has been influential in nursing [37], and I mention it here to emphasize the theoretical and practical benefits to reasoning about illnesses as changing conditions over time. For patients and their caregivers, changing conditions includes changes in goals, in responsibilities, and in self-identity. CaringBridge is a multi-condition online health community designed to facilitate discussion of longer-term treatment and recovery processes. Cancer is the exemplar condition, for which the metaphor of a *journey* during which a patient's needs will shift is widespread [38]—about half of CaringBridge sites are related to cancer [11].

I will focus here on existing theory around cancer journeys, although parallels exist for many other health conditions. Jacobs, Clawson, and Mynatt articulated a cancer journey framework from qualitative interviews with cancer patients [39]. The framework is organized into three dimensions: responsibilities, challenges, and how the cancer journey influenced patients' daily life. Responsibilities are defined by Jacobs, Clawson, and Mynatt as "the multiple tasks that are placed on patients during each of the cancer journey phases", referring to the phases described by Hayes et al. [40]. Responsibilities are purposeful and goal-oriented tasks that are required of the patient because of a cancer diagnosis; for example, one task associated with a responsibility like "Preparation" would be getting a wig fitting in advance of anticipated hair-loss due to treatment.

Responsibilities shift over time, a reality we will demonstrate quantitatively in Chapter 3. Because CaringBridge is designed to support patients' communication with their extended support networks [41], patients discuss their responsibilities in their Journal updates. While there is a tension between managing self-presentation and "sharing information related to specific needs and desires" [42], I treat patients' discussions of their responsibilities on CaringBridge as veridical representations [43] of their real-world responsibilities. In other words, I assume patients may *omit* responsibilities from discussion on CaringBridge but will not *fabricate* them. It is this mediated view of a patient's journey that makes study of CaringBridge appealing.

To study health journeys as processes of transition and shifting roles over time, it is useful to introduce the concept of phases. *Cancer phases* are used by patients to self-characterize their needs [44], in medical research to organize programs of care [45], and as the basis for prior HCI research [46]. Hayes et al. articulated a model of cancer phases to describe commonalities in patients' experiences of their cancer journeys [40], adapted from foundational work by Corbin and Strauss [47] and subsequently adapted by Jacobs, Clawson, and Mynatt [39]. Wen and Rose used an earlier iteration of this phase model to identify cancer disease trajectories by identifying phase boundaries via automatic event extraction [48]. Other established stage/phase models, like the widely used transtheoretical model of health behavior change, are used as the basis for taxonomies that are tweaked by experts [49].

An important component of health journeys is end-of-life situations. OHCs have a specific role to play at end-of-life, as the use of technology to aide in communication and support coordination is important to patients' quality of life during hospice [50]. Online hospice communities have been studied for their role facilitating social support during hospice care [51]. Studying near-end-of-life activities on online communities provides an opportunity to explore the use of technology contemporaneous with the dying experience, in contrast to approaches that rely on retrospective interviews [52]. As communication and decision-making labor passes from the patient to their caregivers near death [53], a similar transition happens with site authorship on CaringBridge.[1]

## 2.3 Motivations for digital communication during health journeys

Patients use the internet to find information and support [54, 55]. For pursuing social connection specifically, patients use the internet to overcome isolation, identify others with similar experiences, reinforce existing relationships, and offset deficits in existing relationships [13]. CaringBridge is designed primarily for reinforcing existing relationships [8]. However, the presence of connections between authors (Chapter 4) and interest in connecting with fellow authors (Chapter 5) indicates that authors are also using CaringBridge to address unmet needs [56] and build supportive connections based on shared concerns [57]. These support-seeking behaviors result in the formation of *peer* connections, which we discuss next.

Connecting with experientially-similar others is a key motivation for patients to participate in online support communities [58, 59, 60]. Experientially-similar others serve as important sources of support to both patients [59] and caregivers [61, 62]. For CaringBridge authors, experiential similarity is entangled with the notion of authors as peers. We adopt the four-point definition of peer suggested by Simoni et al.: (a) sharing personal circumstances i.e. some form of health challenge, (b) obtaining benefits from peer support that derive from their status as peers, (c) lacking professional training or medical credentials, and (d) "intentionally setting out to interact with individuals they may or may not encounter in their everyday life" [63]. Finding peers to communicate with online is complicated by many individual factors [64]. I study connections between peers—rather than other health relationships such as mentor/mentee and medical professional/patient—as an opportunity to faciliate support exchange at scale and outside the structural biases and inequalities in clinical environments [65, 66].

As CaringBridge most resembles individual health blogs, the motivations of CaringBridge authors may differ from users of other kinds of OHCs. Blogging is fundamentally social [67].

[1] We will discuss account sharing in greater detail in Chapter 4.

Health blogging fulfills both communicative and therapeutic roles [68], with patients sharing their illness trajectories and processing their experiences through writing [69]. While blogging may provide benefits to patients due to the expressive self-disclosure involved in writing blog posts [70, 71], having responsive and interactive readers provides additional benefits [67]. McCosker and Darcy argue that connectivity between blog authors has the potential to sustain health bloggers in their writings about their health journeys [68]. This dissertation focuses on communication between blog authors as a potent opportunity for understanding the dynamics that produce the benefits of these interactive connections.

## 2.4 Social support

Social Determinants of Health is a clinically-accepted framework identifying critical aspects of human social life that influence individuals' health outcomes [72].[2] One key social determinant of health in the context of OHCs is *social support*, since people engage with OHCs as a form of support seeking [73, 13]. Meta-reviews reveal consistent associations between social support and a variety of health outcomes, e.g. mortality [74, 75, 76]. There is no single accepted measure of social support; meta-reviews of studies in this area identify two broad types of measure [74]. *Structural support* aims to measure a person's social roles and the interconnections among their social ties. It is quantified via self-reported measures that capture the size and structure of a person's social network [77] and a person's social integration [78] or social isolation [79]. *Functional support* aims to measure "functions provided or perceived to be available by social relationships" [74]. It is quantified via self-reported measures of both provided support [80] and perceived support [81, 82] (including perceived loneliness [83]). Social support is complex, so multifaceted measures that combine both functional and structural support are usually preferred [74].

In HCI, Shumaker and Brownell's description of the potential benefits of social support remains influential [84, 85]. Social support has a variety of types and associated potential mechanisms through which it create benefits for the receiver [85]. While receiving support via direct interaction with a supporter has intuitive appeal, not all interaction is perceived as wanted or useful, and in general there is mixed causal evidence for the benefits of online interaction-based social support [7]. A gap between received and perceived support bedevils designers of social

[2] `https://www.cdc.gov/socialdeterminants/index.htm`

support interventions: increased received support is only weakly correlated with perceptions of that support [67, 59]. *Perceptions* of support are thus critical, and interaction is only relevant given the context of the communicative act. Exposure to communicated health information alone, for example, may not lead to greater support, but—when perceived as timely and relevant [86]—informational support can increase adherence [87] and facilitate self-management of symptoms [88, 89]. Perceived support is defined by self-perceptions and self-narratives [35, 39], implying the benefits of "helper therapy", where providing support to others is more beneficial to the provider than the receiver [90]. Support needs differ over the course of a health journey, and providing support presents an opportunity to "give back" and enact self-efficacy [35, 39]. A wide variety of theories articulate the mechanisms and constructs of social support and support by peers specifically [7, 59, 63, 65]. This dissertation uses theory only sparingly, to support the construct validity of using specific proxy measures of support. I never measure support directly, a limitation that prevents us from directly evaluating clinically-relevant outcomes from use of CaringBridge.

## 2.5 OHCs as sources of interaction and support

OHCs are associated with a variety of positive and negative health outcomes for their users, specifically due to the interaction that occurs on them [13]. Understanding this interaction in more detail creates opportunities to improve the provisioning of support on OHCs. For patients, use of OHCs is associated with greater perceived support [91, 58], writing higher sentiment posts after interacting with others [92], engagement measures such as duration of stay in a community [93, 70, 94], perceptions of control over illness [95], and decreased mortality [74, 76]. Caregivers also benefit from use of OHCs in terms of reduced stress, although evidence linking specific behaviors to outcomes for caregivers is more mixed than for patients [96, 97]. Despite many benefits, making connections in OHCs can also have negative impacts on users, increasing their stress and leading them to leave the community [13]. For patients, directly making comparisons with other patients can be distressing [98], as can the sudden drop-out of key community members [99]. Furthermore, who an OHC user interacts with and the type of their interactions with others mediate both length of stay in the community and the benefits derived from using it [93]. These mixed and contextual outcomes make designing for the formation of new connections risky and motivate a careful analysis of communication preferences. There are two specific behaviors that deserve specific attention: reading about peer experiences and interacting with peers.

*Reading about peer experiences.* Reading about the experiences of peers can be beneficial even in the absence of interaction [7]. In addition to learning from the valuable information contained in peers' writing e.g. coping strategies [100], reading peer experiences can build a sense of community [101]. Further, reading can reduce loneliness [102], contribute to feelings of normalcy and hope [102, 103], reduce uncertainty and anxiety [100], and enable collective sensemaking about one's journey [104]. In general, reading the experiences of others can benefit readers by enabling positive and normalizing social comparisons to the experiences of others [98, 105]. But, making social comparisons is not without risk: the negative experiences of others can produce a sense of helplessness or increase distress [98, 106].

*Interacting with peers.* Interacting with peers offers many potential benefits—among them are membership in a community, acquisition of new information, normalization of one's experiences, and relief from distress [13]. Online peer interaction can take three general forms: providing support to others, receiving support from others, and forming reciprocal relationships. As previously mentioned, the gap between perceived and received support makes *providing* support potentially more beneficial than *receiving* support [90]. Reciprocal peer relationships can offer the best of both worlds, but also present significant risks in health contexts [13]. Stress can increase if online contacts are doing poorly *or* doing well due to social comparisons [98]. The sudden drop-out of a connection, due to churn or patient death, can also increase distress [99]. Further, peers might be unintentionally unsupportive due to differences in communication style or support preferences [107]. Due to the risks of interacting with peers, interventions designed to increase interaction cannot be deployed without careful evaluation of the risks and benefits.

*Network formation in OHCs.* This dissertation builds on foundational OHC research from Bambina examining a health forum's network structure and its impact on the transmission of social support [14]. Bambina analyzed a static snapshot of the posts in a cancer forum with 84 active participants, finding a core of highly supportive participants with a long tail of periphery members in the social network. The CaringBridge dataset enables research that addresses two key limitations of Bambina's work: (a) examining connections as a dynamic process rather than as fixed in a static network snapshot, and (b) including non-interacting "lurkers". With a complete and dynamic view of the interaction network on CaringBridge, I focus on the creation of *new* peer connections by authors (Chapter 4).

Multiple factors are associated with new connection formation in OHCs. New connection formation is often motivated by shared social identity [108] and experience [103]. Meng examined new connection formation in a weight management social networking site, finding substantial homophily effects related to health condition [109]. Centola and van de Rijt find similar strong homophily effects, noting that platform-specific traits—such as exercise preference in a fitness community—were less important in new health contact identification than demographic traits [110]. In general, health-relevant traits contribute to the creation of specific connections [111]. Outside of health, Seering et al. consider the immediate context (e.g. recent messages) that leads to first participation on Twitch.tv [112].

### 2.5.1 Roles in OHCs

Users take varied roles in OHCs [113, 114]. Research examining roles in OHCs has tended to focus on group [113, 115] or social roles [114, 116] that are defined by behaviors. For example, Sharma et al. define a "seeker" role in a mental health forum as a person who makes a new thread [116]. In contrast, I argue for a focus on structural roles that arise from a health event that creates a *patient* and any number of non-professional auxiliary *caregivers*; these structural roles are adopted by patients and caregivers and are defined by accompanying expectations and responsibilities [117, 118]. The expectations of each role are associated with (but not defined by) a set of behaviors that are "characteristic of the person in a particular setting" [119]. Thus, patients and caregivers have different behaviors as they enact their role in an OHC. Patient and caregiver responsibilities and behaviors may change frequently [9], but their role is relatively stable. This stability contrasts with the frequently-shifting behavior roles identified by Yang et al. in an online cancer forum [114]. Note that structural roles are not explicitly afforded via the technical interface, in contrast to e.g. moderator roles on Wikipedia and other explicit roles that have been studied online [120, 121]. People with the same structural role may be more likely to interact with each other; Xu et al. found that online communication was more likely to occur between Twitter users who had the same health role, such as "provider" or "engaged consumer" [122]. We explore the interaction dynamics between patients and caregivers in detail in Chapter 4.

Patients and caregivers communicate differently online, although these differences have received little explicit focus. Lu et al.—a notable exception—identified differences in topic and sentiment in posts written by patients and caregivers in an online health forum [123]. In this dissertation, however, we focus on supportive connections and the *target* of online interactions by health role. OHCs may provide a particular opportunity for caregivers seeking support online [124], as patients are given "interpretive precedence" in dealing with a health condition, leaving caregivers without supportive relationships to understand their own role in a broader health journey [125]. When caregivers can communicate with other caregivers digitally, they develop more effective coping strategies for caregiving stress [126]. Offline, caregiver connections with other caregivers may be more passive than active; Gage suggests an important role of serendipity in the formation of new connections [61]. In Chapter 4, I find that serendipity plays a role in at least some of the connections on CaringBridge. I aim to understand the differences in patient and caregiver communication on OHCs in response to a call for developing a deeper empirical understanding of OHC participation [13], a context in which caregivers are understudied [127].

## 2.6 Social support interventions

The archetypal peer support intervention is the support group [128]. Online support groups offer similar approaches using a different medium, although generally still designed for and managed in a clinical setting [129, 130]. Other clinical approaches bridge the gap to OHCs—for example, Haldar et al. designed an OHC for people in the same hospital [34]. Peer support interventions have potentially many goals in mind: providing social support, providing health information or education, developing self-efficacy (e.g. by vicarious viewing of peer behavior [101]), adjusting social norms (e.g. use of a particular health behavior), or even facilitating social movements for patient advocacy [63, 35]. In 2004, Cohen expressed skepticism of peer support interventions in general, identifying a string of peer support group studies finding null effects and arguing for a focus on forming weak ties and propping up existing support networks [128]. Nearly 20 years later, the challenges associated with designing effective peer support interventions remain [7]. As an intervention into people's online social networks, we examine recommender systems as a mechanism for encouraging initial interactions that can blossom into weak-tie relationships. In general, recommendation is one approach to improving the quality of support received by peers via matching peers by some measure of "fit".

Peer recommendation can be conceptualized as a clinical decision support intervention that provides filtered information to patients in order to improve their care. This dissertation focuses on non-clinical outcomes and evaluation approaches, but frameworks exist for reasoning about the efficacy of clinical decision support interventions. For example, the "Five Rights" framework encourages intervention designers to think carefully about who is being targeted, what information is provided, where patients will encounter the intervention, how the intervention is designed, and when to provide the intervention [131]. We study peer recommendation in a less clinical context than most other clinical decision support interventions, but these five questions remain important for designing and evaluating an effective intervention.

### 2.6.1 Health peer matching

Health peer matching has occurred largely in the context of hospital-attached programs where mentors and mentees are matched by a 3rd-party broker, usually a nurse or program manager [141, 135, 152]. Consider "woman-to-woman", a peer support program for women with gynecologic cancer: when a new participant expresses interest in the program, the program manager selects a match "of similar diagnosis and age" from a pool of volunteer mentors [136]. In contrast, online peer recommendation is not constrained to formal mentor/mentee pairings and can draw from a much larger pool of prospective "volunteers" at the cost of the clear expectations that come with structure and a human coordinator. I aim to seriously evaluate *non-coordinated* peer matching as a social support intervention (Chapter 5).

Little explicit guidance exists for peer matching [133, 63]. Table 2.1 lists peer characteristics identified in prior work as salient or important for effective peer matching. I distinguish these characteristics as either *proposed* as an implication of a particular study, *used* in practice to match peers in a study or support program, or *expressed* by participants as preferences for or barriers to effective peer support. There are even more characteristics I omitted from the table, including abilities/skills [134], specific needs [64], interaction medium preferences (e.g. email) [142], and existing social connections [142].

While not intended to serve as a rigorous meta-review, this existing literature suggests a wide range of potential characteristics to incorporate in a peer recommender system. While several works have collected empirical data on preferred peer characteristics in support settings, minimal *comparative* work exists to identify the most important characteristics [63]. Hartzler et al. are a notable exception, running scenario-based sessions in which participants explicitly

**Table 2.1: Peer characteristics identified in prior work as important for effective peer matching.**

| Peer Characteristics | Proposed | Used | Expressed Preference |
|---|---|---|---|
| **Health** | | | |
| Diagnosis | [10, 132, 133] [64, 134, 102] | [135, 136, 137, 138, 57] | [139, 140] [141, 142] |
| Treatment | [132, 64] | [143, 57] | [139, 141] [142] |
| Symptoms | [144] | | [142] |
| Severity | [64, 102] | | [145] |
| Timeline | [9] | | [139, 142] |
| Health role | [10] | | [139] |
| Relevant knowledge | [134] | [87] | [139, 146] |
| **Demographics** | | | |
| Age | [132, 133, 64, 102] | [139, 135, 137, 143, 87] | [139, 142] |
| Gender | [132, 133, 64, 147] | [148, 143, 138, 139, 87] | [139, 145] |
| Ethnicity | | [149, 87] | [150] |
| Sexual orientation | | | [145] |
| Nationality | | | [139] |
| **Life** | | | |
| Geography/location | [10, 133, 102] | [149, 148, 138, 57] | [139] |
| Cultural values/background | [132, 151] | [135, 136] | [140, 150] [142] |
| Employment | | [87] | [142] |
| Religion | | [135] | |
| Politics | | | [145] |
| Socio-economic status | | | [141] |
| Education level | | | [142] |
| Social role | [114, 102, 63] | [136] | [139] |
| Marital status | [64] | [143, 87] | [139] |
| Has children? | | [135, 137, 143] | [139] |
| Language | | [136] | |
| **Other** | | | |
| Communication style | [42] | [139] | [141] |
| Lifestyle | | [148] | [140, 142] |
| Interests | | [152, 57] | [142] |
| Personality | | [148] | [146, 141] |
| Commitment to support & recovery | [144] | | [140, 146] |

evaluated five potential peer mentors based on provided health information [139]. Boyes surveyed cancer patients about the importance of specific shared characteristics such as gender, age, and cancer type, although this data is currently unpublished [132].

### 2.6.2 Algorithmic recommendation for peer matching

Few published works explicitly discuss computational recommendation systems for online health communities. Hartzler et al. matched peer mentors on the basis of shared health interests, language style, and demographics—as extracted from prior posts made in the CancerConnect OHC—although they evaluated these matches in workshop interviews rather than actual use [139]. The other notable example is described only in Diyi Yang's thesis: Yang developed and deployed a recommendation system in the American Cancer Society's Cancer Survivor Network (CSN) forums "to direct participants to useful and informative threads that they might be interested in" [153]. They evaluated a model based on implicit feedback from prior commenting behavior by presenting recommended threads and users within the CSN interface, reporting greater thread click-through rate compared to a baseline model recommending recently popular threads.

Outside of health, a variety of problem formulations and modeling methods have been used for the problem of recommending people. For the reader familiar with supervised learning, recommendation models follow a similar formulation: optimizing a loss function that compares the model's output to "ground-truth" labels: explicit or implicit feedback provided by users. Xu et al. present a useful review [154]. Use of implicit behavioral feedback is based on relevance assumptions, e.g. that clicked items are relevant while non-clicked items are not relevant [155], which may not hold true in practice [156, 157]. Given historical user/item pairs, one can then optimize a pointwise loss that rewards high scores for assumed-relevant user/item pairs and low scores for assumed-irrelevant user/item pairs. Input features vary from IDs for the user and item—which gives the classic matrix factorization approach to collaborative filtering [158]—to side information about the context where the recommendation was generated (e.g. the time and place) or content (e.g. prior comments) from the user or item [159, 160].

Person-to-person recommendation is typically modeled using similar methods as user/item recommendation. Facebook's deep learning recommendation model (DLRM) represents a common approach, using embeddings for categorical features (including user IDs) and MLPs for creating dense representations of other features, then combining all representations with a final MLP [161]. Less recently, other sites have used approaches based on neighborhoods and similarity of interactions to connect with strangers specifically. Twitter's "who to follow" recommendations used an alternative approach similar to PageRank that uses only the existing follow network to make recommendations [162]. Guy et al. explicitly attempted to recommend strangers in an enterprise setting based on number of shared interests and memberships [163]. The modeling problem closest to peer recommendation may be romantic relationship recommendation, a context that aims to encourage interaction between users and values reciprocity [164, 165].

### 2.6.3 Alternatives to recommendation

Recommendation is not the only available mechanism for facilitating online peer connections. Two notable alternatives are improving search and filter tools and designing enriched profile pages to make it easier to represent one's diagnosis, expertise, and support needs [104]. Search is challenging in situations where a user's needs are known only implicitly to the user or are challenging to express in terms the system will understand [166]. We suggest that peer support finding is an *exploratory* [167] search task (e.g. see Pretorius et al.'s discussion of person-centered help-seekers [168]). Even with rich peer profiles available, it is challenging for searchers to formulate a query that captures their needs and intent [169, 170]. Other search systems for finding people—such as expertise-finding systems—were created based on interfaces designed to capture users' needs in a domain-specific query [171]. *Mindsets* is a recent example of the design work needed to capture domain-specific intents during query formulation [169]; additional research is needed in the peer support context to capture support seekers' and providers' intents. In contrast to search, recommendation offers opportunities to engage with potential peers without explicitly articulating a person's current needs.

# Chapter 3

# Tracing cancer patient behavior in an online health community

How do authors use CaringBridge during their health journeys? In this chapter, we focus on cancer patients in order to understand the experience of authors on CaringBridge, including when they join and leave CaringBridge and the responsibilities they discuss. To do so, we construct user models based on authors' unstructured text data. To classify this text data, we used conceptual categories arranged in a taxonomy. In many contexts, useful taxonomies can be defined via the incorporation of qualitative findings, a mixed-methods approach that offers the ability to create qualitatively-informed user models. But operationalizing taxonomies from the themes described in qualitative work is non-trivial and has received little explicit focus. Thus, we propose a process and explore challenges bridging qualitative themes to user models, for both operationalization of themes to taxonomies and the use of these taxonomies in constructing classification models. For classification of new data, we compare common keyword-based approaches to machine learning models. We demonstrate our process to understand cancer patients on CaringBridge, constructing two user models tracing cancer patient experience over time. We identify patterns in the model outputs for describing the longitudinal experience of cancer patients and reflect on the use of this process in future research. The contents of this chapter were previously published at the ICWSM conference [9].

## 3.1 Introduction

Social media data offers the promise of human behavioral insight that is temporally linked and captured contemporaneously with that behavior [16]. While much of this data is unstructured, methods for identifying patterns—such as supervised machine learning—are increasingly being used to extract structure for further analysis [172]. Developing computational models to do this data extraction requires defining a taxonomy: the explicit structure to extract from the underlying social media data. For example, to identify targets of online hate, Salminen et al. defined a complex taxonomy capturing the nuances of hate speech [173]. Researchers use taxonomies that are created by experts, derived unsupervised from the data, or adapted from prior work. In this chapter, we create taxonomies directly from themes identified in qualitative research.

The incorporation of qualitative research into user modeling is beneficial because mixed methods enable researchers to triangulate their understandings, refine theory, and make use of the strengths of both qualitative and quantitative methods [174]. However, the themes and implications described in qualitative work cannot be taken "as is" as a taxonomy. Even when qualitative themes have an appropriate level of granularity for the research question at hand [29], an explicit mapping of themes to divisions in the data must be constructed to derive quantitative models. In this work, we focus on the problem of *bridging* existing qualitative work to computational user models built from social media text data. We consider bridging as a two-stage process involving (1) *operationalization* of qualitative themes into a taxonomy, and (2) *classification* of the data based on that taxonomy. This process has received little explicit focus in prior research; we argue that developing an operationalization process for qualitative themes can better enable the incorporation of qualitative insights into user model taxonomies.

We implement an operationalization method for identifying taxonomic boundaries for two qualitative frameworks in the cancer domain, identifying critical challenges in this process. These taxonomies seek to support modeling based on user-generated text. Two common approaches to this are identifying keywords that signify inclusion in a particular taxonomic category versus supervised machine learning based on human annotation of text into the taxonomic categories. We conduct empirical comparisons of these two approaches for classification of categories in the derived taxonomies.

Our present study is motivated by research questions related to cancer patients' labor and their use of online health communities (OHCs). Substantial sources of social media data capture the experiences of cancer patients, but no existing operationalizations bridge these data to the extensive qualitative work describing the experiences and needs of cancer patients. By bridging existing qualitative health theories into computational models of patients' OHC use, these models can inform the delivery of digital services [46]. From qualitative frameworks developed by Jacobs, Clawson, and Mynatt [39] and Hayes et al. [40], we iteratively develop taxonomies for classifying cancer patient responsibilities and temporal cancer phases. We use supervised machine learning to construct computational models that trace cancer patients' experiences through their OHC posts.

The contributions of this work are (1) an articulation of a bridging process between qualitative themes and quantitative models, (2) a comparison of two classification methods for taxonomies—supervised machine learning and keyword-based classifiers, and (3) the extension of two existing qualitative frameworks to a novel social media context. Our proposed bridging process builds towards researcher triangulation of findings across methodological approaches to build more robust user models. We describe our application of the two stages of the bridging process in the Operationalization and Classification sections, then reflect on the two models' validity and predictions in the Model Analysis section. In the Discussion, we identify implications for future researchers using this method.

## 3.2 Related Work

Social media data contains traces of human activity that, if structured, can reveal human behavior [16, 17]. The unstructured text of social media data constitutes a trace of human behavior, and those texts can inform us about humans' behaviors and beliefs [18]. Social media text has been used to infer ideology [19], personality [17], nutrition [20], and other aspects of human experience. Behavioral analysis via social media is often used to explore human behavior during periods of change like the birth of a child or a health crisis, as we do here [21, 3]. In the next two sections, we discuss background on taxonomies and classification.

### 3.2.1 Operationalization of Taxonomies

To create user models, researchers define taxonomic categories of behavior from three non-exclusive sources: unsupervised machine learning, experts, and qualitative inquiry. Unsupervised machine learning defines categories and the boundaries between them directly from patterns in the data, but it can be hard to validate automatically-inferred patterns or to determine their relevance to the research question at hand [23]. But, questions can be asked and answered using the resulting taxonomies without strictly adhering to prior expectations [24].

Expert-derived taxonomies are built from close collaboration with domain experts [25, 26], a manual reading of existing literature in the target domain [3, 27], or from codebooks of keywords uncovered from "expert" Internet sources [28]. While these taxonomies gain validity from their basis in expert knowledge, this top-down approach may limit the ability to detect novel categories in the data and in many cases the relevant domain expert may not exist.

An alternative is to operationalize a taxonomy from qualitative work, which is the approach we explore. Zhang, Culbertson, and Paritosh aimed to develop a taxonomy from prior work, but found that existing work was too narrow, instead iteratively developing their own taxonomy with experts [29]. Singer et al. used hand-coded survey responses to construct a taxonomy and validated it with an additional survey [30]. While it is ideal for quantitative researchers to collaborate closely with qualitative ones on the same research questions, requiring that qualitative and quantitative experts work together synchronously limits the community's ability to learn from the existing body of qualitative work [31]. By articulating a process of taxonomy operationalization from qualitative themes, user models benefit from existing bottom-up work.

### 3.2.2 Classification of Social Media Data

Once taxonomies are defined, two primary approaches are used to classify available text data: the use of specific word patterns by lexical analysis of texts through the discovery of words closely related to a desired category [175] (i.e. keyword-based approaches) and supervised machine learning (ML).

Keyword-based approaches are appealing because they are interpretable and require no human annotation of data. These approaches often involve soliciting keywords from an expert [26]. The line between building a taxonomy from "constructs of interest" [175] and selecting keywords to use in that taxonomy is often blurred e.g. in [176]. Such approaches run the risk of missing important variants of the phenomena under study [173] and may need additional human validation [177].

In contrast, supervised ML can result in higher precision than keyword lists on social media data [177] and find patterns that are more generalizable and robust [29]. We compare supervised ML to keyword-based approaches to further articulate the trade-offs of interpretability versus robustness.

## 3.3 Study Design

We investigate the proposed bridging process in the context of cancer patients' OHC posts. In this section, we provide the relevant qualitative background (3.3.1), describe the OHC (3.3.2), and discuss the selected data (3.3.3). Subsequent sections describe the operationalization, classification, and finally analysis of the model outputs, with each section addressing the methods used and our results.

### 3.3.1 Cancer patients and OHCs

Online health communities (OHCs) are used by patients and caregivers to seek social support [178]. We focus on patient use of CaringBridge, an online health community. Responding to the call for catalyzing social support by understanding and enhancing OHCs [179], we use unstructured text of patient posts to model their use of CaringBridge. In contrast, most prior user modeling health research has relied primarily on structured health information like self-reported condition [180]. In the next sections, we discuss the theoretical foundations from which we operationalize taxonomies.

#### Phases and transitions

The concept of cancer *phases* are used by patients to self-characterize their needs [44], in medical research to organize programs of care [45], and as the basis for prior HCI research [46]. In this work, we adopt the phase model of cancer articulated by Hayes et al. [40] and adapted by Jacobs, Clawson, and Mynatt [39] to describe commonalities in patients' experiences of their cancer journeys.

While we are the first to use Hayes et al.'s phases in quantitative modeling, Wen and Rose used an earlier iteration of this phase model to identify cancer disease trajectories, although their emphasis is on phase boundary identification via automatic event extraction [48]. Liu, Weitzman, and Chunara utilized supervised ML of social media posts to identify drinking behavior through a series of discrete stages [25]. Although conceptually similar to phases, their stage taxonomy was developed with the input of domain experts. We utilize a similar modeling approach and follow their lead in the use of active learning. Other established stage/phase models, like the widely used transtheoretical model (TTM) of health behavior changes, are used as the basis for taxonomies that are tweaked by experts [49]. The TTM has been refined through both theory-building and empirical validation over many years [181]; in contrast, the Hayes et al. phase model is based directly on qualitative work and has not yet been explored in diverse contexts. Our operationalization contributes to a broader effort of theoretical refinement [22]. On the quantitative side, concepts similar to phases have been operationalized via discrete observable keyword-patterns e.g. for the identification of recovery events [182].

### Cancer journey framework

Jacobs, Clawson, and Mynatt articulated a cancer journey framework (CJF) from qualitative interviews with cancer patients [39]. The CJF is organized into three dimensions: responsibilities, challenges, and how the cancer journey influenced patients' daily life. We focus only on the responsibilities, defined by Jacobs, Clawson, and Mynatt as "the multiple tasks that are placed on patients during each of the cancer journey phases", referring to the phases described by Hayes et al. [40]. Qualitative exploration of the dataset indicated that the other two dimensions were seldom visible in the details of patients' posts. The responsibilities and their corresponding phase assignments are listed in Table 3.1, along with abbreviated responsibility codes used where space is limited. Responsibilities are purposeful and goal-oriented tasks that are required of the patient because of a cancer diagnosis; for example, one task associated with the Preparation responsibility would be getting a wig fitting in advance of anticipated hair-loss due to treatment.

CaringBridge is designed to support patients' communication with their extended support networks [41]. Therefore, we expected that patients would discuss their responsibilities with their CaringBridge support network. While there is a tension between managing self-presentation and "sharing information related to specific needs and desires" [42], we treat patient's discussions of their responsibilities on CaringBridge as veridical representations [43] of their real-world responsibilities. In particular, we assume patients may *omit* responsibilities from discussion on CaringBridge but will not *fabricate* them, such that our computational models can be taken as a high-precision view of cancer patients' responsibilities. By classifying these responsibilities on CaringBridge, we aim to conceptualize patients' communication of their labor.

We selected the CJF and the Hayes et al. phase model for use in this bridging process based on our broader research question, which was related to understanding patient labor needs over time so that we can design more effective, personalized online interventions to meet or reduce those needs. We offer no guidance on the identification and selection of qualitative frameworks for this bridging process other than alignment with the research question of interest; this is an important theoretical problem that deserves additional attention in future work. We acknowledge a broader tension in qualitative research regarding the generalizability of qualitative work; while not all qualitative work is intended to generalize, we select frameworks that comprise "in-depth analysis of specific, local phenomena, with the intention of generalizing to other sites and other people" [183]. Our bridging process builds on that intention.

**End-of-life**

The CJF was developed through retrospective patient interviews. Thus, one limitation is that it necessarily omits cancer journeys that conclude with the death of the patient. OHCs have a role to play in end-of-life situations, as the use of technology to aide in communication and support coordination is important to patients' quality of life during hospice [50]. Online hospice communities have been studied for their role facilitating social support during hospice care [51], but OHCs like CaringBridge have not been specifically investigated in this context. While most studies of technology use at end-of-life have relied on retrospective interviews [52], CaringBridge provides an opportunity to explore the use of technology at end-of-life contemporaneous with the dying experience. As communication and decision-making labor passes from the patient to their caregivers near death [53], we expect that many aspects cannot be captured via the patient's own writing; however, these data remain a unique opportunity to analyze responsibilities articulated during the end-of-life phase.

### 3.3.2 CaringBridge research collaborative

This work was conducted during a research collaboration between CaringBridge (CB) and the University of Minnesota. CB is a global, nonprofit social network dedicated to helping family and friends communicate with and support loved ones during a health journey.

#### Platform description

CaringBridge.org offers individual *sites* for users—free, personal, protected websites for patients and caregivers to share health updates and gather their community's support. Each site prominently features a *journal*, which is a collection of multiple health *updates* by or about a patient. Updates are comprised of text and are timestamped with a creation date and time. This terminology reflects that used by Ma et al. [70].

#### Data description and ethical considerations

The complete dataset used for this analysis includes de-identified information from 588,210 CaringBridge sites created between June 1, 2005 and June 3, 2016. The site data were acquired through collaboration with CB leadership in accordance with CB's Privacy Policy & Terms of Use Agreement. The study in this chapter was reviewed and deemed exempt from further IRB review by the University of Minnesota Institutional Review Board. We acknowledge the tension in HCI between open data dissemination [184] and the ethical necessity to protect participants' rights and privacy [185]. As CB data are highly sensitive, we opt not to publicly release the dataset used for analysis in this chapter or to use crowdsourcing for annotation. In compromise between replicable science and the ethical protection of participants' privacy, we welcome inquiries about the dataset by contacting the authors. We do release our taxonomy definitions and analysis code.[1]

### 3.3.3 Study data selection

Most sites in the CB dataset are not relevant to this study, as the CJF and phase model articulate themes only for cancer patients. We include only sites that self-reported cancer as the health condition category at the time of site creation. For ethical reasons, we further omit sites deleted

[1] `github.com/levon003/icwsm-cancer-journeys`

| Code | Responsibility | Phase |
|---|---|---|
| CO | Communicating the disease to others | PT |
| IF | Information filtering and organization | PT |
| CD | Clinical decisions | PT |
| PR | Preparation | PT |
| ST | Symptom tracking | T |
| CS | Coordinating support | T |
| SM | Sharing medical information | T |
| CP | Compliance | T |
| MT | Managing clinical transition | T |
| FM | Financial management | T |
| CM | Continued monitoring | NED |
| GB | Giving back to the community | NED |
| BC | Health behavior changes | NED |

**Table 3.1: Patient responsibilities in the CJF and the phase within which that responsibility was organized. Phase is either *pretreatment* (PT), *treatment* (T), or *no evidence of disease* (NED).**

by the site authors. To account for shifts in the design and demographics of CaringBridge over time, we include only sites created in 2009 or later. We focus on completed sites, ones with their final journal updates made before April 1st, 2016 (two months before the end of the dataset's span). We analyze only sites active enough to capture part of the patient's cancer journey, which we define as sites with at least five journal updates spanning at least one month.

Finally, as sites may have multiple authors and we are only interested in sites written by patients themselves, we exclude sites in which fewer than 95% of the updates were authored by the patient. We identify updates as patient-authored or not using a binary Vowpal Wabbit logistic regression classifier with L2 regularization [186]. Hashed unigram and bigram bag-of-words features were used. During data exploration, two researchers annotated updates as evidently patient-authored or not. Agreement was generally high (Cohen's $\kappa = 0.72$), disagreements primarily arising from very short updates. To improve classifier accuracy and address biases potentially introduced via non-random sampling of updates for annotation, we conducted several rounds of uncertainty sampling, resulting in a training set of 1,035 updates. This classifier achieved an accuracy of 92.5% on a held-out validation set of 258 updates, which we determined to be sufficient for the accurate identification of sites primarily authored by patients. During random sampling of sites for the human annotation described subsequently, we observed no sites that were not primarily patient-authored. After the exclusion of sites based on the authorship classifier, we selected 4,946 sites for subsequent analysis (described in Table 3.2) containing 158,597 updates.

| | | | |
|---|---|---|---|
| Journal Updates | Median: 22 updates<br>M=32.1; SD=43.7 |  | |
| Site Visits | Median: 1017 visits<br>M=2099.2; SD=4136.9 |  | |
| Survival Time | Median: 8.2 months<br>M=12.9; SD=13.3 |  | |
| Breast | 2752 (55.6%) | Leukemia | 209 (4.2%) |
| Lymphoma | 597 (12.1%) | Ovarian | 169 (3.4%) |
| Other | 380 (7.7%) | Lung | 168 (3.4%) |
| Not Specified | 257 (5.2%) | Myeloma | 120 (2.4%) |
| Colorectal | 225 (4.5%) | Brain | 69 (1.4%) |

**Table 3.2: Descriptive info about the 4,946 selected CB sites. Survival time is the time elapsed between the first and last journal update on a site.**

## 3.4 Operationalization

### 3.4.1 Operationalization Methods

We define operationalization as the construction of a structured taxonomy from description of themes in existing qualitative theory. Following Zhang et al. [187], we suggest that not all themes may be useful in the target social media context; rather, the operationalization process creates a “shared vocabulary” that identifies conceptually coherent categories. Echoing Figueiredo et al. [188], the qualitative framework is a lens— a “conceptual framework to recognize and compare”—to understand the relationship between patients’ writing on CB and the taxonomic categories.

Tangibly, operationalization involves a mapping between indicators in the data and particular qualitative themes. These mappings define the categories in the taxonomy. We operationalize two taxonomies from the phase and responsibility frameworks discussed in Section 3.3.1. In a social media context, data indicators are units of text that relate to the qualitative framework. For example, we defined a particular responsibility to be present in an update if the author explicitly acknowledges having done a related task or having a need for a related task; a patient's description of a task *indicates* the presence of a responsibility. The taxonomy codebook describes which task descriptions indicate particular categories. We focus on indicators of responsibilities that require human but not specific-domain expertise to identify [187]; in particular, it's not at all clear what if any domain expertise could exist for responsibilities given that the indicators are non-medical.

For both phases and responsibilities, we created initial category descriptions directly from the theme descriptions in the corresponding qualitative work. We conducted multiple rounds of annotation followed by discussion to resolve disagreements, resulting in updates to the taxonomy in the form of examples and guidance for annotators. Such iterative processes are widely used in codebook development [27, 176, 189]. Annotators could assign as many responsibility labels to an update as evidence indicated, while phase labels were initially treated as mutually exclusive. Four researchers participated in codebook development and annotation, all familiar with CaringBridge data but not medically trained. Two of these researchers functioned as primary annotators, together annotating the majority of labeled data. Each round of annotation consisted of the primary annotators independently labeling 20 randomly sampled sites and computing Cohen's $\kappa$ to assess the level of inter-rater reliability (IRR). After taxonomies were defined, we annotated additional sites to provide data for the training of classification models.

| Phase | Occurrence | Disagreement | $\kappa$ |
|---|---|---|---|
| PT | 7.4% | 5.5% | 0.91 |
| T | 69.7% | 7.4% | 0.94 |
| EOL | 1.9% | 0.2% | — |
| NED | 6.4% | 3.6% | 0.95 |
| Overall | 99.62% | 10.2% | 0.93 |

**Table 3.3: Annotated phase occurrence proportions and IRR. Disagreement is the percentage of a phase's occurrence in multi-annotated updates with disagreement. Cohen's $\kappa$ is reported for two coders' annotations of 31 sites containing 619 updates; none of these sites contained EOL updates. Overall stats describe updates annotated with any phase.**

| Responsibility | Occurrence | Disagreement | $\kappa$ |
|---|---|---|---|
| CO | 1.3% | 2.3% | 0.00 |
| IF | 7.5% | 17.0% | 0.06 |
| CD | 3.4% | 6.1% | 0.21 |
| PR | 14.4% | 26.2% | 0.22 |
| ST | 20.4% | 32.9% | 0.15 |
| CS | 9.2% | 12.9% | **0.43** |
| SM | 52.4% | 16.7% | **0.57** |
| CP | 46.6% | 26.8% | **0.45** |
| MT | 12.3% | 22.9% | 0.13 |
| FM | 1.8% | 2.6% | **0.42** |
| CM | 5.0% | 7.4% | 0.32 |
| GB | 2.6% | 4.8% | **0.42** |
| BC | 2.6% | 4.4% | **0.44** |
| Overall | 96.19% | 85.2% | 0.10 |

**Table 3.4: Annotated responsibility occurrence proportions and IRR. Disagreement is the percentage of a responsibility's occurrence in multi-annotated updates with annotator disagreement. Cohen's $\kappa$ is reported for two coders' annotations of 20 sites containing 471 updates; the six emphasized responsibilities are used for future classification. Overall stats describe annotated updates containing at least one responsibility, where $\kappa$ evaluates agreement with the requirement that both annotators agree on all responsibilities for that journal.**

### 3.4.2 Operationalization Results

We identified two challenges common to both phase and responsibility operationalization: interrogating thematic boundaries and mapping the conceptual to the observable.

#### Interrogating thematic boundaries

We experienced challenges developing distinct boundaries between themes from the indicators in the text. For the phase taxonomy, we began our exploration using all five phases described by Hayes et al. [40]: screening and diagnosis, information seeking, acute care and treatment, no evidence of disease, and chronic care and disease management. We observed that "screening and diagnosis" and "information seeking" were intertwined; updates in the first few weeks of a site described experiences with no clear correspondence with exactly one of the phase themes. We merged these themes into a single "pretreatment" phase that encapsulates the qualitative descriptions of both, constructing a taxonomy with four categories: pretreatment (PT), treatment

(T), no evidence of disease (NED), and chronic care and disease management (EOL). Hayes et al. included discussion of the valid transitions between phases (depicted in Figure 3.1 as arrows), which we found to cohere with the data patterns we observed i.e. we observed no transitions other than those indicated. To complete the phase taxonomy, additional rounds of annotation focused on clarifying the most relevant thematic boundaries—PT/T and T/NED—and adding examples to the annotation guidance e.g. identifying medical port insertion as a common transition from PT to T.

For the responsibility taxonomy, we observed two distinct types of indicators that referred to the CJF's Support Management responsibility. The patients' literal descriptions of coordinating support blended with the sharing of medical information by authoring the CB update. We split Support Management into two new responsibilities—Coordinating Support and Sharing Medical Info—each defined from subsets of the CJF's description of Support Management. This split enabled us to disentangle acknowledgements by the patient of support coordination apart from the act of writing updates on CB. Pooling could be used for later analyses, but we embraced the suggestive split in the data. With 13 responsibilities, we had many more boundaries to negotiate and discuss, finding that a single task indicator may correspond with multiple responsibilities in an ambiguous way.

**Mapping conceptual to observable**

In mapping conceptual themes to observable units of data, some indicators were ambiguously linked to one or more categories. For the phase taxonomy, we observed updates that described transitions between phases or for which phase could not be confidently identified. To address this challenge, we allowed annotators to select up to two phases for a single update and introduced an "Unknown" checkbox to the annotation interface to indicate uncertainty.

For the responsibility taxonomy, we observed that many responsibilities were ambiguous within the data, consistently finding low IRR despite multiple rounds of iteration and discussion. In the final round of iteration, we adapted a method described by Schaekermann et al. [190] to conduct a more detailed disagreement discussion process for the seven responsibilities for which we found IRR to be the lowest. This process consisted of (i) an evidence-finding phase in which an annotator was asked to highlight specific textual evidence for a particular responsibility's presence in an update, followed by (ii) a reconsideration phase in which annotators who had not indicated the presence of that responsibility were asked to consider the presence of that responsibility in light of the textual evidence provided by another annotator. 25.7% of 152 updates reconsidered in this discussion process resulted in irresolvable disagreement i.e. the primary annotators continued to disagree. Furthermore, after subsequent annotation of 20 sites to compute IRR (Table 3.4), three of the seven responsibilities involved in the disagreement discussion process achieved lower agreement compared to scores on a prior annotation set. The high amount of irresolvable disagreement indicates high ambiguity in those responsibility's themes. We return to this point in Section 6.

### Complete taxonomies

For phases, we defined taxonomic categories over two rounds of iteration, finding high annotator agreement as shown in Table 3.3. Patterns between the annotated phases are shown in Figure 3.1. For responsibilities, we defined taxonomic categories over five rounds of iteration. Table 3.4 shows low annotator agreement for many responsibilities. Low-agreement responsibilities like Preparation may not be useful in describing patient behavior in a social media context without further qualitative elucidation of those responsibilities; as it stands, the mapping from the conceptual to the observable is too ambiguous. As we turn to classification, we drop the lowest-agreement responsibilities and focus on the six responsibilities with Cohen's $\kappa > 0.4$. This division is arbitrary, but reflects commonly used guidelines indicating $\kappa > 0.4$ as moderate agreement [191].

Total human annotations of sites and updates following the final iteration of both taxonomies are shown in Table 3.5. During random sampling, we observed only a single site that ended in the death of the patient, which ran counter to a finding from Ma et al. that 37% of cancer sites on CB do so [70]. We speculated that patient-centered narratives are less likely to provide clear indicators of patient death. To investigate end-of-life sites more carefully, we identified a high-precision filter to identify candidate sites that may contain such updates. We filtered to 63 sites using the conjunction of predictions from a death classifier developed by Ma et al. [70], a keyword list[2], and sites that ended with a non-patient-authored update. After annotation of these sites for phases, we determined that 82.5% of them contained end-of-life updates.

[2] Keywords used: hospice, funeral, death, passed away, obituary, wake, commemoration

| Sampling: | Random | | Uncertainty | | Death | |
|---|---|---|---|---|---|---|
| | S | U | S | U | S | U |
| Cancer Phases | 109 | 2791 | 28 | 278 | 63 | 3852 |
| Responsibilities | 82 | 1891 | 23 | 34 | — | — |

**Table 3.5: Human annotation counts in terms of number of sites (S) and number of journal updates (U).**

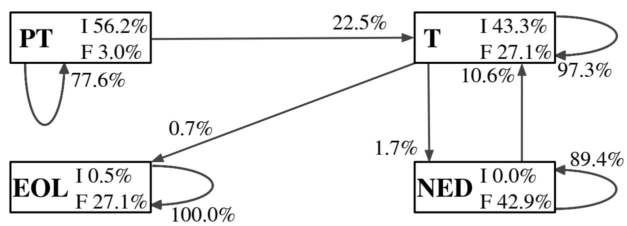



**Figure 3.1: Phase transition probabilities based on human-annotation. Each phase indicates the percentage of sites with an initial (I) and final (F) update in this phase.**

## 3.5 Classification

### 3.5.1 Classification Methods

Using the complete taxonomies operationalized from qualitative work, we classify CB updates by taxonomic category. We train supervised ML models from the data annotated during the development of the taxonomies. We compare the ML classifiers to keyword classifiers that assign a category label to an update if it contains one of the words on a keyword list defined for that category.

#### ML classifier

We formulate both phase and responsibility identification as multilabel classification problems. For phases, the prediction target is a 4x1 vector of labels corresponding to the four phases, whereas for responsibilities the prediction target is a 6x1 vector. To make use of correlations between the classes, we evaluate multilabel models rather than transforming the problem to independent binary classification problems [192]. We remove from consideration all updates with fewer than 50 characters of text content between the title and body text. All models were trained using Vowpal Wabbit [186]. After evaluating several models, we achieved the best performance with cost-sensitive one-against-all (CSOAA) regression models, with human annotations converted to costs in the (0,1) range to be predicted by the regressions. We use hashed unigram, bigram, and skip bigram text features extracted from the title and body text of each journal. CSOAA is a binary logistic regression model per label with weighting applied to minimize false positives [193]. Performance could likely be improved through the use of a state-of-the-art NLP classification model e.g. [194] or through alternative problem formulations e.g. phases as sequences [49]; the models we present here represent a proof of concept and a reasonable ML baseline against which the keyword classifiers may be compared. We found that classification performance was not particularly sensitive to the choice of ML model among the linear classifiers we evaluated.

In the phase model, we make use of both annotator uncertainty and annotator disagreement to increase the cost of human-assigned phases by 0.2 when 'Unknown' is selected and 0.1 when two annotators disagree on an update. We also include contextual information from the two prior updates on the site, adding features from those updates' text and the number of seconds elapsed since those updates.

For training and validation, we used human-annotated journal updates obtained after the final taxonomic iteration. After training initial models, we utilized uncertainty sampling to identify additional updates for annotation (Table 3.5). To improve the phase model, we identified additional sites by averaging uncertainty metrics across all updates on a site.[3] We also selected sites that generated erroneous tags or erroneous transitions, e.g. an update tagged PT and EOL, or a transition from NED to PT. For the responsibility model, we sampled individual journal updates.[4]

We evaluate the performance of the two classifiers using means from fifty executions of 20-fold cross validation. To avoid leaking specific author information into the validation set, CV folds are generated at the site level, with all annotated updates from any specific site appearing in just the training or the validation set.

[3]We used three uncertainty metrics defined by Li and Guo [195]: entropy, distance to the decision threshold, and maximum separation margin.

[4]We used two uncertainty metrics appropriate for multilabel classification defined by Li and Guo [195]: maximum separation margin and label cardinality inconsistency.

**Keyword classifier**

A keyword-based classifier assigns a class label to an update if it contains one of the words on a keyword list defined for that class. While keyword lists are constructed in many ways, we invert the problem and ask: regardless of the keyword selection method, what is the best performance that can be achieved by an optimally-selected keyword list? We develop keyword classifiers to reflect two definitions of "best performance": maximum precision and maximum representativeness. Following a common requirement that keyword lists have near-perfect precision, we first identify a keyword list for each class that ensures perfect precision and the highest possible recall. While identifying the optimal keyword list is NP-hard, we represent the selection of the keyword list for each class as the maximum $k$-cover problem and approximate the optimal lists through a well-known greedy algorithm[5] [196]. In this formulation, for each class label $c$, we identify a set $W_+$ of words appearing only in updates assigned $c$ and a keyword list containing $k$ words in $W_+$ covering the maximum number of updates annotated with $c$. As keyword lists contain only words in $W_+$, each keyword list ensures 100% precision but unknown recall. For this evaluation, we allow the keyword lists' "words" to contain unigrams or bigrams and remove English stopwords from consideration. We evaluate the generalizability of these keyword lists via 10-fold cross validation.

We build a second set of keyword-based classifiers to represent situations where keyword lists are constructed from the words that are most "representative" of each category and for which perfect precision is not a requirement. We identify words that are most associated with each phase and responsibility using frequency-based odds ratio—a measure used in prior OHC work [49]. If $f_c(w)$ is the number of updates assigned class label $c$ that contain word $w$, then $OR(w,c) = (f_c(w) \times f_{\bar{c}}(\bar{w}))/(f_c(\bar{w}) \times f_{\bar{c}}(w))$. For each class label, the keyword list contains the $k$ non-stopword unigrams with the highest odds ratio that appear in at least 10% of updates assigned that class label. These lists contain the words that are most representative of the category relative to the other categories and may better reflect the possible output of an expert-driven keyword identification process. We evaluate the generalizability of these keyword lists by the mean performance over fifty executions of 10-fold cross validation.

[5]The maximum $k$-cover problem's greedy algorithm has an approximation ratio of $1 - 1/e \approx 0.632$ assuming $P \neq NP$, a claim on which this dissertation takes no position. Thus, the identified keyword lists achieve a recall that is at worst 63% of the optimal recall.

### 3.5.2 Classification Results

#### Baseline model results

To contextualize the subsequently presented ML model results, we report two baselines recommended for multilabel classification problems by Metz et al. [197]: (1) *Subset-Accuracy* ($B_{SA}$)—a baseline that predicts the most common multi-label in the dataset, meaning $\{T\}$ for phases and $\{SM,CP\}$ for responsibilities; and (2) *F-Measure* ($B_{FM}$)—a baseline that predicts the set of labels that maximizes F1 score. Results are shown in the final rows of Table 3.6. To reflect an interest in correctly identifying the most-common classes, all mean results in this chapter are computed as weighted macro averages such that per-class performance is weighted by the prevalence of that class.

#### Machine learning model results

Table 3.6 presents the performance of the phase and responsibility ML classifiers. Both models significantly outperform the baselines. Performance is better for the phases than the responsibilities, reflecting the challenges described during operationalization.

We analyzed patterns in the predictions generated by the models. For phases, 7,181 updates (4.7%) are given invalid phase assignments i.e. a combination of labels representing a transition not shown in Figure 3.1. We find a relationship between these erroneous outputs and two primary factors: 69.2% of the invalidly-labeled updates are either less than 500 characters or the first journal on a site. Discounting invalidly-labeled updates, 3.2% of sequential updates are labeled with invalid transitions.

For the responsibility model, we compared the number of responsibilities predicted present in the update to the number of responsibilities in the ground truth for that update. While humans annotated no updates containing all six responsibilities, 4.2% of updates are predicted to contain all six responsibilities. These likely-erroneous predictions are primarily assigned to short updates: 90.4% of updates predicted to contain all six responsibilities are shorter than 500 characters. The model predicts that a higher proportion of updates (+4 percentage points on average) contain each responsibility than the proportions identified by human annotators (Table 3.4). 7.7% of updates are predicted to contain no responsibilities, a decrease of 1.59 percentage points compared to the human-annotated updates.

| Resp. | P | R | F1 | Phase | P | R | F1 |
|---|---|---|---|---|---|---|---|
| CS | 0.75 | 0.83 | 0.80 | PT | 0.91 | 0.95 | 0.93 |
| SM | 0.93 | 0.98 | 0.95 | T | 0.96 | 0.99 | 0.97 |
| CP | 0.90 | 0.97 | 0.93 | EOL | 0.55 | 0.96 | 0.70 |
| FM | 0.47 | 0.92 | 0.58 | NED | 0.86 | 0.86 | 0.86 |
| GB | 0.19 | 0.87 | 0.68 | — | — | — | — |
| BC | 0.32 | 0.41 | 0.34 | — | — | — | — |
| **Mean** | 0.89 | 0.96 | 0.92 | **Mean** | 0.94 | 0.97 | 0.95 |
| $\mathrm{B_{SA}}$ | 0.70 | 0.86 | 0.77 | $\mathrm{B_{SA}}$ | 0.74 | 0.86 | 0.79 |
| $\mathrm{B_{FM}}$ | 0.72 | 0.99 | 0.80 | $\mathrm{B_{FM}}$ | 0.74 | 0.99 | 0.81 |

**Table 3.6: ML classifier performance in terms of Precision (P), Recall (R), and F1 score, along with two baseline measures.**

### Keyword classifier results

Table 3.7 shows the performance of the max-precision keyword classifier for two values of $k$. Even when $k = 10$, the selected keyword lists overfit to the human-annotated data and perform worse than the ML models on held-out sites, demonstrating that these keyword lists fail to capture the salient information of each of the classes under consideration. Table 3.8 shows the performance of the maximally-representative keyword classifier. Note that when $k = 100$ recall is near-perfect in every category, which triggers a corresponding drop in precision and thus F1 score. Performance is significantly better than the max-precision keyword lists, at the cost of low precision. Generalization performance is higher relative to the max-precision keyword lists. Qualitative investigation of the keyword lists generated using both approaches reveals sensible selections. The keywords for the second classifier in particular seem appropriately representative.

Taken together, these results provide evidence that predictive models based on operationalizations from qualitative themes perform better using machine-learning-based approaches rather than keyword lists. However, when generalization to unseen data is not a concern, high-precision lists containing relatively few words can be constructed to achieve high recall, although inconsistent performance across categories may be challenging to identify. For exploratory modeling where precision is less important, small numbers of representative words (as may be revealed during the qualitative operationalization process) can achieve reasonable results and motivate additional data exploration. The use of keyword-based methods may also be seen as a trade-off between interpretability and robustness; the specifics of the modeling application and the need to communicate the prediction process—e.g. to designers—might motivate a preference for keywords over machine learning models. Keywords may also be appropriate when stronger assumptions about the text in a particular domain can be made [198].

| | $k$=10 | | | | $k$=100 | | | |
|---|---|---|---|---|---|---|---|---|
| Class | Train | | Test | | Train | | Test | |
| Label | R | F1 | R | F1 | R | F1 | R | F1 |
| CS | .19 | .32 | .04 | .08 | .87 | .93 | .14 | .16 |
| SM | .34 | .50 | .30 | .46 | .90 | .95 | .73 | .81 |
| CP | .22 | .36 | .20 | .32 | .79 | .88 | .58 | .68 |
| FM | .47 | .64 | .07 | .11 | .95 | .97 | .09 | .09 |
| GB | .39 | .56 | .00 | .00 | .99 | .99 | .05 | .08 |
| BC | .30 | .46 | .02 | .02 | .99 | .99 | .03 | .03 |
| PT | .08 | .15 | .01 | .02 | .45 | .62 | .03 | .04 |
| T | .13 | .23 | .05 | .09 | .49 | .66 | .31 | .46 |
| EOL | .39 | .56 | .21 | .31 | .99 | .99 | .26 | .31 |
| NED | .11 | .20 | .00 | .01 | .52 | .69 | .03 | .04 |

**Table 3.7: Max-precision keyword list performance in terms of Recall (R) and F1 score. Precision is 1 on train.**

## 3.6 Model Analysis

### 3.6.1 Model Validation

To explore the expert validity of the phase and responsibility models, we invited an expert involved in the creation of the qualitative frameworks used in this chapter (an author of the CJF without any affiliation or conflict of interest with this work) to provide feedback on our operationalization. Across the elements of each taxonomy codebook, the expert rated the reasonableness of each definition on a 5-point Likert scale from strongly disagree (-2) to strongly agree (+2). Overall agreement was high for items in both the responsibility (M=1.74) and phase (M=1.83) codebooks. In the expert's qualitative feedback, several comments related to a divergence between what is observed in patient interviews and what patients self-report on CaringBridge, a gap that we leave for future work to better understand the motivations of patient

| Class Label | Train P ($k$=10) | Train R ($k$=10) | Train F1 ($k$=10) | Test P ($k$=10) | Test R ($k$=10) | Test F1 ($k$=10) | Train F1 ($k$=100) | Test F1 ($k$=100) |
|---|---|---|---|---|---|---|---|---|
| CS | .24 | .88 | .37 | .23 | .86 | .36 | .26 | .26 |
| SM | .86 | .98 | .92 | .86 | .98 | .92 | .93 | .93 |
| CP | .77 | .99 | .87 | .77 | .99 | .87 | .87 | .87 |
| FM | .22 | .87 | .35 | .20 | .77 | .30 | .06 | .07 |
| GB | .16 | .65 | .25 | .12 | .50 | .19 | .08 | .08 |
| BC | .14 | .69 | .23 | .08 | .42 | .13 | .08 | .08 |
| PT | .12 | .72 | .21 | .12 | .71 | .20 | .13 | .14 |
| T | .88 | .92 | .89 | .88 | .90 | .88 | .92 | .92 |
| EOL | .10 | .73 | .18 | .11 | .72 | .18 | .03 | .03 |
| NED | .06 | .97 | .12 | .07 | .97 | .13 | .11 | .12 |

**Table 3.8: Maximally-representative keyword list performance in terms of Precision (P), Recall (R) and F1 score.**

sharing in OHCs.

How do we account for poor inter-annotator agreement for responsibilities despite high agreement for phases? The same annotators were involved in both models, which suggests that coder quality is not a primary cause. While annotator domain expertise may be a factor, phase indicators generally require more medical knowledge to identify than responsibility indicators. The assessment of the expert that the operationalizations are reasonable suggests there is no fundamental weakness in the iterative operationalization process used or the resulting taxonomy. Instead, we hypothesize that ambiguity in the identification and mapping of indicators to responsibilities is a critical factor. To probe the role of ambiguity in producing low IRR for the responsibilities, we conducted a qualitative analysis of the primary annotators' comments during the Schaekermann et al.-motivated discussion of disagreements [190].

Looking at the annotator justification in cases of irresolvable disagreement reveals two preliminary themes: (1) disagreement about the directness of supporting evidence needed to assign a responsibility and (2) disagreement about which responsibility a piece of evidence indicates. These themes align with two significant dimensions of ambiguity identified by Chen et al. [199]: (a) data ambiguity, meaning multiple reasonable interpretations, often due to missing or unclear context, and (b) human subjectivity, meaning distinct interpretations resulting from "different levels of understanding or sets of experiences" among annotators. Chen et al. further utilize disagreement between coders as a proxy for ambiguity, with the lower IRR scores relative to the phases indicating a higher degree of ambiguity. The irresolvable cases suggest that data ambiguity is excacerbated by soft boundaries between responsibilities in the codebook, but the supportive external validation of the current codebook and consistently low IRR after five codebook iterations suggest an inherent ambiguity to the classification task. To reduce ambiguity, we view the next reasonable step as conducting additional qualitative work to elucidate the CJF in CB updates specifically (as opposed to additional qualitative inquiry outside the OHC context).

To further investigate the validity of the responsibility model specifically, we tested the expectation that an author mentioning a responsibility in an update is more likely to mention that responsibility in other updates authored in the same week, as most responsibilities in the CJF are more than momentary [39]. For each responsibility $r$, we fit a Poisson regression to predict the number of updates on site $s$ in week $w$ that contain $r$ based on whether a randomly selected journal from $s, w$ contains $r$. We consider only weeks with at least 2 updates and use the total number of updates authored that week on $s$ as the exposure, additionally controlling for the baseline rate of updates on $s$ predicted to contain $r$. Incidence rate ratios are shown in Table 3.9. When an update on a site is predicted to contain a responsibility $r$, other site updates in that week are predicted to contain $r$ at a rate 1.64 times greater than if the update is predicted not to contain $r$. This confirms the hypothesized co-occurrence of responsibilities and provides additional evidence that the responsibility predictions are valid.

| | Contains $r$? | Baseline rate of $r$ | $G^2$ (df=30287) |
|---|---|---|---|
| CS | $1.48 \pm 0.06$ | $1.031 \pm 0.001$ | 22122.01 |
| SM | $1.21 \pm 0.03$ | $1.011 \pm 0.001$ | 4460.91 |
| CP | $1.26 \pm 0.03$ | $1.011 \pm 0.001$ | 7171.64 |
| FM | $2.16 \pm 0.50$ | $1.053 \pm 0.006$ | 11078.42 |
| GB | $1.85 \pm 0.22$ | $1.043 \pm 0.003$ | 15279.27 |
| BC | $1.88 \pm 0.24$ | $1.047 \pm 0.003$ | 14357.79 |
| Mean | 1.64 | 1.033 | — |

**Table 3.9: Within-week responsibility co-occurrence Poisson regression models. Incidence rate ratios with 95% confidence bounds and deviance ($G^2$) are given, demonstrating a greater proportion of site updates contain a responsibility $r$ if another update published in the same week contains $r$. All model coefficients are significant at $p < 0.001$.**

### 3.6.2 Model Integration

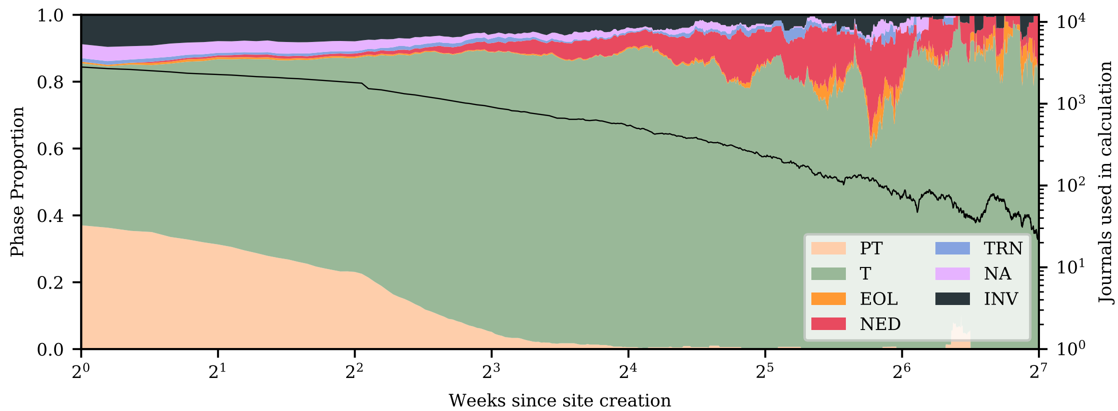



**Figure 3.2: Predicted phases over time on a log scale. Updates across sites are binned to the day and proportions are computed based on a rolling average of 30 days [3]. The right axis and plotted line indicate the number of updates used to compute the proportions; note fewer than 100 updates are available after 40 weeks. Proportions include updates assigned a single phase in addition to updates with multiple phases (TRN), no assigned phases (NA), or an invalid combination of multiple phases (INV).**

By classifying both phases and responsibilities for unannotated updates, we can explore temporal trends and integrate predictions to explore the relationship *between* phases and responsibilities. To mitigate noise introduced by the 4.7% of invalid predicted phases, we reassign the phase prediction of updates surrounded by single-phase updates to match the phase of its neighbors. After reassignment, 2.6% of adjacent updates predict a transition considered invalid in our phase model. Using these reassigned phase predictions, we consider responsibility predictions that co-occur with valid phase predictions to establish baseline responsibility occurrence proportions and the per-phase deviations from that baseline.

#### Phase model predictions over time

Figure 3.2 traces proportions of the phases over time. Few sites have updates in the PT phase past the first 2 months. NED updates are more frequent over time, with a temporal variance that reflects our qualitative observation that NED updates are frequently written on consistent anniversaries (e.g. a spike around one year after initial diagnosis). Few patients continue posting updates in the EOL phase. The vast majority of updates on CB are written during the treatment phase.

**Integrated model predictions**

Figure 3.3 shows the frequency of each responsibility relative to its occurrence in other phases. In contrast to the CJF's categorization of responsibilities into phases (Table 3.1), we find Coordinating Support, Sharing Medical Info, and Compliance appear *less* in treatment updates than in other phases, and Giving Back and Health Behavior Changes appear less in NED updates than in other phases.

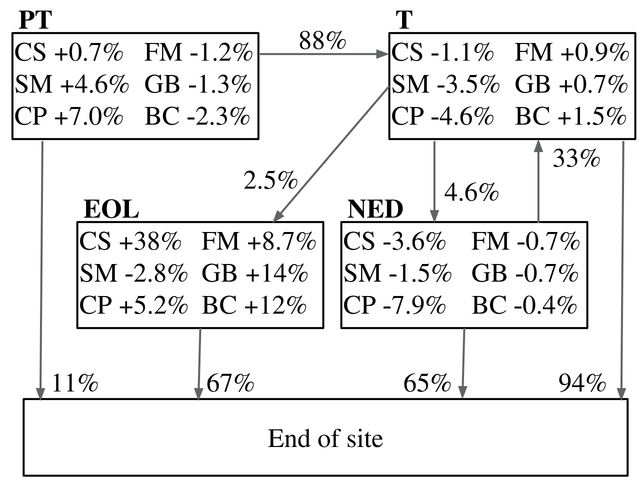



**Figure 3.3: Predicted site phases and responsibilities. Responsibilities are the percentage point change in proportion relative to updates made in other phases. Phase transitions are labeled with the percentage of updates that follow that phase; invalid transitions e.g. NED→PT are not shown but are included in per-phase transition totals.**

## 3.7 Discussion

Bridging qualitative frameworks describing cancer patients to a user model of OHC behavior is an important step towards designing personalized digital services for cancer patients. Practically, we intend to use these models in the design of recommender systems to connect patients based on commonalities in cancer phase and expressed responsibilities, in support of an informed social network with knowledge about the cancer experience [200].

We experienced challenges operationalizing taxonomies from the qualitative frameworks we selected, finding phases easier to operationalize than responsibilities. To generalize this method beyond health-related qualitative frameworks, further study is needed; to facilitate application in other contexts, we discuss three aspects of the qualitative frameworks we selected that made operationalization challenging.

The first aspect is the type of mapping between observable data and conceptual theme. Taxonomic categories will be easier to define if the corresponding indicators in the data form a one-to-one map with the qualitative themes. As the number of indicators that refer to a single theme grows—the Preparation responsibility had many possible referents—the category will be increasingly hard to define and identify reliably. In contrast, the T phase had a limited set of medical indicators that could be reliably identified during operationalization. The second aspect is the temporal scale of behavioral themes. If themes describe behaviors that span lengths of time shorter than the update frequency of the available social media data, a windowed trace of user behavior makes reliable retrieval difficult. Cancer phase changes slowly and could be tracked across multiple updates, but frequent responsibilities were often alluded to without necessary context. The third aspect is the degree to which the qualitative themes are mutually exclusive. Despite periods of transition, cancer phases are largely singular and conflicting indicators within a single update rarely co-occur; responsibilities have no natural exclusivity and a single update may contain many indicators each mapping to many responsibilities.

A risk intrinsic to the bridging process we describe in this work is a perpetuation of the underlying qualitative framework's implicit lens. Our models reproduce the subjectivities of the CJF's source interviews even while mapping to a broader context of social media users. Thus, we risk magnifying or distorting aspects of the patient experience. We suggest that bridging can serve as a compliment to other methods, enabling researchers to triangulate their understandings through the inclusion of user behavior models informed by qualitative themes.

## 3.8 Conclusion

In this chapter, we explored a process for bridging qualitative themes to social media user models. We built two models using taxonomic categories operationalized from two qualitative frameworks to classify unstructured text data and trace behavior over time in an OHC. We identified two primary challenges in the operationalization process along with strategies for managing them. We found that supervised ML outperforms common keyword-based approaches in classification performance. Future work includes developing more sophisticated methods for resolving challenges and understanding ambiguity in the operationalization process in order to broaden the potential scope of qualitative themes and social media contexts in which bridging can be applied.

The models provide insights into CaringBridge users, including the phase in which cancer patients join and leave the community and the responsibilities discussed throughout the course of a cancer journey. The modeling approach is applicable to other health conditions and qualitative frameworks, and represents a reasonable to approach to modeling the longitudinal behavior of OHC users. Models of OHC behavior can be used to improve the delivery of personalized digital health services—including peer recommendations [153]. In the next chapter, I will build on this understanding of the core usage of CaringBridge as a blogging platform to focus on non-journaling behavior by authors: specifically, on peer interaction.

# Chapter 4

# Patterns of patient and caregiver mutual support connections

Chapter 3 described the use of an online health community to communicate with existing supporters, but a key motivation for using online health communities is connecting with *peer* supporters who have similar experiences. However, finding and connecting with peers is challenging and a user's role in a community will influence the formation of peer connections. In this chapter, we study patterns of peer connections between two structural health roles: patient and non-professional caregiver. We examine user behavior in an online health community where finding peers is not explicitly supported. This context lets us use social network analysis methods to explore the growth of such connections in the wild and identify users' peer communication preferences. We investigated how connections between peers were initiated, finding that initiations are more likely between two authors who have the same role and who are close within the broader communication network. Relationships are also more likely to form and be more interactive when authors have the same role. Our results have implications for the design of systems supporting peer communication, e.g. the peer-to-peer recommendation system we build in Chapter 5. The contents of this chapter were previously published at the CSCW conference [10].

## 4.1 Introduction

Online health communities (OHCs) give patients and caregivers the opportunity to mutually support one another. To realize this opportunity, OHCs must be designed to facilitate supportive communication between peers. Social support occurs within the context of an individual's relationships with others [201, 72], so designing for new peer connections requires an understanding of the individual factors associated with the formation and growth of mutually-supportive relationships. In this study, we explore these factors by analyzing an interaction network of peer connections in an OHC that lacks an elaborate technical infrastructure for finding and forming those peer connections.

CaringBridge[1] is an OHC designed to enable patients and caregivers to communicate with their friends and family members about a health event such as an illness or injury [8]. Patients and caregivers on CaringBridge author text posts for their extended support networks on individual blogs called "sites". Interactions between authors—via comments on other authors' sites—constitute peer connections. However, CaringBridge currently does not explicitly support authors in finding other authors to form connections with, providing only limited search functionality and no feed or recommendation system. Despite this lack of affordance, there is substantial inter-author peer interaction on CaringBridge. Thus, studying CaringBridge represents a unique opportunity to observe users' preferences for peer connection when conventional social discovery mechanisms are absent: what connections are sufficiently desirable and accessible to users that they form without explicit design affordances?

We study this appropriative use of CaringBridge to learn about patient and caregiver preferences for peer connections and to understand the factors that lead peer authors to form connections "in the wild". Identifying these factors opens pathways to designing digital interventions that are faithful to user preferences and that provision support more effectively [202]. One key factor is the role adopted by an OHC user. While there are many roles in OHCs, our study focuses on two common health roles: patient and caregiver. Patients and caregivers have different motivations and needs, and the differences between patient and caregiver use of OHCs is understudied [203]. Facilitating peer connections in OHCs is an area of active research [64, 144, 114], and understanding the impact of health role on peer connection formation and growth creates

[1] `https://www.caringbridge.org/`

opportunities to facilitate these connections in a way that is responsive to patients' and caregivers' divergent needs.

We characterize connections between peer authors on CaringBridge with quantitative social network analysis. To understand connection patterns, we analyze relationships between authors as they occur within the network formed by authors' interactions. The network context of an interaction can be as important as the content of that interaction: for example, research predicting cancer stage from users' OHC posts has found that network features were as predictive of cancer stage as the text of the posts themselves [204, 48, 180]. Further, analyzing connections within an interaction network brings focus to the initiators of interactions and not just the support recipients. In addition to forming the basis for supportive relationships, new connections provide benefits to both initiators and receivers [205, 206, 90].

To quantify the impact of authors' health role on connection formation, we use machine learning to identify patient and caregiver authors from their posts and identify differences in interaction behavior between patients and caregivers. In addition to health role, we explore factors related to authors' level of activity, position within the interaction network, and health condition. Our analysis explores these factors to address two broad research questions:

**RQ1 (Initiations):** What factors are associated with the initiation of a new connection by an author?

**RQ2 (Relationships):** What factors are associated with the reciprocation and growth of connections between peer authors?

To address RQ1 and identify factors associated with initiations, we first identify which authors engage in peer connection behavior. Second, we identify when authors' make their first peer connection relative to when they joined the site, as that first initiation is an explicit signal of peer finding behavior. Finally, we identify factors that make an author more likely to be the target of initiations. Thus, we state three sub-questions for RQ1:

- **RQ1a:** Which authors initiate peer connections?
- **RQ1b:** When do authors initiate peer connections?
- **RQ1c:** With whom do authors initiate peer connections?

To address RQ2 and identify factors associated with reciprocation and the growth (or not) of dyadic relationships, we first identify which initiations are likely to be reciprocated. Secondly, we examine which reciprocated dyads are likely to be more interactive and more balanced. Thus, we state two sub-questions for RQ2:

- **RQ2a:** Which authors reciprocate? To which initiators?
- **RQ2b:** What factors lead to more interactive relationships?

The contributions of this chapter are the identification of factors associated with the formation and growth of peer connections and a comparison of connection behavior between two important health roles: patient and caregiver. Specifically, we find that (1) patients are more likely than caregivers to initiate with other authors after receiving interactions, (2) patients who initiate do so earlier than caregivers, (3) authors are more likely to interact with others sharing the same author role, and (4) authors are more likely to reciprocate and form more interactive relationships with others sharing the same author role. These differences in OHC use by author role have implications for the design of online health communities and other digital interventions that benefit both patients and caregivers [114]. We discuss opportunities to integrate these results in peer recommendation systems to foster mutually supportive relationships.

## 4.2 Related Work

In this chapter, we aim to understand the communication preferences of health peers by studying interactions between CaringBridge authors. To understand those preferences, we first discuss motivations for use of digital communications, specifically use of OHCs for peer support. Second, we discuss research on health roles, laying out the conceptual groundwork for a focus on patients and caregivers. Finally, we discuss connection dynamics on OHCs, with a focus on factors previously identified as important for the formation of new connections and their impact on the formation of supportive peer relationships.

### 4.2.1 Motivations for digital communication during health journeys

Patients use the internet to find information and support [54, 55]. For pursuing social connection specifically, patients use the internet to overcome isolation, identify others with similar experiences, reinforce existing relationships, and offset deficits in existing relationships [13]. CaringBridge is designed primarily for reinforcing existing relationships [8]. However, the existence of connections between authors indicates that authors also are using CaringBridge to address unmet needs [56] and build supportive connections based on shared concerns [57]. These support-seeking behaviors result in the formation of *peer* connections, which we discuss next.

Connecting with experientially-similar others is a key motivation for patients to participate in online support communities [58, 59, 60]. Experientially-similar others serve as important sources of support to both patients [59] and caregivers [61, 62]. For CaringBridge authors, experiential similarity is entangled with the notion of authors as peers. We adopt the four-point definition of peer suggested by Simoni et al.: (a) sharing personal circumstances i.e. some form of health challenge, (b) obtaining benefits from peer support that derive from their status as peers, (c) lacking professional training or medical credentials, and (d) "intentionally setting out to interact with individuals they may or may not encounter in their everyday life" [63]. Finding peers to communicate with online is complicated by many individual factors [64], which we model quantitatively in this study. Peer-finding behaviors have been implicated as an opportunity for technological innovation in the context of OHCs [114, 42] and more broadly [128, 207]. We study connections between peers—rather than other health relationships such as mentor/mentee and medical professional/patient—as an opportunity for exchanging support with experientially-similar others.

As CaringBridge most resembles individual health blogs, the motivations of CaringBridge authors may differ from users of other kinds of OHCs. Blogging is fundamentally social [67]. Health blogging fulfills both communicative and therapeutic roles [68], with patients sharing their illness trajectories and processing their experiences through writing [69]. While blogging may provide benefits to patients due to the expressive self-disclosure involved in writing blog posts [70, 71], having responsive and interactive readers provides additional benefits [67]. McCosker and Darcy argue that connectivity between blog authors has the potential to sustain health bloggers in their writings about their health journeys [68]. Our study focuses on communication between "blog" authors as a potent opportunity for understanding the dynamics that produce the benefits of these interactive connections.

### 4.2.2 Patient & caregiver: Structural health roles

Users take varied roles in OHCs [113, 114]. Research examining roles in OHCs has tended to focus on group [113, 115] or social roles [114, 116] that are defined by behaviors. For example, Sharma et al. define a "seeker" role in a mental health forum as a person who makes a new thread [116]. In contrast, we examine structural roles that arise from a health event that creates a *patient* and any number of non-professional auxiliary *caregivers*; these structural roles are adopted by patients and caregivers and are defined by accompanying expectations and responsibilities [117, 118]. The expectations of each role are associated with (but not defined by) a set of behaviors that are "characteristic of the person in a particular setting" [119]. Thus, patients and caregivers have different behaviors as they enact their role in an OHC. Patient and caregiver responsibilities and behaviors may change frequently [9], but their role is relatively stable. This stability contrasts with the frequently-shifting behavior roles identified by Yang et al. in an online cancer forum [114]. Note that structural roles are not explicitly afforded via the technical interface, in contrast to e.g. moderator roles on Wikipedia and other explicit roles that have been studied online [120, 121]. People with the same structural role may be more likely to interact with each other; Xu et al. found that online communication was more likely to occur between Twitter users who had the same health role, such as "provider" or "engaged consumer" [122]. We explore the interaction dynamics between patients and caregivers in detail on CaringBridge.

Patients and caregivers communicate differently online, although these differences have received little explicit focus. Lu et al.—a notable exception—identified differences in topic and sentiment in posts written by patients and caregivers in an online health forum [123]. In this work, however, we focus on supportive connections and the *target* of online interactions by health role. OHCs may provide a particular opportunity for caregivers seeking support online [124], as patients are given "interpretive precedence" in dealing with a health condition, leaving caregivers without supportive relationships to understand their own role in a broader health journey [125]. When caregivers can communicate with other caregivers digitally, they develop more effective coping strategies for caregiving stress [126]. Offline, caregiver connections with other caregivers may be more passive than active; Gage suggests an important role of serendipity in the formation of new connections [61]. We find that serendipity plays a role in at least some of the connections on CaringBridge. We aim to understand the differences in patient and caregiver communication on OHCs in response to a call for developing a deeper empirical understanding of OHC participation [13], a context in which caregivers are understudied [127].

### 4.2.3 OHCs as sources of interaction and support

OHCs are associated with a variety of positive and negative health outcomes for their users, specifically due to the interaction that occurs on them [13]. Understanding this interaction in more detail—particularly with respect to author role—creates opportunities to improve the provisioning of support on OHCs. For patients, use of OHCs is associated with greater perceived support [91, 58], writing higher sentiment posts after interacting with others [92], engagement measures such as duration of stay in a community [93, 70, 94], perceptions of control over illness [95], and decreased mortality [74, 76]. Caregivers also benefit from use of OHCs in terms of reduced stress, although evidence linking specific behaviors to outcomes for caregivers is more mixed than for patients [96, 97]. Despite many benefits, making connections in OHCs can also have negative impacts on users, increasing their stress and leading them to leave the community [13]. For patients, directly making comparisons with other patients can be distressing [98], as can the sudden drop-out of key community members [99]. Furthermore, who an OHC user interacts with and the type of their interactions with others mediate both length of stay in the community and the benefits derived from using it [93]. These mixed and contextual outcomes make designing for the formation of new connections risky. We address this risk by examining the specific connections made between OHC users in order to identify user communication preferences.

We build on foundational OHC research from Bambina examining a health forum's network structure and its impact on the transmission of social support [14]. Bambina analyzed a static snapshot of the posts in a cancer forum with 84 active participants, finding a core of highly supportive participants with a long tail of periphery members in the social network. Our dataset enables research that addresses two key limitations of Bambina's work: (a) examining connections as a dynamic process rather than as fixed in a static network snapshot, and (b) including non-interacting "lurkers", which in our research is the population of CaringBridge authors who never interact with a fellow author. With a complete and dynamic view of the interaction network, we are able to focus on the *initiation* of new peer connections by authors.

Multiple factors are associated with new connection formation in OHCs. New connection formation is often motivated by shared social identity [108] and experience [103]. Meng examined new connection formation in a weight management social networking site, finding substantial homophily effects related to health condition [109]. Centola and van de Rijt find similar strong homophily effects, noting that platform-specific traits—such as exercise preference in a fitness community—were less important in new health contact identification than demographic traits [110]. In general, health-relevant traits contribute to the creation of specific connections [111]. Outside of health, Seering et al. consider the immediate context (e.g. recent messages) that leads to first participation on Twitch.tv [112]; their examination of the factors that affect initiation with others broadly inspires our approach, although we focus less on immediate context and more on socio-cultural factors such as author role.

By studying initiations between users of an OHC without technical infrastructure for finding peers, we learn about the factors associated with the formation of supportive peer connections.

## 4.3 Dataset & CaringBridge Research Collaborative

This work was conducted during a research collaboration between CaringBridge (CB) and the University of Minnesota. CB is a global, nonprofit social network dedicated to helping family and friends communicate with and support loved ones during a health journey.[2]

CaringBridge.org offers individual *sites* for users—free, personal, protected websites for patients and caregivers to share health updates and gather their community's support. Each site prominently features a *journal*, which is a collection of timestamped, textual health *updates* by or about a patient. We use this terminology in concordance with previous CB research [70, 9]. *Authors* are CB users who have posted one or more updates on one or more sites. One author may publish updates on multiple sites, and multiple authors may publish updates on the same site. Other registered users can comment on authors' sites but do not have sites of their own and are omitted from analysis, as we study peer interaction specifically and non-author users are likely not peers: they are primarily the friends, family, and acquaintances of site authors [8].

To motivate our focus on CB specifically, we briefly discuss the affordances CB offers for connection with peer authors. Rains argues that four primary affordances of communication technologies are most relevant for health-related social connection: visibility, availability, control, and reach [13]. The design of CB offers reach—"potential to contact specific individuals, groups, or communities"—in only a limited way. The search function on CB retrieves only sites with matching titles, which in most cases means a patient's full name is required to find a site.

[2] Some of the text in this section is identical to that appearing in other works that result from this collaboration [70, 9, 8].

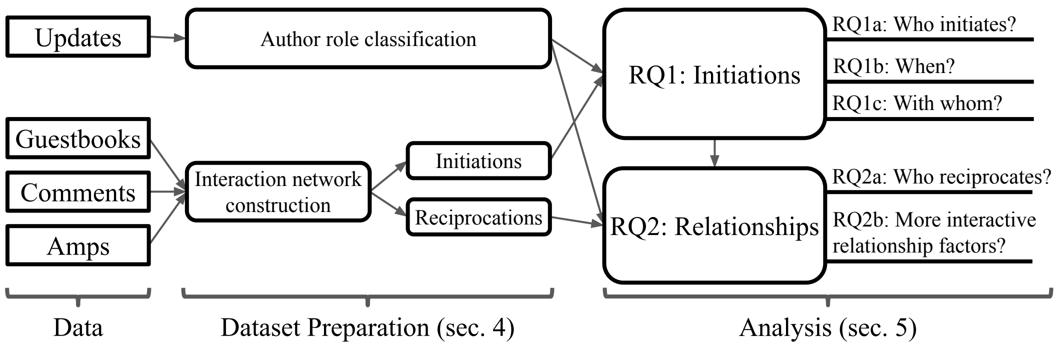



**Figure 4.1: Research dependency map. Background data preparation methods are necessary to address RQ1, which is built upon by RQ2.**

This barrier to social discovery means that achieving reach and the formation of a broader community requires additional effort. Thus, we study peer connections in an environment where "finding and interacting with peers facing the same health issue" is not specifically supported [13]. CB authors' appropriative use of CB for peer connection provides an opportunity to understand the ways that peers connect with each other during their health journeys without the mediation of an explicit social discovery mechanism.

The CB dataset used in this work includes de-identified information from 535,481 authors and 588,210 sites created between June 1, 2005 and June 3, 2016, collectively containing 19M journal updates. The data were acquired through collaboration with CB leadership in accordance with CB's Privacy Policy & Terms of Use Agreement. This study was reviewed and deemed exempt from further IRB review by the University of Minnesota Institutional Review Board. We opt not to publicly release the dataset used for analysis in this chapter in order to protect participants' privacy [185]. We welcome further inquiries about the dataset by contacting the authors.

## 4.4 Methods: Dataset Preparation

To study peer connections and address our research questions, we first classified the role of individual CB authors using machine learning (sec. 4.4.2) and constructed the network of interactions from the log data (sec. 4.4.3). Figure 4.1 provides a high-level overview of the dependencies between these data preparation methods (sec. 4.4) and the analysis methods that address the research questions (sec. 4.5). In constructing the network, we identified both initiations between authors and reciprocated dyads, which we used to study initiations (RQ1) and relationships (RQ2).

We release our code on GitHub.[3] Analysis code makes primary use of Python's scikit-learn [208], NetworkX [209], StatsModels [210], NumPy [211], Pandas [212], and Matplotlib [213] packages and R's mlogit [214] and stargazer [215] packages.

### 4.4.1 Terminology

For ease of reference, we list here the key terms we use in this chapter:

- **Valid authors** are CB authors who meet basic requirements for being included in the study, such as publishing at least two journal updates over at least a 24-hour period. Authors are the subset of registered CB users who have published at least one journal update on a CB site. (See section 4.4.2)
- **Author role** is the perspective from which an author account writes and publishes updates—either *patient*, *caregiver*, or *mixed*. (See section 4.4.2)
- **Interactions** are one of three types of directed engagement (i.e. guestbooks, amps, and comments, introduced in Section 4.4.3) between an initiating author and a receiving author. Figure 4.3 shows the interaction UI.
- **Initiations** are the subset of interactions that compose the first interaction between an initiating author and a receiving author. A *first initiation* is the first interaction an author makes on CB to any receiving author. An *initiating author* or initiator has made at least one initiation. A *non-initiating author* is an author who has made no initiations. A *non-receiving author* is an author who has not been the target of any interaction.
- **Reciprocations**, or reciprocal initiations, are the subset of initiations that reciprocate an earlier initiation from another author. In a dyad, the reciprocation is the initiation that comes second and closes the loop.
- A **connection** exists between two different authors if at least one interaction has occurred between them.

[3] `https://github.com/levon003/cscw-caringbridge-interaction-network`

- **Relationships** are dyads with reciprocated initiations and the associated history of interactions between the two authors. The minimum number of interactions in a relationship is two: the original initiation and the reciprocal initiation.

- **Initiations period**—January 1st, 2014 to June 1st, 2016: the 2.5 year period of interactions used for RQ1. (See section 4.4.3.) RQ2 uses data from the full data collection period: January 1st, 2005 to June 1st, 2016.

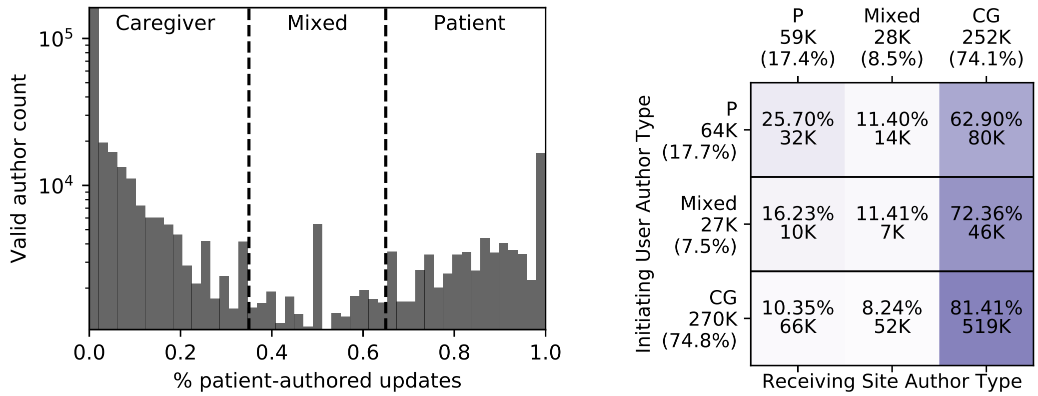
### 4.4.2 Authorship on CB

In order to study inter-author interactions, we first select a sample of valid authors and then develop a machine learning classifier to identify author role (patient or caregiver) from their text updates.

#### Valid author selection

Our dataset contains 535,481 total author users. In this chapter, we analyze 362,345 *valid* authors with at least 2 updates published more than 24 hours apart. We exclude authors who have written any posts on sites determined to be spam by CB internal tooling ($-10{,}776$ authors), who we manually identified as either spammers ($-17$) or CB-internal test accounts ($-9$), or who only published updates within a 24-hour period ($-162{,}334$). We selected the 24 hour threshold for author exclusion based on the distribution of author tenure (see Appendix A.1).

As one author may publish updates on multiple sites and one site may have updates published by multiple authors, identifying which authors interact with which other authors is challenging. We use the term *valid sites* to refer to the 340,414 sites (57.9% of total) on which a valid author has written at least one update. We identified 18,691 *multi-site authors* (5.2%) publishing on 2+ sites and 79,115 *mixed-site authors* (21.8%) publishing on at least one site on which other valid authors have (co-)authored updates. This estimate broadly aligns with estimates of multi-authorship from early studies of group blogs [216]. As we discuss in the next section, the percentage of mixed-site authors is likely a conservative lower bound since author account sharing is common on CB.




**(a) Distribution of the proportion of each author's updates that are classified as patient-authored. Vertical lines indicate the thresholds used to assign a role to each author.**

**(b) Initiation counts broken down by the author role assigned to the initiating author and the receiving site.**

**Figure 4.2: Classification of author role based on the text of an author's updates (a) enables a comparison of author initiations by assigned role (b).**

### Author role classification

Authors on CB take the role of either *patients* or *caregivers*.[4] However, authors can take on multiple roles or switch roles in three cases: (1) multi-site authors may take a caregiver role on one site and a patient role on a second, (2) an author may use a single site for recounting two health journeys from the perspective of both a patient and a caregiver, or (3) account sharing may result in updates published from the same author account but written by both a patient and one or more of that patient's caregivers. To tease apart these factors, we first classify authorship at the journal update level. We then classify an author's role as either Patient, Caregiver, or Mixed based on the classification of the updates published by that author. *Mixed* indicates one of the three observed cases above.

We trained a machine learning classifier to predict the author role of 15,850,052 updates that were authored on valid sites and contained at least 50 characters. We combined human annotations of updates' author role from two previous CB studies [9, 8] with additional annotations created while doing exploratory data analysis and using active sampling on earlier iterations of the classifier. Two of the authors independently annotated 429 updates, resulting in a Cohen's $\kappa$ of 0.829, which indicates sufficient reliability for this study [191]. Combined with

[4] In figures and tables, we use the abbreviations P and CG respectively.

the annotations from previous studies, we had 6,932 human-annotated updates from 305 sites.[5]

We used a linear SVM classifier on TFIDF-transformed unigram and bigram features, a common approach to binary text classification [217]. As the training data are not identically distributed due to sampling differences across the three annotation efforts, we train with balanced classes by randomly downsampling the majority-class updates, an approach we found to outperform other training regimens (e.g. training with all annotated data). We used hold-one-out cross validation to evaluate the performance of the model.[6] Accuracy was 95.08% and micro-averaged F1 score was 0.95. Patient-annotated updates ($n$ = 5,938, precision = 0.99, recall = 0.95, F1 = 0.97) were classified correctly at a greater rate than caregiver-annotated updates ($n$ = 994, precision = 0.77, recall = 0.93, F1 = 0.84).

We categorized individual authors as either Patient, Caregiver, or Mixed. To assign a role to an individual author, we aggregated from the author role predictions of updates published by that author. We used the same approach to categorize sites. Through manual investigation of 30 sites, we identified a variety of usage patterns, including a high frequency of sites with both patient- and caregiver-classified updates. To assess the general patterns in author role and to allow for error introduced by the classifier, we use a consistent set of thresholds to define author role: *Caregiver* sites/authors have less than one third of their updates classified as patient-authored, *Mixed* sites/authors have between one third and two thirds of their updates classified as patient-authored, and *Patient* sites/authors have more than two thirds of their updates classified as patient-authored. Figure 4.2a show the distribution of the proportion of each author's updates that are classified as patient-authored, along with the thresholds. The use of permissive thresholds to assign a role label captures the general perspective from which an author writes and keeps cascading classifier error to a minimum. In the cross-validated ground truth data, 87.87% of sites were classified at least two-thirds correctly. Thus, in the case of sites with all-Patient or all-Caregiver updates, the site-level error rate using these thresholds is at most 12.12%, which we deem acceptable.

[5]Note: 44 ground-truth updates (0.63%) were assigned ambiguous or mixed labels that were reassigned to Caregiver for training, reflecting a predominant interest in patients: the binary classifier is trained to predict if updates are patient-authored or not.

[6]Since a shared site/author could leak information about the held-out data and give an overly optimistic view of classifier performance, we hold out at the site level rather than holding out individual updates.

| | All interactions | From valid authors | Non-self-interactions | Initiations |
|---|---|---|---|---|
| Guestbooks | 82,980,359 | 5,864,304 | 5,212,720 | 654,192 |
| Amps | 63,314,738 | 3,536,819 | 3,037,844 | 148,787 |
| Comments | 31,052,715 | 1,094,435 | 881,781 | 111,623 |
| **Total** | **177,347,812** | **10,495,558** | **9,132,345** | **914,602** |

**Table 4.1: Interaction counts broken down by type. This study considers only interactions from valid authors to valid sites. Self-interactions are interactions on sites where the interacting author has published an update and are excluded when building the network. Initiation counts are the number of non-self author/site pairs that were initiated by each type of interaction.**

Applying the definitions above, we find 74.77% of author accounts are classified as Caregiver, 17.74% as Patient, and 7.49% as Mixed. The distribution is similar for sites. Mixed-author accounts may indicate either a single author embodying multiple roles or multiple people sharing the same account credentials. 96% of Mixed-author accounts are shared by a patient and a caregiver (see Appendix A.2), which complicates analyses treating interactions between accounts as interactions between two people and suggests caution when interpreting Mixed-author results. Overall, using the classifier predictions, we estimate that 22.06% of updates are patient-authored (see Appendix A.3).

### 4.4.3 Author interactions & network structure

#### Inter-author interaction types & analysis period

To study the interactions between authors, we construct a network from the log data in the CB dataset. Direct messaging is not supported on CB; instead, all interactions are by an author on a site. *Guestbooks* are text posts left by a CB user on a site.[7] *Comments* are text posts left by a CB user on a specific journal update on a site. *Amps*[8] are "likes" represented with a small heart icon and left by a CB user on a specific journal update on a site. The interface for these interactions is shown in Figure 4.3. In this chapter, we consider only interactions from valid authors to valid sites, with counts as shown in Table 4.1. Each interaction is associated with a unique identifier for the user and the site, as well as a timestamp. Amps lack timestamp information, so we assume that amps occur at the publication time of the associated journal update. (We analyze

[7]Guestbooks were renamed "Well Wishes" by CB, but we exclusively use the older name.
[8]"Amps" represent the idea of "amplifying" the love, hope, and compassion of the visitor.

this assumption in Appendix A.4.) Other interactions are possible (see Appendix A.7), but we focus on guestbooks, comments, and amps as they are publicly visible, identifiable to a specific author/site pair, and result in a notification for the receiving author(s).

Figure 4.4 shows the number of each interaction type on CB over time. Note the introduction of amps and comments as features on CB. In order to avoid irregularities related to the introduction of new interaction types and to analyze a more established version of the network, all RQ1 analyses will focus on the state of the network from January 1, 2014 to June 1, 2016, which we refer to as the *initiations period*. When models are fit to the initiations data, Jan 2014 – Jan 2016 (80% of the initiations period) is treated as training and inference data, and Jan 2016 – Jun 2016 (the remaining 20%) is treated as the test data and target for prediction. For RQ2 we use data from the full period (January 2005 to June 2016) because reciprocations and relationship interactions are susceptible to being right-censored i.e. interactions within the relationship occur after the end of the data collection period.

### Constructing the author interaction network

We constructed a directed author interaction network in which initiations form edges between author nodes. We analyzed this network, which includes reciprocal initiations, for RQ1. We analyzed the interactions between reciprocated dyads for RQ2.

Interactions occur by an author on a *site*. Therefore, to construct a directed one-mode network containing only author nodes requires assumptions about the intended target *user* when an author interacts with a site. However, no assumptions are needed to construct a two-mode network [218] containing both site and author nodes; simply make an edge between an author and a site if any interaction exists between that author and that site. When constructing our network, we exclude interactions from authors to sites on which they have published any update as *self-interactions* (12.99% of all interactions). Table 4.1 shows that the resulting network has 915K initiations that form edges. To convert this two-mode network into a one-mode network, we assume that each interaction links the interacting author to *all authors* who have previously written an update on that site. In addition, we noted during data exploration that many guestbooks and comments are directed to both the caregiver author(s) of a site and the patient themselves, even if the patient had not yet published an update on the site for which they are the subject (or perhaps had not even yet created a CaringBridge account). Bloom et al. made a similar observation during their study of caregiver-authored CB sites. They found

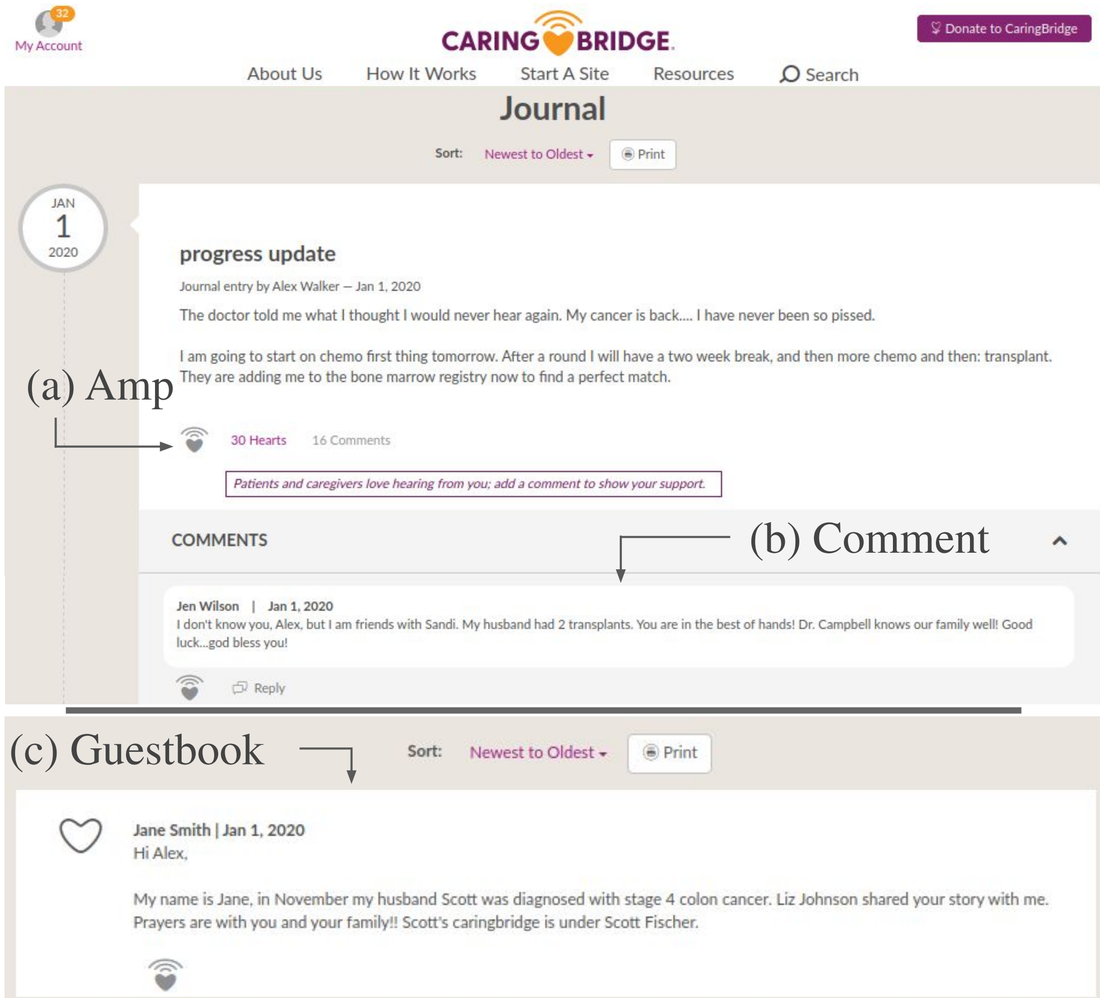



**Figure 4.3: CaringBridge interactions used in this study. Amps (a) and comments (b) are associated with a specific journal update, while guestbooks (c) are free-standing text posts left at the site level. Names and dates changed and texts paraphrased and anonymized [4].**

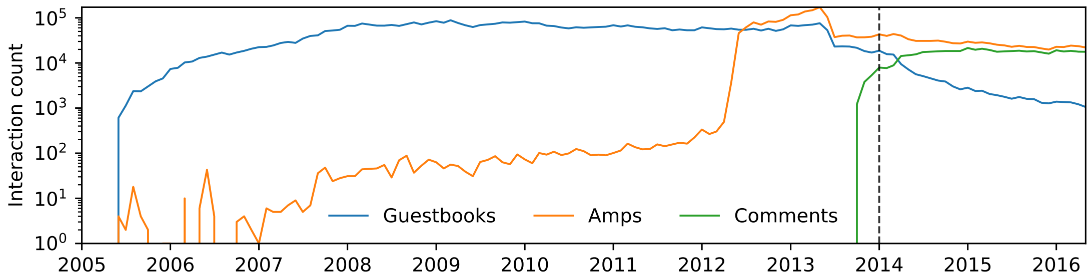



**Figure 4.4: Counts of each interaction type over time, on a log scale. The vertical dashed line indicates the beginning of the initiations period.**

| | | | |
|---|---|---|---|
| Active Authors | 66,440 | SCC Count | 2,590 |
| Connected Authors | 55,655 (83.8%) | WCC Count | 2,335 |
| Isolates | 10,785 (16.2%) | Largest SCC Size | 16,946 (25.5%) |
| Max In-degree | 612 | Largest WCC Size | 45,038 (67.8%) |
| Max Out-degree | 409 | Largest SCC Diameter | 38 |

**Table 4.2: Summary statistics for the interactions network at the end of the dataset (2016-06-01).**

that support is not directed solely at the caregiver and instead is almost always directed to the patient in combination with the caregiver [127]. Thus, we also draw an edge between the interacting author and any authors who publish a patient-classified update on the site at a later time. Such patient-specific edges are uncommon; only 4.9% of all edges in our network are drawn to patient authors who had not yet published at the time of the interaction.[9]

Our construction process resulted in 1,144,492 edges in the one-mode author interaction network. Using the assumptions above, 9.1M author→site interactions results in 14.8M author→author interactions, of which 1.1M are initiations and thus form the edges within the interactions network.

### Interaction network structure

We offer a brief description of the overall structure of the network formed by author interactions on CaringBridge in order to understand the context in which interactions are occurring. Given that the network is directed, components of connected authors can be identified as *strongly connected*—meaning a set of authors all reachable following the directed edges in the network—or

[9]We also observe similar quantitative results when these patient-specific edges are not included for the three models (RQ1c, RQ2a, RQ2b) that are affected by this assumption.

*weakly connected*—meaning a set of authors all reachable following edges in any direction in the network. The simplest strongly-connected component (SCC) is two authors who have interacted with each other. The simplest weakly-connected component (WCC) is two authors where one author has interacted with the second, but that interaction has not been reciprocated.

Table 4.2 provides summary statistics for the network at the end of the data collection period. These statistics summarize the subgraph consisting of *active* authors only—the subset of valid authors that were active on CB within 6 months of the end of the data collection period—in order to capture only the recent connections.[10] At the end of the data collection period, 30.4% of active authors are in one or more of 21,910 total reciprocated dyads. The network is dominated by a single large WCC, in which a large SCC is embedded. This pattern is consistent with the structure of other online health groups [13, 219]. We observe a lack of large isolated subnetworks, echoing findings in Urbanoski et al. [220]; the second-largest SCC and WCC contain only 14 and 18 authors respectively.[11] The vast majority of initiations (new edges) either grow or occur within the largest weakly-connected component (see further analysis in Appendix A.6). The number of active authors is generally decreasing during the initiations period (from 79.7K to 66.4K), as is the proportion of active authors in the largest connected component. See Appendix Figure A.3 for a temporal view of connectivity within the network. Author indegree and outdegree are positively correlated ($r=0.468$, $p < 0.001$), consistent with prior work suggesting that online support-giving is highly reciprocal [221].

## 4.5 Methods: Analysis

Having classified authors by role and built the author interaction network, we now present the methods used to address our research questions. To address RQ1, we isolate for analysis the initiations that form the edges in the interactions network. Because initiations between authors on CaringBridge are unexpected, we need to understand "what initiations look like". To determine if and how connected authors know each other, we conducted a content analysis of comment and guestbook initiations (sec. 4.5.1). Next, to address the three subquestions of RQ1, we fit three different quantitative models:

1. To identify the factors associated with which authors do any initiation, we used logistic

[10] Statistics are similar for the full network without inactive authors removed.
[11] The full distribution of the connected components is shown in Appendix Figure A.2.

regression to predict whether an author initiates or not (sec. 4.5.3).

2. To identify the factors associated with when authors initiate relative to their first published update, we used linear regression to predict the amount of time between an author's first update and their first initiation (sec. 4.5.4).

3. To identify the factors associated with whom authors choose to initiate, we used conditional multinomial logistic regression to predict the target of each initiation (sec. 4.5.5).

The factors we consider during modeling are represented in four types of features: network, author role, activity, and health condition. We motivate these features (sec. 4.5.2) prior to introducing the models.

To address RQ2, we isolated for analysis the reciprocated dyads within the network. To address the two subquestions of RQ2, we fit three quantitative models:

1. To identify the factors associated with which authors reciprocate initiations, we used logistic regression to predict if an initiation will be reciprocated (sec. 4.5.6).

2. To identify the factors that result in more interactive relationships, we used negative binomial regression to predict the total number of interactions in a relationship. Additionally, we used logistic regression to predict if a relationship is balanced or not (sec. 4.5.7).

### 4.5.1 RQ1 Methods: Initiations within the network

We conducted a content analysis to characterize the relationship between the initiator and the receiver. We randomly sampled 400 comment initiations and 400 guestbook initiations made by valid authors in the initiations period. Two annotators independently coded the 800 initiations. Annotators used the text of the initiation—and the text of the associated journal update in the case of comment initiations—to identify two aspects of the initiation: (1) whether the initiator's tie with the receiver existed before the health event that resulted in the creation of the CB site, and (2) the relation between the initiator and the receiver. Aspect (1) utilized a closed code set to identify the initiator-receiver tie as pre-health-event, post-health-event, or unknown. Aspect (2) was coded in an open manner, allowing for any relationship descriptor that could be identified from the context of the text e.g. friend, fellow patient, one-time site visitor, etc. The two annotators met to discuss and resolve disagreements. Reliability was established through these disagreement discussion meetings [222]. We direct readers to Appendix A.8 for additional details regarding the annotation process. By annotating these two aspects of initiations, the content analysis contextualizes the RQ1 quantitative results: it characterizes the relationship between interacting authors on CB and surfaces genuine first-time interactions of the type that digital interventions may try to facilitate.

### 4.5.2 RQ1 Methods: Features for modeling

In identifying factors that are associated with the initiation of new connections by authors, we consider four sets of features that prior work suggests are associated with the initiation of connections. Here, we introduce the four feature sets as motivated by prior work and discuss operationalization in the CB context at a high level. The specific features used in the models are introduced later in the relevant model description.

- **Network features**—Interaction between users is affected by the network context of the initiator and the receiver [13, 72]. Since initiations between authors on CB are unexpected, we explore the impact of CB network context on predicting who initiates with whom. We include network features that capture the current position of the initiator and the receiver. For example, "triadic closure" is a well-known network phenomena in which two previously unconnected people with a mutual contact are likely to connect [223]; on CB, we can explore triadic closure using a binary feature that indicates if two not-yet-connected authors share a mutual connection. Simpler features describe if an author has ever been interacted with or if an author has ever initiated with others.

- **Author role features**—Structural health role may affect user interaction online. For example, Hartzler et al. found that role (as patient, survivor, or caregiver) was important in finding peer mentors [139]. We include role as a categorical variable (Patient, Caregiver, or Mixed, see sec. 4.4.2) to quantify the importance of health role to connection on CB. Where appropriate, we also include features for site author configuration e.g. binary indicators for mixed-site authors, as mixed-site authorship may suggest multiple intended recipients of an interaction.

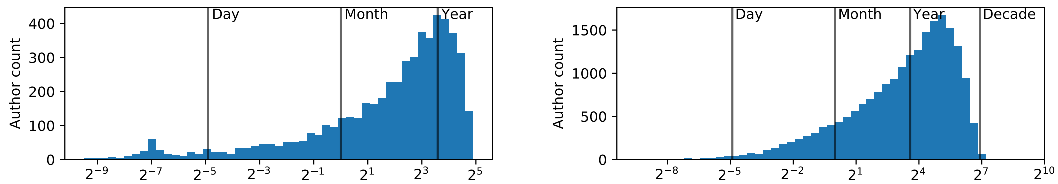
- **Activity features**—An author's level of activity is associated with their engagement with others. In general, receiving replies online is associated with retention [116]. On CB specifically, posting frequency is correlated with receiving comments [70]. Therefore, an author's update frequency is a key confounder for understanding initiations behavior; highly active authors may have different connection patterns than less active authors.

- **Health condition features**—Health condition and health status are important predictors in the formation of new connections in OHCs [109, 111]. On CB, we operationalize health condition from author-reported data. When CB authors create a site, they have the option to self-report a broad health condition category such as Cancer or Injury. While only 59.1% of valid sites self-report a health condition and such single-label categorization may be overly reductive [11], the sites that do report a condition can inform us about differences in inter-author communication behavior based on the health condition under question. This data enables us to confirm expected patterns in increased interaction between authors who have the same health condition [224, 58] and compare the importance of health condition to other factors. We assign health conditions to authors based on the reported health condition of the sites they've authored or 'None' when no condition is reported. For the 446 (3.1%) authors with multiple sites that report different health conditions, we assign the first non-None/'Condition Unknown' health condition reported. Ten total health condition categories were assigned, with counts shown in Appendix Table A.1. When used as a feature, health condition is included as a ten-level categorical variable and abbreviated "HC".

As this work is exploratory, we had no specific hypotheses to test; instead, we fit the most parsimonious model that still enabled us to explore the relative importance of the four factors described above.

### 4.5.3 RQ1a Methods: Who? Initiating authors

Which authors initiate during their time on CB, and which do not? To understand which factors predict initiation, we fit a logistic regression model predicting whether an author has made any initiation. To avoid bias in the model from right-censoring—some authors will initiate but only after the end of our data collection period—we predict an outcome of initiating within 1 year of first authorship. Similarly, we include a feature describing if an author was interacted with by another author in the first year. Only 11.7% of authors make their first initiation more than 1 year from their first update. Only 2.4% of authors are first interacted with more than 1 year




**(a) Pre-authorship initiators ($n$=5,439): Time between first initiation and first authored update**

**(b) Post-authorship initiators ($n$=20,687): Time between first authored update and first initiation**

**Figure 4.5: Distribution of time between first authorship and first initiation. For pre-authorship initiators, the median time to authorship is 5.8 months (mean 7.8mos). For post-authorship initiators, the median time to initiation is 13.6 months (mean 22.3mos).**

from their first update. Models with non-bounded outcomes and predictors produced similar results.

#### Features

Author role was included as a categorical variable. Binary indicators for the mixed-site and multi-site author designations were included as potential confounders. We included update count—the total number of updates published by that author on CB—as well as update frequency—the ratio of that count to the author's tenure in months—as indicators of author activity level. We included health condition as a categorical variable.

### 4.5.4 RQ1b Methods: When? Initiation timing

Given that an author is going to initiate, when is their first initiation likely to occur? We aim to understand the lifecycle of authors on CB by modeling when authors transition to peer-seeking behavior relative to their activities as authors. Thus, we differentiate between *pre-authorship initiators* who first interact with another author before publishing their first journal update and *post-authorship initiators* who first interact after publishing their first journal update. 20.82% ($n$=5,439) of valid authors are pre-authorship initiators, with the remaining being post-authorship initiators. We treat the pre-authorship initiator and post-authorship initiator cases separately, fitting linear regression models to predict the number of months between first authorship and initiation. Figure 4.5 shows the distribution of this interval.

#### Features

To understand the relationship of the time between first initiation and first published update to the total time spent on CB, we compute "total active time"—the number of months between an author's first published update or interaction and their last recorded update or interaction. We include total active time in order to develop an understanding of initiator life-cycle and capture the relationship of initiation to the update-writing activities of authors. For post-authorship initiators, we add a binary feature "Is interacted with?"–1 if that author was interacted with by any author pre-initiation and 0 otherwise. Pre-authorship initiators cannot be interacted with by definition, as they lack sites to interact with until they become authors.

### 4.5.5 RQ1c Methods: With whom? Initiation target

We now turn to the question of whom an author initiates with given that they are initiating with someone. We model initiations between authors as discrete choices to add a new directed edge to the graph, following Overgoor et al. [223]. In this paradigm, we fit a model to compute the conditional probability of a particular initiating author choosing the targeted receiving author—as opposed to all other authors who have sites on CB—given that the initiator is making a new initiation at this particular moment in the lifecycle of CB. Fitting a model to estimate this probability enables us to evaluate the relative importance of author traits such as health condition and role on the choice of a new connection target.

#### Conditional multinomial logit models

We used the Conditional Multinomial Logit Regression model, or conditional mlogit model, to estimate the probability of an author initiating with another author. Specifically, the conditional mlogit model estimates the probability $P_{i,t}(j, C)$ of author $i$ initiating with author $j$ from among the set of candidates $C$ at time $t$. Given features $x_t$ for each author, we learn coefficients $\theta$ such that $P_{i,t}(j, C) = \exp(\theta^T x_{j,t}) / \big( \sum_{\ell \in C} \exp(\theta^T x_{\ell,t}) \big)$ [223]. As it is not computationally feasible to compare the initiation target against all other possible authors (i.e. $C =$ "all authors"), one can employ negative sampling to select a subset of *candidate* authors that were *not* initiated with as a comparison group, without biasing the coefficient estimates [223]. We sample 24 candidate authors from the set of all valid authors with sites at the time of the initiation who have not previously been initiated with by the initiating author.[12] Thus, for each initiation, the model selects the true target from one of 25 candidate receiving authors. A model performing no better than random would achieve only 4% accuracy at identifying the correct target.

### Features

For each initiation, features are computed for the pairs formed by the initiating author and each of the 25 candidate authors (which includes the target author). We use features from all four sets:

- **Network features**—Following [223], we include (1) a binary feature for any non-zero indegree, as well as the log of the (2) outdegree and (3) indegree of each candidate, using a censored log that returns 0 when degree is zero. We include additional binary features for (4) reciprocation, which is 1 if the candidate has previously interacted with the initiator, (5) weak connection (recommended in [121]), which is 1 if the candidate is already in the same weakly-connected component as the initiator, and (6) friend-of-friend, which is 1 if the initiator is connected to a neighboring author that is already connected to the candidate (i.e. a feature for triadic closure [225]). All network features are computed from the state of the network at the time of the initiation.

- **Author role features**—Includes (1) author role of the candidate, (2) a binary feature that is 1 if author role is the same between the initiator and the candidate, and binary features for if the candidate is (3) a multi-site author or (4) a mixed-site author.

- **Activity features**—Includes (1) count of updates made by candidate at the time of the initiation, (2) frequency of updates i.e. update count divided by author tenure in months, (3) number of days since the candidate's most recent published update prior to the initiation, and (4) number of days since the candidate's first published update.

- **Health condition features**—Includes only (1) a binary feature that is 1 if the initiator and the candidate author are assigned the same non-None health condition. (See section 4.5.2.)

[12]Negative candidates are sampled from the state of the network at the time of the initiation and so authors who have not yet posted an update cannot be negative candidates.

We fit and present results for a full model containing all feature sets as well as models with each feature set independently. We checked inputs for colinearity, finding no two features were highly correlated.[13] While the feature sets we use contain no demographic information, prior work suggests that demographic homophily plays only a small role in OHCs [121]. However, we do fit a model using a proxy of author geography, discussed next.

### Geographic analysis

One potentially important factor in predicting initiations on CB is the geographic proximity of CB authors. Geographic proximity may indicate existing offline social relationships or post-diagnosis connections made in-person rather than on CaringBridge. While a fine-grained investigation of geographic effects on the relationship between the CB interactions network and the social networks of P/CG is out-of-scope for this work given the available data, we generate a rough proxy for geographic proximity by assigning US states to authors based on the IP addresses of their journal updates and guestbooks.[14] To evaluate the impact of geography proximity on initiations, we fit a separate full mlogit model including this proxy as a feature.

We use the Maxmind GeoLite2 City database (from Aug 13, 2019) to do IP geolocation lookups.[15] We refer to journal updates and guestbooks that are assigned identifiable geographic coordinates as *geo-identifiable posts*. 93% of CB authors' geo-identifiable posts are entirely based in the United States. Among these authors, we attempt to assign US states as our proxy for geographic proximity. We avoid the direct use of latitude/longitude estimates to reduce the bias introduced through the use of IP geolocation [226].

A US state is assigned to an author if that author has at least 10 geo-identifiable posts and among those posts the most-frequently-occurring state holds a plurality with at least a 20% margin above the second most frequent state, with the intent of creating a high-precision, low-recall state assignment. We fit a conditional mlogit model that includes a dummy variable for same state assignment—1 when the initiating author and the candidate author have the same state assignment, and 0 otherwise—in order to assess the importance of geographic co-location.

[13] The greatest correlation is $r = 0.44$ between total update count and days since first published update.
[14] Our data do not capture IP addresses for amps or comments.
[15] `https://dev.maxmind.com/geoip/geoip2/geolite2/`

### 4.5.6 RQ2a Methods: Reciprocation

RQ2 addresses relationships—reciprocated dyads of valid authors. Such relationships start with an initiation between an initiating author and a receiving author, followed by a reciprocal initiation from the receiving author to the initiating author after some period of time. To understand which non-reciprocal initiations will result in a reciprocal initiation, we fit a logistic model to predict if the receiving author will reciprocate. As with RQ1a (sec. 4.5.3), to account for potential right-censoring of reciprocations i.e. reciprocations that occurred after the end of the data collection period, we predict if a reciprocation will occur *within one year* of the original initiation and train only on author pairs with an originating initiation occurring no later than one year before the end of the dataset.[16] Thus, to broaden the scope of the reciprocations and relationships considered, we use initiations from the full eligible range (Jan 2005 to June 2015). Across all pairs of authors with at least one directed initiation between them in this time period, 12% are reciprocated.

To understand which initiations will result in a reciprocal initiation, we fit a logistic regression model predicting if an author pair with one initiation will be reciprocated within one year. As features, we utilize initiator author role and receiver author role, including a full interaction term, in order to tease apart the impact of author role on reciprocation. In addition, we include the number of months (log transformed) between the receiver's first published update and the original initiation in order to understand when authors are most likely to reciprocate a connection in their time on CB. We include the same duration (log transformed) for the initiator, although as the initiator may not yet be a published author themselves we include a binary indicator variable that is 1 if the initiator published their first update before the initiation.

### 4.5.7 RQ2b Methods: Relationships

A relationship is any reciprocated dyad between valid authors *and* their associated history of interactions. Our analysis includes dyadic relationships from the full timeline in order to compensate for bias introduced by right censoring, as some interactions in a relationship will occur after the end of the data collection period (see Appendix A.11 for additional analysis). We fit two quantitative models in order to understand what factors—especially author role—are associated with more interactive and more balanced relationships.

[16] 74.8% of reciprocations occur within one year.

### Number of interactions

To explore the impact of author role on total number of interactions in a relationship, we fit a Negative Binomial regression model to estimate this count. As the data are counts and significantly over-dispersed, a negative binomial model is appropriate; empirically, we observe a better fit with a negative binomial model than with Poisson or log-linear regression. To improve the fit and reduce sensitivity to outliers, only relationships with fewer than 345 interactions (99th percentile) were included in the training data.[17] We included an interaction term between initiator author role and reciprocator author role to tease apart the impact of role for both authors. We also included a feature for the duration of the relationship in months, which is measured from the first initiation to the last observed interaction within that relationship. Finally, to control for the level of balance in the relationship, we include a binary feature "Is balanced?" using the definition of balance introduced in the next section.

### Relationship balance

Balanced relationships are potentially desirable because giving and receiving support are actions that reinforce each other but play distinct roles in realizing positive effects [205]. We wanted to contrast one-sided relationships to relationships where support is mutually exchanged. We operationalized relationship balance as the percentage of interactions in a relationship made by the original initiator vs the reciprocator. For all relationships, we classify a relationship as balanced if no author made more than 75% of the interactions in the relationship. We fit a binary logistic regression model to predict if a relationship is balanced or not. As in the interaction counts model, we include features for the relationship duration and a full interaction between the initiator and receiver author role. We include the total number of interactions as an additional control.

## 4.6 Results

Results follow the same structure as the methods in Section 4.5. We present the results for RQ1 in sections 4.6.1-4.6.4, followed by the results for RQ2 in sections 4.6.5 and 4.6.6.

[17] Results were similar excluding outliers at the 98th and 99.9th percentile. The max number of interactions in one relationship was 18,340.

### 4.6.1 RQ1 Results: Initiations within the network

We annotated 800 initiations to identify (1) whether the initiator's relationship with the receiver existed before the health event that led to the creation of the CB site and (2) the relation between the initiator and the receiver. We provide quotes from comments (marked 'C') and guestbooks (marked 'G') to illustrate relation categories, paraphrased and renamed to preserve anonymity and reduce traceability [4, 227]. The majority of initiations are unidentifiable in both respects e.g. *"I am praying for you! May the love and support of family and friends be of comfort to you."* (G). 57.4% ($n$=459) of initiations were coded with an unknown timing relative to the health event, while 63.9% ($n$=511) were coded with an unknown relation.

The non-unknown initiations provide insight about initiators' goals and relationships to the receiver. 28.6% ($n$=229) of initiations were coded as pre-health-event, while only 12.1% ($n$=97) were coded as post-health-event. For each, we list the most common relations in order to demonstrate what pre-health-event and post-health-event initiations look like on CB. The most common pre-health-event relations were:

- Friend ($n$=118) e.g. *"My heart is heavy from hearing the news. .... Remember to search for answers and ask questions so that you understand everything. Hugs to all the family."* (C).

- Family ($n$=20) e.g. *"Hi Uncle Steve, A good day for scans, I will remember to sigh an extra prayer! Think of u often! Love, Anna"* (C).

The most common post-health-event relations were:

- Third-party connections ($n$=35) e.g. *"Hello, I came across your site because a mutual friend commented on facebook. I am sorry you are going through this. I am battling breast cancer (and also a mom of 6) and I wish you well."* (C).

- CG of similar patient ($n$=24) e.g. *""Hi Johnson family, we met at Parents of Preemies day. I loved reading the stories of strong little Timmy and especially the last update that he is home! Congratulations!"* (G).

- One-time visitor ($n$=15) e.g. *"My heart aches with your familiar story. I've never met you, but you and my husband now share a similar story. My husband is a STAGE IV PROSTATE CANCER SURVIVOR. You can do this. We are here for you guys and just like*

> *everyone else- we want to help. We are praying hard that you get some answers that give you HOPE. "* (C).

Some post-health-event initiations suggest that the initiation is a genuine "first contact" between these two authors, e.g. *"I stumbled across a link to Terry's CaringBridge page and read through your loving entries. I took care of my mother's CaringBridge page. A labor of love and a nice way to keep loved ones informed."* (G). Others include explicit links to their CG sites: *"Hi, my name is Kaylee and i came across your page! i look forward to following your story. My website is: (CB site link)"* (G). More detailed results, including the counts of each type of relation identified and additional quotes, are presented in Appendix A.8.

This content analysis suggests that **a small but important percentage of peer connections are formed between authors post-health-event, although a larger percentage are from existing connections re-established on CB as a result of the health event**. We now have an idea of what these initiations look like as we move into the quantitative modeling. This content analysis also surfaces several interesting phenomena that are not addressed by our quantitative work. How were post-health-event one-time visitors getting links to the receiving CB site? What is the role of relation in the (re)forming of mutually supportive relationships? These questions could be explored in future qualitative work.

### 4.6.2 RQ1a Results: Who initiates?

Which authors initiate peer connections? We fit logistic regression models to identify the factors that differentiate authors who will never interact with a fellow author (42.7% of authors, $n$=154,811) from *initiating authors*—authors who have made at least one initiation (57.3%). The models are trained on the 53,335 authors who published their first update in the initiations period. Table 4.3 shows three models predicting author initiation within their first year on CB. In exploratory modeling, we observed a strong impact of being interacted with on initiation behavior; being interacted with is associated with a 182% increase in the odds of initiating. Thus, we fit two additional models, splitting the data by whether they had been interacted with, shown as models (2) and (3) in Table 4.3. When an author is not interacted with, being a patient rather than a caregiver is associated with a 30% decrease in the odds of initiation. When an author *is* interacted with, being a patient is associated with a 21% increase in the odds of initiation. This disparity suggested a statistical interaction effect between being interacted with and author role.

| | (1) Full | (2) Int Received | (3) No Int Received |
|---|---|---|---|
| Intercept | -0.323* | 0.242* | -0.845* |
| | (0.017) | (0.025) | (0.026) |
| Role = Mixed | -0.104* | 0.048*** | -0.311* |
| | (0.038) | (0.052) | (0.064) |
| Role = P | -0.024*** | 0.189* | -0.357* |
| | (0.024) | (0.032) | (0.040) |
| Update count | 0.005* | 0.004* | 0.001* |
| | (0.000) | (0.000) | (0.000) |
| Update frequency (updates/month) | -0.011* | -0.010* | -0.006* |
| | (0.001) | (0.001) | (0.001) |
| Is a mixed-site author? | 0.074* | -0.210* | 0.014*** |
| | (0.019) | (0.025) | (0.033) |
| Is a multi-site author? | 0.012*** | 0.275* | 0.012*** |
| | (0.042) | (0.062) | (0.063) |
| HC = Cancer | 0.104* | 0.098* | 0.108* |
| | (0.021) | (0.029) | (0.036) |
| HC = Cardiovascular/Stroke | 0.048*** | 0.071*** | 0.040*** |
| | (0.048) | (0.065) | (0.077) |
| HC = Congenital/Immune Disorder | 0.087*** | 0.135*** | -0.067*** |
| | (0.112) | (0.146) | (0.200) |
| HC = Infant/Childbirth | -0.105*** | -0.026*** | -0.455* |
| | (0.074) | (0.095) | (0.141) |
| HC = Injury | -0.116* | -0.221* | -0.056*** |
| | (0.056) | (0.071) | (0.097) |
| HC = Neurological Condition | 0.261* | 0.174* | 0.466* |
| | (0.062) | (0.086) | (0.093) |
| HC = Other | 0.074*** | 0.042*** | 0.190*** |
| | (0.078) | (0.108) | (0.118) |
| HC = Surgery/Transplantation | 0.030*** | 0.110*** | -0.012*** |
| | (0.055) | (0.076) | (0.088) |
| Observations | 53,335 | 28,870 | 24,465 |
| Log Likelihood | -35923.278 | -19434.315 | -14391.070 |
| Test Accuracy | 58.99% | 60.95% | 73.23% |

**Table 4.3: Three logistic regression models for predicting if an author will initiate with other authors. Model (1) includes all authors. Model (2) includes only authors who receive at least one interaction from another author in their first year. Model (3) includes only authors who are not interacted with in their first year. Table 4.4 explores the interaction between author role and being interacted with. Note: $^{*}$p<0.05, $^{**}$p<0.01, $^{***}$p<0.001.**

| Feature | Coef. | Std.Err. | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -0.9049 | 0.0163 | -55.4107 | 0.0000 | -0.9369 | -0.8729 |
| Role = Mixed | -0.2979 | 0.0648 | -4.5992 | 0.0000 | -0.4248 | -0.1709 |
| Role = P | -0.2962 | 0.0400 | -7.3971 | 0.0000 | -0.3747 | -0.2177 |
| Is interacted with? | 1.0857 | 0.0212 | 51.2408 | 0.0000 | 1.0441 | 1.1272 |
| Role = Mixed : Int'ed with? | 0.4356 | 0.0825 | 5.2807 | 0.0000 | 0.2739 | 0.5972 |
| Role = P : Int'ed with? | 0.5932 | 0.0504 | 11.7622 | 0.0000 | 0.4943 | 0.6920 |

**Table 4.4: Logistic regression model predicting if an author will initiate with other authors. P authors are less likely than CG authors to initiate in the absence of interactions. Both P and CG become much more likely to initiate if interacted with (182% increase in the odds of initiating), but this effect is stronger for P than for CG. (Observations = 53,335, model d.f. = 5, log-likelihood = $-$34,147, test accuracy = 67.9%)**

Table 4.4 shows this significant interaction effect, which demonstrates that patient authors are much more likely to be initiators after being interacted with compared to caregivers, although both patients and caregivers are more likely to initiate after being interacted with. **Among non-receivers, caregivers are more likely to initiate than patients; among receivers, patients are more likely to initiate than caregivers.**

We observe differences in initiation probability by health condition. For example, compared to reporting no health condition, an author self-reporting Cancer is associated with an 11% increase in the odds of initiating. However, we caution against over-interpretation of the less-common health conditions categories such as Congenital/Immune Disorder. Activity level and authorship configuration have small effects on probability of initiation.[18]

### 4.6.3 RQ1b Results: When?

When do authors initiate peer connections? To understand the lifecycle of CB users and the relationship between intended use (publishing updates) and appropriative use (peer connection), we use linear regression to model the time between first published update and first peer initiation. This interval is shown in Figure 4.5. We conduct analyses of author timing for the 5,439 pre-authorship initiators and the 20,687 post-authorship initiators separately. Only 3.72% of

[18] In a separate model predicting if an author is interacted with, rather than if they initiate, these effects are more relevant, as could be expected. Each additional update published is associated with a 1% increase in the odds of being interacted with, and being a mixed-site author is associated with a 123% increase in the odds of being interacted with (versus a 15% increase in the odds of initiating).

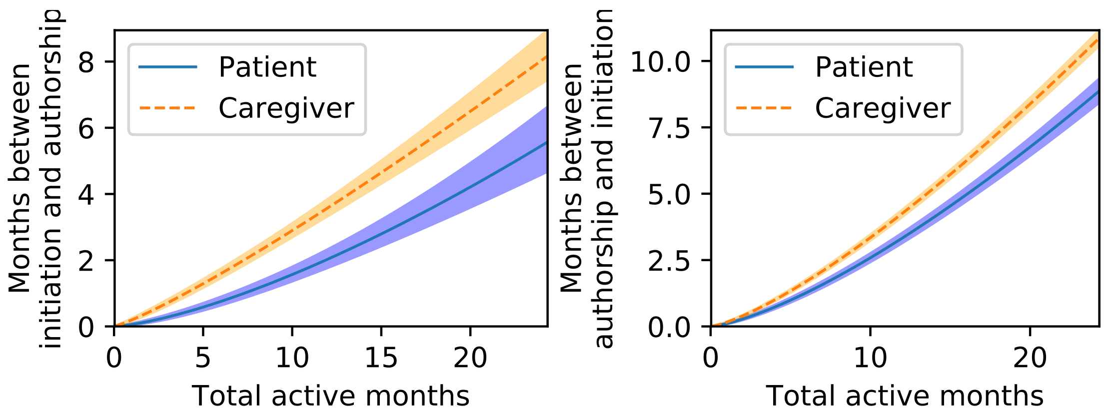
| Feature | (1) Pre-authorship Coef. | SE | (2) Post-authorship Coef. | SE |
|---|---|---|---|---|
| Intercept | -1.6215*** | 0.078 | -1.8437*** | 0.032 |
| Role = Mixed | -0.6928** | 0.264 | -0.0897 | 0.108 |
| Role = P | -1.2148*** | 0.166 | -0.3996*** | 0.071 |
| Total Active Time (log months) | 1.1660*** | 0.030 | 1.3251*** | 0.010 |
| Total Active Time : Role = Mixed | 0.1657 | 0.102 | 0.0187 | 0.033 |
| Total Active Time : Role = P | 0.2607*** | 0.064 | 0.0616** | 0.023 |
| $R^2$ | 0.310 | | 0.542 | |
| F-stat | 489.2*** | | 4,899*** | |
| Observations | 5,439 | | 20,687 | |
| Log-likelihood | -10,699 | | -33,771 | |

**Table 4.5: Linear regression models (d.f.=5) predicting the time between an author's first published update and their first initiation (log months). Model (1) includes only pre-authorship initiators, whereas model (2) includes only post-authorship initiators. Note: *p<0.05, **p<0.01, ***p<0.001.**

initiating authors do so post-authorship but before receiving an interaction, with 75.46% of initiating authors doing so after authoring their first update *and* receiving an interaction. Among pre-authorship initiators, patients publish their first update 1.28 months sooner after first initiation than caregivers (t=−5.31, p¡0.001). Among post-authorship initiators, patients initiate 4.7 months sooner after first authorship than caregivers (t=−11.91, p¡0.001). Controlling for total active time on CB by fitting a linear model to predict the number of months between first authorship and first initiation (log transformed), we observe the same pattern. We find a significant interaction between total active time and author role for both pre-authorship initiators (ANOVA F=3.72, p¡0.01) and post-authorship initiators (ANOVA F=8.91, p¡0.001). Both pre- and post-authorship model details are given in Table 4.5.

An effects plot of the author role interaction among initiators is shown in Figure 4.6. **Compared to caregivers, patients initiate sooner after becoming an author.** The effects plot shows a positive but sublinear trend, indicating that initiating earlier is associated with pre-initiation time forming a smaller percentage of an author's total time on CB.[19] Furthermore, the gap between patients and caregivers widens among users active on CB for a longer total period

[19]If the trend were one-to-one, the percentage of total time that is between authorship and initiation would be the same on average for all authors, irrespective of their total time on CB. Instead, we see that total time is associated with a smaller percent of total time in the interval between authorship and initiation.




**(a) Pre-authorship initiators ($n = 5439$)** **(b) Post-authorship initiators ($n = 20687$)**

**Figure 4.6: Effects plot of interaction between author role and total active time on CB pre- and post-initiation. The sublinear trend indicates that a longer time active on CB is associated with a shorter time between first authorship and first initiation. Shading indicates the 99% confidence interval. Mixed author role is not shown, as the difference from the CG condition is not significant (Table 4.5).**

of time. For pre-authorship initiators who are only active for one month, patients are predicted to become authors 4.2 days sooner than caregivers. But among pre-authorship initiators who are active for two years, patients are predicted to become authors 77.2 days sooner than caregivers. For post-authorship initiators, the gap at one month is 1.6 days, widening to 59 days at two years. This widening gap suggests a "lifecycle" model of CB use in which the authors active for a longer period of time are more likely to initiate earlier as a percentage of their total active time than those authors active for a shorter period of time, although we make no attempt to untangle the causal directionality of this effect.

We also fit linear regression models using a larger set of confounding features—but without total active time—in order to assess the predictability of time between first authorship and first initiation. For space reasons, model coefficients are presented in Appendix Table A.4. These two models demonstrate that pre-authorship initiation is intrinsically high variance ($R^2 < 0.01$). For post-authorship initiators, the author role, health condition, and "interaction received" features are significant predictors at the 99% confidence level and so the proportion of variance explained is higher ($R^2 = 0.13$). For post-authorship initiators, receiving an interaction is associated with, on average, initiating 10.5 months sooner. This large difference extends the results from RQ1a: **Not only are receiving authors more likely to initiate if interacted with, but**

**also they will initiate much sooner than authors who are not receivers.**

### 4.6.4 RQ1c Results: With whom?

With whom do authors initiate peer connections? We fit conditional mlogit models to predict the probability of an author being the target of an initiation, as determined by both the traits of the initiator and the traits of the target. Table 4.6 shows the model coefficients and test accuracies. With the exception of the health condition model, all models predict the correct initiation target significantly above chance (i.e. 4%), with the full model predicting the accurate target from among 25 candidates 77.3% of the time. In isolation, the single most important feature is the number of days since the candidate's most recent update; a model that exclusively predicts as the target the candidate who has updated most recently actually achieves 81% accuracy in the testing period, which explains the strong prediction performance of model #4. However, we are interested in inference: the relative importance of the features. Two features have a directionality difference between the focused models vs the full model: (a) the author being a patient rather than a caregiver, which overall makes an author more likely to be the target of an initiation but not when controlling for non-role factors; and (b) candidate update count, for which more updates is associated with a greater overall likelihood of being the initiation target, but not when controlling for non-activity factors. We thus focus on analysis of the full model with all features.

The most important binary feature is the reciprocation indicator—i.e. if choosing this candidate would result in a reciprocated connection—which is consistent with the importance of reciprocity in online interactions [221]. In general, network features are more predictive of initiation than having the same author role. We see strong triadic closure effects i.e. "Is friend-of-friend?" has a large positive coefficient, and even being weakly connected with a candidate increases the likelihood of a connection. Authors who have already received at least one initiation are more likely to be selected for subsequent initiations by other authors, even when controlling for author tenure and number of updates published by that author. **An author's network position is an important factor in receiving new initiations,** with authors who have been interacted with by other authors being the most likely target for new initiations.

Author role is also important in the selection of a target author. **Initiations are more likely to occur between two authors with the same role.** This effect is stronger for caregivers than

| | (1) All | (2) N | (3) R | (4) A |
|---|---|---|---|---|
| Candidate out-degree (log) | $-0.191^{*}$ | $-0.510^{*}$ | | |
| | (0.005) | (0.004) | | |
| Has in-degree? | $0.756^{*}$ | $0.995^{*}$ | | |
| | (0.017) | (0.013) | | |
| Candidate in-degree (log) | $0.649^{*}$ | $0.674^{*}$ | | |
| | (0.005) | (0.003) | | |
| Is reciprocal? | $20.016^{*}$ | $13.068^{*}$ | | |
| | (0.460) | (0.174) | | |
| Is weakly connected? | $1.767^{*}$ | $2.454^{*}$ | | |
| | (0.021) | (0.021) | | |
| Is friend-of-friend? | $5.220^{*}$ | $4.881^{*}$ | | |
| | (0.097) | (0.050) | | |
| Candidate Role = Mixed | 0.020 | | $0.095^{*}$ | |
| | (0.018) | | (0.012) | |
| Candidate Role = P | $-0.242^{*}$ | | $0.124^{*}$ | |
| | (0.012) | | (0.008) | |
| Same author role? | $0.299^{*}$ | | $0.371^{*}$ | |
| | (0.012) | | (0.008) | |
| Candidate multi-site author? | $0.315^{*}$ | | $0.249^{*}$ | |
| | (0.015) | | (0.010) | |
| Candidate mixed-site author? | $0.474^{*}$ | | $1.365^{*}$ | |
| | (0.008) | | (0.005) | |
| Candidate update count | $-0.0003^{*}$ | | | $0.001^{*}$ |
| | (0.00004) | | | (0.00002) |
| Candidate update frequency | $0.007^{*}$ | | | $0.004^{*}$ |
| | (0.0002) | | | (0.0002) |
| Days since recent update | $-0.011^{*}$ | | | $-0.013^{*}$ |
| | (0.00005) | | | (0.00005) |
| Days since first update | $-0.001^{*}$ | | | $-0.001^{*}$ |
| | (0.00001) | | | (0.00001) |
| Same health condition? | $0.213^{*}$ | | | |
| | (0.009) | | | |
| Observations | 155,141 | 155,141 | 155,141 | 155,141 |
| Log Likelihood | −133,747 | −353,610 | −465,556 | −206,744 |
| Test accuracy | 77.3% | 32.4% | 9.8% | 73.3% |

**Table 4.6: Conditional mlogit models predicting initiation probability for an initiating author and an arbitrary candidate author. The first model includes all features sets; models 2-4 include only one of the feature sets: Network, Role, and Activity respectively. The model fit with only the health condition feature is not shown: the model's single coefficient is 0.411 (s.e. 0.006), and it has log likelihood -496,923.3 and test accuracy 4.9%. Note: $^{*}$p$<$0.01.**

for patients. We observe no significant difference between caregiver-role candidates and mixed-role candidates. Authors of multiple sites and authors on mixed sites are also more likely to be the target of initiation, perhaps due to a larger set of interested readers that may include peer authors.

The activity features demonstrate that initiation is more likely when a candidate has more recently become an author, having written fewer updates but at a high frequency. As discussed above, authors are much more likely to be the target of initiations shortly after they publish an update, which may suggest update dissemination effects and textual content factors that are not captured in this analysis.

We also fit models using only the subset of the 66,616 first initiations, in order to evaluate whether the first initiation made by an author is somehow different. Feature directionality and relative magnitude remain the same, with the exception that first-time initiators are less likely to initiate with mixed-site authors.

### Geographic analysis

32.7% (118,534) of authors were assigned US states using the state-assignment procedure. As a face validity check, the most frequent state assignments are Minnesota (6.3%; CaringBridge was launched in Minnesota), California (2.8%; the most populous US state), and Texas (2.8%; the second most populous US state). Authors assigned states are more active, more likely to have a plurality of their updates in a single US state, and more likely to initiate with other authors (2.41 vs 2.26 mean initiations, $p < 0.001$) than the average CB author.

In the initiations period, initiations between state-assigned authors account for only 4.5% (7,007) of the total initiations. 49.5% of these initiations were between two authors that have the same US state assignment, a percentage significantly above chance. Fitting a full multinomial logit model that includes a dummy variable when the initiating author and the candidate author have the same state assignment confirms the importance of this feature: sharing a state assignment increases the odds of initiating with an author, holding other variables fixed. (The model details and full comparison is shown in Appendix A.10, Table A.5.) Fitting a model with only that feature results in a test accuracy (on 668 initiations in the test period) of 27.8%. This analysis suggests the importance of geographic co-location, although given the biased nature of the proxy used it is hard to reason about the magnitude of this effect relative to the other contextual factors. At a minimum, **geographic co-location is an important predictor of initiation**

| Feature | Coef. | Std.Err. |
|---|---|---|
| Intercept | -2.320 | 0.008 |
| Initiator Role = Mixed | -0.090 | 0.016 |
| Initiator Role = P | -0.332 | 0.013 |
| Receiver Role = Mixed | -0.178 | 0.017 |
| Receiver Role = P | -0.110 | 0.015 |
| Init. Role = Mixed : Rcvr. Role = Mixed | 0.597 | 0.043 |
| Init. Role = P : Rcvr. Role = Mixed | 0.687 | 0.034 |
| Init. Role = Mixed : Rcvr. Role = P | 0.630 | 0.036 |
| Init. Role = P : Rcvr. Role = P | 1.204 | 0.025 |
| Was init. author? | 1.115 | 0.010 |
| Was init. author? : Months after first update | -0.310 | 0.002 |
| Months after first rcvr. update | 0.036 | 0.002 |

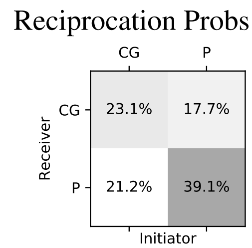



**Table 4.7: Logistic regression model predicting if an author pair with one initiation will be reciprocated within one year. All coefficients are significant at p<0.001. Reciprocation is more likely between two authors that have the same role. Predicted reciprocation probabilities are given for various initiating and receiving author roles fixing the other variables such that both the initiator and the receiver became authors one month before the original initiation. (Observations = 737,747, model d.f. = 8, log-likelihood = $-257{,}750$).**

**for some authors**.

### 4.6.5 RQ2a Results: Reciprocations

Which authors reciprocate and to which initiators? Among all pairs of authors with at least one directed initiation, only 12% are reciprocated within one year. We fit a model to predict if an author will reciprocate an initiation from another author within one year, as a function of both the initiator's and the receiver's author role. Table 4.7 shows the coefficients for the logistic regression model. **Authors are more likely to reciprocate an initiation when the initiator and the receiver have the same role.** Patients are particularly likely to reciprocate in general, with patients who initiate with caregivers receiving the lowest reciprocation rates. Unsurprisingly, initiations from pre-authorship initiators are much less likely to be reciprocated; the initiator having already published their first update at the time of initiation is associated with a 205% increase in the odds of reciprocation. When the initiator is already an author, reciprocation is less likely the longer that author has been on CB. In contrast, receivers are *more* likely to initiate the longer they have been on CB. To the right of Table 4.7, we show estimated

| Feature | IRR | Coef. | Std.Err. | z | p |
|---|---|---|---|---|---|
| Intercept | 17.479 | 2.8610 | 0.0061 | 471.7680 | 0.0000 |
| Initiator AR = Mixed | 0.981 | -0.0194 | 0.0125 | -1.5531 | 0.1204 |
| Initiator AR = P | 0.928 | -0.0748 | 0.0099 | -7.5329 | 0.0000 |
| Reciprocator AR = Mixed | 0.967 | -0.0338 | 0.0128 | -2.6322 | 0.0085 |
| Reciprocator AR = P | 0.946 | -0.0553 | 0.0108 | -5.1382 | 0.0000 |
| Init. AR = Mixed : Recip. AR = Mixed | 1.044 | 0.0430 | 0.0328 | 1.3112 | 0.1898 |
| Init. AR = P : Recip. AR = Mixed | 1.187 | 0.1717 | 0.0254 | 6.7680 | 0.0000 |
| Init. AR = Mixed : Recip. AR = P | 1.083 | 0.0798 | 0.0268 | 2.9724 | 0.0030 |
| Init. AR = P : Recip. AR = P | 1.261 | 0.2321 | 0.0179 | 12.9835 | 0.0000 |
| Is balanced? | 0.702 | -0.3545 | 0.0056 | -63.0274 | 0.0000 |
| Duration (months) | 1.020 | 0.0198 | 0.0001 | 150.5271 | 0.0000 |
| Alpha | — | 0.9337 | 0.0035 | 265.0840 | 0.0000 |

**Table 4.8: Negative Binomial Regression model predicting dyadic relationship interaction counts. Incidence rate ratios (IRR) are given in the first column. Alpha is the estimated dispersion parameter, which is assumed non-zero. (Observations = 125,629, model d.f. = 10, log-likelihood = $-$525,510)**

reciprocation probabilities based on author role, given that both the initiating and receiving author published their first journal update one month before the initiation. While only 12% are initiated among all pairs, the probability of initiation among patient/patient pairs in which the initiator has already published their first update is more than twice that baseline.

### 4.6.6 RQ2b Results: Relationships

What factors lead to more interactive relationships? We present results for models predicting the total number of interactions in a relationship and the degree to which the interactions are balanced between both authors in a relationship.

#### Number of interactions

We identified 125,629 relationships for analysis. The median relationship has 13 interactions, and 93.5% of relationships have 100 or fewer interactions. We evaluated the impact of having the same author role on the number of interactions in a relationship. Table 4.8 shows the incidence rate ratios and coefficients for the negative binomial regression model predicting the total number of interactions in a relationship. Features with incidence rate ratios greater than 1 are

| Feature | Coef. | Std.Err. | z | p |
|---|---|---|---|---|
| Intercept | 0.3867 | 0.0097 | 39.7880 | 0.0000 |
| Initiator AR = Mixed | -0.0940 | 0.0253 | -3.7129 | 0.0002 |
| Initiator AR = P | -0.1148 | 0.0201 | -5.7070 | 0.0000 |
| Reciprocator AR = Mixed | -0.1349 | 0.0260 | -5.1841 | 0.0000 |
| Reciprocator AR = P | -0.1170 | 0.0218 | -5.3628 | 0.0000 |
| Init. AR = Mixed : Recip. AR = Mixed | 0.2020 | 0.0666 | 3.0319 | 0.0024 |
| Init. AR = P : Recip. AR = Mixed | 0.1186 | 0.0515 | 2.3035 | 0.0212 |
| Init. AR = Mixed : Recip. AR = P | 0.1306 | 0.0544 | 2.4004 | 0.0164 |
| Init. AR = P : Recip. AR = P | 0.3156 | 0.0363 | 8.6853 | 0.0000 |
| Interaction count | -0.0029 | 0.0001 | -26.0014 | 0.0000 |
| Duration (months) | -0.0060 | 0.0002 | -25.6988 | 0.0000 |

**Table 4.9: Logistic regression model predicting dyadic relationship balance. Dyadic relationships in which both authors have the same author role (AR) are more likely to be balanced. (Observations = 124,377, model d.f. = 10, log-likelihood = $-$85,749)**

associated with an increased total number of interactions, while an incidence rate ratio less than 1 is associated with a decreased total. **Relationships have more interactions when authors have the same role.** Compared to caregiver/caregiver relationships, relationships where the initiator or the receiver is a patient is associated respectively with a 7.2% and a 5.4% decrease in the rate of interactions. However, *both* authors being patients is associated with a 10.7% increase in interactions relative to caregiver/caregiver relationships.

### Relationship balance

Balance refers to the difference in the number of interactions made by the original initiator vs the original reciprocator. To control for the noise introduced by short relationships, we train the model using only relationships with at least 10 interactions ($n$=124,377). The initiator of a relationship tends to interact more: 63% of relationships involve a majority of the interactions coming from the initiator. A majority (52.47%) of relationships are balanced, with an additional 32.82% dominated by the initiator and the final 14.71% dominated by the reciprocator. Appendix Figure A.5 shows the distribution of relationships by the percentage of interactions coming from the initiator.

We predict relationship balance as a function of author role while controlling for total interactions and duration. Table 4.9 show the coefficients for the logistic regression model predicting

| RQ | Result | Sec. |
|---|---|---|
| RQ1a | Among non-receivers, caregivers are more likely to initiate than patients; among receivers, patients are more likely to initiate than caregivers. | 4.6.2 |
| RQ1b | Compared to caregivers, patients initiate sooner after becoming an author. | 4.6.3 |
| RQ1c | Initiations are more likely to occur between two authors who have the same role. | 4.6.4 |
| RQ2a | Authors are more likely to reciprocate an initiation when the initiator and the receiver have the same role. | 4.6.5 |
| RQ2b | Reciprocated relationships have more interactions and are more balanced when authors have the same role. | 4.6.6 |

**Table 4.10: Summary of results.**

relationship balance. **Reciprocated relationships are more balanced when authors have the same role.** Having the same author role is associated with a 12% increase in the odds of a relationship being balanced. As with number of interactions, patient/patient relationships are more likely to be balanced compared with caregiver/caregiver relationships.

## 4.7 Discussion

In the analyses presented in section 4.6, we identified a variety of behavioral patterns and differences between patient and caregiver authors on CB. Table 4.10 highlights our key findings. Why do we observe differences in connection behavior between patients and caregivers? We suggest two primary interpretations for the observed differences: First, the differences might indicate diverging *preferences* of patients and caregivers for specific connection behaviors. Second, the differences might indicate the existence of communicative or technical *barriers* influencing the observed behaviors. These interpretations are not mutually exclusive, with both aspects together contributing to the communication patterns observed in this study.

We emphasize that future qualitative work is needed to detangle these two interpretations. For example, we found that while patients are more likely to initiate with peers than caregivers, caregivers are more likely to initiate after they have received an interaction. Is this result due to patients' greater desire to actively seek out a support community (preference), to caregivers' lower knowledge of peer authors on CB (barrier), to caregivers' greater need for support alongside a stigma against asking for that support (preference+barrier), or to other factors? Suggestively, Smith et al. found that—in the case of instrumental support—a greater proportion of caregivers do not ask for needed support compared to patients [8]. To identify the cause of the initiation gap, researchers could qualitatively study how these unmet caregiver support needs manifest in OHC peer connection behaviors.

This section explores our results through the lens of preferences and barriers. First, we explore the RQ2 results with an eye toward fostering supportive online relationships, highlighting connections to prior work and unknowns that could be addressed in future work. Second, we discuss the implications of these findings for recommendation systems designed to facilitate new relationships. Third, we discuss the implications of our results for future research that incorporates structural health roles, including plausible factors that affect caregivers' behavioral preferences for connection.

### 4.7.1 Fostering online relationships

Our results show that patient/patient and caregiver/caregiver author dyads are more likely to have highly interactive and balanced relationships than patient/caregiver dyads. Does this observed gap reveal a preference for interactions amongst authors who have the same role? This interpretation is supported by Thoits' theory that experientially-similar others are important sources of support because they provide active coping assistance beyond the instrumental support provided by offline caregivers [59]. The prevalence of same-role dyads thus reflects a preference for authors who have had or are having similar experiences. This preference could be supported through designs that aid authors in identifying experientially-similar others to engage for support. For example, Ruthven proposes "narrative retrieval" as a novel IR task—one that could be applied in this context to identify similar-narrative authors who are sharing or who have shared experiences with the seeking author [228]. In addition to supporting same-role dyads, future qualitative work could investigate the specific types of support provided in *different*-role dyads in order to identify the strategies used to make [currently rare] patient/caregiver relationships mutually beneficial and verify Thoits' theory about active coping assistance.

The observed gap could also indicate social barriers making it harder for patient/caregiver dyads to form relationships on CB. Two salient barriers are that caregivers may not "know what to say" to patients [229] and that patients may perceive support offers as unhelpful [230]. Tools such as MepsBot represent an opportunity to intervene during the comment-drafting process to increase the confidence of caregivers that they are writing comments that will be perceived as supportive to the receiving patient [107]. Given the shifting needs of patients, sometimes non-response or a non-text response such as an amp may be the appropriate interaction for the relationship; designs for intervening in an on-going communcative process may benefit from incorporating and surfacing elements of the existing relationship's context in order to adapt to the complicated norms around response in sensitive health contexts [231, 232]. In summary, the gap between patient/caregiver and same-role author dyads indicates both preference for communication with experientially-similar others and socio-technical barriers to cross-role communication. OHCs should consider addressing this gap in pursuit of mutually supportive relationships.

### 4.7.2 Designing peer recommendation systems

Much research on OHCs is motivated by the goal of designing recommendation systems to form new relationships [128, 144, 64, 42, 233, 139]. Such systems have a goal of facilitating mutually supportive communication. However, creating a recommendation system to facilitate supportive communication is hard; many online support interventions do not work as intended, producing minimal positive changes [234, 235]. Peer recommendation systems that are faithful to the preferences of users have a better chance to succeed. Our work reveals the preferences of users "in the wild" and suggests three types of features that could be incorporated into peer recommendation systems in order to support the communication behaviors that OHC users are already doing.

- *Author role features.* People benefit from and want relationships with experientially-similar others [13, 59]. That people with the same author role tend to form these types of connections on CB provides evidence that the incorporation of author role information as additional recommendation system features could facilitate connections that have the shared experience qualities authors are seeking [114, 236]. It is particularly notable that having the same author role is associated with more interactive relationships, a desirable outcome of recommendation to increase supportive social engagement. Author role should be considered alongside shared demographic or health condition traits as strong correlates with supportive connections [109]. Our recommendation echoes calls to connect caregiver family members of cancer patients with others in a similar position [125].

- *Network features.* Network features play an important role in who users initiate with; therefore, recommendation should also take network characteristics into account [121]. Since authors who have initiated with others are also more likely to be the target of future initiations, one goal of such a system may be to encourage a first interaction in order to facilitate the formation of future connections. Such an outcome could be encouraged through the recommendation of popular, central authors in the largest weakly-connected component with many existing connections. Once connected with at least one other, friends-of-friends and others in the same component are natural choices given the importance of the network features in the RQ1c analysis.

- *Temporal features.* Recommendation should also consider the impact of the timing of recommendations given; recommendations given in a particular period of a health journey may be more impactful in terms of positive health outcomes [72]. We find that patients initiate sooner than caregivers do, which may suggest differential “readiness” for forming peer connections by author role. Both patients and caregivers will initiate more quickly after being interacted with, so recommendation of previously-uninteracted-with authors may result in more communication overall. Reciprocation is most likely when contacted by newer authors, so creating opportunities for newer authors to interact may be particularly beneficial for creating more reciprocated connections. We also observe that initiating early is associated with a longer total time on CB. While longer-term use of CB is not necessarily beneficial for authors, it may indicate both need and opportunity for the cultivation of the longer-term mutually supportive relationships that form the foundation of a self-sustaining health community [219].

Our results suggest that incorporating these features could facilitate more supportive connections. However, subsequent experimental work is necessary to verify the effectiveness of these features for the peer recommendation problem in OHCs.

### 4.7.3 Incorporating structural roles in future research

We highlight two implications of our results for research that incorporates structural roles.

*Integrate structural and behavioral roles.* Our results show that the structural roles *patient* and *caregiver* are associated with differences in OHC connection behavior. While these roles have been used in previous literature [61], they have functioned mainly as descriptors without a clear definition. Our results suggest an important research opportunity: understanding the relationship between behavioral roles and structural roles. Behavioral roles describe patterns of user behavior. For example, Yang et al. define role as “a set of interaction patterns regulated by explicit or implicit expectations and adopted by people in a social context to achieve specific social goals” [114]. Using this definition, they label patterns of behavior with names like “story sharer”. Structural roles function on a higher level, where adopting the role of *patient* or *caregiver* represents a personal transition: in terms of motivation, responsibilities, and relationships with others [118, 36]. Identifying and tracking a person’s behavioral roles during their transition into a structural role such as *patient* could link motivations for use of OHCs with specific behaviors e.g. sharing stories. Such linking enables supporting a variety of structural roles by designing for the motivations that lead to the behaviors associated with those roles. Tensions between structural role and behavioral enactment of that role are relevant outside the health context as well. In contexts where structural roles have explicit technical support, such as content moderation [237], separately examining moderator motivations and behaviors could motivate changes in the technical support provided for that role.

*Focus on caregiver motivations.* Our results about caregivers’ use of CB has implications for understanding their motivations for connection. First, we find that—in the absence of receiving an interaction—caregivers are more likely than patients to initiate with other CB authors. Caregivers’ greater propensity to connect with others may reflect a lack of offline support for caregivers that creates stronger motivations for caregivers to initiate, particularly in the search for experientially-similar others [62]. Alternately, the observed appropriative use of CB for inter-caregiver communication may reflect a lack of appropriate channel for this communication in the readily-available communication technologies already in use [238]. However, further research is necessary to untangle the degree to which this gap between patients and caregivers is indicative of unmet support needs versus a simple homophily preference. Second, our results show that patients are particularly receptive to interactions from other authors and that caregivers are less affected by receiving an interaction. Caregivers may be less likely to view themselves as the relevant recipient of the message [125, 127] or may view reciprocation as an inappropriate articulation of personal concerns and challenges [239]. In general, the appropriative use of CB for peer communication rather than health blogging presents opportunities to meet unmet needs. Qualitative work is needed to understand the motivations that lead caregivers to initiate and to understand the particular importance these online peer relationships play in acquiring support for caregivers.

## 4.8 Limitations & Future Work

In this section, we discuss some limitations of our approach and sketch opportunities for future work to address these limitations.

The CB interaction network we study here is a partial view of the true social network, which includes both offline connections and online connections established or developed on other platforms. Furthermore, authors that do interact may differ from authors who do not interact on a variety of demographic and psychosocial factors [240, 58]. In studying the connection behavior of authors already on CB, we are engaging with a non-random sample of patients and caregivers, so application of these patterns to offline contexts should be done cautiously [241]. Cross-platform and online/offline studies would contribute greatly to an understanding of the online health support ecosystem and the applicability of these findings to patients and caregivers more generally.

In studying peer connections, connections formed on CB between two strangers are the most similar to those created via hypothetical recommendation systems. However, identifying and isolating only these connections is challenging. Detailed content analysis or other qualitative approaches to both identify and understand the formation of these connections would be valuable.

In reasoning about the importance of author features for peer connections, we note the risk of unobserved confounds [242]. While we attempted to address key confounders (e.g. geographic location) through additional analyses, the inclusion of additional likely confounding factors (e.g. existing offline relationships) would increase confidence in our findings. Furthermore, while we found an increased likelihood of connection between authors with the same role or health condition, it is impossible to differentiate homophily effects from contagion effects in observational data [243]. Future experimental work would enable exploration of these differing causes. In examining the full network, we also did not account for tie strength [244], which may be possible to estimate from the specific interactions between two authors. Incorporating tie strength would enable the comparison of weak ties to strong ties formed on OHCs, which are known to have different supportive functions [201].

In this study, we did not explore the impact of forming peer connections on specific outcomes such as engagement, perceived support, or stress. Further understanding of these outcomes is an important area for further research before the implementation of systems that facilitate the creation of these connections [13]. While explorations of web-based social support have found correlations between social support and positive outcomes such as decreased stress, experimental interventions have not always found a decrease in stress even as received support increased for participants [245, 13, 98]. Causal work is needed to understand the contexts in which peer connections are beneficial to participants.

Overall, author connections on OHCs provide a fruitful ground for further inter-disciplinary multi-method research. We continue our investigation of author connections—including new connections among strangers—in Chapter 5.

## 4.9 Conclusion

In this chapter, we explored the formation of peer connections in an OHC without explicit peer finding mechanisms. By examining the peer connections that CaringBridge authors did form, we learned about their preferences. We found significant differences in the initiation, reciprocation, and maintenance of these connections between two important structural roles: patients and caregivers. This work indicates the importance of structural health roles to behavior in online health communities and suggests opportunities for the design of systems to actively facilitate or recommend these connections. A focus on author roles opens up multiple opportunities for future research applying these results to the dynamics and design of mixed-role systems. In particular, experimental work is needed to integrate author role into peer recommendation systems designed to facilitate interaction and foster mutually supportive relationships. In Chapter 5, we discuss the feasibility of that experimental work for facilitating peer connections.

# Chapter 5

# Peer recommendation as a support intervention

In Chapters 3 and 4, we developed an understanding of CaringBridge authors' journaling and peer interaction behavior. In this chapter, we synthesize that understanding to develop a system for peer recommendation. Peer recommendation systems are a computational approach to peer finding that encourage users to evaluate specific, recommended others. I conceptualize peer recommendation as a health intervention into the online social networks of OHC users. As an intervention, the peer recommendation system encourages users to read about the experiences of and interact with peers: two behaviors linked with potential health benefits. I evaluated the CaringBridge peer recommender system by conducting a 12-week field study in which authors received weekly peer recommendations via email. Promising field study results support the usefulness of and demand for peer recommendation and suggest benefits to evaluating larger peer recommendation interventions. The primary contribution of this study is practical guidance on the development and evaluation of peer recommendation interventions for OHCs.

## 5.1 Introduction

Social support helps people in health crises cope with stressful circumstances. Access to emotional, informational, and instrumental support is associated with increased quality of life [246], improved psychosocial health [247], and physical health [74]. Support from *peers*—people who have had similar health experiences—is particularly useful [63, 7]. However, a person's existing offline support community may lack peers [59, 65]. Online communities provide a place for people to support each other in ways that their existing offline support communities cannot by offering opportunities to connect with peers. While health support is exchanged online on diverse platforms [182, 248, 249], online health communities (OHCs) are specifically designed for health-related discussion and support [7, 250].

Even in OHCs, however, finding supportive peers can be time-consuming [104, 64]. Algorithmic matching systems could enable design features that help OHC users to find peers, but existing approaches are limited or have remained entirely theoretical [139, 10]. For example, a common approach to peer finding requires a user to explicitly search or filter for people or topics of interest, a process that support seekers and providers may find labor-intensive and discouraging in health-related contexts [104, 168]. *Recommendation systems* for peer finding can incorporate the user's past behaviors as an *implicit* signal to identify potential matches—people with valuable similar experiences [114]. However, no existing recommendation system has been evaluated for peer matching in OHCs: no experimental evidence links the availability of peer recommendations with hypothesized increases in beneficial peer connection behaviors.

The goal of this chapter is to explore the idea of connecting peers in online health communities for mutual support. We accomplish this by designing, developing, and evaluating an email-based peer recommendation system for users of the online community CaringBridge. Despite recent research arguing for the potential utility of peer recommendation [144, 64, 42, 233, 139, 114, 134, 10, 9, 104], no study yet describes use of peer recommendation in practice. We conducted a *feasibility study* to assess the usefulness of and demand for peer recommendation. Feasibility studies are designed to determine if an intervention should be evaluated in a larger or more comprehensive study [1]. We argue that peer recommendation should be conceptualized as a health intervention that can change specific user behaviors—behaviors that are linked to psychosocial or physical health. Our proposed intervention is designed to increase two behaviors: reading about peers' health experiences and increasing interaction among peers. In the rest of this introduction, we summarize the proposed intervention (sec. 5.1.1), its evaluation in a 12-week field study (sec. 5.1.2), and the encouraging findings for our core contribution: evidence for the feasibility of using recommendation to connect peers in OHCs (sec. 5.1.3).

### 5.1.1 What is the proposed intervention?

The proposed intervention is displaying recommended peers to current OHC users. This design intervention has two aspects: an *interface* that displays details about recommendations and an *algorithm* that selects peers to recommend in the interface. The interface shows "profiles"—summaries of each individual peer being recommended [251]. By including previews of and links to a user's recent activity within the OHC, we enable prospective peers to evaluate the relevance of the recommended user [152]. Second, the algorithm identifies "relevant" peers for a specific user by incorporating their prior activity in the OHC and ranking potential peers based on their mutual activity.

We ground our feasibility study in the context of a large existing OHC—CaringBridge.org. CaringBridge users write blogs to describe their health experiences to their broader support networks. In the CaringBridge context, we will be recommending blogs written by peers to the authors of existing blogs. A recommendation intervention offers the potential for CaringBridge blog authors to form connections with peers that they may not be explicitly seeking but for whom they can give or receive meaningful support. We use a weekly email as the interface to display recommendations, producing textual profiles that contain previews of recent blog activity. We use a machine learning-based recommendation system to rank potential peers based on historical interaction behavior on CaringBridge.

As an intervention, recommending the blogs of peers to OHC users aims to directly alter those users' natural social networks. Such manipulations are necessarily complex [252], so a focus on specific user behaviors that the manipulation will induce is important. The intervention is expected to increase two behaviors: reading about the experiences of peers and interacting with peers. Both of these behaviors are associated with benefits, such as reduced stress, useful coping information, and a sense of community [13]—discussed further in sec. 5.2.1. However, the *experimental* evidence linking peer connection with benefits is mixed [7], and includes potential risks such as increased distress [13]. These mixed outcomes motivate us to carefully evaluate the feasibility of an intervention to increase these behaviors and produce benefits for participants.

**Table 5.1: Feasibility assessment of a peer recommendation intervention. We collected evidence in five focus areas (originally described by Bowen et al. [1]). Each point of evidence is associated with a corresponding section (in parentheses).**

| Feasibility Area | Description | Evidence |
|---|---|---|
| Demand | Interest in the intervention | Prior use (5.3.1), expressed interest (5.5.1), actual use (5.5.2) |
| Implementation | Tangible design and engineering to implement the intervention in a particular context | The system design: both interface (5.3.2) and model (5.3.3) |
| Practicality | Requirements for administering the intervention | Model quality (5.3.3), required data (5.3.3), compute time (5.3.3) |
| Acceptability | How participants react to the intervention | Explicit participant preferences (5.5.1) and feedback (5.5.2) |
| Efficacy | How much the intervention affects the desired behaviors | Reading behavior (5.5.3), interaction behavior (5.5.3), second-order effects (5.5.3) |

### 5.1.2 How did we assess feasibility of the intervention?

We conducted a field study to assess the feasibility of a peer recommendation intervention and to identify requirements for running a larger randomized controlled trial. *Feasibility* refers to the ability to use an intervention in reality: a broad and necessarily multifaceted construct. Thus, we collected evidence of feasibility in five focus areas, summarized in Table 5.1: Demand, Implementation, Practicality, Acceptability, and Efficacy. The names of these focus areas are adapted from Bowen et al.'s discussion of feasibility assessment for health interventions [1].

For each focus area, we collected converging lines of evidence to understand the feasibility of the proposed intervention. We derived this evidence from three primary data sources: a log data analysis of CaringBridge user data, surveys of CaringBridge users, and participant feedback during a 12-week field study. We used log data to develop a content-based recommendation model based on implicit feedback from historical peer interactions. We used survey data to identify motivations for peer connection, to characterize interesting peers, and to understand how field study participants engaged with recommendations. We observed usage of our recommendation system during a field study to assess the impact on reading and interaction behavior—as well as to identify any second-order impacts on the usage behavior of both the participants receiving recommendations and the blog authors receiving potentially-unwanted attention from peer strangers.

### 5.1.3 Contributions

The primary contribution of this chapter is a determination that peer recommendation interventions designed to facilitate OHC user connections are feasible. We offer this feasibility assessment in terms of demand for the intervention, implementation challenges, practicality of administration, acceptability to participants, and efficacy. We present our system design as a model for future peer recommender systems. During the 12-week field study, 79 participants received weekly peer recommendations via email, leading to hundreds of repeat visits to blogs and hundreds of extra peer interactions. Participants clicked 5% of recommendations, although less than half of the participants clicked any recommendation and fewer still chose to visibly interact with recommended blogs. We find no evidence of second-order harms or benefits, and overall find an interest in and willingness to engage with blogs written by peer strangers. We conclude with implications for the further development of peer recommendation systems for OHCs, including considerations for both the context of deployment and design trade-offs. We believe that peer recommendation systems can facilitate connections that authors may not be explicitly seeking—and that those connections can facilitate meaningful support. We offer the first substantive evaluation of a real peer-matching recommendation intervention and its impact on OHC user behavior.

## 5.2 Related Work

To design a peer recommendation system for users of online health communities (OHCs), we drew from existing work on OHCs, on peer matching, and on algorithmic recommendation. The most direct precursor to our current work is Hartzler et al.'s study of peer mentor recommendations for cancer patients and caregivers [139], which we discuss as an algorithmic approach to peer matching.

### 5.2.1 Online health community use

People use OHCs in the hope that they will obtain useful support. Both people experiencing a health condition and people caring for a loved one are motivated to use OHCs to help overcome isolation by offsetting deficits in existing relationships and identifying people who have had similar experiences [13]. While the effects of using OHCs are somewhat unclear, feeling socially supported is a key determinant of health linked to OHC use [7, 74]. Meta-reviews reveal consistent associations between social support and a variety of health outcomes, e.g. mortality [74, 75, 76]. OHCs have diverse interfaces and affordances, including forums, listservs, blogs, chatrooms, Q&A sites, and update feeds [13]. The social support available from OHCs is also diverse, including informational support from discussions of treatments and symptoms, emotional support from sympathetic others and being a part of a community, and other forms of support including the instrumental and the spiritual [7, 85]. Support can come from many types of users [114], but one particular benefit of OHCs is that they expose you to many *peers*.

Peers are people who have similar experiences [63]. Peer relationships differ from professional/patient and mentor/mentee relationships in that no formal role or expectation structures the relationship [63]. Compared to a person's existing offline support networks, peer can provide more useful support [59, 63]. A wide variety of theoretical models support the potential benefits of peer connection [7, 59, 63, 65]. We avoid adopting a specific theoretical basis for our current study, as we do not operationalize or measure social support directly, instead focusing on behaviors—reading and interaction—that are compatible with multiple theoretical models.

#### Reading about peer experiences

Reading about the experiences of peers can be beneficial even in the absence of interaction [7]. In addition to learning from the valuable information contained in peers' writing e.g. coping strategies [100], reading peer experiences can build a sense of community [101]. Further, reading can reduce loneliness [102], contribute to feelings of normalcy and hope [102, 103], reduce uncertainty and anxiety [100], and enable collective sensemaking about one's journey [104]. In general, reading the experiences of others can benefit readers by enabling positive and normalizing social comparisons to the experiences of others [98, 105]. But, making social comparisons is not without risk: the negative experiences of others can produce a sense of helplessness or increase distress [98, 106].

**Interacting with peers**

Interacting with peers offers many potential benefits—among them are membership in a community, acquisition of new information, normalization of one's experiences, and relief from distress [13]. Online peer interaction can take three general forms: providing support to others, receiving support from others, and forming reciprocal relationships. While receiving support from others has obvious appeal, not all support is perceived as wanted or useful, and in general there is mixed causal evidence for the benefits of online interaction-based social support [7]. A gap between received and perceived support bedevils designers of social support interventions: increased received support is only weakly correlated with perceptions of that support [67, 59]. *Providing* support to peers, on the other hand, may be more beneficial to the provider than the receiver [90]. Support needs differ over the course of a health journey, and providing support presents an opportunity to "give back" and enact self-efficacy [35, 39]. Reciprocal peer relationships can offer the best of both worlds, but also presents significant risks in health contexts [13]. Stress can increase if online contacts are doing poorly *or* doing well due to social comparisons [98]. The sudden drop-out of a connection, due to churn or patient death, can also increase distress [99]. Further, peers might be unintentionally unsupportive due to differences in communication style [107]. Due to the risks of interacting with peers, interventions designed to increase interaction cannot be deployed without careful evaluation of the risks and benefits—which motivates us to conduct an initial feasibility study for peer recommendation specifically.

### 5.2.2 Social support interventions

The archetypal peer support intervention is the support group [128]. Online support groups offer similar approaches using a different medium, although generally still designed for and managed in a clinical setting [129, 130]. Other clinical approaches bridge the gap to OHCs—for example, Haldar et al. designed an OHC for people in the same hospital [34]. Peer support interventions have potentially many goals in mind: providing social support, providing health information or education, developing self-efficacy (e.g. by vicarious viewing of peer behavior [101]), adjusting social norms (e.g. use of a particular health behavior), or even facilitating social movements for patient advocacy [63]. In 2004, Cohen expressed skepticism of peer support interventions in general, identifying a string of peer support group studies finding null effects and arguing for a focus on forming weak ties and propping up existing support networks [128]. Nearly 20 years later, the challenges associated with designing effective peer support interventions remain [7]. As an intervention into people's online social networks, we examine recommender systems as a mechanism for encouraging initial interactions that can blossom into weak tie relationships. In general, recommendation is one approach to improving the quality of support received by peers via matching peers by some measure of "fit". We discuss prior approaches to peer matching next.

### 5.2.3 Health peer matching

Health peer matching has occurred largely in the context of hospital-attached programs where mentors and mentees are matched by a 3rd-party broker, usually a nurse or program manager [141, 135, 152]. Consider "woman-to-woman", a peer support program for women with gynecologic cancer: when a new participant expresses interest in the program, the program manager selects a match "of similar diagnosis and age" from a pool of volunteer mentors [136]. In contrast, online peer recommendation is not constrained to formal mentor/mentee pairings and can draw from a much larger pool of prospective "volunteers" at the cost of the clear expectations that come with structure and a human coordinator. In this study, we aim to seriously consider non-coordinated peer matching as a health-related social support intervention.

Little explicit guidance exists for peer matching [133, 63]. Table 2.1 provided an overview of peer characteristics identified in prior work as salient or important for effective peer matching. We distinguish these characteristics as either *proposed* as an implication of a particular study, *used* in practice to match peers in a study or support program, or *expressed* by participants as preferences for or barriers to effective peer support. There are characteristics we don't represent in the table, such as abilities/skills [134], specific needs [64], interaction medium preferences (e.g. email) [142], existing social connections [142], etc. While not intended to serve as a rigorous meta-review, this existing literature suggests a wide range of potential characteristics to incorporate in a peer recommender system. A general finding from HCI research is that a shared health condition is not a requirement for peer communication; people perceive value to learning from and communicating with others based on many factors that shift throughout a health journey as needs and communication goals change [64].

While several works have collected empirical data on preferred peer characteristics in support settings, minimal *comparative* work exists to identify the most important characteristics [63]. Hartzler et al. are a notable exception, running scenario-based sessions in which participants explicitly evaluated five potential peer mentors based on provided health information [139]. Boyes surveyed cancer patients about the importance of specific shared characteristics such as gender, age, and cancer type, although this data is currently unpublished [132]. We extend this body of work by conducting a survey in which participants indicated their preference for specific peer characteristics. While these specific characteristics could be incorporated in future systems, for the purpose of our system we consider machine learning approaches that learn valued peer characteristics from prior peer interactions, discussed next.

### 5.2.4 Algorithmic recommendation for peer matching

Few published works explicitly discuss computational recommendation systems for online health communities. Hartzler et al. matched peer mentors on the basis of shared health interests, language style, and demographics—as extracted from prior posts made in the CancerConnect OHC—although they evaluated these matches in workshop interviews rather than actual use [139]. The other notable example is described only in Diyi Yang's thesis: Yang developed and deployed a recommendation system in the American Cancer Society's Cancer Survivor Network (CSN) forums "to direct participants to useful and informative threads that they might be interested in" [153]. They evaluated a model based on implicit feedback from prior commenting behavior by presenting recommended threads and users within the CSN interface, reporting greater thread click-through rate compared to a baseline model recommending recently popular threads. In contrast to the CSN forums, CaringBridge is a blogging platform without existing interface recommendation features for discovering other blogs. We present an evaluation for a new CaringBridge recommender that extends beyond click rate into explicit feedback and the effect of the recommendations on a variety of user behaviors.

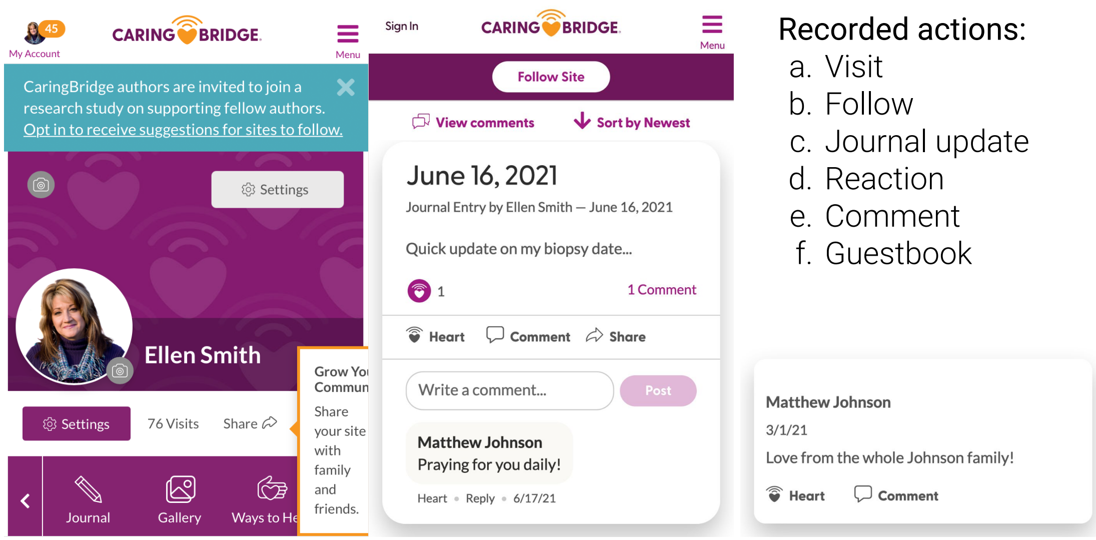
Outside of health, a variety of problem formulations and modeling methods have been used for the problem of recommending people. Recommendation models are generally supervised machine learning models that optimize a loss function comparing the model's output to "ground truth" labels: explicit or implicit feedback provided by users. Xu et al. present a useful review [154]. Use of implicit behavioral feedback is based on relevance assumptions, e.g. that clicked items are relevant while non-clicked items are not relevant [155], which may not hold true in practice [156, 157]. We assume that interactions between OHC users indicates relevance, discussed further in sec. 5.3.3. Given historical user/item pairs, one can then optimize a pointwise loss that rewards high scores for assumed-relevant user/item pairs and low scores for assumed-irrelevant user/item pairs. Input features vary from IDs for the user and item—which gives the classic matrix factorization approach to collaborative filtering [158]—to side information about the context where the recommendation was generated (e.g. the time and place) or content (e.g. prior comments) from the user or item [159, 160].

#### Alternatives to recommendation

Recommendation is not the only available mechanism for facilitating online peer connections. Two notable alternatives are improving search and filter tools and designing enriched profile pages to make it easier to represent one's diagnosis, expertise, and support needs [104]. Search is challenging in situations where a user's needs are known only implicitly to the user or are challenging to express in terms the system will understand [166]. We suggest that peer support finding is an *exploratory* [167] search task (e.g. see Pretorius et al.'s discussion of person-centered help-seekers [168]). Even with rich peer profiles available, it is challenging for searchers to formulate a query that captures their needs and intent [169, 170]. Other search systems for finding people—such as expertise-finding systems—were created based on interfaces designed to capture users' needs in a domain-specific query [171]. *Mindsets* is a recent example of the design work needed to capture domain-specific intents during query formulation [169]; additional research is needed in the peer support context to capture support seekers' and providers' intents. In contrast to search, recommendation offers opportunities to engage with potential peers without explicitly articulating a person's current needs.

#### Machine learning for matching people

Person-to-person recommendation is typically modeled as similar to the conventional user/item recommendation problem. Facebook's deep learning recommendation model (DLRM) represents a common approach, using embeddings for categorical features (including user IDs) and MLPs for creating dense representations of other features, then combining all representations with a final MLP [161]. Less recently, other sites have used approaches based on neighborhoods and similarity of interactions to connect with strangers specifically. Twitter's "who to




**(a) Site home page, as viewed by a logged-in author of that site. The study recruitment banner is visible at the top of the page.**

**(b) Journal page and entry, as viewed by a logged-out visitor.**

**(c) The six logged-in user actions, above a guestbook on the Well Wishes page.**

**Figure 5.1: The CaringBridge interface. We record six logged-in user actions: visits (to any of these site pages), follows (clicks on "Follow Site" and others, see sec. 5.3.1), Journal updates, reactions (on Journal updates, comments, or guestbooks), comments (on Journal updates or guestbooks), and guestbooks.**

follow" recommendations used an alternative approach similar to PageRank that uses only the existing follow network to make recommendations [162]. Guy et al. explicitly attempted to recommend strangers in an enterprise setting based on number of shared interests and memberships [163]. The modeling problem closest to peer recommendation may be romantic relationship recommendation, a context that aims to encourage interaction between users and values reciprocity [164, 165]. We drew on modern deep learning approaches to recommendation—the model we use during the field study is a simplified form of DLRM—to design a system appropriate for the peer support context.

## 5.3 System Design

### 5.3.1 Observed prior use of CaringBridge for peer connection

CaringBridge.org is an online health community and blogging platform that has been the focus of prior HCI research e.g. [10, 8, 70]. In collaboration with the CaringBridge organization, we were given access to usage data from the CaringBridge website. Registered users on CaringBridge can create blogs called *sites*, on which they can publish blog posts called *Journal updates*. An *author* is a user who has published at least one Journal update. Much usage of CaringBridge is based around notification emails (an example is shown in Figure 5.2). Visitors to a site can follow that site to be notified via email when an author of that site publishes a new Journal update. Follows are somewhat complex on CaringBridge. Clicking the Follow Site button (depicted in Fig. 5.1b) has the effect of subscribing the visitor to one of several types of notification emails, but users can accomplish the equivalent by turning on notifications for that site under the list of "Sites You Visit" contained on the user's Notifications page. Simply *visiting* a site will add it to the Sites You Visit list. Further complicating matters, removing a site from the Sites You Visit list is unrelated to Follows, and a visitor will automatically Follow a site if they visit *and* interact with that site.

The majority of visitors to a site will be a patient's existing support network [10]. Registered visitors can leave *reactions* to Journal updates (a reaction labeled "Heart" is the default) to show their support. Alternately, they can leave text-based *comments* on individual Journal updates or write *guestbooks* which are comments that appear on a special Well Wishes page. These interface components, and the actions that we track for this study, are shown in Figure 5.1.

We are interested here in *interactions* (via reactions, comments, and guestbooks) between registered CaringBridge authors. Adopting the terminology used by Levonian et al. [10], an *initiation* is the first interaction between an author and a site on which they are not an author. In assessing demand for a peer recommendation intervention, we consider the volume of inter-author communication on CaringBridge, studied in detail by Levonian et al. [10]. Between 2010 and 2021, 275K authors initiated with other authors, more than 32.3% of all 852K authors active during that period; this interaction occurs despite minimal existing discovery features to facilitate finding peers. While CaringBridge offers a search feature, authors use this tool to find *specific* authors, not for general searches for e.g. particular health conditions (see Appendix B.1).

**Table 5.2: Associational differences between sites that receive interactions (ints) from at least one peer author and sites that only receive interactions from visitors. Site tenure is the number of months between the first and last Journal update on a site. # updates is the number of Journal updates published more than 30 days after the first update. The common language effect size is 62.4% for site tenure and 62.1% for number of updates. All comparisons are significant at the 99.5% significance level.**

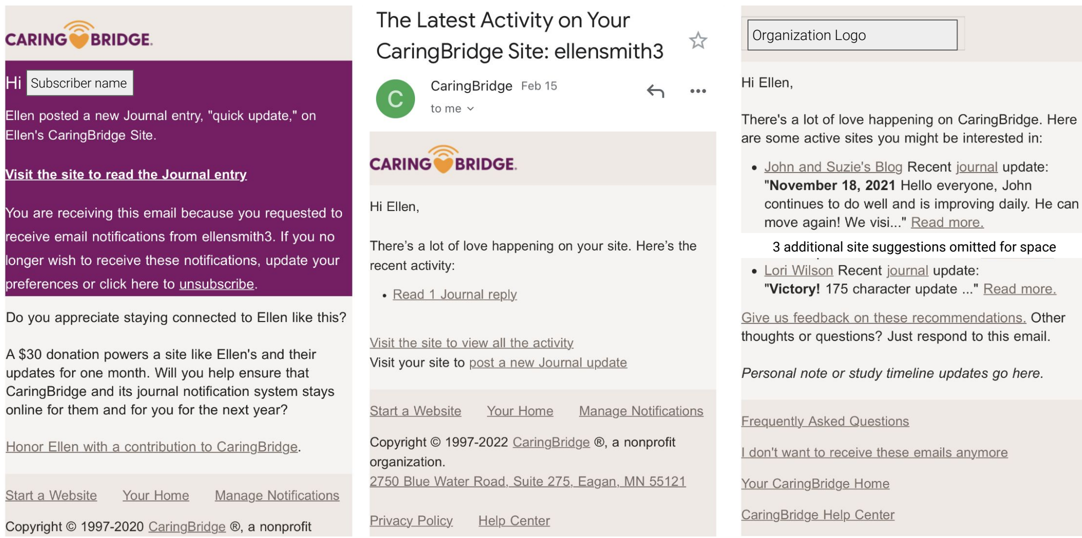
| | 1+ peer int within 30 days | Non-peer ints only | Difference |
|---|---|---|---|
| Site Count | 92,352 sites | 100,424 sites | -8,072 sites |
| Site Tenure (Median) | 3.9 months | 0.8 months | +3.2 months |
| # updates (M; SD) | 20.5 (47.6) | 12.4 (35.7) | +8.1 updates |
| # updates (Median) | 8 updates | 2 updates | +6 updates |
| % sites with 2+ updates | 78.1% | 59.5% | +18.7pp |

Based on sec. 5.2.1, we would expect that peer interaction is associated with behavior change for authors. Prior work demonstrates that receiving interactions from visitors—peers and non-peers—is associated with retention on CaringBridge [12], but are peer interactions more impactful than non-peer interactions? Table 5.2 shows associational differences between CaringBridge sites based on the interactions (reactions, comments, and guestbooks) received from visitors within 30 days of a site's first published Journal update. Sites for which at least one visitor interaction is left by a peer author will publish on CaringBridge for a median of 3.2 additional months (with 6 additional updates) compared to sites that receive only interactions from visitors.[1] This positive correlation suggests that peer interaction could change author behavior in ways that are themselves correlated with author benefits [70]. A peer recommendation intervention intervenes in this existing ecosystem of peer interaction: we propose a design appropriate for the CaringBridge context in the next section.

### 5.3.2 Adapting a recommendation system to CaringBridge

For a peer recommender intervention to be successful, the system must be adapted to the specifics of the context. We summarize three adaptations in our design for the CaringBridge context: recommending blogs rather than peers, using email as the recommendation medium, and designing the email interface.

[1]This analysis was conducted on 192,776 sites created between January 1, 2014 and September 1, 2021 (the start of the study) that received at least 1 visitor interaction in the first 30 days. These general results hold when adjusting for publishing rate, number of interactions received, number of visitors, and number of peer visitors. We omit full modeling results (Poisson for # updates and Cox's proportional hazards model for site tenure) for brevity.




**(a) Existing design: Visitor notification email** **(b) Existing design: Author notification email** **(c) New design: Site Suggestion email**

**Figure 5.2: Existing design of email notifications on CaringBridge and the Site Suggestion email interface designed for this study. Data shown is a representative fabrication.**

### Peer recommendation or blog recommendation?

While we are motivated by peer-to-peer connection, we are restricted by the interface affordances provided by CaringBridge. CaringBridge offers only a limited public profile view that is minimally exposed in the interface; instead, the sites themselves present a better view into an author's activity. For this reason, we recommend *sites* rather than *authors*. The system optimizes for peer connections, but presents those recommendations as site summaries.

### Email as recommendation medium

As much usage of CaringBridge is motivated by email notifications, we chose to deliver recommendations via email. Author notification emails (depicted in Figure 5.2b) are sent to site authors when a visitor interacts on a site or a co-author publishes a Journal update. As our recommendations are author-centric and authors are familiar with the design of the author notification emails, we used the same layout and CSS style for our Site Suggestion emails.

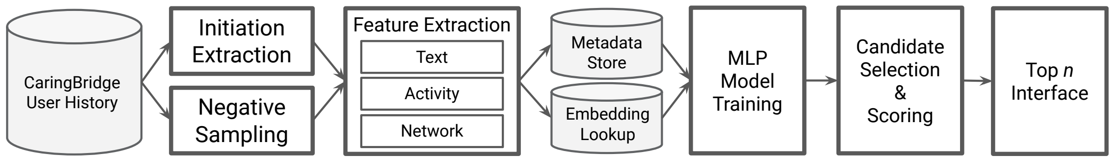



**Figure 5.3: CaringBridge peer recommendation system overview**

#### Design of the Site Suggestion email

The email design can be seen in Figure 5.2c. Site Suggestion emails contain 5 bullet-pointed site suggestions, a request for feedback and link to a feedback survey, a link to the study FAQ, and an unsubscribe link. Each site recommendation is presented with the site's title (usually the patient's name) and a link to the site's Journal page. We follow Hartzler et al.—who found the most important aspect of evaluating peer mentors is sample posts—by including a preview of the most recent Journal update for each recommended site [139]. We chose to present 5 recommendations as a trade-off between email length and multiple options. We discuss the model used to generate site recommendations in the next section.

### 5.3.3 Model development

To present recommended sites to a recommendation-seeking author, we include the top 5 sites as scored by a recommendation model. We adhered to two key design requirements: First, recommendations should be *personalized*, focusing on the right connection rather than a popular connection. We should not recommend any one site to a lot of people as that could create a negative experience for that site. Second, recommendations should be available even for authors who have never visited or interacted with another CaringBridge site. Because support-based reading and interaction is our goal, an author's initial "cold start" recommendations should be of similar quality to recommendations for long-time authors.

To provide recommendations that meet these two criteria, we implemented a content-based recommendation system. Social matching systems require users to disclose sensitive personal information [253], and CaringBridge is a context where authors are already making those sensitive disclosures in the content of the Journal updates they publish [9]. Our modeling task is to use implicit feedback from user activity (sec. 5.3.3) and content-based features (sec. 5.3.3) to create a recommendation model that predicts historical peer connections and, eventually, new site recommendations (sec. 5.3.3). We discuss and compare several recommendation models (sec. 5.3.3), as well as what features (sec. 5.3.3) and resources (sec. 5.3.3) are required to generate peer recommendations using our system.

### Implicit feedback

We processed all historical CaringBridge author interactions since January 1st, 2010. We train our recommendation models on *implicit feedback*—the behavioral signals that indicate an author is interested in reading a site or interacting with the author of that site [254]. Three potential sources of implicit feedback are available: (1) *first visits*—when an author visits another author's site for the first time, (2) *initiations*—when an author interacts with a site for the first time, (3) *reciprocations*—when an author initiates with a site and an author of that site subsequently interacts on the initiator's site, or (4) *follows*—for example as used by Twitter as an implicit signal for some user recommendations [252]. Due to the unusual nature of Follows on CaringBridge (see sec. 5.3.1) and lack of clarity around this feature's usage, we omit it from further consideration. Choice of implicit feedback signal can have a significant impact both on what recommendations are shown to users and how those users engage with those recommendations [255]. We choose initiations as our implicit feedback signal, as it is less noisy than site visits (i.e. a visit may not indicate interest) while more plentiful than reciprocations (i.e. reciprocations are less common than unreciprocated initiations [10], which means less data available for model training). By selecting initiations, we assume that leaving a reaction, comment, or guestbook on another author's site indicates a preference for reading that site and interacting with that author relative to other sites. Optimizing for recommendations that increase initiations is likely to increase actual initiations [256], but we acknowledge a semantic gap between the implicit feedback metric and metrics of interest; a construct like "perceived social support" may not increase despite receiving more peer interaction [234, 257].

For each initiation between a source author and a target site, we generate training data. Using initiations presents three complexities: (1) a user may initiate with a site before they become an author, (2) authors may write multiple blogs, and (3) authors may co-write with other authors on a single blog. We address these complexities by ranking *author/site pairs* rather than sites alone. For each initiation, we generate one positive sample for each author on the target site (target author/site pairs) *and* each site written by the source author (source author/site pairs). An author/site pair is *eligible* if the author has published at least three Journal updates on that site—a minimum activity threshold that ensures sufficient data is available during feature extraction.

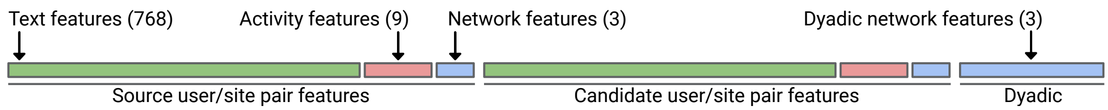



**Figure 5.4: Features available for recommendation, including 780 features for the recommendation-seeking *source* author/site pair and the same for each *candidate* author/site pair. Three dyadic features capture the relationship between the source and candidate within the initiation network. Total features: 1563**

We include in the training data only initiations between eligible author/site pairs. Eligibility is required at the time of the initiation; if an initiating author has not (yet) published three Journal updates on a site, no training samples will be produced for that initiation. At prediction time, given a recommendation-seeking source author, we score all *candidate* author/site pairs: eligible author/site pairs that has been active on CaringBridge in the last week and have not previously been interacted with by the source author. An active author is one that has created a Journal update, comment, guestbook, or reaction on any site within the last week. By focusing only on recommending active authors, we ensure that a recommendation-seeking author could receive a response from a site's author if they leave a comment.

As we ultimately recommend sites to authors, we convert the ranking of candidate author/site pairs into a ranking of sites. If a site appears among candidate pairs twice (because that site has two authors), we remove all but the highest score from that site. If a recommendation-seeking author is an author of multiple sites, we merge the scores for each source author/site pair by averaging the scores for each candidate pair. These merging strategies reflect the intuition that a good match with any one author of a site makes the site relevant. In practice, these strategies have minimal impact on rankings. For each positive sample generated by initiation extraction, we sample an assumed-negative author/site pair to add to the training data. While more complex approaches to negative sampling have been recently proposed [258], we adopt uniform sampling of a single negative as a widely-used baseline. The negative author/site pair is randomly selected from among the candidates. As a history of previous initiations are not required for eligibility, we avoid selection bias issues that arise in models that require user feedback on negative samples [259].

### Model features

Given two author/site pairs, we use the recent behavior of the authors on CaringBridge to construct a feature representation of the two, summarized in Figure 5.4. Different models might use subset of these features, but we describe here the full set included in the model deployed during the field study.

- **Text** ($768 \times 2$ features): We incorporate context from the three most recent Journal updates written on a site. For each Journal update, we use the pre-trained RoBERTa [260] model available in the HuggingFace Transformers package [261] to compute size-768 contextualized word embeddings. Then, we mean pool the token embeddings and then the update embeddings to produce a single vector representation, an effective general approach to using word embeddings [262, 263, 264]. Analysis during model development suggested minimal impact of the pooling strategy (mean vs max vs concatenation) on model performance.

- **Activity** ($9 \times 2$ features): For each of Journal updates, reactions, comments, and guestbooks, we include the count of that action within the last week and the time elapsed to the most recent action (in hours). In addition, we include the author's current tenure (time elapsed to the author's first published journal update).

- **Network** ($6 \times 2 + 3$ features): During initiation extraction, we maintain the interaction network between authors as described by Levonian et al., in which new edges are created between authors when an initiation occurs [10]. For each author/site pair, we include: indegree, outdegree, and weakly-connected component size. In addition, we include three dyadic features: whether the two authors are weakly connected, whether the candidate author is the friend-of-a-friend of the source author (i.e. this initiation would create triadic closure), and whether the candidate author has previously initiated with the source author (i.e. this initiation would be a reciprocation to a prior initiation).

### Offline evaluation

Given an initiation, the goal for our recommender system is to rank the site that was actually initiated with as high as possible: ideally in rank 1, above other candidates. We evaluate various modeling approaches by splitting initiations chronologically into a training, validation,

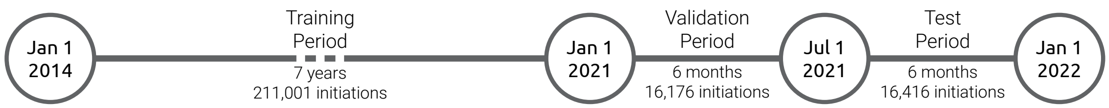



**Figure 5.5: Dataset splits and associated author initiation totals. All reported metrics are from the test period.**

and test set—see Figure 5.5 for initiation counts. Offline evaluations can diverge substantially from the usage preferences expressed by recommendation consumers [265]; we conduct one here to compare recommendation algorithms in terms of accuracy, coverage, and diversity of predictions on historical initiations.

During evaluation of each initiation, features are captured at the millisecond before the initiation actually occurred: for the source author/site pair (who did the initiation), for the target (who is being initiated with), and for the candidates (who are eligible, active author/site pairs that the source had not previously initiated with). This evaluation approach intends to reward models that rank a site well if the source author/site pair actually initiated with that site at the time the recommendation was made. The number of eligible, active candidates varies according to the time of the initiation; the median number of eligible, active authors during the test period was 13,252.

We used two primary evaluation metrics: mean reciprocal rank (MRR) and hit rate (HR). Mean reciprocal rank is computed based on the rank $r$ assigned to the target site, where 1 is the best rank given to the highest score. We punish ties by using competition ranking e.g. two sites scored the same both get assigned rank 2 and no site is assigned rank 1. At prediction time, we break ties randomly. MRR is the mean of the reciprocal ranks $(1/r)$ for every test initiation. Hit rate is the proportion of the time that $r$ is less than some threshold. As we provide 5 recommendations in a Site Suggestion email, we report HR@5 (how often would the model recommend the target site in a five-site set) and HR@1 (how often would the model rank the target site first). When we compare multiple hyperparameter configurations for a model, we select the best on the basis of validation MRR and report test metrics as the median from 3 random seeds.

Beyond accuracy-style metrics, we consider *coverage* as an important secondary goal [266]. Coverage has two aspects: (1) how many users can receive recommendations or have their sites recommended and (2) the diversity of sites that are recommended in practice. The first aspect of coverage is model agnostic and heavily affected by our decision to require three Journal updates to be eligible. Of 124,051 new authors in 2020, only 51.4% will publish three updates. Further, of those authors that do publish 3 updates, recommendations cannot be delivered until the publication of the 3rd update—a median wait of 2.5 days, but at least 72.4 days for the slowest-publishing 10% of authors. Beyond authors, a hypothetical system that served recommendations to any visitor with an interaction could reach an additional 1.3M new visitors who first interacted in 2020.

The second aspect of coverage—the diversity of recommended sites—is model dependent. To evaluate the coverage of recommended sites, we randomly sampled 1000 eligible, active authors at the end of the training period and produced recommendation sets with the 5 highest-scoring sites for each author. This approach emulates daily model retrains, a common approach to recommendation that we use in the field study. We then compared the sites that were actually recommended (R) to the sites that were not recommended (N), among the 12,432 candidate sites available at noon UTC on January 1, 2021. We consider three plausible goals for diverse recommendations, with an associated metric for each goal. The first goal is that a large number of unique sites are recommended rather than a few sites recommended many times; we compute the percentage of unique recommendations given ($|\mathrm{R}|/5000$). The second goal is that newer sites are recommended, a period during which support may be particularly impactful [267]; we compute the minimum site tenure for each recommendation set and report the mean. The third goal is that authors without previous interactions from other authors are recommended; we compute the proportion of recommended sites that have no prior connections ($|\mathrm{S_R}|/|\mathrm{R}|$ for "siloed" sites $\mathrm{S_R}$) and report the ratio to the proportion for non-recommended sites ($|\mathrm{S_N}|/|\mathrm{N}|$). If the ratio is less than 1, that model recommends fewer siloed sites than would be expected by chance. Existing initiations are with siloed sites only 2.8% of the time during the test period.

### Model comparison

We describe the recommendation models we implemented and compared, starting with the model we used during the field study. Table 5.3 presents the performance of these models in the order we describe them.

**MLP.** We used a multi-layer perceptron (MLP) with two hidden layers as a parsimonious

**Table 5.3: Offline test performance for various accuracy and coverage metrics. Coverage is reported in terms of number of unique recommended sites ($|\mathbf{R}|$), the percentage of unique recommendations made (out of 5000 total recommendations), and the mean of the minimum site tenure in each rec set (MMST). The final column displays the percentage of recced sites that have no prior connections (i.e. are "siloed", $|\mathbf{S_R}|/|\mathbf{R}|$) and the same percentage for non-recced sites. The bolded model was used during the field study.**

| Network | MRR | HR@1 | HR@5 | $\|R\|$ | %Unique | MMST | $\frac{\|S_R\|/\|R\|}{\|S_U\|/\|U\|}$ |
|---|---|---|---|---|---|---|---|
| $\text{MLP}_{\text{Tuned}}$ | 0.173 | 13.37% | 20.27% | 665 | 13.3% | 6.2 weeks | 11% / 25% |
| $\mathbf{MLP_{Study}}$ | 0.163 | 12.76% | 19.00% | 735 | 14.7% | 5.6 weeks | 11% / 25% |
| PeopleYouKnow | 0.102 | 7.86% | 13.07% | 3972 | 79.4% | 29.2 weeks | 15% / 28% |
| CosSim | 0.002 | 0.05% | 0.25% | 3403 | 68.1% | 27.1 weeks | 25% / 23% |
| MostInits | 0.035 | 1.32% | 4.90% | 12 | 0.2% | 113 weeks | 0% / 24% |
| Random | 0.001 | ¡0.0% | 0.04% | 4159 | 83.2% | 17.6 weeks | 24% / 24% |

yet effective deep recommender model. All source, candidate, and dyadic features are concatenated into a single input vector. The output layer uses a sigmoid activation to score the inputs and we optimize the standard pointwise binary cross-entropy loss [154]. We trained the model for 1000 epochs over the full training data, holding out a random 1% of the training initiations to compute hold-out loss. Further modeling and optimization details can be found in Appendix B.2. We report results for two MLP models: $\text{MLP}_{\text{Study}}$ and $\text{MLP}_{\text{Tuned}}$. $\text{MLP}_{\text{Study}}$ was the result of manual tuning before the study, and best reflects the model configuration used during the field study. $\text{MLP}_{\text{Tuned}}$ was the result of more systematic hyperparameter tuning (App. B.2) and adds additional hidden units, weight decay, and dropout.

We compare the MLP model to two personalized baselines. **PeopleYouKnow** uses only the dyadic network features, scoring highest the sites on which authors have already initiated with the source, then sites that are friends-of-friends with the source, then sites that are in the same network component as the source, and finally all other sites. **CosSim** scores author/site pairs by computing the cosine similarity between source and candidate. Cosine similarity was used by Hartzler et al. to match peers by health interest [139].

We include two non-personalized baselines. **MostInits** scores each site by the number of initiations received by that site within the last week—following the recommendation of Ji et al. to use a popularity baseline that takes into account the time point when a user interacts with the system [268]. In other recommendation contexts, popularity baselines like MostInits are strong contenders; not so on CaringBridge, where popularity follows a flatter distribution.

**Table 5.4: Offline performance of an MLP model trained with combinations of the activity (A), network (N) and text (T) feature sets. The model using all feature sets (A+N+T) is $\mathrm{MLP}_{\mathrm{Tuned}}$ from Table 5.3.**

| Feature sets | MRR | HR@1 | HR@5 | $\lvert R \rvert$ | % Unique | Site Age | $\frac{\lvert S_R \rvert / \lvert R \rvert}{\lvert S_U \rvert / \lvert U \rvert}$ |
|---|---|---|---|---|---|---|---|
| A+N | 0.204 | 16.12% | 23.62% | 598 | 12.0% | 0.9 weeks | 12.0% / 24.4% |
| A+N+T | 0.173 | 13.37% | 20.27% | 665 | 13.3% | 6.2 weeks | 10.8% / 24.6% |
| N+T | 0.144 | 11.75% | 16.58% | 803 | 16.1% | 5.4 weeks | 12.5% / 24.6% |
| N (Network) | 0.136 | 11.70% | 15.42% | 842 | 16.8% | 56.4 weeks | 11.6% / 24.7% |
| A (Activity) | 0.058 | 3.23% | 6.99% | 11 | 0.2% | 0.8 weeks | 9.1% / 23.9% |
| A+T | 0.043 | 1.73% | 5.25% | 181 | 3.6% | 0.8 weeks | 16.0% / 24.0% |
| T (Text) | 0.017 | 0.58% | 1.85% | 98 | 2.0% | 0.6 weeks | 13.3% / 23.9% |

MostInits has the best MRR of several plausible non-personalized activity baselines; others are discussed in Appendix B.2. **Random** ranks sites randomly, and is included as a useful point of comparison.

Table 5.3 compares the models by accuracy and coverage metrics. The best model ($\mathrm{MLP}_{\mathrm{Tuned}}$) would recommend the site that was actually initiated with more than 20% of the time; the study model ($\mathrm{MLP}_{\mathrm{Study}}$) performs slightly worse. Both MLP models are more likely to recommend newer sites than the baseline models, but less likely to recommend a variety of sites and sites that had not previously received an initiation. PeopleYouKnow achieves impressive MRR and HR metrics, demonstrating the strong utility of the dyadic features; however, when the source and target are not connected, it predicts randomly (hence the high percentage of unique recommendations). CosSim recommends the largest number of siloed candidates, although low MRR indicates that similarity alone is a poor predictor of historical initiation behavior.

### Feature ablations

Providing personalized recommendations requires some form of sensitive data collection from users [253], but privacy and other ethical concerns necessitate collecting as little sensitive data as is required to produce useful recommendations (or in some cases opting not to provide recommendations at all). Pervasive data collection contributes to perceptions of recommendation services as invasive “little brothers” [160], changing behavior and undermining the potential benefits of recommendation. Therefore, we compared the relative importance of the different

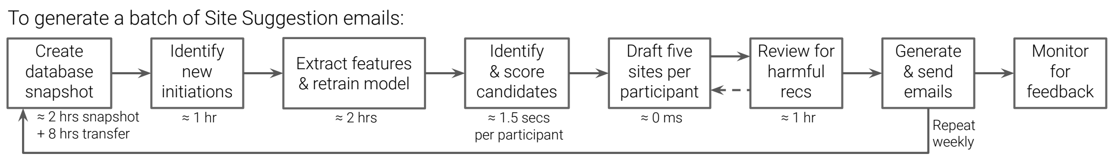



**Figure 5.6: Site Suggestion email generation during the field study. Recommendations were typically sent around noon, 36 hours after the database snapshot was taken.**

feature sets in order to understand the utility of collecting particular types of sensitive data—results are shown in Table 5.4. Surprisingly, the model trained without text features performs better on all accuracy metrics and similarly in terms of coverage, and it is also more likely to recommend newer sites. This analysis suggests that the collection of textual data may not be necessary and peer recommendation may still be practical if interaction network data is available: reasonable recommendations could be generated purely based on usage metadata. We still chose to include text features in the model we evaluated during the field study, as otherwise new authors with minimal activity and no peer interactions would all receive the same recommendations, conflicting with our design goal of personalization even in the cold-start setting. We discuss these results and this decision further in Appendix B.2.3.

### Model deployment & required resources

Figure 5.6 shows the steps required to generate a batch of Site Suggestion emails. The most time-intensive step was the anonymization and manual transfer of a nightly database snapshot, which led to a 36-hour lag time between the snapshot and the associated email batch. This 36-hour lag time could mean that the Journal update previewed in the email was no longer the most recent when the emails were sent, but we deemed this lag acceptable as it gave us time to do weekly robustness checks on the trained model and the identified recommendations. We created the Site Suggestion emails by retraining the recommender model on the most recent snapshot and scoring all eligible, active candidate sites. Rather than recommending the top 5 sites by score, we decided to limit the total number of times a site was recommended in a single batch to at most 10 times. We made this decision for two reasons: first, as the perception of stranger visits by site authors is unknown, a large influx of strangers might create a negative experience for the recommended site's authors; second, this restriction ensures that a diversity of sites are recommended each week. To achieve this limit, we randomly drafted sites by score until each participant had five recommendations. In the first batch, we applied the limiting only to the bottom quarter of recommended sites by total visit count. See details of the drafting process and an analysis of the small impact on recommendation quality in Appendix B.5. For ethical reasons, the first author manually inspected all recommended sites, manually excluding sites that would be inappropriate to send to participants. Two sites were removed this way: one for spam and one for COVID-19-related health misinformation. This small percentage of potentially harmful recommendations suggests that after some initial validation of quality, individual recommendations need not be inspected in practice.

For future studies, the actual resource requirements could be reduced by computing recommendations in-house. Disk usage for the weekly snapshot and the feature databases was approximately 96GB. The most RAM-intensive processing is maintaining the author interaction network during feature extraction; all other processing tasks are parallelizable and generally IO-bound. The CPU-based training and inference of the recommendation model could be accelerated with the use of GPUs, although at 1.5s per participant we found inference to be fast enough to serve even a much larger participant population.

## 5.4 Methods

To evaluate our recommendation system, we conducted a field study. The previous section summarized aspects of feasibility related to the system design, while the next section summarizes aspects of feasibility related to the field study. Table 5.5 summarizes these two phases and the associated components, alongside the relevant feasibility aspect. We discuss the assembled evidence by feasibility area in the Discussion (sec. 5.6).

The field study consisted of recruiting CaringBridge authors and sending them 11 weekly Site Suggestion emails. The full analysis timeline is shown in Figure 5.7. We analyzed field study data using both qualitative and quantitative methods, discussed in subsequent sections. System implementation and analysis code are available on GitHub.[2] Analysis code makes primary use of Python's scikit-learn [269], statsmodels [270], transformers [261], NumPy [271], pandas [272], and Matplotlib [273] packages. The recommendation model was trained using PyTorch [274].

[2] `https://github.com/levon003/HealthBlogRec`

**Table 5.5: Study phases and corresponding outputs. Each output provides evidence toward one or more aspects of feasibility.**

| Phase | Component | Analysis output | Feasibility Aspect | Section |
|---|---|---|---|---|
| System Design | Interface | Observed prior use | Demand | 5.3.1 |
| | | CB-adapted rec intervention | Implementation | 5.3.2 |
| | Model | Offline evaluation | Practicality + Impl. | 5.3.3 |
| | | Feature ablations | Practicality | 5.3.3 |
| Field Study | Survey | Self-reported prior use | Demand | 5.5.1 |
| | | Interest & motivations | Demand | 5.5.1 |
| | | Peer characteristics | Acceptability | 5.5.1 |
| | Rec Email | Click rate | Demand | 5.5.2 |
| | | Explicit feedback | Acceptability | 5.5.2 |
| | | Rec characteristics | Practicality | 5.5.2 |
| | Website | Reading behavior | Efficacy + Acceptability | 5.5.3 |
| | | Interaction behavior | Efficacy + Acceptability | 5.5.3 |
| | | Second-order effect estimates | Efficacy | 5.5.3 |
| | | Effect size estimates | Efficacy | 5.5.3 |

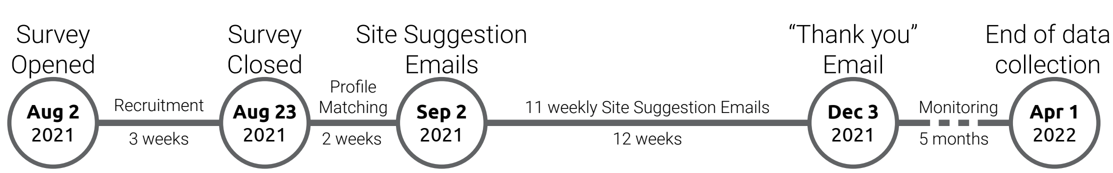



**Figure 5.7: Study timeline**

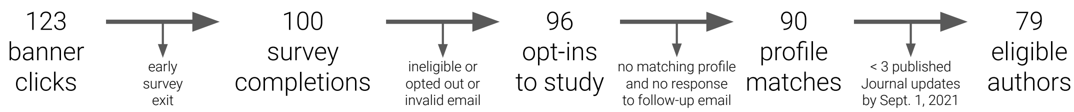



**Figure 5.8: Recruitment pipeline.**

### 5.4.1 Recruitment Survey

We recruited active CaringBridge authors by displaying a banner ad with a link to an opt-in survey. The banner appeared only to logged-in CaringBridge users on the home page of sites on which they are an author, as shown in Figure 5.1a. The recruitment banner being visible only on the home page of an authored site means that it will not be seen by users who authored a site in the past but are not regularly visiting their site(s). However, this recruitment method also excludes active authors who (a) visit a site sub-page like the Journal page (Fig. 5.1b) directly e.g. via a bookmark or (b) access CaringBridge using the mobile app. The recruitment banner and opt-in survey were active for three weeks in August 2021. The survey asked authors to opt-in to the study and included three optional questions on peer connection: on prior use of CaringBridge for connecting with strangers, on motivations for peer connection, and on characteristics that make a peer connection appealing. Full survey text is available in Appendix B.3.1.

#### Participant matching

Figure 5.8 shows the recruitment pipeline. Of the 100 survey completions, 96 opted-in and met the three study consent criteria: being at least 18 years old, being a current CaringBridge author, and consenting to provide the email address associated with their CaringBridge account in order to receive Site Suggestion emails.

We matched survey responses to a specific CaringBridge account based on their provided email. For the 8 cases where we couldn't find an associated account, we sent a follow-up email asking for profile information, matching 2 additional profiles. Finally, we excluded 11 participants who had published less than 3 Journal updates by September 1, 2021 (see sec. 5.3.3 for discussion of the minimum update requirement). Ultimately, 79 participants were sent Site Suggestion emails.

**Table 5.6: Pre-study CaringBridge usage by participants enrolled in the field study and by a pseudo-control group of eligible non-enrolled authors. Author tenure is the number of days between a user's first published Journal update and Sept. 1, 2021. Density histograms indicate distribution shape for the participants (black) and the pseudo-control group (gray). *Indicates a significant difference at the 99.5% threshold, for a Welch's $t$-test on the mean difference and for a Mann-Whitney $U$ test (reported as the common language effect size aka ROC AUC).**

| | Participants ($n_{\mathrm{P}}$=79) | | Pseudo-Control ($n_{\mathrm{C}}$=1759) | | | |
|---|---|---|---|---|---|---|
| | Med. | M (SD) | Med. | M (SD) | $\mathrm{M_P}$ - $\mathrm{M_C}$ | $U_{\mathrm{P}}/(n_{\mathrm{P}}n_{\mathrm{C}})$ |
| Author tenure | 179 | 709 (1147) | 3894 | 4079 (983) | -3369* | 3.2%* |
|  | | | | | | |
| Journal updates | 28 | 98.6 (262.4) | 77 | 167.1 (307.7) | -68.5 | 31.1%* |
|  | | | | | | |
| Peer site visits | 3 | 6.5 (12.3) | 10 | 32.0 (84.0) | -25.5* | 23.9%* |
|  | | | | | | |
| Peer site inits. | 1 | 2.5 (4.4) | 5 | 11.5 (35.2) | -9.0* | 23.8%* |
|  | | | | | | |
| Peer site ints. | 2 | 39.7 (94.4) | 30 | 201.2 (1384.5) | -161.5* | 29.0%* |
|  | | | | | | |

### Observed prior use

The median participant had been writing Journal updates on CaringBridge for fewer than 6 months at the time of enrollment. But, consistent with observations by Levonian et al. [10], a majority of the 79 participants both visited and interacted with at least one fellow author's site. This evidence suggests a demand for peer connection, although the field study extends beyond known contacts to peer strangers. To quantify the differences between authors that chose to enroll in the study and other CaringBridge authors, we identified 30K authors who visited their own site while the banner survey was live and thus could have seen the banner, filtering down to 1,759 who had at least 3 Journal updates and thus could have received Site Suggestion emails. Table 5.6 compares participants to this set of non-enrolled authors, which we term a *pseudo-control* group, revealing that participants are generally newer to CaringBridge than other eligible authors. While the field study uses an uncontrolled design, we will use this pseudo-control group to estimate the potential impact of the peer recommendation intervention on author behavior.

### 5.4.2 Recommendation Emails

After we went an initial Site Suggestion email on September 2, 2021, we conducted an initial assessment of interest and determined the study could proceed; we resumed sending emails on September 17, 2021 at a weekly pace. After 11 Site Suggestion emails, a "thank you" email was sent with a final request for feedback. To evaluate the effectiveness of the Site Suggestion emails as a recommendation interface, we analyzed the rate of clicks on recommendations, the explicit feedback we received from participants, and the characteristics of the site previews we included in the emails.

Click estimates are based on the fusion of multiple data sources, although the primary source was via UTM tags embedded in each email's links. Appendix B.4 presents additional details. Explicit participant feedback was collected from four sources: direct responses to the emails, responses to a feedback survey linked in each email, an unsubscription survey linked in each email, and responses to the final "thank you" email sent on December 3, 2021. The Site Suggestion email was previously described in sec. 5.3.2. The survey texts are presented in Appendix B.3.2 and B.3.3. The "thank you" email was sent in two versions. To participants that had clicked on none of the recommendations, we asked for feedback on whether they had seen the emails and why they chose not to visit any of the recommended sites. To participants that had clicked on at least one recommendation, we listed up to three random sites they clicked and asked for reflections.

The primary information available to participants while they were deciding to click on a recommendation was the content of the recent Journal update preview. To capture the preview characteristics that our participants could see and respond to, we conducted a thematic content analysis on the text of these previews. Two researchers generated open and axial codes independently, then used an affinity mapping process based on the Grounded Theory method [275]. Based on our thematic analysis and on existing work with CaringBridge Journal updates [70, 8], we isolated a set of 6 categories in order to produce quantitative prevalence estimates for each of the categories and to identify which categories were associated with clicks—further details in Appendix B.6.

### 5.4.3 Observed behavior on CaringBridge

#### Reading and interaction behavior

To analyze the two primary behavioral outcomes—reading and interaction behavior—we identified measured proxies as indicators. We used repeated visits and site Follows as proxies for interest in reading a site. While a single site visit may still indicate value for the reader, we would need either explicit feedback or some measure of dwell time on the site. Thus, we focus on instances where a participant returns to a site at least once. Follow actions are harder to interpret (see 5.3.1), but at a minimum indicate an interest in continuing to read updates on the followed site. We use reactions, comments, and guestbooks left on a recommended site as evidence of interaction.

We analyzed the textual content of the comments and guestbooks created by participants. Three researchers conducted a qualitative analysis of the participant/site dyads where interaction occurred, writing axial codes and memos based on the Journal updates and comments in which participants and authors of recced sites interacted [275].

#### Second-order effects of recommendations on behavior

In addition to the reading and interaction behavior of participants, receiving Site Suggestion emails and visiting strangers' CaringBridge sites might have second-order effects: harms or benefits that accrue to both participants and the authors of the sites they visit. While we cannot draw firm causal conclusions from our non-experimental study, we can check for potential harms by estimating the effect of recommendations on two secondary behavior outcomes: (a) publishing Journal updates—the "primary" use of CaringBridge by authors—and (b) visits to and interactions with fellow authors' sites—a hypothesized outcome of peer interaction. We estimate the effects of two "treatments": for authors, we estimate the *participation effect* of receiving 11 weeks of Site Suggestion emails by comparing participants to the pseudo-control group of eligible unenrolled authors (introduced in sec. 5.4.1); for sites, we estimate the *visit effect* of receiving site visits from peer strangers by comparing visited recommended sites to both *non-visited* recommended sites and a pseudo-control group of *non-recommended* sites. The non-recommended pseudo-control sites are the five highest-scoring sites for each participant and each batch after removing the sites we actually recommended—additional details and comparison to recommended sites in Appendix B.5.

We quantify the difference between the "treated" authors/sites and the untreated authors/sites, producing three estimates: associational, model-adjusted, and causal. The associational difference for a behavioral outcome $Y$ is $\mathrm{E}[Y|T=1] - \mathrm{E}[Y|T=0]$; the raw observed difference in the mean between treated and untreated authors/sites, where $\mathrm{E}[Y|T=1]$ is the mean outcome for the treated group i.e. participant authors or visited sites. The model-adjusted estimate uses linear regression (OLS) to adjust for activity variables $A$, adding assumptions about model misspecification to compute the quantity $\mathrm{E}[Y|T=1,A] - \mathrm{E}[Y|T=0,A]$. The causal estimate requires us to make untestable assumptions about the modeled relationship between the treated and untreated groups [276]—see Appendix B.7 for a detailed discussion of these assumptions. Using potential outcomes notation, we define $\mathrm{E}[Y^{t=1}]$ as the mean behavioral outcome that would have been observed if all authors in the participant and pseudo-control group had been participants in the study. The true causal effect $\mathrm{E}[Y^{t=1}] - \mathrm{E}[Y^{t=0}]$ is the influence on $Y$ that would occur if Site Suggestion emails were sent to all eligible authors. We use the Bang-Robins doubly robust estimator to compute the causal effect, a modeling approach which combines inverse probability weighting and standardization [277] (see also [276], Ch. 13).

The author and site outcomes are measured as number of actions in the 13 weeks post-study and post-visit respectively. As the non-visited comparison sites were not visited by participants, we fabricate a visit time by sampling a random visit time from among the sites that *were* visited in that same batch. The activity variables $A$ are computed based on an equivalent time window before the event of interest—13 weeks before the study for participants and 5 weeks before the site visit respectively. (We include only 5 weeks of activity context due to a lack of available repeat visit data before August 2021.) The analysis is not sensitive to the time window over which the pre-study features and post-study outcomes were measured (see Appendix B.7). The associational and causal estimates require specification of a model that includes all relevant confounds. For example, as we saw in Table 5.6, pseudo-control authors have been active on CaringBridge longer than participants. Activity variables used are shown in Appendix Table B.8, although we cannot measure important confounds such as "interest in receiving Site Suggestion emails", which is unobserved and potentially unexplained by the activity variables we do observe.

#### Sample sizes needed for a powered RCT

We use observed participant behavior during our field study to estimate the effect sizes of the peer recommendation intervention's impact on reading and interaction behavior. The structure of our data let us consider two potential future interventions: a one-time recommendation email and a recurring, weekly recommendation email. We estimate the effects of a one-time recommendation email by including only visits and interactions resulting from the first batch of recommendation emails. We compute sample sizes for both a replication (no control group) and an RCT (with control group) from the estimated standardized effect sizes at 80% power with $\alpha$=0.05 using G*Power's one-tailed point biserial model [278]. Additional details are available in Appendix B.8.

## 5.5 Results

We recruited authors via a survey (sec. 5.5.1). Then, we sent 79 participants weekly emails (sec. 5.5.2). Finally, we observed the subsequent usage of CaringBridge by both the participants and the authors of recommended sites (sec. 5.5.3).

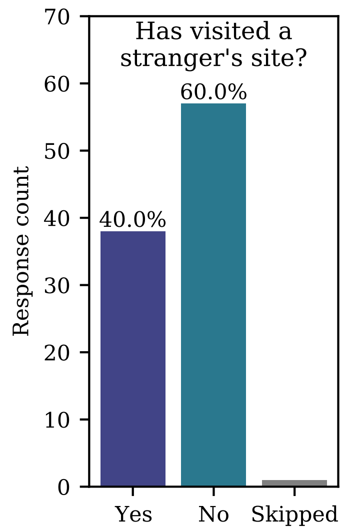
### 5.5.1 Recruitment Survey

Quantitative survey responses are summarized in Table 5.7.

#### Self-reported prior use

40% of participants said they visited the CaringBridge site of an author who they did not know personally. If we assume that participants visited CaringBridge sites with the account they used to fill the survey, we can compare self-reported use to actual use: 76% of participants had made a logged-in visit to a CaringBridge they do not author—considering only the 79 participants we could link to an existing CaringBridge account (see sec. 5.4.1). Thus, among authors who have visited one or more CaringBridge sites, 65% report visiting the site of a stranger. This prior usage provides evidence that the visits and interactions among peers includes strangers, although most interactions are likely between people who already know each other offline [10].

**Table 5.7: Recruitment survey responses.**




| | # checked | % |
|---|---|---|
| **Motivations** | | |
| Learn from others | 75 | 79.8% |
| Communicate with peers | 43 | 45.7% |
| Receive experienced support | 41 | 43.6% |
| Mentor newer authors | 27 | 28.7% |
| Not interested, but maybe in future | 6 | 6.4% |
| Never interested | 2 | 2.1% |
| Not interested, but maybe in past | 0 | 0.0% |
| Something else | 8 | 8.5% |
| **Characteristics** | | |
| Similar diagnosis or symptoms | 79 | 84.0% |
| Similar treatment | 51 | 54.3% |
| High-quality writing | 48 | 51.1% |
| Same caregiver relationship | 30 | 31.9% |
| Lives near me | 23 | 24.5% |
| Similar cultural background | 13 | 13.8% |
| Something else | 8 | 8.5% |

### Interest & motivations

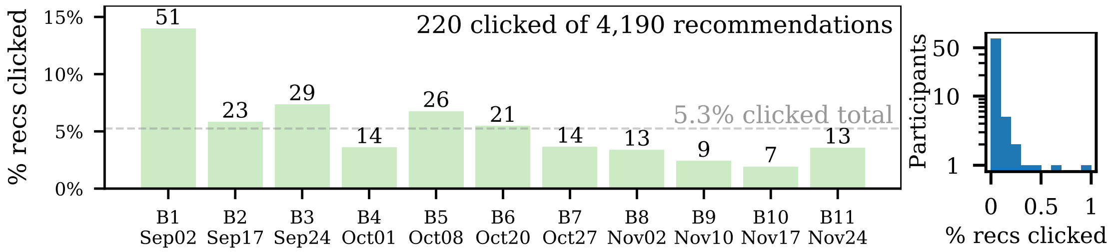
We asked participants why they might visit a fellow author's CaringBridge site, pre-populating a variety of options based on prior work. The top motivation was to learn from others (80%), then to communicate with peers (46%) and to receive experienced support (44%). A smaller percentage (29%) were motivated by mentoring newer authors. 8 respondents (9%) indicated they were not interested in visiting stranger's sites, although 6 suggested they could be interested in the future. Free response motivations centered on learning from others, including finding inspiration, hearing how others handled similar issues, and learning "ideas on how to engage readers" of their own site. *"I would like to see how other CaringBridge authors articulate their reactions and feelings as they undergo the medical treatments for and rehabilitation from whatever medical conditions they experience."* Two free responses described motivations for not engaging with others' sites, both describing it as a distractor during a busy time. *"it's hard to think about joining in someone else's journey. I can focus on writing this blog discussing our journey because it is letting friends and family know what is happening so it is narrow enough* *that it doesn't take away from the other things needed to be accomplished during the rest of the day."* These responses indicate that demand for peer connection is goal-driven and subject to constraints.

#### Peer characteristics

We asked participants what characteristics of peer sites would make them want to read and engage with that site. The most selected characteristic was a similar diagnosis or symptoms (84%), a theme echoed in more specific free responses e.g. "neurological conditions" and a participant calling shared diagnosis the *most* important characteristic. A majority of respondents indicated that similar treatments (54%) and high-quality writing (51%) were important. Free responses indicated a desire for "inspirational", "honest" writing with specific details that shows "positive and negative aspects". One participant wanted to see *"multiple posts all the way through death. I wanted to see what I would probably be writing as time progressed. A glimpse into the future if you will."* Fewer respondents indicated that having the same caregiver relationship with the patient was important, although that checkbox was only relevant for non-patient respondents. 25% of respondents indicated that geography was important; one specified sharing the same hospital as a relevant characteristic. Few selected similar cultural background (14%): a divergence from prior work [140], although the question may have been phrased too euphemistically to elicit specific preferences about e.g. age, gender, and ethnicity. Three participants specifically identified age of the author or patient as a relevant detail. Two participants listed shared social context as important characteristics, such as already knowing the person or having "common friendships". One participant said they sought Spanish-language updates, while another said they were looking for sites authored by healthcare professionals. These responses provide evidence for the types of peer recommendations that are perceived as most useful and acceptable by CaringBridge users—as well as indicating that preferences are diverse, necessitating personalization.

### 5.5.2 Recommendation Emails

In total, we sent 4,190 recommendations to 79 participants, with 526 unique sites recommended. To evaluate the effectiveness of the Site Suggestion emails as a recommendation interface, we analyzed the rate of clicks on recommendations (sec. 5.5.2), the explicit feedback we received




**(a) Recommendation clicks by email batch**

**(b) By participant**

**Figure 5.9: Recommendation click counts and percentages by email batch and participant. Participants are shown grouped by decile; 49 participants never clicked a recommendation, while 2 clicked more than half.**

from participants (sec. 5.5.2), and the characteristics of the site representations we included in the emails (sec. 5.5.2).

### Click rate

Over the 11 weekly batches of Site Suggestion emails, participants clicked 220 (5.3%) of the 4,190 recommendations. Figure 5.9 breaks down observed clicks by email batch and by participant. Clicks are approximately log-normal; only 30 of 79 (38%) participants clicked any recommendations, and the most active participant, who we designate P1, was responsible for 54 (24.5%) of the recommendation clicks—and the majority of the interactions, as we will see in sec. 5.5.3. We would like to estimate click rate conditional on opening the email (i.e. the click-through rate) independently from the base click rate, but for logistical and ethical reasons we did not use email trackers in the email HTML so we assume every email was received and opened.

### Explicit feedback

8 participants provided explicit feedback on the recommendations in 13 responses. No participants provided rec feedback after batch 7, so responses reflect initial impressions. Table 5.8 summarizes the quantitative feedback received: only 4 of 10 responses found the recommendations interesting in general, while 46% of responses to specific recommendations were deemed relevant compared to 39% irrelevant. As the sample is small and the textual feedback is more informative, we conducted no further statistical analysis on rec relevance feedback. Five of the

**Table 5.8: Explicit quantitative feedback to 11 Site Suggestion emails from 8 total participants.**

| Recommendations interesting in general? | | Specific recommendations relevant? | |
|---|---|---|---|
| | | Very Relevant | 20 |
| Yes | 4 | Somewhat Relevant | 6 |
| Unsure/Neutral | 3 | Unsure/Neutral | 8 |
| No | 3 | Somewhat Irrelevant | 17 |
| | | Very Irrelevant/Offensive | 5 |

recommendations were deemed "very irrelevant or offensive", and associated text feedback can tell us why. Participants objected to sites due to having "choppy, poor grammar and name calling", describing the death of a patient (*"don't want to read about ppl dying of cancer"*), being "too religious", and describing patients of a very different age than the participant. Objections to Christian religious content may reflect belief misalignment between reader and author [85]; one participant requested the ability to filter sites *"by the number of times Christ is mentioned"*. Less severe objections focused on aspects of the writing: "boring", "wordy", "no reflection", "too much philosophizing", "bragging". One participant desired sites that "go beyond simply updating the reader", describing their own efforts as an author to inspire readers and to include "silver linings".

On the positive side, participants valued the "wide range of experiences" present in the recommended sites. One participant was surprised by recommendations for different medical conditions but describes realizing that *"what is important to me is to see and be inspired by how others deal with* any *difficult situation."* Two participants valued the recency of the displayed updates. *"I'm finding myself following a lot of these sites. I enjoy the sites that update you often ... I feel like I'm abreast of what's going on and take a personal interest into the person and their health battle."* Participants valued common ground—such as a patient being treated at the same hospital—and familiarity with issues faced. *"I was more interested in the ones I was kind of familiar with. Understood more."* One patient gave us a general summary of how they engaged with recommendations: *"Even with all I'm going through ... I've come to care about these people who sites I am following. And I leave a comment every time I login on 95% of them because I know what it's like when nobody comments. So I'm really really grateful you guys have [sent me site recommendations] because it is just helps me personally to take my mind off of things when I can go and pray for some other poor souls problems. So good job!"*

6 participants unsubscribed during the course of the study. 5 provided unsubscription motivations: 2 participants indicated a disinterest in receiving recommendations (*"I thought it was something else"*). 1 indicated that they were no longer using CaringBridge due to the death of their loved one. 2 indicated a lack of time to engage with the recommendations, with one adding in addition: *"it's kind of depressing. Since we're already going through cancer treatments, it's hard to look at what other people are going through."*

We received four replies to our final "thank you" email. Only one non-clicking participant replied, indicating they had seen a few of the Site Suggestion emails but focused on using CaringBridge to get support from their existing network. *"CaringBridge filled its purpose and functioned as expected. ... I received a lot of care and support from friends."* We received three replies from participants who clicked at least once, with all responses emphasizing the importance of some kind of common ground. One participant connected first with a site due to common ground—the patient and participant had previously lived in the same small suburb—and then kept following the site due to its use of poetry and song lyrics to process grief. One participant described reading others' sites as "helpful and informative", due both to feeling comforted seeing a similar person navigate treatment and to learning strategies when writing their own Journal updates. One participant decided not to follow some of the recommended sites because they had a lot of existing followers. *"Having a Caringbridge site I know what it's like when you don't have a lot of supporters. ... So I try to help by leaving comments 99% of the time on the sites I follow to try and help the caregivers stay positive and give them kudos for all they are handling."*

### Recommendation characteristics

Our thematic analysis identified three high-level themes. We report the first-batch prevalence of categories identified from these themes in Table 5.9—see further thematic description and quantitative summary in Appendix B.6.

**Disclosing health status, symptoms, and treatment.** Previews that address the question, "what or how is the patient doing?" Authors report on both recent heath updates and future plans or expectations. *"Friday is a huge milestone for me. My chemo port is going to be removed."* Disclosures vary from reporting on processes (*"Sally is getting better day by day"*), to discrete events (*"Bryan met with his neurosurgeon yesterday."*), to transitions (*"Kristen is finally off the*

**Table 5.9: Prevalence of identified content categories in the Journal update previews. Only the presence of expressive writing was significantly associated with clicks (47.1% of first-batch recommendations with expressive writing were clicked vs 28.7% without, $p = 0.017$).**

| Preview Category | First batch prevalence |
|---|---|
| Reporting health status | 85.2% |
| Neutral disclosures | 31.5% |
| Positive disclosures only | 25.8% |
| Negative disclosures only | 22.2% |
| Positive & negative disclosures | 5.8% |
| Expressive Writing | 31.2% |
| Managing Author/Audience Relationship | 17.5% |
| Expressions of Appreciation | 5.8% |

*ventilator.*").

**Communicating emotion and reflection.** Previews that express the author's attitudes or reflect on author or patient experiences. Emotions are either linked to specific experiences during the health journey (*"we are immensely happy to finally be home"*) or characterize the author's current mental state (*"I wish you could experience this wonderful euphoria I am feeling"*). Reflection is often used as the introduction to a narrative (*"It's been 12 weeks since Beth's accident. I cannot believe it's been that long..."*).

**Managing author/audience relationship.** Previews that engage with the reader. Includes requests, expressions of gratitude, and context about the update, author, or writing process. Requests acknowledge the reader as an active member in the patient's health journey. *"Please send all your prayers as I start my journey through chemotherapy. I am nervous, but with your help I know I can get through this."* Expressions of gratitude acknowledge received support from readers. Previews that provide additional context make the author's writing work visible to the reader e.g. by discussing update frequency. *"Sorry it's been a while since I last posted an update. I've been finding it hard to work up the energy to write, but I'll do my best to recap the last few weeks."*

### 5.5.3 Observed behavior on CaringBridge

We describe the observed impact of the recommendation intervention on behaviors: reading (sec. 5.5.3), interaction (sec. 5.5.3), and second-order effects on other behaviors (sec. 5.5.3).

**Table 5.10: Observed participant behavior in response to Site Suggestion emails, from the start of the study (September 2021) to the end of data collection (April 2022).**

| | Recommendations | | Participants | | Recced sites | |
|---|---|---|---|---|---|---|
| Behavior | $n$ | % (of 4190 total) | $n$ | % (of 79 total) | $n$ | % (of 526 total) |
| First Visits/Clicks | 220 | 5.5% | 30 | 38.0% | 158 | 30.0% |
| Second Visits | 86 | 2.1% | 17 | 21.5% | 76 | 14.4% |
| Repeat Visits | 589 | - | 17 | 21.5% | 76 | 14.4% |
| Follows | 24 | 0.1% | 5 | 6.3% | 23 | 4.4% |
| Initiations | 36 | 0.9% | 9 | 11.4% | 33 | 6.3% |
| Interactions | 948 | - | 9 | 11.4% | 33 | 6.3% |
| Text Interactions | 268 | - | 4 | 5.1% | 20 | 3.8% |
| Relationships | 1 | 0.0% | 1 | 1.3% | 1 | 0.2% |

Participant reading and interaction behaviors are the primary outcomes and are summarized in Table 5.10.

### Reading behavior

Table 5.10 presents counts for both repeat visits and site Follow actions. Follows were rare; only five participants ever followed a site. Repeat visits were more common: 86 (39.1%) of 220 clicked recommendations were visited a second time, and collectively accrued 589 repeat visits during and after the study.

### Interaction behavior

Nine participants interacted with recommended sites, which is 30% of those who visited at least one recommended site. Collectively, 33 sites received 948 additional interactions as a result of this study, although the majority of these accrue to just a few sites: median interactions with a single site was 6 (M=28.7; SD=56.9). Participants initiated with 38 non-recommended sites during the course of the study as well—likely with sites authored by people they already knew [10]. Thus, participation was associated with a 94.7% increase in total initiations during the study period. At an average of 0.37 initiations per author during the study period, participants also initiated more often than the 0.05 initiation average for non-participating active authors. For context, participant initiations comprised 0.2% of all the author initiations that occurred on CaringBridge during the study period, from among 39.5K active authors. 71.7% of the interactions were reactions. This proportion is similar to the percentage of reaction interactions for all users (71.6%) and all participant interactions with non-recommended sites (71.2%) during the same period. We focus on text-based interactions next.

Only four participants interacted using comments or guestbooks, resulting in 20 participant/site dyads in which interaction occurred. Our qualitative analysis identified two areas of interest: *first-contact strategies* and *potential norm violations*. We observed a diversity of first contact strategies, varying from formally introducing the self (*"I am managing a CaringBridge site for my sister. ... Your site was mentioned in an email from CaringBridge, I took interest in your story."*), to general expressions of support (*"Hope you get some rest soon"*), to establishing common ground by sharing their personal health experiences (*"I kind of know the road you are traveling. [personal health history]"*). A larger study could investigate first-contact strategies that are particularly effective, either at encouraging a response or at providing useful support. In addition, we observed "potential norm violations": situations where participant comments diverge from other visitor comments. These divergences include specific behaviors: first-time sympathetic comments on posts announcing the death of the patient, being the first and only commenter on a Journal update, bringing in potentially-unwanted religious messaging (*"God bless you"*), and asking explicit questions of the update author. The commenting norms perceived by "power user" peers may diverge from most visitors, enabling them to provide support in situations others may not, but running the risk of producing uniquely negative experiences. P1 provided the only comment on a death announcement update: *"I am so very, very sorry for the passing of your loved one."* We have no evidence of how these comments were received by authors of recced sites, with no specific objections either in reply to comments or to CaringBridge help/support lines.

The author of a recced site responded to a participant comment in only a single instance, and their subsequent interactions expanded into a relationship. While one relationship is insufficient for generalization, we sketch it here as a rich example. P1's first contact with R1 shared personal experiences and offered support. *"You don't know me but you sound like you're handling this all very well. .... I know what it's like to get chemo and I know what it's like after."* Within weeks, R1 left supportive comments on P1's recent Journal updates (*"I can't even begin to imagine how strong you are"*). Support exchange continued, with evidence that off-CaringBridge communication had been established. Two months into their relationship, R1 explicitly referenced P1 in a Journal update (*"I have a friend, I think she'd agree with that title.*

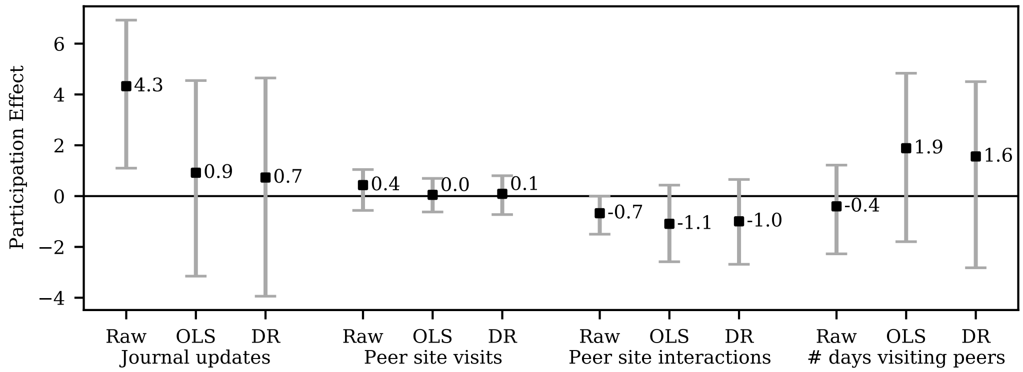



**Figure 5.10: Participation effect: Estimated impacts of receiving Site Suggestion emails on author behavior in the 90 days after the study. Raw estimate is the mean difference between the participants and the pseudo-control group of unenrolled but eligible authors. The OLS estimate adjusts for pre-study activity on CaringBridge while the doubly robust (DR) estimate adds additional causal identification assumptions. 95% confidence intervals are computed via bootstrapping (1000 iterations).**

*She somehow found me through Caring Bridge, and has been a constant support to me and my family since."*) As a frequent commenter, P1 occasionally generated further discussion (Visitor: *"I agree with [P1]."*) and on one occasion was thanked for their support by R1's relative. We take the creation of a new relationship as an important existence proof for the benefits of peer recommendation, although it is difficult to quantify the expected number of relationships per recommendation from this single example. P1 is much more active than other authors, being on the 97th percentile by number of Journal updates and the 99th percentile by number of initiations and interactions. Will authors that are less active still form peer relationships given the opportunity? We estimate the sample sizes needed for a larger trial in sec. 5.5.3.

### Second-order effects on participant and recommended site behavior

Estimates of the participation effect are shown in Figure 5.10. Additional outcomes for both participants and clicked sites are shown in Appendix B.7. While the associational effect of recommendation indicates that participants publish more Journal updates than non-participants, we suspect this is primarily due to participants' shorter tenure, and the effect disappears after adjusting for author tenure. We present a selection of outcomes related to visiting and interacting with others' sites (specifically excluding any sites recommended during the study); none of these estimates suggest a significant positive or negative effect at the 95% significance level.

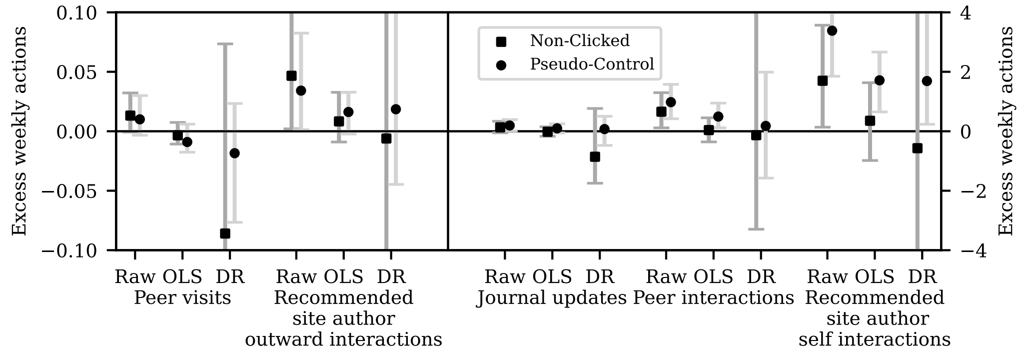



**Figure 5.11: Visit effect: Estimated impacts of receiving a stranger visit on recommended site behavior in the 90 days after the visit when compared to non-clicked recommendations (square, dark-gray) and pseudo-control recommendations (circle, light-gray). Peer visits and interactions count *non*-participant author behavior directed at the site. Recommended site author interactions count behavior of authors of the site. The OLS estimate adjusts for pre-study activity on CaringBridge while the doubly robust (DR) estimate adds additional causal identification assumptions. 95% confidence intervals are computed via bootstrapping (1000 iterations).**

We conclude that receiving recommendation emails had a small or high-variance effect on participant behavior—and no clear harms.

Estimates of the visit effect are shown in Figure 5.11. The impact of an individual visit should be small, so we expect and observe in practice small effect sizes. We observe that peer visits are linked to additional author interactions on their own site and on peer sites, although these effects are smaller after adjustment. Absent clear evidence of a harmful effect on associated behaviors, and given the successful primary behavior manipulation, we recommend proceeding to an RCT for the intervention.

### Sample sizes needed for a powered RCT

Figure 5.12 shows the sample sizes needed for an uncontrolled replication of our feasibility study, based on the observed visit and interaction behavior. If sending a one-time recommendation email, at least 314 authors should be included in order to obtain a reliable estimate of total peer interactions with (and visits to) recommended sites. Designing a trial to estimate the effect of recommendation on relationships is more challenging, as no relationships formed due to the first batch of Site Suggestion emails and only one relationship formed during the entire study. Based on that one relationship, a recurring email study would need 478 participants to

| **Sample sizes needed to detect beneficial effect** (at 80% power and 5% sig. threshold) | Unique repeat visits | Total Interactions |
|---|---|---|
| One-time email | 75 ($d$=0.28) | 314 ($d$=0.14) |
| Recurring email | 81 ($d$=0.26) | 444 ($d$=0.12) |

**Figure 5.12: Sample sizes needed to detect effects of the magnitude we observed during the feasibility study. Field study effect sizes are shown parenthetically as Cohen's $d$.**

detect at least 1 relationship per 79 participants ($d$=0.11, $\alpha = 0.05$, $\beta$=0.2), although 10 times that number of participants would be needed if the proportion of participants that form relationships is closer to 1 in 1000 than 1 in 100—and tens of thousands are needed if relationship formation is uniformly probable across participant/recommendation pairs (as only 1 in 4190 recommendations led to a relationship). We also estimated effect sizes relative to the pseudo-control group (see full results in Appendix B.8). Based on unadjusted effect size differences between those groups, a 300-participant recurring-email RCT—with 50% not receiving recommendation emails—should be sufficient to estimate the relative impact of recommendation on total peer visits and interactions. Assuming the same click rate we observed in this study (5.3%), such an RCT would also be sufficiently powered to investigate any potential negative impact of peer visits on recommended site update frequency.

## 5.6 Discussion: Feasibility of a peer recommendation intervention

In this chapter, we implemented a peer recommendation system in order to assess its feasibility as a behavior-change intervention. We collected evidence for feasibility in five areas (Table 5.1): Demand, Implementation, Practicality, Acceptability, and Efficacy. Here, we summarize the evidence in each feasibility area and the implications for the future development of peer recommendation interventions.

### 5.6.1 Demand

Demand refers to interest in the intervention. Specifically, we collected evidence around demand via prior use (sec. 5.3.1), expressed interest (sec. 5.5.1), and actual use (sec. 5.5.2). Prior use indicates both a large number of peer author interactions and correlations between interaction with peer authors and retention. Peer interaction includes interactions with peer strangers as well: 40% of participants reported previously visiting the site of a stranger. Participants indicated a motivation to connect with peers, although less interest in interaction specifically; 80% reported an interest in learning from the experiences of others, compared to 46% reporting an interest in communication "with a peer who understands". These expressions of interest were matched in practice with relatively high recommendation click rates—5% overall, 14% in the first batch of emails.

### 5.6.2 Implementation

Implementation refers to the tangible design and engineering required to implement the intervention. Specifically, we collected evidence around implementation requirements for both the recommendation interface (sec. 5.3.2) and the recommendation model (sec. 5.3.3). We used email as the recommendation medium to adapt peer recommendation to the CaringBridge context, due to the ubiquity of email notifications on the platform. However, a text-centric email interface may be inappropriate on platforms that emphasize non-text content or want to provide recommendations in "always-available" recommendation interfaces that may be more familiar to users. Future studies should adapt the interface design to the platform—integrating recommendations into OHC interfaces as appropriate—although we found that email recommendations were reasonable and accepted by participants. Representing peer profiles in an interface remains an important open question for future work [251]; text previews were effective, but their focus on recency is a trade-off compared to curated profiles or previews that attempt to highlight the expected utility of the recommended site to the viewer e.g. via explanations [114]. The model underlying the interface was based on implicit feedback from historical peer interactions. Future peer recommender systems should carefully consider the available implicit feedback and the implications of optimizing for historical patterns when new interaction patterns are desired. For example, learning from the experiences of others was the most common participant motivation, but learning was likely not the primary motivation for authors' historical interactions—collecting explicit feedback could supplement or replace this implicit feedback to better align the training objective and user motivations.

### 5.6.3 Practicality

Practicality refers to requirements for administering the intervention in practice. Specifically, we collected evidence of practicality by investigating model quality (sec. 5.3.3), required data (sec. 5.3.3), and compute time (sec. 5.3.3). We presented an offline evaluation based on historical data that enables model comparison. In our context, none of the non-personalized approaches we considered were effective at capturing historical peer interaction dynamics, but such approaches may still be appropriate in other contexts. For example, if only activity data is available, it may be reasonable to form peer cohorts based on sign-up time and activity level [279]. If interaction data is available, our results suggest that traditional interaction-based recommendation models may be effective at recreating historical patterns without the need for elaborate disclosures. The RoBERTa-based similarity approach we considered was generally ineffective; the utility of sensitive health disclosures to peer recommendations needs further consideration, such as using feature extraction approaches targeted to the domain e.g. activity or health role classifiers [114, 10]. We did not attempt to compare multiple models during our field study; future investigations will need to link the specific peer connection benefits sought with the type of modeling approach used and conduct appropriate online evaluations. Given our interest in interaction, we recommended only recent sites in order to make candidate scoring more practical—$\approx$13K active authors vs 1 million total authors—but a system focused on recommending historical information or completed journeys might consider other candidate-generation approaches [262].

### 5.6.4 Acceptability

Acceptability refers to how participants react to the intervention. Specifically, we collected evidence of acceptability via explicit participant preferences (sec. 5.5.1) and feedback (sec. 5.5.2). Pre-study, the characteristics identified as most important for potential peer connections were a similar diagnosis, treatment, and engaging writing. These were the most frequently mentioned characteristics in recommendation feedback as well, with the addition of similar age. In general, the recommendations were only modestly acceptable to participants; 39% of specific recommendations were deemed irrelevant, frequently due to a lack of common ground between the participant and the recommended site. We chose a modeling approach that optimizes for interactions and does not explicitly reward similarity. Future peer recommenders might consider including features that capture the specific aspects of similarity deemed most important to users. For example, users might volunteer information about health condition, or it might be inferred from existing disclosures [11].

We received several explicit requests from users to alter the types of recommendations they were receiving, and one request from a user for the ability to filter the recommendations. Functionally, these requests speak to the coordination role played by a recommendation system. In clinical contexts, a human coordinator can incorporate this feedback to alter their peer matching approach on an individual level [141]. An algorithmic recommendation system faces a much greater challenge soliciting useful feedback on social matches [253]. By offering tooling to provide feedback, recommendation users can *self*-coordinate. Designing for self-coordination introduces trade-offs between user learning, system explanations, and ease of use [280]. An alternative might be designing peer matching systems with a human coordinator "in the loop"; the recommendation system could function as an assistance tool to facilitate matching thousands of peers simultaneously.

### 5.6.5 Efficacy

Efficacy refers to how much the intervention affects the desired behaviors. We collected evidence of efficacy by examining both reading and interaction behavior effects (secs. 5.5.3 and 5.5.3) as well as second-order behavior effects (sec. 5.5.3). Given the small sample and the uncontrolled design, we can only evaluate efficacy in a limited way, although results are promising. 39% of visited sites were visited a second time and a majority (57%) of participants who clicked at least once went on to visit at least one recommended site twice, suggesting that participants were interested in reading about ongoing health journeys. A smaller percentage (30%) of participants interacted with at least one recommended site, and only one reciprocal relationship formed as a result of the recommendations. Unclear social norms around interaction with strangers may present a barrier to greater interaction, such as whether an author is open to receiving unsolicited support [10]. Future design work could explore soliciting indicators of openness to supporter interaction and first-contact writing assistance for peer supporters who are not sure what to say [85, 107]. We observed no evidence of second-order harms due to recommendations. These results suggest that efficacy should be evaluated in a larger, controlled trial.

All of the interactions we observed in this study—excepting the one relationship—saw our participants providing support to others. Being a peer supporter may by more attractive and accessible than being a support recipient [141]. On one hand, a recommender system that encourages a supporter role runs the risk of generating self-fulfilling prophecies and preventing exploration of other roles [104]—leading users to deprioritise their own needs [7]—as well as contributing to exclusion of new users and reinforcement of existing cliques [281]. On the other hand, exposure to peer recommendations might lead to *more* exploration than what users choose without this scaffolding [282], and supporting others might serve as the endpoint for a personal transition to working in broader service of a health community [35, 9]. In general, the impacts of recommendation on community social dynamics are hard to predict [283]—and it is for that reason that future trials are necessary. Our feasibility study increases our confidence that peer recommendation will have positive benefits, including for equity. While only 1 in 40 of pre-recommender initiations are with sites without prior peer interactions, 1 in 10 of the model recommendations are for these siloed sites. In future studies, recommendation and visit dynamics should be carefully monitored for harmful social dynamics or intervention-generated inequalities [284].

## 5.7 Ethical Considerations

This research was reviewed by the University of Minnesota's Institutional Review Board. All usage data was collected in compliance with the CaringBridge terms of service and privacy policy. To mitigate potential negative impacts from unwanted negative contact, we restricted the pool of recommended sites to include only sites with the lowest privacy settings i.e. those that are indexable by search engines and visible to all visitors.

## 5.8 Limitations & Future Work

Our initial feasibility study of peer recommendation has significant limitations, although we are excited by the prospect of additional research developing the systems and interventions we consider. We developed and evaluated only a single interface and a single model. Our interface follows existing design patterns in CaringBridge notification emails, but a more iterative design process could produce more useful representations of recommendations [139]. The model we use is not "state-of-the-art", but represents a reasonable modern approach. Future studies should conduct more extensive offline and online model comparisons [265]—including with models that are focused on specifically facilitating reciprocal matching [285].

We looked at the potential impact of the recommendations for 12 weeks during the field study and 13 weeks post-study. We saw no influence of participants' visits on the journaling behavior of recommended sites in the 13 weeks post-study, but our design prevents us from analyzing any longer-term influence on the authors of recommended sites. In particular, authors that perceive a change in their audience after visits or interactions from peers might change the *content* of their Journal updates. Writing updates involves self-disclosure, and self-disclosure can lead to vulnerability and potential disappointment in the absence of reciprocation [286]. For an audience of peers, authors may be less likely to disclose negative information [287], which could decrease the potential benefits of expressive writing and self-reflection [70] and, unintentionally, lead to posts that are less useful for both peer and non-peer readers. Future studies should collect explicit feedback from recommended authors to understand the perception of authors toward peer visitors. The relationship between perceived audience and self-disclosure behavior remains an important open question to consider before deployment of peer recommendation systems that may change that audience. As an intervention, the peer recommender system we evaluated was intended to target behaviors correlated with social support benefits. We proposed an RCT focused on quantifying the increase in reading and interaction as a result of exposure to peer recommendations. However, due to the gap between received and perceived support [13], an increase in these behaviors may not produce corresponding increases in perceptions of social support. Future experimental designs could include pre-post self-report measures, either for social support directly (e.g. using the social connectedness scale [288]) or for desired impact on downstream health (e.g. using the perceived stress scale [289] or measures for health-related quality of life [290]).

We designed and evaluated the peer recommender system only with users of a single OHC. We argue these results generalize most to OHCs with similar affordances to CaringBridge. Rains argues that communication technologies for social connection have four primary affordances: visibility, availability, control, and reach [13]. Control, for example, is the potential to manage interactions, while availability is the potential to interact at particular times or places, so other text-based asynchronous communities provide similar potential. Visibility is the potential to make one's self known to others or to observe others' behavior; on CaringBridge, visibility is linked to specific blogs and communities that form around those blogs. Forums or listservs offer very different trade-offs around visibility of user-generated content, as do OHCs with rich user profiles [139, 104]. Reach is the potential to contact specific individuals, groups, or communities: CaringBridge provides reach only to specific individuals known to the user by name [10]. Peer recommendation might function differently in an ecosystem where other social discovery features are already present [153]. Future work should explore peer recommendation in OHCs with diverging affordances from those provided by CaringBridge—such as Q&A communities. We hope that our specific implementation serves to promote future investigation of peer recommendation: including models, interfaces, and outcomes for both individuals and communities.

## 5.9 Conclusion

In this chapter, I designed and evaluated a recommendation system to facilitate peer interactions. The evaluation centered on feasibility: can the intervention be implemented as described? Results were promising: I observed evidence of demand for peer recommendation, I produced an implementation with modest data requirements and design complexity that was nevertheless appropriate for CaringBridge, I determined it was practical to administer the intervention via email, participants reported that recommendations were generally acceptable, and the system did produce meaningful additional activity, including one relationship. I identified several promising avenues for future work grounded in our results. In the final chapter of this dissertation, I will turn to identifying more speculative opportunities for future work.

# Chapter 6

# Discussion and Conclusion

## 6.1 Summary: what have we learned about peer communication in OHCs?

In Chapter 3, I developed models and modeling techniques for extracting health behaviors from CaringBridge Journal updates, motivated in part by a desire to help authors find timely information and support from peers. In Chapter 4, I characterized the peer interaction that was already happening on CaringBridge, demonstrating the growth of the author interaction network and the importance of author characteristics such as health role. Chapter 5 combines these insights to design a peer recommendation system appropriate for the CaringBridge context. Separately, the contributions of this thesis are:

- A quantitative summary of cancer patient joining and leaving behavior in terms of their cancer phase.
- A quantitative assessment of the responsibilities that cancer patients discuss online, operationalized from a qualitative framework.
- Factors associated with the formation and growth of peer connections.
- A comparison of peer connection behavior between patients and caregivers.
- A system design for peer recommendation in OHCs.
- An assessment of the feasibility of a peer recommendation system.

Together, these contributions suggest the complexity of usage behaviors in online health communities. This dissertation grapples with that complexity, ultimately contributing a vision of quantitative modeling, built on qualitative insight, to support the disparate needs, goals, and identities that motivate OHC usage. A key practical limitation of this work is revealed in the lack of integration: cancer phase and author role predictions could both be used in a peer recommendation system. However, in designing a peer recommendation system I omitted both of these considerations. One reason is the lack of mature understanding of the relative costs and benefits of these modeling goals. A second is the complexity of integration: even without author role and cancer phase classifiers, the recommendation system and its potential influence on the social ecosystem is enormously complex.

This complexity suggests a value to studying recommenders that are already deployed in platforms, where full-lifecycle exposure to recommendations can be assumed and A/B comparisons can be feasibly conducted. I suggest, however, that studying recommendation for users of platforms *without* recommendation serves an important role: it is a window into the hopes and disappointments experienced by users facing a deficit. Peer recommendation is appealing to me and other researchers in part because these technologies can address deficits in social support and medical information without a need to directly address the immense structural and societal inequities that produce these deficits [284]. However, the actual usage of and feedback to peer recommendation demonstrates that complexities necessitate continual development of the conceptual assumptions underlying an implementation. In the rest of this chapter, I engage these complexities by suggesting important open challenges and opportunities in research around peer communication and peer recommendation in OHCs.

## 6.2 Peer recommendation systems for OHCs: challenges and opportunities

Seaver conceptualizes recommendation systems as "traps", designed around metrics and mindsets that encourage frequent and enduring usage [291]. If recommendations are traps, what does it mean to be trapped in an online health community? Haimson positions OHCs as "social transition machinery" that facilitate transitions [292], building on observations that leaving an OHC after accomplishing one's desired social transition and associated informational and interaction goals should not be treated as a bad outcome [56]. Thus, designing peer recommenders based on metrics like total clicks or interactions implicitly imagines a social transition that involves continued engagement with the OHC. Other metrics could better reflect peer recommender systems as facilitators of specific support goals—possibly including preference for an algorithm that encourages initial use but decreases subsequent use as e.g. informational support needs are met from reading other sites. The ideal peer recommender system may make support tasks more "efficient", decreasing not just usage of the recommender system but of the OHC as a whole.

Peer recommendation is designed to change the makeup of the audience that an OHC user interacts with. On CaringBridge, perceptions of audience will affect the content of site updates [67]. Site authors disclose sensitive information with a particular audience in mind [293], altering their disclosures to perform tensions between an idealized and authentic online self [294]. These dynamics need further investigation, and these effects may not be unidirectional—peer interactions may decrease update frequency or suppress negative self-disclosure, but such interactions could also create social pressure to *continue* updating beyond the point that it is useful for the author. In general, peer recommendation—somewhat uneasily—manipulates a communication environment with unclear social expectations, as roles and goals shift over time [35].

An important consideration for any health intervention is "who benefits?" [284]. For peer support interventions, a key question is whether the resulting supportive communication benefits primarily people who already have access to quality social support or people who lack that access. These are the social compensation ("poor get richer") and the social enhancement ("rich get richer") benefit models [295]. Similar to prior work, our observations suggest that the "poor get richer" [67]; sites without existing connections are more likely to be recommended than the sites that authors currently interact with, which suggests disproportionate benefits could flow to sites without lots of existing support. However, we do not know if these sites lack offline support, nor do we know if access to social support is the most salient equity issue in the deployment of peer recommenders. I avoided an analysis of the impact of recommendation on demographic traits such as gender, largely because these are hard to assess without a dedicated study. (Absent self-report data, identity characteristics might be classified from Journal updates, although such a process is fraught with methodological and ethical challenges [296].) The intersection of health inequalities and peer recommender development is ripe for future work, including both qualitative—such as an investigation of how OHC users evaluate recommendations and perceive strangers—and quantitative—such as a careful evaluation of the trade-off between offline accuracy, coverage, and diversity metrics.

## 6.3 Limitations

I already discussed two limitations: a focus on a single OHC makes reasoning about generalizability challenging (sec. 2.1.2) and a lack of modeling integration due to complexity (sec. 6.1). Both represent an opportunity for future researchers to engage specific groups of users to understand the benefits of providing recommendations that are sensitive to a specific set of needs. I suggest two additional limitations that challenge my own understanding of recommendation interventions for online peer communication.

### 6.3.1 "Peerness" and perceptions of support

A key assumption of this work is that authoring health updates on CaringBridge is a sufficient condition to be a "peer" to other authors. The peer definition I used from Simoni et al. includes the characteristics (a) obtaining benefits from peer support that derive from their status as peers (b) "intentionally setting out to interact with individuals they may or may not encounter in their everyday life" [63]. In the context of interaction between CaringBridge authors, these conditions hold; commenting on another author's post is both intentional and—based on my qualitative investigation of commenting behavior—often based on peer status i.e. common ground. A recommendation system that facilitates interactions among strangers strains this definition; when a visitor leaves a supportive comment induced by a recommendation, the receiving author may be unaware of the visitor's peer status and did not necessarily intend to receive such interactions. Thus, the recommender facilitates peer support only insofar as authors articulate the "critical commonality" that the authors share during interaction [65]. Optimal matching theory suggests that particular illnesses require support appropriate to the specifics of the situation [65]; peers are more likely to provide that support due to their experiential similarity [59], but the specific importance of the supporter and/or receiver of support perceiving "peerness" remains unclear. Other OHCs specifically include formal peer mentors or health professionals [34], and the implications of peer matching with those users may be quite different. For example, health professionals may be more likely to provide prescriptive information than descriptive information [297].

### 6.3.2 Interventions or existing ecosystems?

In Chapter 5, I argued that peer recommendation should be conceptualized as a health intervention and evaluated accordingly: with a feasibility study followed by larger controlled trials. While this framing correctly places focus on the manipulated behaviors and other outcomes, it misrepresents the way that recommendation systems function as a part of broader community designs. It is challenging to evaluate a single interface feature in isolation, and especially so for features that facilitate communication. One obvious objection to concerns about the impact of recommendation on individual behavior or on community dynamics is that recommendation is ubiquitous on social media platforms, including many OHCs. While further study is therefore valuable, it may be that a focus on how existing recommendation systems function will be more immediately impactful then a focus on prospective harms from recommendation "interventions"—interventions that are already widely used in practice. One reason to conduct RCTs and other focused trials is that the research community's understanding of peer interaction dynamics in OHCs continues to be poor [219], as does our understanding of social media interaction in general [298]. But the specific behaviors of interest, such as peer interaction, rest on shaky theoretical ground [63]. Thus, the specific intervention I propose is poorly defined and disconnected from potential mechanisms. Future peer recommendation interventions should be based on specific mechanisms of behavior change and benefit, such as exposure to relevant symptom management information leading to greater self-efficacy.

## 6.4 Conclusion

Communication with peers online can be a valuable source of social support. I studied an online health community where peer communication occurred adjacent to primary communication with a non-peer support network. In that context, I designed and evaluated a peer recommendation system for users of online health communities. My work contributes an understanding of peer communication patterns and suggests benefits to the further development of peer recommendation interventions.

# References

[1] Deborah J. Bowen, Matthew Kreuter, Bonnie Spring, Ludmila Cofta-Woerpel, Laura Linnan, Diane Weiner, Suzanne Bakken, Cecilia Patrick Kaplan, Linda Squiers, Cecilia Fabrizio, and Maria Fernandez. How We Design Feasibility Studies. *American Journal of Preventive Medicine*, 36(5):452–457, May 2009.

[2] Sanford Weisberg. *Applied Linear Regression*. Wiley, Hoboken, NJ, 3rd edition edition, December 2013.

[3] Michael J. Paul, Ryen W. White, and Eric Horvitz. Diagnoses, Decisions, and Outcomes: Web Search as Decision Support for Cancer. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15*, WWW'15, pages 831–841, Florence, Italy, 2015. ACM Press.

[4] Annette Markham. Fabrication as Ethical Practice. *Information, Communication & Society*, 15(3):334–353, April 2012. Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2011.641993.

[5] Ryan J. Gallagher, Elizabeth Stowell, Andrea G. Parker, and Brooke Foucault Welles. Reclaiming Stigmatized Narratives: The Networked Disclosure Landscape of #MeToo. *Proceedings of the 2019 conference on Computer supported cooperative work - CSCW '19*, May 2019.

[6] Anna Fang and Haiyi Zhu. Matching for Peer Support: Exploring Algorithmic Matching for Online Mental Health Communities. *Computer Supported Cooperative Work (CSCW)*, 2022.

[7] Kimberley R. Allison, Pandora Patterson, Daniel Guilbert, Melissa Noke, and Olga Husson. Logging On, Reaching Out, and Getting By: A Review of Self-reported Psychosocial Impacts of Online Peer Support for People Impacted by Cancer. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1):95:1–95:35, April 2021.

[8] C. Estelle Smith, Zachary Levonian, Haiwei Ma, Robert Giaquinto, Gemma Lein-Mcdonough, Zixuan Li, Susan O'conner-Von, and Svetlana Yarosh. "I Cannot Do All of This Alone": Exploring Instrumental and Prayer Support in Online Health Communities. *ACM Trans. Comput.-Hum. Interact.*, 27(5):38:1–38:41, August 2020.

[9] Zachary Levonian, Drew Richard Erikson, Wenqi Luo, Saumik Narayanan, Sabirat Rubya, Prateek Vachher, Loren Terveen, and Svetlana Yarosh. Bridging Qualitative and Quantitative Methods for User Modeling: Tracing Cancer Patient Behavior in an Online Health Community. In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media*, volume 14, page 12, Atlanta, GA, May 2020. AAAI.

[10] Zachary Levonian, Marco Dow, Drew Erikson, Sourojit Ghosh, Hannah Miller Hillberg, Saumik Narayanan, Loren Terveen, and Svetlana Yarosh. Patterns of Patient and Caregiver Mutual Support Connections in an Online Health Community. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3):275:1–275:46, January 2021.

[11] Changye Li, Zachary Levonian, Haiwei Ma, and Svetlana Yarosh. Condition Unknown: Predicting Patients' Health Conditions in an Online Health Community. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18, pages 281–284, New York, NY, USA, October 2018. Association for Computing Machinery.

[12] Ruyuan Wan, Zachary Levonian, and Svetlana Yarosh. How much is a "like" worth? Engagement and Retention in an Online Health Community. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pages 413–417. Association for Computing Machinery, New York, NY, USA, October 2020.

[13] Stephen A. Rains. *Coping with Illness Digitally*. MIT Press, 2018. Google-Books-ID: EbPkuQEACAAJ.

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

[15] Judith S. Olson and Wendy A. Kellogg. *Ways of Knowing in HCI*. Springer Publishing Company, Incorporated, 2014.

[16] Alexandra Olteanu, Onur Varol, and Emre Kiciman. Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 370–386, Portland, Oregon, USA, 2017. ACM Press.

[17] Vivek Kulkarni, Margaret L. Kern, David Stillwell, Michal Kosinski, Sandra Matz, Lyle Ungar, Steven Skiena, and H. Andrew Schwartz. Latent human traits in the language of social media: An open-vocabulary approach. *PLOS ONE*, 13(11):e0201703, November 2018.

[18] James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. Psychological aspects of natural language. use: our words, our selves. *Annu Rev Psychol*, 54:547–577, 2003.

[19] Amy X. Zhang and Scott Counts. Modeling Ideology and Predicting Policy Change with Social Media: Case of Same-Sex Marriage. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2603–2612, Seoul, Republic of Korea, 2015. ACM Press.

[20] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, pages 1155–1168, San Francisco, California, USA, 2016. ACM Press.

[21] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting Depression via Social Media. *ICWSM*, 13:1–10, 2013.

[22] Robert Adcock and David Collier. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3):529–546, 2001.

[23] Niharika Sachdeva, Ponnurangam Kumaraguru, and Munmun De Choudhury. Social Media for Safety: Characterizing Online Interactions Between Citizens and Police. 2016.

[24] Shauna Julia Concannon, Madeline Balaam, Emma Simpson, and Rob Comber. Applying Computational Analysis to Textual Data from the Wild: A Feminist Perspective. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 226:1–226:13, New York, NY, USA, 2018. ACM.

[25] Jiaxin Liu, Elissa R. Weitzman, and Rumi Chunara. Assessing Behavior Stage Progression From Social Media Data. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1320–1333, New York, NY, USA, 2017. ACM.

[26] Emre Kiciman, Scott Counts, and Melissa Gasser. Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use. *ICWSM*, May 2018.

[27] Shaodian Zhang, Edouard Grave, Elizabeth Sklar, and Noémie Elhadad. Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. *Journal of Biomedical Informatics*, 69(Supplement C):1–9, May 2017.

[28] Tom Huang, Anas Elghafari, Kunal Relia, and Rumi Chunara. High-resolution Temporal Representations of Alcohol and Tobacco Behaviors from Social Media Data. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):54:1–54:26, December 2017.

[29] Amy X Zhang, Bryan Culbertson, and Praveen Paritosh. Characterizing Online Discussion Using Coarse Discourse Sequences. *ICWSM*, page 10, 2017.

[30] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, pages 1591–1600, Perth, Australia, 2017. ACM Press.

[31] David L. Morgan. Practical Strategies for Combining Qualitative and Quantitative Methods: Applications to Health Research. *Qual Health Res*, 8(3):362–376, May 1998.

[32] Kenya Mejia and Svetlana Yarosh. A Nine-Item Questionnaire for Measuring the Social Disfordance of Mediated Social Touch Technologies. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):1–17, December 2017.

[33] Richard Meyes, Melanie Lu, Constantin Waubert de Puiseau, and Tobias Meisen. Ablation Studies in Artificial Neural Networks, February 2019. Number: arXiv:1901.08644 arXiv:1901.08644 [cs, q-bio] version: 2.

[34] Shefali Haldar, Yoojung Kim, Sonali R. Mishra, Andrea L. Hartzler, Ari H. Pollack, and Wanda Pratt. The Patient Advice System: A Technology Probe Study to Enable Peer Support in the Hospital. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW2):112:1–112:23, October 2020.

[35] Joshua Introne, Isabel Munoz, and Bryan Semaan. The Narrative Tapestry Design Process: Weaving Online Social Support from Stories of Stigma. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):389:1–389:35, October 2021.

[36] Afaf lbrahim Meleis. Transitions Theory. In Marlaine C. Smith and Marilyn E. Parker, editors, *Nursing theories & nursing practice*, pages 361–380. F.A. Davis Company, Philadelphia, PA, 4 edition, 2015.

[37] Karen L. Schumacher and Afaf lbrahim Meleis. Transitions: A Central Concept in Nursing. *Image: the Journal of Nursing Scholarship*, 26(2):119–127, June 1994.

[38] Lila J. Finney Rutten, Neeraj K. Arora, Alexis D. Bakos, Noreen Aziz, and Julia Rowland. Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). *Patient Education and Counseling*, 57(3):250–261, June 2005.

[39] Maia Jacobs, James Clawson, and Elizabeth D. Mynatt. A Cancer Journey Framework: Guiding the Design of Holistic Health Technology. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '16, pages 114–121, ICST, Brussels, Belgium, Belgium, 2016. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[40] Gillian R. Hayes, Gregory D. Abowd, John S. Davis, Marion L. Blount, Maria Ebling, and Elizabeth D. Mynatt. Opportunities for Pervasive Computing in Chronic Cancer

Care. In *Pervasive Computing*, Lecture Notes in Computer Science, pages 262–279. Springer, Berlin, Heidelberg, May 2008.

[41] Michael Massimi, Jill P. Dimond, and Christopher A. Le Dantec. 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, pages 719–728, New York, NY, USA, 2012. ACM.

[42] Mark W. Newman, Debra Lauterbach, Sean A. Munson, Paul Resnick, and Margaret E. Morris. 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*, CSCW '11, pages 341–350, New York, NY, USA, 2011. ACM.

[43] Susan Leigh Star and Anselm Strauss. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1):9–30, March 1999.

[44] Jordan Eschler, Zakariya Dehlawi, and Wanda Pratt. Self-Characterized Illness Phase and Information Needs of Participants in an Online Cancer Forum. In *Ninth International AAAI Conference on Web and Social Media*, April 2015.

[45] Michele O'Brien, Carrie Tompkins Stricker, Jacqueline D. Foster, Kimberly Ness, Aimee Ginsburg Arlen, and Rowena N. Schwartz. Navigating the Seasons of Survivorship in Community Oncology. *Clinical Journal of Oncology Nursing*, 18(s1):9–14, February 2014.

[46] Maia Jacobs, James Clawson, and Elizabeth D. Mynatt. Cancer navigation: opportunities and challenges for facilitating the breast cancer journey. pages 1467–1478. ACM, 2014.

[47] J. M. Corbin. The Corbin and Strauss Chronic Illness Trajectory model: an update. *Sch Inq Nurs Pract*, 12(1):33–41, 1998.

[48] Miaomiao Wen and Carolyn Penstein Rose. Understanding Participant Behavior Trajectories in Online Health Support Groups Using Automatic Extraction Methods. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12, pages 179–188, New York, NY, USA, 2012. ACM.

[49] Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. Forum77: An Analysis of an Online Health Forum Dedicated to Addiction Recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1511–1526, New York, NY, USA, 2015. ACM. event-place: Vancouver, BC, Canada.

[50] Daren K. Heyland, Peter Dodek, Graeme Rocker, Dianne Groll, Amiram Gafni, Deb Pichora, Sam Shortt, Joan Tranmer, Neil Lazar, Jim Kutsogiannis, and Miu Lam. What matters most in end-of-life care: perceptions of seriously ill patients and their family members. *CMAJ*, 174(5):627–633, February 2006.

[51] Lorraine R. Buis. Emotional and Informational Support Messages in an Online Hospice Support Community. *CIN: Computers, Informatics, Nursing*, 26(6):358–367, November 2008.

[52] Robert Douglas Ferguson, Michael Massimi, Emily Anne Crist, and Karyn Anne Moffatt. Craving, Creating, and Constructing Comfort: Insights and Opportunities for Technology in Hospice. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 1479–1490, New York, NY, USA, 2014. ACM.

[53] Thomas J. Prendergast and Kathleen A. Puntillo. Withdrawal of Life Support: Intensive Caring at the End of Life. *JAMA*, 288(21):2732–2740, December 2002.

[54] Mette Terp Høybye, Christoffer Johansen, and Tine Tjørnhøj-Thomsen. Online interaction. Effects of storytelling in an internet breast cancer support group. *Psycho-Oncology*, 14(3):211–220, 2005.

[55] Wen-ying Sylvia Chou, Benmei Liu, Samantha Post, and Bradford Hesse. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003–2008. *J Cancer Surviv*, 5(3):263–270, September 2011.

[56] Michael Massimi, Jackie L. Bender, Holly O. Witteman, and Osman H. Ahmed. 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*, CSCW '14, pages 1491–1501, New York, NY, USA, 2014. ACM. event-place: Baltimore, Maryland, USA.

[57] Jeana Frost and Michael Massagli. Social Uses of Personal Health Information Within PatientsLikeMe, an Online Patient Community: What Can Happen When Patients Have Access to One Another's Data. *J Med Internet Res*, 10(3):e15, 2008.

[58] Stephen A. Rains and Kevin B. Wright. Social Support and Computer-Mediated Communication: A State-of-the-Art Review and Agenda for Future Research. *Annals of the International Communication Association*, 40(1):175–211, January 2016.

[59] Peggy A. Thoits. Mechanisms Linking Social Ties and Support to Physical and Mental Health. *J Health Soc Behav*, 52(2):145–161, June 2011.

[60] Susannah Fox. *The social life of health information, 2011*. Pew Internet & American Life Project Washington, DC, 2011.

[61] Elizabeth A. Gage. Social networks of experientially similar others: Formation, activation, and consequences of network ties on the health care experience. *Social Science & Medicine*, 95:43–51, October 2013.

[62] Marén Schorch, Lin Wan, David William Randall, and Volker Wulf. Designing for Those Who Are Overlooked: Insider Perspectives on Care Practices and Cooperative Work of Elderly Informal Caregivers. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 787–799, New York, NY, USA, 2016. ACM. event-place: San Francisco, California, USA.

[63] Jane M. Simoni, Julie C. Franks, Keren Lehavot, and Samantha S. Yard. Peer interventions to promote health: Conceptual considerations. *American Journal of Orthopsychiatry*, 81(3):351–359, 2011.

[64] Jordan Eschler and Wanda Pratt. "I'm So Glad I Met You": Designing Dynamic Collaborative Support for Young Adult Cancer Survivors. In *Proceedings of the 2017 ACM*

*Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1763–1774, New York, NY, USA, 2017. ACM. event-place: Portland, Oregon, USA.

[65] J. W. Turner, J. A. Grube, and J. Meyers. Developing an optimal match within online communities: an exploration of CMC support communities and traditional support. *Journal of Communication*, 51(2):231–251, 2001. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.2001.tb02879.x.

[66] James S. House. Understanding social factors and inequalities in health: 20th century progress and 21st century prospects. *J Health Soc Behav*, 43(2):125–142, June 2002.

[67] Stephen A. Rains and David M. Keating. The Social Dimension of Blogging about Health: Health Blogging, Social Support, and Well-being. *Communication Monographs*, 78(4):511–534, December 2011.

[68] Anthony McCosker and Raya Darcy. Living with Cancer. *Information, Communication & Society*, 16(8):1266–1285, October 2013.

[69] Jina Huh and Mark S. Ackerman. Collaborative Help in Chronic Disease Management: Supporting Individualized Problems. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 853–862, New York, NY, USA, 2012. ACM.

[70] Haiwei Ma, C. Estelle Smith, Lu He, Saumik Narayanan, Robert A. Giaquinto, Roni Evans, Linda Hanson, and Svetlana Yarosh. Write for Life: Persisting in Online Health Communities Through Expressive Writing and Social Support. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):73:1–73:24, December 2017.

[71] Ko Hsiu-Chia and Kuo Feng-Yang. Can Blogging Enhance Subjective Well-Being Through Self-Disclosure? *CyberPsychology & Behavior*, 12(1):75–79, February 2009.

[72] Lisa F. Berkman, Ichirō Kawachi, and M. Maria Glymour. *Social Epidemiology*. Oxford University Press, 2014. Google-Books-ID: qHpYCwAAQBAJ.

[73] Gary L. Kreps. Online Information and Communication Systems to Enhance Health Outcomes through Communication Convergence. *Hum Commun Res*, 43(4):518–530, October 2017.

[74] Julianne Holt-Lunstad, Timothy B. Smith, and J. Bradley Layton. Social Relationships and Mortality Risk: A Meta-analytic Review. *PLOS Medicine*, 7(7):e1000316, July 2010.

[75] Bert N. Uchino. Social Support and Health: A Review of Physiological Processes Potentially Underlying Links to Disease Outcomes. *J Behav Med*, 29(4):377–387, August 2006.

[76] Bert N. Uchino, Kimberly Bowen, Robert Kent de Grey, Jude Mikel, and Edwin B. Fisher. Social Support and Physical Health: Models, Mechanisms, and Opportunities. In Edwin B. Fisher, Linda D. Cameron, Alan J. Christensen, Ulrike Ehlert, Yan Guo, Brian Oldenburg, and Frank J. Snoek, editors, *Principles and Concepts of Behavioral Medicine: A Global Handbook*, pages 341–372. Springer New York, New York, NY, 2018.

[77] Barton J. Hirsch. Psychological dimensions of social networks: A multimethod analysis. *American Journal of Community Psychology*, 7(3):263–277, 1979.

[78] B. S. Hanson, P. O. Ostergren, S. Elmståhl, S. O. Isacsson, and J. Ranstam. Reliability and validity assessments of measures of social networks, social support and control–results from the Malmö Shoulder and Neck Study. *Scand J Soc Med*, 25(4):249–257, December 1997.

[79] Tom K. Greenfield, Jürgen Rehm, and John D. Rogers. Effects of depression and social integration on the relationship between alcohol consumption and all-cause mortality. *Addiction*, 97(1):29–38, January 2002.

[80] Alan Vaux, Sharon Riedel, and Doreen Stewart. Modes of social support: The social support behaviors (SS-B) scale. *American Journal of Community Psychology*, 15(2):209–232, 1987.

[81] Irwin G Sarason, Henry M Levine, Robert B Basham, and Barbara R Sarason. Assessing social support: The social support questionnaire. *Journal of personality and social psychology*, 44(1):127, 1983.

[82] T. E. Seeman and L. F. Berkman. Structural characteristics of social networks and their relationship with social support in the elderly: who provides support. *Soc Sci Med*, 26(7):737–749, 1988.

[83] Dan Russell, Letitia A. Peplau, and Carolyn E. Cutrona. The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, 39(3):472–480, 1980.

[84] Sally A Shumaker and Arlene Brownell. Toward a theory of social support: Closing conceptual gaps. *Journal of social issues*, 40(4):11–36, 1984.

[85] C. Estelle Smith, Avleen Kaur, Katie Z. Gach, Loren G. Terveen, Mary Jo Kreitzer, and Susan O'Conner-Von. What is Spiritual Support and How Might It Impact the Design of Online Communities? *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW):43:1–43:42, April 2021.

[86] Maia Jacobs, Galina Gheihman, Krzysztof Z. Gajos, and Anoopum S. Gupta. "I think we know more than our doctors": How Primary Caregivers Manage Care Teams with Limited Disease-related Expertise. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):1–22, November 2019.

[87] Jane Hughes, Eleri Wood, and Gaynor Smith. Exploring kidney patients' experiences of receiving individual peer support. *Health Expectations*, 12(4):396–406, December 2009. Publisher: John Wiley & Sons, Ltd.

[88] Mary P. Gallant. The Influence of Social Support on Chronic Illness Self-Management: A Review and Directions for Research. *Health Educ Behav*, 30(2):170–195, April 2003. Publisher: SAGE Publications Inc.

[89] Alexander S. Young, Edmund Chaney, Rebecca Shoai, Laura Bonner, Amy N. Cohen, Brad Doebbeling, David Dorr, Mary K. Goldstein, Eve Kerr, Paul Nichol, and Ruth Perrin. Information Technology to Support Improved Care For Chronic Illness. *J Gen Intern Med*, 22(Suppl 3):425–430, December 2007.

[90] Frank Riessman. The “Helper” Therapy Principle. *Soc Work*, 10(2):27–32, April 1965.

[91] Hyun Jung Oh, Elif Ozkaya, and Robert LaRose. How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. *Computers in Human Behavior*, 30:69–78, January 2014.

[92] B. Qiu, K. Zhao, P. Mitra, D. Wu, C. Caragea, J. Yen, G. E. Greer, and K. Portier. Get Online Support, Feel Better – Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 274–281, October 2011.

[93] Yi-Chia Wang, Robert Kraut, and John M. Levine. 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, pages 833–842, New York, NY, USA, 2012. ACM.

[94] Diyi Yang, Robert Kraut, and John M. Levine. Commitment of Newcomers and Oldtimers to Online Health Support Communities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6363–6375, New York, NY, USA, 2017. ACM.

[95] Gül Seçkin. Satisfaction with health status among cyber patients: testing a mediation model of electronic coping support. *Behaviour & Information Technology*, 32(1):91–101, January 2013.

[96] Chunling Hu, Simon Kung, Teresa A. Rummans, Matthew M. Clark, and Maria I. Lapid. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. *J Am Med Inform Assoc*, 22(e1):e194–e209, April 2015.

[97] Donna J. Kaltenbaugh, Mary Lou Klem, Lu Hu, Eleanor Turi, Alice J. Haines, and Jennifer Hagerty Lingler. Using Web-Based Interventions to Support Caregivers of Patients With Cancer: A Systematic Review. *Oncology Nursing Forum*, 42(2):156–164, February 2015.

[98] Sumaira H. Malik and Neil S. Coulson. Computer-mediated infertility support groups: An exploratory study of online experiences. *Patient Education and Counseling*, 73(1):105–113, October 2008.

[99] Angelica Attard and Neil S. Coulson. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. *Computers in Human Behavior*, 28(2):500–506, March 2012.

[100] Tiina Yli-Uotila, Anja Rantanen, and Tarja Suominen. Online Social Support Received by Patients With Cancer. *CIN: Computers, Informatics, Nursing*, 32(3):118–126, March 2014.

[101] Herman Saksono. Algorithmic Patient Matching in Peer Support Systems for Hospital Inpatients. page 6, May 2021.

[102] Matthieu Tixier and Myriam Lewkowicz. "Counting on the Group": Reconciling Online and Offline Social Support among Older Informal Caregivers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3545–3558, New York, NY, USA, May 2016. Association for Computing Machinery.

[103] Sarah Holbrey and Neil S. Coulson. A qualitative investigation of the impact of peer to peer online support for women living with Polycystic Ovary Syndrome. *BMC Women's Health*, 13(1):51, December 2013.

[104] Doğa Gatos, Aslı Günay, Güncel Kırlangıç, Kemal Kuscu, and Asim Evren Yantac. How HCI Bridges Health and Design in Online Health Communities: A Systematic Review. In *Designing Interactive Systems Conference 2021*, DIS '21, pages 970–983, New York, NY, USA, June 2021. Association for Computing Machinery.

[105] Zoë C. Skea, Sara J. MacLennan, Vikki A. Entwistle, and James N'Dow. Enabling mutual helping? Examining variable needs for facilitated peer support. *Patient Education and Counseling*, 85(2):e120–e125, November 2011.

[106] Naveena Karusala, David Odhiambo Seeh, Cyrus Mugo, Brandon Guthrie, Megan A Moreno, Grace John-Stewart, Irene Inwani, Richard Anderson, and Keshet Ronen. "That courage to encourage": Participation and Aspirations in Chat-based Peer Support for

Youth Living with HIV. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–17, New York, NY, USA, May 2021. Association for Computing Machinery.

[107] Zhenhui Peng, Qingyu Guo, Ka Wing Tsang, and Xiaojuan Ma. Exploring the Effects of Technological Writing Assistance for Support Providers in Online Mental Health Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–15, Honolulu, HI, USA, April 2020. Association for Computing Machinery.

[108] Na Sun, Xiying Wang, and Mary Beth Rosson. How Do Distance Learners Connect? Shared Identity, Focused Work and Future Possibilities. page 13, 2019.

[109] Jingbo Meng. Your Health Buddies Matter: Preferential Selection and Social Influence on Weight Management in an Online Health Social Network. *Health Communication*, 31(12):1460–1471, December 2016.

[110] Damon Centola and Arnout van de Rijt. Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, 125:19–31, January 2015.

[111] Kirsten P. Smith and Nicholas A. Christakis. Social Networks and Health. *Annual Review of Sociology*, 34(1):405–429, 2008.

[112] Joseph Seering, Jessica Hammer, Geoff Kaufman, and Diyi Yang. Proximate Social Factors in First-Time Contribution to Online Communities. *CHI*, page 14, 2020.

[113] Diane Maloney-Krichmar and Jenny Preece. A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community. *ACM Trans. Comput.-Hum. Interact.*, 12(2):201–232, June 2005.

[114] Diyi Yang, Robert Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '19, page 12, 2019.

[115] Kenneth D. Benne and Paul Sheats. Functional Roles of Group Members. *Journal of Social Issues*, 4(2):41–49, 1948.

[116] Ashish Sharma, Monojit Choudhury, Tim Althoff, and Amit Sharma. Engagement Patterns of Peer-to-Peer Interactions on Mental Health Platforms. In *arXiv:2004.04999 [cs]*, April 2020. arXiv: 2004.04999.

[117] Helen Rose Fuchs Ebaugh. *Becoming an Ex: The Process of Role Exit*. University of Chicago Press, 1988.

[118] Orville G Brim and Carol D Ryff. On the properties of life events. *Life-span development and behavior*, 3, 1980.

[119] Greg L. Stewart, Ingrid S. Fulmer, and Murray R. Barrick. An Exploration of Member Roles as a Multilevel Linking Mechanism for Individual Traits and Team Outcomes. *Personnel Psychology*, 58(2):343–365, 2005.

[120] Moira Burke and Robert Kraut. Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, page 27, San Diego, CA, USA, 2008. ACM Press.

[121] Lu (Lucy) Yan, Jianping Peng, and Yong Tan. Network Dynamics: How Can We Find Patients Like Us? *Information Systems Research*, 26(3):496–512, August 2015.

[122] Weiai Wayne Xu, I-Hsuan Chiu, Yixin Chen, and Tanuka Mukherjee. Twitter hashtags for health: applying network and content analyses to understand the health knowledge sharing in a Twitter-based community of practice. *Qual Quant*, 49(4):1361–1380, July 2015.

[123] Yingjie Lu, Yang Wu, Jingfang Liu, Jia Li, and Pengzhu Zhang. Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community. *Journal of Medical Internet Research*, 19(4):e109, 2017. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[124] Jan Colvin, Lillian Chenoweth, Mary Bold, and Cheryl Harding. Caregivers of Older Adults: Advantages and Disadvantages of Internet-based Social Support. *Family Relations*, 53(1):49–57, 2004.

[125] Ulrika Sandén, Fredrik Nilsson, Hans Thulesius, Maria Hägglund, and Lars Harrysson. Cancer, a relational disease. *International Journal of Qualitative Studies on Health and Well-being*, 14(1):1622354, January 2019.

[126] Kang Namkoong, Lori L. DuBenske, Bret R. Shaw, David H. Gustafson, Robert P. Hawkins, Dhavan V. Shah, Fiona M. McTavish, and James F. Cleary. Creating a Bond between Caregivers Online: Impact on Caregivers' Coping Strategies. *J Health Commun*, 17(2):125–140, 2012.

[127] Rosaleen Duggan Bloom. *Patients' and Caregivers' Experience of Social Support on CaringBridge*. Ph.D., The University of Utah, United States – Utah, 2017. ISBN: 9780438889811.

[128] Sheldon Cohen. Social Relationships and Health. *American Psychologist*, 59(8):676, November 2004.

[129] Sarah D. Kowitt, Katrina R. Ellis, Veronica Carlisle, Nivedita L. Bhushan, Kristin Z. Black, Kaitlyn Brodar, Nicole M. Cranley, Kia L. Davis, Eugenia Eng, Michelle Y. Martin, Jared McGuirt, Rebeccah L. Sokol, Patrick Y. Tang, Anissa I. Vines, Jennifer S. Walker, and Edwin B. Fisher. Peer support opportunities across the cancer care continuum: a systematic scoping review of recent peer-reviewed literature. *Support Care Cancer*, 27(1):97–108, January 2019.

[130] Catherine Walshe and Diane Roberts. Peer support for people with advanced cancer: a systematically constructed scoping review of quantitative and qualitative evidence. *Current Opinion in Supportive and Palliative Care*, 12(3):308–322, September 2018.

[131] Jerome A Osheroff, Jonathan M. Teich, Donald Levick, Luis Saldana, Ferdinand T. Velasco, Dean F. Sittig, Kendall M. Rogers, and Robert A. Jenders. *Improving Outcomes with Clinical Decision Support : An Implementer's Guide*. CRC Press, 2nd edition, February 2012.

[132] Allison Boyes, Heidi Turon, Alix Hall, Rochelle Watson, Anthony Proietto, and Robert Sanson-Fisher. Preferences for models of peer support in the digital era: A cross-sectional survey of people with cancer. *Psycho-Oncology*, 27(9):2148–2154, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pon.4781.

[133] Anna M. Anderson, Elizabeth C. Lavender, Esther Dusabe-Richards, Teumzghi F. Mebrahtu, Linda McGowan, Philip G. Conaghan, Sarah R. Kingsbury, Gerry Richardson, Deborah Antcliff, and Gretl A. McHugh. Peer mentorship to improve self-management of hip and knee osteoarthritis: a randomised feasibility trial. *BMJ Open*, 11(7):e045389, July 2021. Publisher: British Medical Journal Publishing Group Section: Rheumatology.

[134] Haley MacLeod, Grace Bastin, Leslie S. Liu, Katie Siek, and Kay Connelly. ”Be Grateful You Don’t Have a Real Disease”: Understanding Rare Disease Relationships. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, pages 1660–1673, New York, NY, USA, May 2017. Association for Computing Machinery. event-place: Denver, Colorado, USA.

[135] Arden Moulton, Amy Balbierz, Stephanie Eisenman, Elizabeth Neustein, Virginia Walther, and Irwin Epstein. Woman to Woman: A Peer to Peer Support Program for Women With Gynecologic Cancer. *Social Work in Health Care*, 52(10):913–929, November 2013. Publisher: Routledge _eprint: https://doi.org/10.1080/00981389.2013.834031.

[136] Arden Moulton. Woman to Woman: a hospital-based support program for women with gynecologic cancer and their families. In *Social Work in Health Settings*, page 12. Routledge, 4 edition, 2016. Num Pages: 12.

[137] Hannah Kang Moran, Joanna Veazey Brooks, and Lori Spoozak. “There’s somebody like me”: perspectives of a peer-to-peer gynecologic cancer mentorship program. *Support Care Cancer*, 29(12):7679–7686, December 2021.

[138] Catherine Walshe, Diane Roberts, Lynn Calman, Lynda Appleton, Robert Croft, Suzanne Skevington, Mari Lloyd-Williams, Gunn Grande, and Guillermo Perez Algorta. Peer support to maintain psychological wellbeing in people with advanced cancer: findings from a feasibility study for a randomised controlled trial. *BMC Palliative Care*, 19(1):129, August 2020.

[139] Andrea L Hartzler, Megan N Taylor, Albert Park, Troy Griffiths, Uba Backonja, David W McDonald, Sam Wahbeh, Cory Brown, and Wanda Pratt. Leveraging cues from person-generated health data for peer matching in online communities. *J Am Med Inform Assoc*, 23(3):496–507, May 2016.

[140] Gayathri Embuldeniya, Paula Veinot, Emma Bell, Mary Bell, Joyce Nyhof-Young, Joanna E. M. Sale, and Nicky Britten. The experience and impact of chronic disease peer support interventions: A qualitative synthesis. *Patient Education and Counseling*, 92(1):3–12, July 2013.

[141] Francesca Taylor, Robin Gutteridge, and Carol Willis. Peer support for CKD patients and carers: overcoming barriers and facilitating access. *Health Expectations*, 19(3):617–630, 2016. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.12348.

[142] Andrea Civan, David W. McDonald, Kenton T. Unruh, and Wanda Pratt. Locating patient expertise in everyday life. In *Proceedings of the ACM 2009 international conference on Supporting group work*, GROUP '09, pages 291–300, New York, NY, USA, May 2009. Association for Computing Machinery.

[143] S.c. Ieropoli, V.m. White, M. Jefford, and D. Akkerman. What models of peer support do people with colorectal cancer prefer? *European Journal of Cancer Care*, 20(4):455–465, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2354.2010.01218.x.

[144] Kathleen O'Leary, Arpita Bhattacharya, Sean A. Munson, Jacob O. Wobbrock, and Wanda Pratt. Design Opportunities for Mental Health Peer Support Technologies. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1470–1484, New York, NY, USA, 2017. ACM. event-place: Portland, Oregon, USA.

[145] Sharron Sandhu, Paula Veinot, Gayathri Embuldeniya, Sydney Brooks, Joanna Sale, Sicong Huang, Alex Zhao, Dawn Richards, and Mary J. Bell. Peer-to-peer mentoring for individuals with early inflammatory arthritis: feasibility pilot. *BMJ Open*, 3(3):e002267, January 2013. Publisher: British Medical Journal Publishing Group Section: Rheumatology.

[146] Jeremy Heyer, Zachary Schmitt, Lynn Dombrowski, and Svetlana Yarosh. Opportunities for Enhancing Access and Efficacy of Peer Sponsorship in Substance Use Disorder Recovery. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Honolulu HI USA, April 2020. ACM.

[147] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media. *ACM Trans. Comput.-Hum. Interact.*, 25(5):28:1–28:35, October 2018.

[148] Elizabeth C. Lavender, Esther Dusabe-Richards, Anna M. Anderson, Deborah Antcliff, Linda McGowan, Philip G. Conaghan, Sarah R. Kingsbury, and Gretl A. McHugh. Exploring the feasibility, acceptability and value of volunteer peer mentors in supporting self-management of osteoarthritis: a qualitative evaluation. *Disability and Rehabilitation*, 0(0):1–11, September 2021. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09638288.2021.1964625.

[149] C-L Dennis, E Hodnett, L Kenton, J Weston, J Zupancic, D E Stewart, and A Kiss. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. *BMJ*, 338(jan15 2):a3064–a3064, January 2009.

[150] Phyllis N. Butow, Laura T. Kirsten, Jane M. Ussher, Gerard V. Wain, Mirjana Sandoval, Kim M. Hobbs, Katharine Hodgkinson, and Annie Stenlake. What is the ideal support group? Views of Australian people with cancer and their carers. *Psycho-Oncology*, 16(11):1039–1045, 2007. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pon.1172.

[151] Kathryn Rehberg, Adina Fleischmann, Elana Silber, Suzanne C O'neill, Frances Marcus Lewis, and Kenneth P Tercyak. Facilitators of peer coaching/support engagement and dissemination among women at risk for and surviving with breast cancer. *Translational Behavioral Medicine*, 11(1):153–160, January 2021.

[152] Nazanin Andalibi and Madison K. Flood. Considerations in Designing Digital Peer Support for Mental Health: Interview Study Among Users of a Digital Support System (Buddy Project). *JMIR Mental Health*, 8(1):e21819, 2021. Company: JMIR Mental Health Distributor: JMIR Mental Health Institution: JMIR Mental Health Label: JMIR Mental Health Publisher: JMIR Publications Inc., Toronto, Canada.

[153] Diyi Yang. *Computational Social Roles*. PhD Thesis, Carnegie Mellon University, Language Technologies Institute, School of Computer Science, 2019.

[154] Jun Xu, Xiangnan He, and Hang Li. Deep Learning for Matching in Search and Recommendation. *INR*, 14(2–3):102–288, July 2020. Publisher: Now Publishers, Inc.

[155] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, pages 133–142, New York, NY, USA, July 2002. Association for Computing Machinery.

[156] Hongyu Lu, Min Zhang, and Shaoping Ma. Between Clicks and Satisfaction: Study on Multi-Phase User Preferences and Satisfaction for Online News Reading. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 435–444, Ann Arbor MI USA, June 2018. ACM.

[157] Hongyi Wen, Longqi Yang, and Deborah Estrin. Leveraging post-click feedback for content recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems*, RecSys '19, pages 278–286, New York, NY, USA, September 2019. Association for Computing Machinery.

[158] Steffen Rendle, Walid Krichene, Li Zhang, and John Anderson. Neural Collaborative Filtering vs. Matrix Factorization Revisited. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, pages 240–248, New York, NY, USA, September 2020. Association for Computing Machinery.

[159] Harald Steck, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, and Justin Basilico. Deep Learning for Recommender Systems: A Netflix Case Study. *AI Magazine*, 42(3):7–18, November 2021. Number: 3.

[160] K. E. Goldschmitt and Nick Seaver. Shaping the Stream: Techniques and Troubles of Algorithmic Recommendation. In David Trippett, Monique M. Ingalls, and Nicholas Cook, editors, *The Cambridge Companion to Music in Digital Culture*, Cambridge Companions to Music, pages 63–81. Cambridge University Press, Cambridge, 2019.

[161] Maxim Naumov, Dheevatsa Mudigere, Hao-Jun Michael Shi, Jianyu Huang, Narayanan Sundaraman, Jongsoo Park, Xiaodong Wang, Udit Gupta, Carole-Jean Wu, Alisson G. Azzolini, Dmytro Dzhulgakov, Andrey Mallevich, Ilia Cherniavskii, Yinghai Lu, Raghuraman Krishnamoorthi, Ansha Yu, Volodymyr Kondratenko, Stephanie Pereira, Xianjie

Chen, Wenlin Chen, Vijay Rao, Bill Jia, Liang Xiong, and Misha Smelyanskiy. Deep Learning Recommendation Model for Personalization and Recommendation Systems. *arXiv:1906.00091 [cs]*, May 2019. arXiv: 1906.00091.

[162] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. WTF: the who to follow service at Twitter. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 505–514, New York, NY, USA, May 2013. Association for Computing Machinery.

[163] Ido Guy, Sigalit Ur, Inbal Ronen, Adam Perer, and Michal Jacovi. Do you want to know? recommending strangers in the enterprise. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 285–294, New York, NY, USA, March 2011. Association for Computing Machinery.

[164] Yi Tay, Luu Tuan, and Siu Hui. CoupleNet: Paying Attention to Couples with Coupled Attention for Relationship Recommendation. *ICWSM*, 12(1), June 2018. Number: 1.

[165] Luiz Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. RECON: a reciprocal recommender for online dating. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 207–214, New York, NY, USA, September 2010. Association for Computing Machinery.

[166] Nicholas J. Belkin. Anomalous States of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science*, 5:133–143, 1980.

[167] Anne Aula. Query Formulation in Web Information Search. In *Proceedings of the IADIS International Conference WWW/Internet 2003*, page 9, 2003.

[168] Claudette Pretorius, Darragh McCashin, Naoise Kavanagh, and David Coyle. Searching for Mental Health: A Mixed-Methods Study of Young People's Online Help-seeking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery, New York, NY, USA, April 2020.

[169] Sruthi Viswanathan, Behrooz Omidvar-Tehrani, and Jean-Michel Renders. What is Your Current Mindset? In *CHI Conference on Human Factors in Computing Systems*, pages 1–17, New Orleans LA USA, April 2022. ACM.

[170] Nina Wacholder. Interactive query formulation. *Ann. Rev. Info. Sci. Tech.*, 45(1):157–196, 2011.

[171] Omayma Husain, Naomie Salim, Rose Alinda Alias, Samah Abdelsalam, and Alzubair Hassan. Expert Finding Systems: A Systematic Review. *Applied Sciences*, 9(20):4250, January 2019. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[172] Katie Kuksenok, Michael Brooks, John J. Robinson, Daniel Perry, Megan K. Torkildson, and Cecilia Aragon. Automating large-scale annotation for analysis of social media content. In *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media"(TextVis' 12)*, 2012.

[173] Joni Salminen, Hind Almerekhi, Milica Milenković, Soon-gyo Jung, Jisun An, Haewoon Kwak, and Bernard J Jansen. Anatomy of Online Hate: Developing a Taxonomy and Machine Learning Models for Identifying and Classifying Hate in Online News Media. *ICWSM*, page 10, 2018.

[174] Sonali K. Shah and Kevin G. Corley. Building Better Theory by Bridging the Quantitative-Qualitative Divide. *Journal of Management Studies*, 43(8):1821–1835, 2006.

[175] Ethan Fast, Binbin Chen, and Michael S. Bernstein. Empath: Understanding Topic Signals in Large-Scale Text. pages 4647–4657. ACM Press, 2016.

[176] R. Stuart Geiger and Aaron Halfaker. Operationalizing Conflict and Cooperation Between Automated Software Agents in Wikipedia: A Replication and Expansion of 'Even Good Bots Fight'. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):49:1–49:33, December 2017.

[177] Michael L Birnbaum, Sindhu Kiranmai Ernala, Asra F Rizvi, Munmun De Choudhury, and John M Kane. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. *J Med Internet Res*, 19(8), August 2017.

[178] Xinning Gui, Yu Chen, Yubo Kou, Katie Pine, and Yunan Chen. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):50:1–50:19, December 2017.

[179] Meredith M. Skeels, Kenton T. Unruh, Christopher Powell, and Wanda Pratt. Catalyzing social support for breast cancer patients. page 173. ACM Press, 2010.

[180] Acar Tamersoy, Munmun De Choudhury, and Duen Horng Chau. Characterizing Smoking and Drinking Abstinence from Social Media. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, pages 139–148, New York, NY, USA, 2015. ACM.

[181] James O Prochaska and Wayne F Velicer. The transtheoretical model of health behavior change. *American journal of health promotion*, 12(1):38–48, 1997.

[182] Stevie Chancellor, Tanushree Mitra, and Munmun De Choudhury. Recovery Amid Pro-Anorexia: Analysis of Recovery in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2111–2123, New York, NY, USA, 2016. ACM. event-place: San Jose, California, USA.

[183] Michael Muller, Shion Guha, Eric P.S. Baumer, David Mimno, and N. Sadat Shami. Machine Learning and Grounded Theory Method: Convergence, Divergence, and Combination. In *Proceedings of the 19th International Conference on Supporting Group Work*, GROUP '16, pages 3–8, New York, NY, USA, 2016. ACM.

[184] Kasper Hornbæk, Søren S. Sander, Javier Bargas-Avila, and Jakob Grue Simonsen. Is Once Enough? On the Extent and Content of Replications in Human-Computer Interaction. 2014.

[185] Amy S. Bruckman, Casey Fiesler, Jeff Hancock, and Cosmin Munteanu. CSCW Research Ethics Town Hall: Working Towards Community Norms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17 Companion, pages 113–115, New York, NY, USA, 2017. ACM.

[186] John Langford, Lihong Li, and Alex Strehl. Vowpal Wabbit online learning project, 2007.

[187] Amy X. Zhang, Martin Robbins, Ed Bice, Sandro Hawke, David Karger, An Xiao Mina, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, and Jennifer Lee. A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles. In

*Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, pages 603–612, Lyon, France, 2018. ACM Press.

[188] Mayara Costa Figueiredo, Clara Caldeira, Tera L. Reynolds, Sean Victory, Kai Zheng, and Yunan Chen. Self-Tracking for Fertility Care: Collaborative Support for a Highly Personalized Problem. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):36:1–36:21, December 2017.

[189] Robert Adcock and David Collier. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3):529–546, 2001.

[190] Mike Schaekermann, Joslin Goh, Kate Larson, and Edith Law. Resolvable vs. Irresolvable Disagreement: A Study on Worker Deliberation in Crowd Work. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):154:1–154:19, November 2018.

[191] J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977.

[192] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier Chains for Multi-label Classification. In Wray Buntine, Marko Grobelnik, Dunja Mladenić, and John Shawe-Taylor, editors, *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science, pages 254–269. Springer Berlin Heidelberg, 2009.

[193] Alina Beygelzimer, John Langford, and Bianca Zadrozny. Weighted one-against-all. In *Proc. of AAAI'05*, pages 720–725, 2005.

[194] Jeremy Howard and Sebastian Ruder. Universal Language Model Fine-tuning for Text Classification. *arXiv:1801.06146 [cs, stat]*, January 2018. arXiv: 1801.06146.

[195] Xin Li and Yuhong Guo. Active Learning with Multi-Label SVM Classification. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, page 7, 2013.

[196] Uriel Feige. A Threshold of Ln N for Approximating Set Cover. *J. ACM*, 45(4):634–652, July 1998.

[197] Jean Metz, Luís F. D. de Abreu, Everton A. Cherman, and Maria C. Monard. On the Estimation of Predictive Evaluation Measure Baselines for Multi-label Learning. In Juan Pavón, Néstor D. Duque-Méndez, and Rubén Fuentes-Fernández, editors, *Advances in Artificial Intelligence – IBERAMIA 2012*, Lecture Notes in Computer Science, pages 189–198. Springer Berlin Heidelberg, 2012.

[198] Brendan O'Connor, David Bamman, and Noah A. Smith. Computational text analysis for social science: Model assumptions and complexity. In *Proc. of the NeurIPS Workshop on Computational Social Science and the Wisdom of Crowds*, Sierra Nevada, Spain, December 2011.

[199] Nan-Chen Chen, Margaret Drouhard, Rafal Kocielnik, Jina Suh, and Cecilia R. Aragon. Using Machine Learning to Support Qualitative Coding in Social Science: Shifting the Focus to Ambiguity. *ACM Trans. Interact. Intell. Syst.*, 8(2):9:1–9:20, June 2018.

[200] Meredith McLain Skeels. *Sharing by Design: Understanding and Supporting Personal Health Information Sharing and Collaboration within Social Networks*. ProQuest LLC, 2010.

[201] Terrance L. Albrecht and Mara B. Adelman. *Communicating social support*. Sage Publications, July 1987. Google-Books-ID: nhtHAAAAMAAJ.

[202] Rob Cross, Andrew Parker, Laurence Prusak, and Stephen P. Borgatti. *Knowing What We Know: Supporting Knowledge Creation and Sharing in Social Networks*. Oxford University Press, January 2004.

[203] Rosaleen D. Bloom, Susan Beck, Wen-Ying Sylvia Chou, Maija Reblin, and Lee Ellington. In Their Own Words: Experiences of Caregivers of Adults With Cancer as Expressed on Social Media. *Oncology Nursing Forum*, 46(5):617–630, September 2019. Publisher: Oncology Nursing Society.

[204] Mukund Jha and Noemie Elhadad. Cancer Stage Prediction Based on Patient Online Discourse. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, ACL 2010*, pages 64–71, Sweden, 2010.

[205] Eunkyung Kim, Jeong Yeob Han, Tae Joon Moon, Bret Shaw, Dhavan V. Shah, Fiona M. McTavish, and David H. Gustafson. The process and effect of supportive message expression and reception in online breast cancer support groups. *Psycho-Oncology*, 21(5):531–540, May 2012.

[206] Thomas A. Wills. Downward comparison principles in social psychology. *Psychological Bulletin*, 90(2):245–271, 1981.

[207] Christine Dunkel-Schetter. Social Support and Cancer: Findings Based on Patient Interviews and Their Implications. *Journal of Social Issues*, 40(4):77–98, January 1984.

[208] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[209] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[210] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[211] Stefan van der Walt, S. Chris Colbert, and Gael Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2):22–30, March 2011.

[212] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[213] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[214] Yves Croissant. *mlogit: Multinomial Logit Models*, 2019. R package version 1.0-2.

[215] Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia, 2018. R package version 5.2.2.

[216] Marti A. Hearst and Susan T. Dumais. Blogging Together: An Examination of Group Blogs. In *Third International AAAI Conference on Weblogs and Social Media*, March 2009.

[217] Sida Wang and Christopher Manning. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 90–94, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[218] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, January 2008.

[219] Joshua Introne, Bryan Semaan, and Sean Goggins. A Sociotechnical Mechanism for Online Support Provision. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3559–3571, New York, NY, USA, 2016. ACM. event-place: San Jose, California, USA.

[220] Karen Urbanoski, Trevor van Mierlo, and John Cunningham. Investigating Patterns of Participation in an Online Support Group for Problem Drinking: a Social Network Analysis. *Int.J. Behav. Med.*, 24(5):703–712, October 2017.

[221] Wenjing Pan, Cuihua Shen, and Bo Feng. You Get What You Give: Understanding Reply Reciprocity and Social Capital in Online Health Support Forums. *Journal of Health Communication*, 22(1):45–52, January 2017.

[222] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and Inter-rater Reliability in Qualitative Research. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), November 2019.

[223] Jan Overgoor, Austin R Benson, and Johan Ugander. Choosing to Grow a Graph: Modeling Network Formation as Discrete Choice. page 12, 2019.

[224] Peggy A. Thoits. Stress and Health: Major Findings and Policy Implications. *J Health Soc Behav*, 51(1_suppl):S41–S53, March 2010.

[225] Mark S. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

[226] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. IP Geolocation Databases: Unreliable? *SIGCOMM Comput. Commun. Rev.*, 41(2):53–56, April 2011.

[227] Amy Bruckman. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology*, 4(3):217–231, September 2002.

[228] Ian Ruthven. Making Meaning: A Focus for Information Interactions Research. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, CHIIR '19, pages 163–171, Glasgow, Scotland UK, March 2019. Association for Computing Machinery.

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

[230] Nicholas Iannarino. *Social Support in Young Adult Cancer Survivors and Their Close Social Network Members*. PhD Thesis, University of Kentucky, January 2014.

[231] Pamara F. Chang, Janis Whitlock, and Natalya N. Bazarova. "To Respond or not to Respond, that is the Question": The Decision-Making Process of Providing Social Support to Distressed Posters on Facebook:. *Social Media + Society*, February 2018. Publisher: SAGE PublicationsSage UK: London, England.

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

[233] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. Make New Friends, but Keep the Old: Recommending People on Social Networking Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 201–210, New York, NY, USA, 2009. ACM. event-place: Boston, MA, USA.

[234] Brian Lakey and Catherine J. Lutz. Social Support and Preventive and Therapeutic Interventions. In Gregory R. Pierce, Barbara R. Sarason, and Irwin G. Sarason, editors, *Handbook of Social Support and the Family*, The Springer Series on Stress and Coping, pages 435–465. Springer US, Boston, MA, 1996.

[235] Gunther Eysenbach, John Powell, Marina Englesakis, Carlos Rizo, and Anita Stern. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ*, 328(7449):1166, May 2004.

[236] Kevin B. Wright and Claude H. Miller. A Measure of Weak-Tie/Strong-Tie Support Network Preference. *Communication Monographs*, 77(4):500–517, December 2010.

[237] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. Moderator engagement and community development in the age of algorithms. *New Media & Society*, 21(7):1417–1443, July 2019.

[238] Manya Sleeper, William Melicher, Hana Habib, Lujo Bauer, Lorrie Faith Cranor, and Michelle L. Mazurek. Sharing Personal Content Online: Exploring Channel Choice and Multi-Channel Behaviors. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 101–112, Santa Clara, California, USA, 2016. ACM Press.

[239] Yunan Chen, Victor Ngo, and Sun Young Park. Caring for Caregivers: Designing for Integrality. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 91–102, New York, NY, USA, 2013. ACM. event-place: San Antonio, Texas, USA.

[240] Jeong Yeob Han, Jiran Hou, Eunkyung Kim, and David H. Gustafson. Lurking as an Active Participation Process: A Longitudinal Investigation of Engagement with an Online Cancer Support Group. *Health Commun*, 29(9):911–923, 2014.

[241] Eszter Hargittai. Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*, 38(1):10–24, July 2018.

[242] Matthew O. Jackson. *Social and Economic Networks*. November 2010.

[243] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, 40(2):211–239, May 2011.

[244] Moira Burke and Robert E. Kraut. Growing Closer on Facebook: Changes in Tie Strength Through Social Network Site Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 4187–4196, New York, NY, USA, 2014. ACM.

[245] Wade Hill, Clarann Weinert, and Shirley Cudney. Influence of a Computer Intervention on the Psychological Status of Chronically Ill Rural Women. *Nurs Res*, 55(1):34–42, 2006.

[246] Fredrick D Ashbury, Cathy Cameron, Shawna L Mercer, Margaret Fitch, and Eleanor Nielsen. One-on-one peer support and quality of life for breast cancer patients. *Patient Education and Counseling*, 35(2):89–100, October 1998.

[247] Debra Umberson and Jennifer Karas Montez. Social Relationships and Health: A Flashpoint for Health Policy. *J Health Soc Behav*, 51(1_suppl):S54–S66, March 2010.

[248] Eva Sharma and Munmun De Choudhury. Mental Health Support and Its Relationship to Linguistic Accommodation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 641:1–641:13, New York, NY, USA, 2018. ACM. event-place: Montreal QC, Canada.

[249] Jacqueline L Bender, Maria-Carolina Jimenez-Marroquin, and Alejandro R Jadad. Seeking Support on Facebook: A Content Analysis of Breast Cancer Groups. *Journal of Medical Internet Research*, 13(1):e16, February 2011.

[250] Alyson L. Young and Andrew D. Miller. "This Girl is on Fire": Sensemaking in an Online Health Community for Vulvodynia. In *Proceedings of the 2019 CHI Conference*

*on Human Factors in Computing Systems*, CHI '19, pages 129:1–129:13, New York, NY, USA, 2019. ACM. event-place: Glasgow, Scotland Uk.

[251] Andrea L. Hartzler, Bridget Weis, Carly Cahill, Wanda Pratt, Albert Park, Uba Backonja, and David W. McDonald. Design and Usability of Interactive User Profiles for Online Health Communities. *ACM Trans. Comput.-Hum. Interact.*, 23(3):15:1–15:33, June 2016.

[252] Jessica Su, Krishna Kamath, Aneesh Sharma, Johan Ugander, and Sharad Goel. An Experimental Study of Structural Diversity in Social Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 14:661–670, May 2020.

[253] Loren Terveen and David W. McDonald. Social matching: A framework and research agenda. *ACM Trans. Comput.-Hum. Interact.*, 12(3):401–434, September 2005.

[254] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, USA, June 2009. AUAI Press.

[255] Zahra Nazari, Praveen Chandar, Ghazal Fazelnia, Catherine M. Edwards, Benjamin Carterette, and Mounia Lalmas. Choice of Implicit Signal Matters: Accounting for User Aspirations in Podcast Recommendations. In *Proceedings of the ACM Web Conference 2022*, WWW '22, pages 2433–2441, New York, NY, USA, April 2022. Association for Computing Machinery.

[256] Qian Zhao, F. Maxwell Harper, Gediminas Adomavicius, and Joseph A. Konstan. Explicit or implicit feedback? engagement or satisfaction?: a field experiment on machine-learning-based recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 1331–1340, Pau France, April 2018. ACM.

[257] Rafael del Pino-Casado, Antonio Frías-Osuna, Pedro A. Palomino-Moral, María Ruzafa-Martínez, and Antonio J. Ramos-Morcillo. Social support and subjective burden in caregivers of adults and older adults: A meta-analysis. *PLOS ONE*, 13(1):e0189874, January 2018. Publisher: Public Library of Science.

[258] Jingtao Ding, Yuhan Quan, Quanming Yao, Yong Li, and Depeng Jin. Simplify and Robustify Negative Sampling for Implicit Collaborative Filtering. *Advances in Neural Information Processing Systems*, 33:1094–1105, 2020.

[259] Ji Yang, Xinyang Yi, Derek Zhiyuan Cheng, Lichan Hong, Yang Li, Simon Xiaoming Wang, Taibai Xu, and Ed H. Chi. Mixed Negative Sampling for Learning Two-tower Neural Networks in Recommendations. In *Companion Proceedings of the Web Conference 2020*, pages 441–447, Taipei Taiwan, April 2020. ACM.

[260] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [cs]*, July 2019. arXiv: 1907.11692.

[261] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics.

[262] Jimmy Lin, Rodrigo Nogueira, and Andrew Yates. Pretrained Transformers for Text Ranking: BERT and Beyond. *arXiv:2010.06467 [cs]*, October 2020. arXiv: 2010.06467.

[263] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.

[264] Zachary Levonian, Chia-Jung Lee, Vanessa Murdock, and F. Maxwell Harper. Trade-offs in Sampling and Search for Early-stage Interactive Text Classification. In *27th International Conference on Intelligent User Interfaces*, IUI '22, pages 566–583, New York, NY, USA, March 2022. Association for Computing Machinery.

[265] Joeran Beel and Stefan Langer. A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems. In Sarantos Kapidakis, Cezary Mazurek, and Marcin Werla, editors, *Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science, pages 153–168, Cham, 2015. Springer International Publishing.

[266] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 257–260, New York, NY, USA, September 2010. Association for Computing Machinery.

[267] Ido Guy, Inbal Ronen, and Eric Wilcox. Do You Know? Recommending People to Invite into Your Social Network. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, page 10, Sanibel Island, Florida, USA, February 2009. ACM.

[268] Yitong Ji, Aixin Sun, Jie Zhang, and Chenliang Li. A Re-visit of the Popularity Baseline in Recommender Systems. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1749–1752, July 2020. arXiv: 2005.13829.

[269] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[270] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[271] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. Publisher: Springer Science and Business Media LLC.

[272] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[273] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. Publisher: IEEE COMPUTER SOC.

[274] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[275] Kathy Charmaz. *Constructing grounded theory: a practical guide through qualitative analysis*. Sage, London, 2006. OCLC: 898127973.

[276] Miguel A. Hernán and J. M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC., Boca Raton, 1 edition, 2020.

[277] Heejung Bang and James M. Robins. Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, 61(4):962–973, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2005.00377.x.

[278] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007.

[279] F. Maxwell Harper, Shilad Sen, and Dan Frankowski. Supporting social recommendations with activity-balanced clustering. In *Proceedings of the 2007 ACM conference on Recommender systems*, RecSys '07, pages 165–168, New York, NY, USA, October 2007. Association for Computing Machinery.

[280] Krzysztof Z. Gajos and Lena Mamykina. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *27th International Conference on Intelligent*

*User Interfaces*, IUI ’22, pages 794–806, New York, NY, USA, March 2022. Association for Computing Machinery.

[281] Drashko Nakikj and Lena Mamykina. A Park or A Highway: Overcoming Tensions in Designing for Socio-emotional and Informational Needs in Online Health Communities. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’17, pages 1304–1319, New York, NY, USA, 2017. ACM.

[282] Jessica Su, Aneesh Sharma, and Sharad Goel. The Effect of Recommendations on Network Structure. In *Proceedings of the 25th International Conference on World Wide Web*, WWW ’16, pages 1157–1167, Republic and Canton of Geneva, CHE, April 2016. International World Wide Web Conferences Steering Committee.

[283] Drashko Nakikj and Lena Mamykina. Lost in Migration: Information Management and Community Building in an Online Health Community. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, pages 1–14, Montreal QC, Canada, April 2018. Association for Computing Machinery.

[284] Tiffany C Veinot, Hannah Mitchell, and Jessica S Ancker. Good intentions are not enough: how informatics interventions can worsen inequality. *Journal of the American Medical Informatics Association*, 25(8):1080–1088, August 2018.

[285] Iván Palomares, James Neve, Carlos Porcel, Luiz Pizzato, Ido Guy, and Enrique Herrera-Viedma. Reciprocal Recommender Systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion*, 69:103–127, May 2021.

[286] Ekaterina R. Stepanova, John Desnoyers-Stewart, Kristina Höök, and Bernhard E. Riecke. Strategies for Fostering a Genuine Feeling of Connection in Technologically Mediated Systems. In *CHI Conference on Human Factors in Computing Systems*, CHI ’22, pages 1–26, New York, NY, USA, April 2022. Association for Computing Machinery.

[287] Stephen J. Lepore, Joanne S. Buzaglo, Morton A. Lieberman, Mitch Golant, Judith R. Greener, and Adam Davey. Comparing Standard Versus Prosocial Internet Support

Groups for Patients With Breast Cancer: A Randomized Controlled Trial of the Helper Therapy Principle. *J Clin Oncol*, 32(36):4081–4086, December 2014.

[288] Richard M. Lee and Steven B. Robbins. Measuring belongingness: The Social Connectedness and the Social Assurance scales. *Journal of Counseling Psychology*, 42(2):232–241, 1995.

[289] Eun-Hyun Lee. Review of the Psychometric Evidence of the Perceived Stress Scale. *Asian Nursing Research*, 6(4):121–127, December 2012.

[290] Centers for Disease Control and Prevention. Measuring Healthy Days. Technical report, CDC, Atlanta, GA, November 2000.

[291] Nick Seaver. Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 24(4):421–436, December 2019. Publisher: SAGE Publications Ltd.

[292] Oliver Haimson. Social Media As Social Transition Machinery. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):63:1–63:21, November 2018.

[293] Bernd Ploderer, Wally Smith, Steve Howard, Jon Pearce, and Ron Borland. Patterns of support in an online community for smoking cessation. In *Proceedings of the 6th International Conference on Communities and Technologies - C&T '13*, pages 26–35, Munich, Germany, 2013. ACM Press.

[294] Oliver L. Haimson, Tianxiao Liu, Ben Zefeng Zhang, and Shanley Corvite. The Online Authenticity Paradox: What Being "Authentic" on Social Media Means, and Barriers to Achieving It. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):423:1–423:18, October 2021.

[295] Jeong Yeob Han, Jung-Hyun Kim, Hye Jin Yoon, Minsun Shim, Fiona M. McTavish, and David H. Gustafson. Social and Psychological Determinants of Levels of Engagement with an Online Breast Cancer Support Group: Posters, Lurkers, and Non-Users. *J Health Commun*, 17(3):356–371, 2012.

[296] Morgan Klaus Sheuerman, Katta Spiel, Oliver L. Haimson, Foad Hamidi, and Stacy M. Branham. HCI Guidelines for Gender Equity and Inclusivity, v1.1, May 2020.

[297] Olivia K. Richards, Adrian Choi, and Gabriela Marcu. Shared Understanding in Care Coordination for Children's Behavioral Health. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1):1–25, April 2021.

[298] Dean Eckles and Eytan Bakshy. Bias and High-Dimensional Adjustment in Observational Studies of Peer Effects. *Journal of the American Statistical Association*, 116(534):507–517, April 2021. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2020.1796393.

[299] Aaron Halfaker, Os Keyes, Daniel Kluver, Jacob Thebault-Spieker, Tien Nguyen, Kenneth Shores, Anuradha Uduwage, and Morten Warncke-Wang. User Session Identification Based on Strong Regularities in Inter-activity Time. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 410–418, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee. event-place: Florence, Italy.

[300] Maia Jacobs, Henriette Cramer, and Louise Barkhuus. Caring About Sharing: Couples' Practices in Single User Device Access. In *Proceedings of the 19th International Conference on Supporting Group Work*, GROUP '16, pages 235–243, New York, NY, USA, 2016. ACM. event-place: Sanibel Island, Florida, USA.

[301] Airi M I Lampinen. Account Sharing in the Context of Networked Hospitality Exchange. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 499–504, New York, NY, USA, 2014. ACM. event-place: Baltimore, Maryland, USA.

[302] Zachary C. Lipton, Yu-Xiang Wang, and Alex Smola. Detecting and Correcting for Label Shift with Black Box Predictors. *arXiv:1802.03916 [cs, stat]*, February 2018. arXiv: 1802.03916.

[303] R. Stuart Geiger, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. Garbage in, garbage out? do machine learning application papers in social computing report where human-labeled training data comes from? In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, pages 325–336, Barcelona, Spain, January 2020. Association for Computing Machinery.

[304] Leslie N. Smith and Nicholay Topin. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates, May 2018. Number: arXiv:1708.07120 arXiv:1708.07120 [cs, stat].

[305] Prabhu Teja Sivaprasad, Florian Mai, Thijs Vogels, Martin Jaggi, and François Fleuret. Optimizer Benchmarking Needs to Account for Hyperparameter Tuning, August 2020. Number: arXiv:1910.11758 arXiv:1910.11758 [cs, stat].

[306] Casey Fiesler and Nicholas Proferes. "Participant" Perceptions of Twitter Research Ethics:. *Social Media + Society*, March 2018. Publisher: SAGE PublicationsSage UK: London, England.

[307] Jean Garcia-Gathright, Christine Hosey, Brian St. Thomas, Ben Carterette, and Fernando Diaz. Mixed methods for evaluating user satisfaction. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 541–542, Vancouver British Columbia Canada, September 2018. ACM.

[308] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. Click Models for Web Search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(3):1–115, July 2015. Publisher: Morgan & Claypool Publishers.

[309] Thorsten Joachims, Ben London, Yi Su, Adith Swaminathan, and Lequn Wang. Recommendations as Treatments. *AI Magazine*, 42(3):19–30, November 2021. Number: 3.

# Appendix A

# Author connection details

This appendix contains details that provide additional context and analysis of potential confounders.

## A.1 Valid author identification

We omitted authors without at least two journal updates published more than 24 hours apart based on an analysis of author *tenure*—the amount of time between an author's first and last published updates. Figure A.1 shows the distribution of tenure for all authors. We observe a bimodal lognormal distribution of author tenure and fit a two-component log-normal GMM in an approach adapted from Halfaker et al. [299]. The lines overlayed on the histogram show the fit of the two Gaussian components to the tenure data. We use the approximate visual intersection of the two GMM components, 24 hours, as a criterion for being a valid author. Self-reported health conditions by valid authors—which are used as features in the models fit to address RQ1—are shown in Table A.1.

## A.2 Account sharing

Author accounts classified as Mixed (7.49% of authors, see Section 4.4.2) may indicate either a single author embodying multiple roles or multiple people sharing the same account credentials. Such account sharing generally occurs for convenience in the case of a trusted relationship

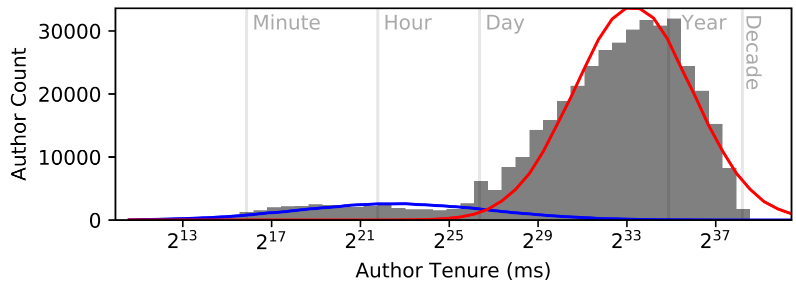



**Figure A.1: Distribution of author tenure—the time between first and last journal update written by that author—for all 572,309 author users. The 161,799 (28.3%) authors with only a single update (and thus 0 tenure) are not shown. Median author tenure is 83 days (mean 311 days).**

| Health-condition category | Count | % |
|---|---|---|
| None (not reported) | 152,818 | 42.17% |
| Cancer | 109,339 | 30.18% |
| Other | 37,556 | 10.36% |
| Surgery/Transplantation | 15,415 | 4.25% |
| Injury | 12,910 | 3.56% |
| Cardiovascular/Stroke | 12,685 | 3.50% |
| Neurological Condition | 9,376 | 2.59% |
| Infant/Childbirth | 7,952 | 2.19% |
| Condition Unknown | 2,252 | 0.62% |
| Congenital/Immune Disorder | 2,042 | 0.56% |

**Table A.1: Health condition assignments to valid authors based on site-level self-reports.**

between the patient and the caregiver [300]. Account sharing is potentially problematic for analyses treating interactions between accounts as interactions between two people, but particularly so if an account is shared by both a patient and a caregiver. Thus, we classify an author account as Shared if, on any site, between one third and two thirds of that author's updates are classified as patient-authored. Using this conservative definition, we find only 7.46% of accounts are Shared.[1] Author account sharing is closely linked to the Mixed classification: 95.9% of Mixed authors are Shared, suggesting that authors only rarely embody multiple roles e.g. writing as a patient author on one site and as a caregiver author on another. Due to the high-proportion of Mixed-author accounts that are shared, in subsequent modeling we include as an author feature only the Mixed author role (as a dummy-coded categorical variable) and not a separate indicator variable of account sharing. We note that authors classified as Mixed are likely multiple people using the same account.

One implication of this analysis is that user accounts are nuanced and the assumption of one account being associated with one person or even one role is frequently mistaken. Further work on roles must grapple with the reality of user account sharing and the challenges it presents to both analysts and users [301]. For designers in particular, a "one person, one account" assumption may undermine the effectiveness of designed interventions, e.g. recommended articles to edit on Wikipedia or personalized social media feeds.

## A.3 Computing patient-authored update proportion

We computed the proportion of patient-authored updates on valid sites using a random sample of 5,000 unlabeled journal updates. We used the 305 models trained during hold-one-out cross validation in order to compute standard error as an estimate of the variability of this proportion. The model predicted that 24.84% (s.e. 0.03%) of unlabeled journal updates were patient-authored. As the label distribution in the training data is different from the label distribution over the target updates, we need to correct for this distribution shift as it will bias the estimate towards the balanced training distribution. We use Black Box Shift Estimation [302] to quantify the shift in distribution and produce a revised estimate, finding that 22.06% (s.e. 0.11%) of unlabeled journal updates were patient-authored.

[1] A more permissive definition labels author accounts as Shared if on any site that author has published both a Patient-classified and a Caregiver-classified update. While this definition—which labels 53.4% of accounts as Shared—likely captures primarily classifier noise, it can be treated as an upper bound on author account sharing.

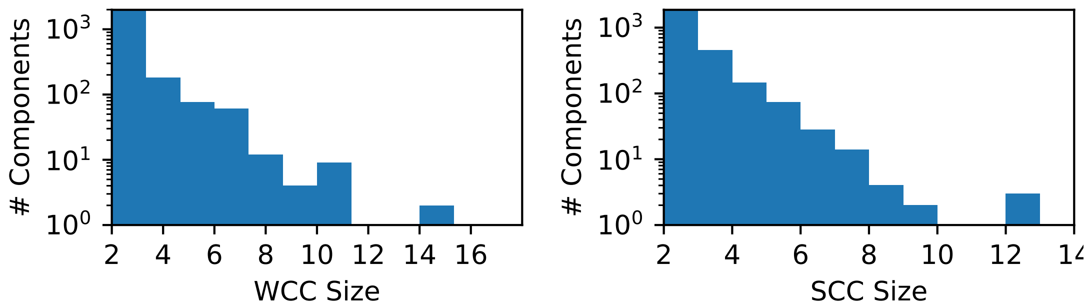



**Figure A.2: Distribution of the sizes of the 2335 weakly connected components (WCCs) and 2590 strongly connected components (SCCs) composed of two or more active authors at the end of the analysis period. The largest WCC (size=45038) and largest SCC (size=16946) are not shown. Active authors are valid authors who were active on CB within six months of the end of the analysis period.**

## A.4 Assumption analysis: Amp timestamps

Amps ("likes" on CB) lack timestamp information, so we assume that amps occur at the publication time of the associated journal update. To assess whether this assumption is reasonable, we examine the moment when the amps feature was introduced, reasoning that amps on journal updates published before the amps feature launched indicate a lag time between the update publication and the amp interaction. Only 0.32% of amps occur on updates published before the launch of the amps feature, and comparing updates published the week before the launch date to the updates published in the week after, only 23.1% of amps are recorded pre-launch. This analysis suggests that the majority of amps are given in the week the journal update was published.

## A.5 Interaction network details

At the end of the initiations period, Figure A.2 shows the distribution of the connected components, excluding the largest. Figure A.3 shows the proportions of active authors in various network positions throughout the data range.

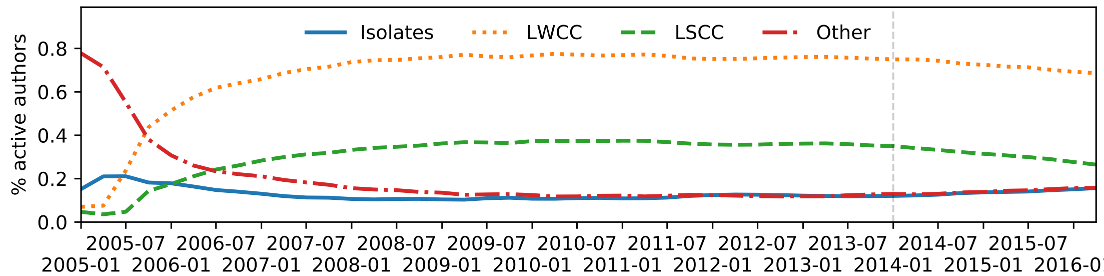



**Figure A.3: Proportion of active authors—authors with public activity on CB within 6 months of the sampled date—based on their position within the network. "Isolates" is the proportion of authors unconnected to any other author. "LWCC" is the proportion of authors in the largest weakly connected component. "Other" is the proportion of authors weakly connected to at least one other user but not in the LWCC. Together, these three account for all active authors. For comparison, we also show the proportion of authors in the largest strongly connected component ("LSCC"). The vertical dashed line indicates the beginning of the initiations period.**

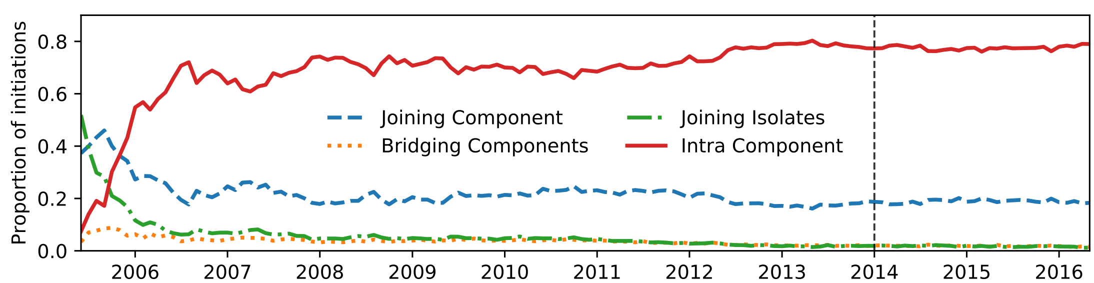



**Figure A.4: Proportion of each initiation type over time. Initiation types are as defined by Gallagher et al. [5]. The vertical dashed line indicates the beginning of the initiations period.**

## A.6 Initiation type classification & network growth

In order to understand how initiations relate to the growth of the network over time, we classified initiations according to the initiator's and the receivers' position within the network. We identify four different initiation types, adapting definitions from Gallagher et al. to our context [5]: (1) *Joining Component* is an initiation connecting an unconnected author to an existing weakly-connected component. (2) *Bridging Component* is an initiation connecting two weakly-connected components, merging them. (3) *Joining Isolates* is an initiation that connects two previously unconnected authors (i.e. two "isolates"). (4) *Intra Component* is an initiation between two authors who were already in the same weakly-connected component. By definition, all reciprocations are Intra Component initiations. We classified all initiations as one of these four types.

Figure A.4 shows the proportion of each type of initiation over time. Before the start of the initiations period in 2014, the network formed through Joining Isolates and Joining Component initiations before the majority of initiations became Intra Component. Within the initiations period, 201,188 initiations were made, and over this period the proportion of each type remained quite consistent. Bridging Component and Joining Isolates combined make up only 3.62% of the initiations in the initiations period. The vast majority of initiations are between authors in the largest weakly-connected component (WCC) and other members of the largest WCC or previously unconnected authors. 10.79% of initiations in the initiations period are reciprocal, which is 13.91% of the Intra Component initiations. The vast majority of Intra Component initiations are within the largest WCC specifically; only 3.5% of Intra Component initiations involved components other than the largest. Of the Joining Component initiations, 24.08% of initiations-period initiations are initiated by the unconnected author and not someone in the component. These proportions suggest that the majority of initiations grow the largest WCC or occur within the largest WCC. Overall, the network factor most associated with new initiations is the largest WCC.

## A.7 Other interaction types

In the author interaction network, we include only guestbooks, amps, and comments, as described in Section 4.4.3. However, we omit three types of interaction that may behave differently than the included interaction types: (1) author visits to other sites, (2) comments on other comments, and (3) explicit or text-based links to other sites in journal updates. First, we omit visits because our visit data is incomplete and can be only tenuously linked to specific authors at specific times and is invisible to the receiving site's authors. Second, we omit comments left in response to guestbooks and update comments as the feature is relatively recent and minimally used. Third, inter-site HTML links can be published in the text of journal updates. To assess the impact of excluding such text-based inter-site links on the validity of the resulting network, we conducted a high-precision analysis of existing inter-site links. Direct hyperlinks were extracted from the 19M journal updates. 101,923 valid links to CB sites were identified, of which 32.2% (32,767) were determined to be self-links i.e. they linked to the site on which the update was authored. Of the remaining 69,156 inter-site link interactions, 100% were found to be redundant with existing interactions recorded via the other interaction types between the update author and the linked site. Thus, we conclude that the three interaction types used provide a sufficiently detailed view of the inter-author interaction network.

## A.8 Initiation annotation details

Initiation annotation (see section 4.5.1) proceeded in four rounds. Both annotators are authors of our published paper [10] and familiar with the CaringBridge dataset. In the first round, 30 guestbooks and 30 comments were coded and discussed to establish a codebook (or "coding scheme" [303]); these 60 initiations were discarded from further consideration. Each subsequent round consisted of sampling an equal number of guestbooks and comments, two coders independently annotating them, and meeting to discuss disagreements and update the codebook. 200 initiations were sampled in the second and third rounds and 400 in the final round for a total of 800 initiations.

As the codebook was not intended to be generalizable beyond this specific context and the relation codes were an open set, we did not compute a statistical measure of IRR and instead resolved all disagreements via discussion [222]. Due to the inherent subjectivity in the annotation

| Relation Category | Post-health-event | Pre-health-event | Unknown | Total |
|---|---|---|---|---|
| Unknown | 7 | 67 | 437 | 511 |
| Friend | 1 | 118 | 10 | 129 |
| Third-party connection | 35 | 3 | 9 | 47 |
| CG of similar patient | 24 | 0 | 1 | 25 |
| Family | 1 | 20 | 1 | 22 |
| Other | 7 | 10 | 1 | 18 |
| One-time visitor | 15 | 1 | 0 | 16 |
| Coworker or Schoolmate | 0 | 10 | 0 | 10 |
| Fellow patient | 7 | 0 | 0 | 7 |
| (No text) | - | - | - | 15 |
| Total | 97 (12.1%) | 229 (28.6%) | 459 (57.4%) | 800 |

**Table A.2: Annotated initiation counts, broken down by the two annotation types: "Relation Category", meaning the relation between the initiator and the receiver, and whether this tie existed before the health event that is the focus of the CB site. Fifteen initiations containing only whitespace characters were not annotated but are included in the total.**

task, disagreements were relatively common; for the annotation of tie formation timing relative to the health event, raw agreement at the end of the second, third, and fourth round of annotation was 77.5%, 70.0%, and 76.8% respectively. All disagreements were resolved quickly and centered on when sufficient evidence is present in the initiation to assign a non-Unknown label. Due to the lack of context, guestbooks are notably harder to annotate than comments. The codebook is available in the GitHub repository.[2]

Result details beyond those in section 4.6.1 are presented here. Table A.2 shows the annotated initiation counts, broken down by high-level relationship category. Categories were selected after annotation to summarize the data at a higher level than the raw codes e.g. a "listserv contact" becomes "Other" and multiple subtypes become "Friend". Table A.3 shows representative initiations along with their annotated values. Quotes are paraphrased to preserve poster anonymity and reduce traceability, which we deem ethically appropriate given the sensitive context [4, 227].

[2] `https://github.com/levon003/cscw-caringbridge-interaction-network`

| Relation Category | Pre/post? | Initiation Text |
|---|---|---|
| CG of similar patient | Post | Hi John, Our son Tommy, 16, is in the room next door. He is day+14. We are sending you lots of love, prayers and positive thoughts through the walls. We hope everyday you get a little stronger and feel better. |
| Third-party connection | Post | Hi Don. I am a friend of Ben's and through him I've been following your journey since last June. I just want you to know that countless prayers have been said for you, your family and the doctors treating you. I am thankful that Danny started this site so that we can all encourage you every step of the way. |
| One-time visitor | Post | Diya, I know we have never met but tonight my heart and prayers go out to you. I traveled the road you are on seven years ago. Ella and the breast cancer support group, friends and family were my strength. Be strong. My story is under (CB site name) in Caring Bridge. If you ever need anyone to talk to, please call me. Anytime. |
| Fellow patient | Post | Congratulations on Day +2! I am the friend of Jenna's who also has multiple myeloma. Today is Day +84 for me. The next 30 days will be the toughest for you but try to walk and eat as much as you can to encourage all those little stem cells to grow! Sara Jones ((CB site name) on CaringBridge) |
| Unknown | Post | Patel family, I just read about your son in an article written by (local journalist). I had no idea. I am sending you prayers and positive thoughts. God bless you all. |
| Friend | Pre | Sarah, Dan and I are so grateful to have this connection to you through Caring Bridge. Our prayers have been winging your way since we heard the news on your hospitalization. We are traveling home tomorrow and will be back in church this Sunday. We are holding you in our hearts. We love you, Molly and Dan |
| Unknown | Unknown | You are all in our thoughts on this wonderful day. Abigail |

**Table A.3: Annotated initiation exemplars. A sample of paraphrased initiations and the annotations: (a) relation between the initiator and the receiver and (b) if the initiator was deemed to be a pre-health-event or a post-health-event connection. All names are aliases.**

## A.9 RQ1b model details

Table A.4 presents two linear regression models predicting the time between an author's first update and first initiation. Model #1, for pre-authorship initiators, demonstrates that reasoning about the high-variance relationship between initiations and going on to become an author is extremely high variance, perhaps because the health event that precipitates the creation of a site has not yet occurred. The post-authorship initiation model demonstrates the importance of receiving an interaction and also that multi-site authors initiate much later than single-site authors.

## A.10 Geographic model

Table A.5 show the full model details comparing the conditional mlogit model for initiations (see section 4.5.5) with a model fit using only the subset of initiations between users that are assigned US states (see section 4.5.5). Model (1) is the full model on all the initiations in the initiations period. Model (2) includes only the subset of authors with state assignments. Model (3) is that same subset with an additional dummy variable indicating matching state assignment between the initiator and the candidate. Comparing (1) and (2) demonstrates that this author subset is broadly similar in initiation factors compared to the full author sample, while (3) demonstrates the importance of matching state assignments. Note that a matching state assignment is less important than the network-based features.

## A.11 Right-censoring and survival analysis

In addressing RQ2b, we predict the number of interactions in a relationship, rather than relationship duration (section 4.5.7). We avoid doing a survival analysis due to long gaps between author interactions on CB making it hard to predict right-censoring. Simulating an end of the dataset 6 months earlier than the true end and assuming that any author with a published update or interaction within 6 months of the simulated dataset end is right-censored, we miss more than half of the authors that are actually censored (recall = 0.495); this level of inaccuracy occurs despite using a censorship threshold that is twice that used in prior work [70].

This difficulty leads us to avoid fitting survival models to predict relationship duration.

| | (1) Pre-authorship | (2) Post-authorship |
|---|---|---|
| Intercept | 8.049** | 25.696** |
| | (0.139) | (0.355) |
| Role = Mixed | -0.568 | 1.572* |
| | (0.415) | (0.654) |
| Role = P | -1.33** | -3.926** |
| | (0.248) | (0.412) |
| HC = Cancer | 0.241 | 3.449** |
| | (0.234) | (0.374) |
| HC = Cardiovascular/Stroke | 0.597 | 3.172** |
| | (0.526) | (0.859) |
| HC = Condition Unknown | -3.106 | 20.102** |
| | (7.176) | (2.574) |
| HC = Congenital/Immune Disorder | -1.197 | -3.44 |
| | (1.275) | (1.997) |
| HC = Infant/Childbirth | -1.356 | 4.551** |
| | (0.886) | (1.089) |
| HC = Injury | 0.38 | 5.433** |
| | (0.611) | (0.936) |
| HC = Neurological Condition | 0.405 | 6.293** |
| | (0.657) | (1.038) |
| HC = Other | -1.403 | 19.134** |
| | (0.813) | (0.695) |
| HC = Surgery/Transplantation | 0.426 | 8.003** |
| | (0.609) | (0.86) |
| Will become multi-site author? | -0.845 | |
| | (0.438) | |
| Is multi-site author? | | 12.671** |
| | | (0.979) |
| Int received? | | -10.475** |
| | | (0.35) |
| Int received? : Time to first received int | | 0.732** |
| | | (0.024) |
| Observations | 5,438 | 20,687 |
| $R^2$ | 0.008 | 0.128 |
| Residual SE | 7.175(df = 5425) | 23.191(df = 20672) |
| F Statistic | 3** (df = 12; 5425) | 216** (df = 14; 20672) |

**Table A.4: Linear regression models predicting the time between an author's first update and first initiation. Model (1) includes only pre-authorship initiators, whereas model (2) includes only post-authorship initiators. "Time to first received int" is the number of months between an author's first update and first received interaction. Note: $^{*}$p$<$0.05; $^{**}$p$<$0.01.**

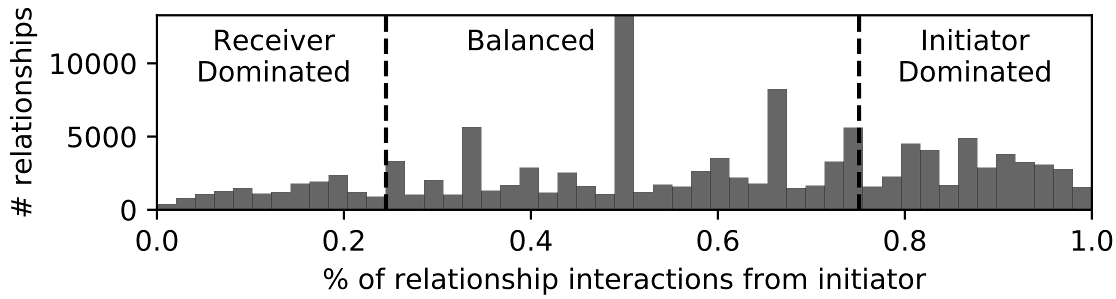



**Figure A.5: Distribution of the proportion of a relationship's interactions made by the initiator of that relationship and the thresholds used to identify balanced relationships. 52.47% of relationships are balanced using the indicated thresholds.**

Empirically, relationships initiated by caregivers are longer than those initiated by patients (27.8 months vs 26.7 months respectively, t=5.64, p¡0.001), although as discussed above this effect may be due to caregivers staying on CB longer or some other effect introduced by the right-censored nature of the data.

| | (1) | (2) | (3) |
|---|---|---|---|
| Candidate out-degree (log) | $-0.191^{***}$ | $-0.225^{***}$ | $-0.231^{***}$ |
| | (0.005) | (0.025) | (0.029) |
| Has in-degree? | $0.756^{***}$ | $1.907^{***}$ | $1.795^{***}$ |
| | (0.017) | (0.170) | (0.182) |
| Candidate in-degree (log) | $0.649^{***}$ | $0.960^{***}$ | $0.974^{***}$ |
| | (0.005) | (0.027) | (0.030) |
| Is reciprocal? | $20.016^{***}$ | $8.458^{***}$ | $7.999^{***}$ |
| | (0.460) | (0.574) | (0.568) |
| Is weakly connected? | $1.767^{***}$ | $4.647^{***}$ | $4.281^{***}$ |
| | (0.021) | (0.430) | (0.435) |
| Is friend-of-friend? | $5.220^{***}$ | $3.196^{***}$ | $2.801^{***}$ |
| | (0.097) | (0.201) | (0.207) |
| Candidate Role = Mixed | 0.020 | 0.110 | 0.111 |
| | (0.018) | (0.085) | (0.098) |
| Candidate Role = P | $-0.242^{***}$ | $-0.187^{***}$ | $-0.247^{***}$ |
| | (0.012) | (0.063) | (0.072) |
| Same author role? | $0.299^{***}$ | $0.331^{***}$ | $0.295^{***}$ |
| | (0.012) | (0.059) | (0.068) |
| Same health condition? | $0.213^{***}$ | $0.366^{***}$ | $0.400^{***}$ |
| | (0.009) | (0.049) | (0.056) |
| Candidate multi-site author? | $0.315^{***}$ | $0.752^{***}$ | $0.772^{***}$ |
| | (0.015) | (0.058) | (0.066) |
| Candidate mixed-site author? | $0.474^{***}$ | $0.367^{***}$ | $0.352^{***}$ |
| | (0.008) | (0.059) | (0.067) |
| Candidate update count | $-0.0003^{***}$ | $-0.001^{***}$ | $-0.0004^{***}$ |
| | (0.00004) | (0.0001) | (0.0001) |
| Candidate update frequency | $0.007^{***}$ | $0.013^{***}$ | $0.013^{***}$ |
| | (0.0002) | (0.002) | (0.002) |
| Days since recent update | $-0.011^{***}$ | $-0.006^{***}$ | $-0.005^{***}$ |
| | (0.00005) | (0.0001) | (0.0001) |
| Days since first update | $-0.001^{***}$ | $-0.001^{***}$ | $-0.0005^{***}$ |
| | (0.00001) | (0.00003) | (0.00004) |
| Same U.S. state assignment? | | | $2.723^{***}$ |
| | | | (0.069) |
| Observations | 155,141 | 7,007 | 7,007 |
| Log Likelihood | −133,746.600 | −4,830.810 | −3,743.011 |
| Test Accuracy | 77.2% | 84.3% | 87.1% |

**Table A.5: Conditional mlogit models for initiation with the subset of authors given US state assignments via IP geolookup.** ***Note:*** $^{***}$ **indicates p$<$0.01**

# Appendix B

# Author recommendation details

## B.1 Use of search feature on CaringBridge

Do CaringBridge users use search to attempt to find information or supporters? To address this question, we collected a dataset of user-initiated search queries on CaringBridge. These queries were extracted from internal logs collected between July 4, 2021 and July 10, 2021. Our dataset contained 103,830 searches comprising 32,722 unique query strings. We preprocessed the query strings by splitting them into tokens based on whitespace.

Based on a random sample of 100 queries with 1, 2, or 3 tokens and visual inspection of additional samples, all queries corresponded to either person names or existing site URL strings. Thus, we conclude that a very high percentage of searches are for a specific CaringBridge site. Queries with 4 or more tokens comprise $< 1\%$ of the query strings. Visual inspection suggests that the majority of queries with 4+ tokens are help requests or open-domain queries including spam. We could identify no instances of users searching for e.g. a particular health condition, treatment, or symptom; if search is used in this way, it occurs at a low prevalence. Results were identical when isolating to queries conducted by authors rather than by all users.

## B.2 Modeling and optimization details

This section provides additional details beyond those provided in sec. 5.3.3.

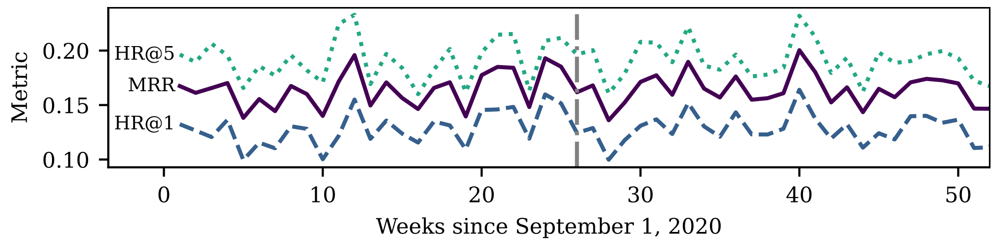



**Figure B.1: $\text{MLP}_{\text{Study}}$ offline performance throughout the validation and test period. No evidence of a non-zero relationship between time since model training and offline performance metrics. (All significance tests have $p$¿0.05.) This flat relationship suggests that (a) the offline analysis is minimally impacted by training set leakage and (b) frequent model retrains may not be necessary in practice, although we retrained the model weekly during the field study.**

### B.2.1 MLP model details

Both $\text{MLP}_{\text{Study}}$ and $\text{MLP}_{\text{Tuned}}$ use two hidden layers, ReLU activation functions, and dropout. Optimization occurs over 1000 epochs with a one-cycle learning rate scheduler proposed by Smith and Topin [304] and the Adam optimizer with default hyperparameters. Following best practices to prevent overfitting, we used for prediction the weights from the epoch with the lowest hold-out loss. Hold-out loss was computed over a random 1% of training initiations. We conducted random search experiments (not reported) using the learning rate and Adam hyperparameter distributions given by Sivaprasad et al., finding minimal differences [305]. $\text{MLP}_{\text{Study}}$ uses 100 hidden units per layer with a dropout $p$ of 0.1. $\text{MLP}_{\text{Tuned}}$ uses 300 hidden units per layer with a dropout $p$ of 0.5 and adds a weight decay of 0.0001. Hyperparameter tuning occurred using grid search to set hidden units $\in \{100, 300, 500\}$, dropout $p \in \{0.1, 0.5, 0.9\}$, weight decay $\in \{0, 0.0001, 0.01\}$, and maximum learning rate $\in [0.008, 0.016]$. We fit models from 3 random seeds at each hyperparameter combination and selected parameters based on the median of the 3 models' MRR. We did not explore additional model architectures in depth, and offline metrics could be improved through the use of a model closer to the state-of-the-art or through a larger hyperparameter sweep.

One downside to our chronological validation and testing sets is that this setup deviates from the common modeling practice of retraining on regular intervals. Fortunately, Figure B.1 indicates that this affect has a minimal impact on our offline evaluation.

**Table B.1: Offline test performance for various baselines. This table shows additional baselines beyond those presented in Table 5.3.**

| Network | MRR | HR@1 | HR@5 | $\lvert R \rvert$ | %Unique | MMST | $\frac{\lvert S_R \rvert / \lvert R \rvert}{\lvert S_U \rvert / \lvert U \rvert}$ |
|---|---|---|---|---|---|---|---|
| PeopleYouKnow | 0.102 | 7.86% | 13.07% | 3972 | 79.44% | 29.2 weeks | 15% / 28% |
| CosSim | 0.002 | 0.05% | 0.25% | 3403 | 68.1% | 27.1 weeks | 25% / 23% |
| MF | 0.002 | 3.05% | 15.23% | 13 | 0.26% | 40.3 weeks | 0% / 24% |
| RecentInits | 0.036 | 1.27% | 4.81% | 6 | 0.12% | 0.4 weeks | 0% / 24% |
| MostInits | 0.035 | 1.32% | 4.90% | 12 | 0.24% | 113 weeks | 0% / 24% |
| NewestAuthor | 0.025 | 0.97% | 3.19% | 6 | 0.12% | 0.1 weeks | 83% / 24% |
| RecentJournals | 0.016 | 0.19% | 1.58% | 6 | 0.12% | 3.1 weeks | 17% / 23% |
| MostInteractive | 0.007 | 0.17% | 0.68% | 5 | 0.10% | 0.9 weeks | 0.0% / 24% |
| MostJournals | 0.004 | 0.13% | 0.38% | 5 | 0.10% | 0.4 weeks | 60% / 24% |
| Random | 0.001 | ¡0.0% | 0.04% | 4159 | 83.18% | 17.6 weeks | 24% / 24% |

### B.2.2 Baselines

Table B.1 compares additional baselines beyond those shown in Table 5.3. We explored CosSim with other feature sets; they all perform worse in terms of MRR and Coverage metrics than using all features. **MF** is the conventional matrix factorization approach to collaborative filtering, using the dot product to compute similarity and selecting hyperparameters as described by Rendle et al. [158];[1] note the high hit-rate but low MRR, indicating a strong popularity bias. The new non-personalized baselines are **RecentInits**, which ranks sites by the amount of time since the last initiation with that site, **MostJournals/RecentJournals**, which mirror MostInits and RecentInits but count published Journal updates rather than initiations, **NewestAuthor**, which ranks newest sites first, and **MostInteractive**, which ranks sites by the number of recent interactions made by authors on that site.

### B.2.3 Feature ablation discussion

The results presented in Table 5.4 suggest that RoBERTa text features are not important for peer recommendation, and that recommendations based solely on text data would require a different

[1] All matrix factorization models were trained for 100 epochs with the same optimizer as the MLP model; the best model (by validation MRR) used embeddings of size 128 and weight decay of 0.0001. Embeddings were trained for the 41,567 authors and 79,588 sites that appear at least twice in training-period initiations; a single embedding each was reserved for previously unseen authors and sites.

approach than the one we use here. We suspect that the relative unimportance of text data reflects a bias in our implicit feedback signal and offline evaluation metrics toward authors already known to the source. Chen et al. observed that social network information was more effective for discovering known contacts, while content similarity was more effective for discovering new connections [233]—a dynamic that may be recurring here, as most historical initiations are between known contacts [10]. Even if text-based recommendations are less useful, users may perceive them to be less invasive than network-based recommendations (or vice-versa). Careful assessment of the perceived acceptability of this data collection for peer recommendation is necessary [306].

We chose to deploy the model that included the text data ($\mathrm{MLP}_{\mathrm{Study}}$) in part because it will still provide personalized recommendations, even to authors who have not yet interacted with peers. We can observe this effect in the coverage predictions: while recommended site diversity is higher for authors who had previously initiated (342 unique sites for 439 source users, each site recommended on average 6.4 times) compared to authors who had no previous interactions (303 unique sites for 561 source users, each site recommended on average 9.3 times), these differences are relatively small. Further, these recommendations are of a similar quality: MRR on test initiations is actually higher for people with no previous initiations than previous initiators (0.190 vs 0.147). Thus, we were satisfied that deploying $\mathrm{MLP}_{\mathrm{Study}}$ satisfied our cold-start design requirement.

## B.3 Survey Materials

All questions other than the eligibility, consent, and email questions are optional. All surveys were hosted on the Qualtrics platform.

### B.3.1 Preference Survey

All registered CaringBridge authors (18+ years old) are invited to take this 1-minute survey, which is part of a research collaboration by CaringBridge and a team of technology researchers at the University of Minnesota.

The purpose of CaringBridge is to help people get the support they need during health journeys—and support comes from many places! Many authors choose to make their CaringBridge sites open for anyone to read. **We want to send you emails with links to CaringBridge sites**

**that we think you'll want to follow.**

If you are interested in participating in our study, we'll send you personalized emails with links to CaringBridge sites. (As always, [privacy comes first][2] on CaringBridge: no one can read your site unless you want them to.) Opting in and sharing your opinion will help CaringBridge and the research community design features that make giving support easy and make a difference in the lives of the patients and caregivers you care about. Complete information and an FAQ for this study are available by clicking here [link].

***Click the arrow in the lower right corner of your screen to take the survey.***

—Page Break—

**Is Adult**

I am 18 years or older.

- Yes, I am 18 years or older.
- No, I am less than 18 years old.

**Is Registered**

Do you have a registered CaringBridge account?

- Yes, I have a CaringBridge account.
- No, I don't have a CaringBridge account.

**Is Author**

Are you can author or co-author of a CaringBridge site? You're an author or co-author if you've ever written a [Journal][3] update on a CaringBridge site.

- Yes, I am an author or co-author on a CaringBridge site.
- No, I am a visitor to CaringBridge and I haven't helped author a site.

[2] Link to `https://www.caringbridge.org/what-we-offer/the-privacy-you-choose`
[3] Link to `https://www.caringbridge.org/what-we-offer/a-journal-for-your-journey`

**Opt-in**

Do you agree to receive personalized follow-up emails with links to CaringBridge sites that we suggest?

- Yes, I want to participate in the study and receive follow-up emails with site suggestions.
- No, I am not interested in participating in the study.

—Page Break—

**Email**

**We just need the email address that you use with CaringBridge so that we can send you personal site suggestions.**

We won't share this email address with anyone outside of CaringBridge or the research team; we'll only use this email address to contact you and to connect to your CaringBridge user account.

- The email address I use with CaringBridge is: [Free Response]
- I'm not sure which email address is associated with my CaringBridge account, but the email I use most is: [Free Response]
- I changed my mind about participating in the study.

**Wants study result follow-up**

**Are you interested in reading about the results of this study when they become available in the future?** If you select yes, we will send you an email update with information about our results.

- Yes, please email me with the results of this study.
- No, thanks.

—Page Break—

Thanks for providing your info, you'll get emails with site suggestions from us soon. The following completely optional questions will help us create better site suggestions for you.

**Previous visit to stranger site**

(Optional) Have you ever visited the CaringBridge site of an author who you did not know personally?

- Yes, I have visited the CaringBridge site of someone I didn't know.
- No, I have never visited a stranger's CaringBridge site.

**Motivations**

(Optional) Which of the following might motivate you to visit a fellow author's CaringBridge site, even if you didn't personally know them? Check all that apply.

- To learn from the journeys of other CaringBridge authors.
- To help mentor or support newer CaringBridge authors.
- To receive advice or support from more experienced authors.
- To communicate with a peer who understands.
- I'm not interested in visiting other authors' CaringBridge sites right now, but I might want to in the future.
- I'm not interested in visiting other authors' CaringBridge sites right now, but I would have wanted to in the past.
- I'm never interested in visiting other authors' CaringBridge sites.
- Something else: [Free Response]

**Characteristics**

(Optional) What characteristics of an author or their site would make you want to read & engage with that person's CaringBridge site? Check all that apply.

- High-quality writing or photos
- Similar diagnosis or symptoms to you or the loved one you care for

- Similar treatments to you or the loved one you care for
- Lives near me
- Similar cultural background to you or the loved one you care for
- For caregivers: Sharing the same relationship (e.g. spouse, child) to the person they care for
- Something else: [Free response]

**Free Response – General**

(Optional) Anything else you want to share with us about visiting the CaringBridge site of a fellow author? [Free Response]

### B.3.2 Feedback Survey

Thank you for providing feedback. All questions on this page are optional: submit your feedback by clicking the right arrow at the very bottom of this survey form.

Not sure why you received this email & survey? You agreed to receive site suggestion emails in a survey you took in August. See the FAQ [link]. If you don't want to receive these emails anymore, you can unsubscribe [link].

**Overall Interest**

Overall, did the suggested sites seem interesting to you?

- Yes, the sites were generally interesting to me.
- Unsure or neutral.
- No, the sites were generally uninteresting to me.

**Free Response – Interesting**

Briefly, what seemed interesting to you about the suggested sites? [Free Response]

**Free Response – Uninteresting**

Briefly, what seemed uninteresting to you about the suggested sites? [Free Response]

**Specific Recommendation Relevance**

Specifically, how relevant did you find each of the suggested sites? [5-item response matrix with levels: Very Relevant; Somewhat Relevant; Unsure or Neutral; Somewhat Irrelevant; Very Irrelevant, Offensive, or Spam]

- 1st suggested site
- 2nd suggested site
- 3rd suggested site
- 4th suggested site
- 5th suggested site

**Free Response – General**

Any other thoughts you want to share about these site suggestions or your experience so far in this study? [Free Response]

### B.3.3 Unsubscribe Survey

**Page 1**

To unsubscribe: just enter your email address on the line below and click the right arrow.

Or, maybe you're looking for a feedback form [link] or the study FAQ [link] (including contact details for study coordinators) instead.

**Email**

- Unsubscribe me from additional emails from cb-suggestions@umn.edu. My email address is: [Free response]

—Page Break—

**Free Response – Unsubscribe**

- Tell us why you unsubscribed? [Free response]
- Anything else you'd like to share with us? Thank you for participating in our study! [Free response]

## B.4 Click data

We identify recommendation clicks in the Site Suggestion emails via three data sources: (1) Google Analytics counts, (2) CloudFront logs, and (3) logged-in user visits. Links to the recommended sites contain UTM tracking information, including the site, the batch ID, and a unique participant ID.

Each data source has limitations. The Google Analytics summary can only provide a total count and won't be incremented if JavaScript is disabled. An event won't be captured in the CloudFront log if UTM tracking tags are stripped and in unknown other cases.[4] Logged-in visits will only be recorded at or near the time of a recommendation click if the participant is already logged in or logs in while on the site.

Google Analytics reports 270 total recommendation clicks, while we have timestamped CloudFront log entries for only 232 clicks, suggesting that we are missing timestamps, participant, and site information for 14.1% of clicks. Of the 232 CloudFront clicks, 198 correspond to unique participant/site clicks. We can recover 22 clicks missing from the CloudFront request logs via the logged-in user visits, for a total of 220 total unique participant/site clicks. For subsequent statistical analyses, we assume that any additional clicks are missing at random, although in practice we are more likely to be missing clicks from participants who do not log in. If we assume that the Google Analytics count is authoritative and 14.1% of clicks are missing at random from the CloudFront data, then the expected number of missing participant/site clicks is 10. Thus, we expect only a small impact on our statistical analyses. Logged-in visit data are extracted from daily snapshots of the CaringBridge database. Thus, we only captured the

[4] "We recommend that you use the logs to understand the nature of the requests for your content, not as a complete accounting of all requests. CloudFront delivers access logs on a best-effort basis. The log entry for a particular request might be delivered long after the request was actually processed and, in rare cases, a log entry might not be delivered at all." `https://web.archive.org/web/20211206215502/https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/AccessLogs.html`

most recent logged-in visit within a 24-hour period, sufficient to identify daily repeat visits but insufficient for fine-grained analysis of browsing behavior.

## B.5 Site Recommendation Analysis

In sec. 5.5.2, we describe characteristics of the recommended set of sites. In this section, we provide additional pre-click details and use them to create a pseudo-control comparison set of non-recommended sites.

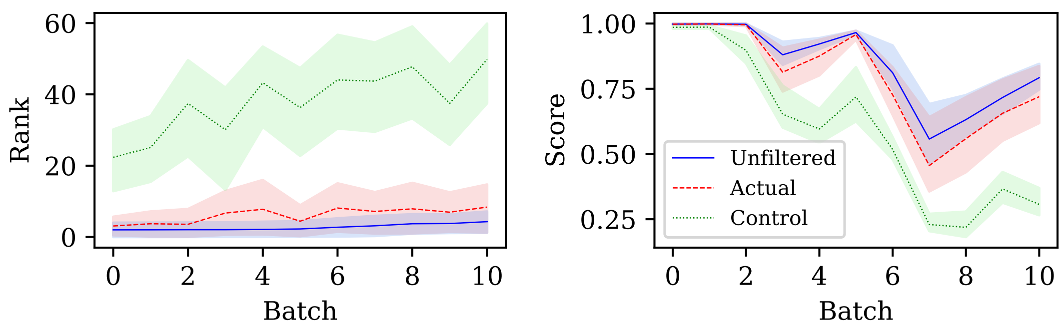



**Figure B.2: Distribution rank and the corresponding scores assigned by the recommendation model for unfiltered (blue, solid), actually recommended (red, dashed), and pseudo-control (green, dotted) sets by weekly batch. Across all batches, the unfiltered group contains $n$=397 unique sites, the actual contains $n$=526 unique sites, and the pseudo-control contains $n$=511 unique sites. Each line shows the mean rank and score respectively, while the shaded region indicates the max and min value within each set.**

### B.5.1 Site filtering

In sec. 5.3.3 we discuss a decision limiting the total number of times a site was recommended in a single batch to at most 10 times. Specifically, we conducted a five-round draft. In each round, a random participant ordering is chosen and participants take turns selecting their highest-scored site that has been picked fewer than 10 times. We analyze the impact of this decision in Figure B.2 by plotting the mean of the rank and score distributions for hypothetical recommendation sets generated without filtering. We observe that the maximum model score varies between batches as the model was retrained weekly, and that the filtered sites were similar in score to the sites that would have been recommended without filtering—all filtered sites were still in the

**Table B.2: Inter-rater reliability as Cohen's $\kappa$ and percent agreement (%A) between the two annotators in the final two coding rounds. Post-discussion codebook updates resulted in greater agreement during round 3.**

| | Round 2 ($n$=143) | | Round 3 ($n$=376) | |
|---|---|---|---|---|
| Preview Category | $\kappa$ | %A | $\kappa$ | %A |
| Reporting Health | 0.56 | 83.9% | 0.62 | 87.8% |
| Positive Disclosures | 0.72 | 87.4% | 0.78 | 89.4% |
| Negative Disclosures | 0.61 | 88.1% | 0.76 | 92.8% |
| Managing Audience Relationship | 0.66 | 90.2% | 0.74 | 91.0% |
| Expression of Appreciation | 0.78 | 96.5% | 0.90 | 98.1% |
| Expressive Writing | 0.38 | 69.2% | 0.48 | 73.7% |
| All | 0.33 | 37.8% | 0.45 | 48.9% |

top 0.1% of the model's scores. The use of filtering resulted in a 32.5% increase in the number of unique recommended sites, a fair trade-off for the modest decrease in score and rank induced by filtering.

### B.5.2 Psuedo-control set of non-recommended sites

The pseudo-control set is composed of the top five sites for each participant within a batch of recommendations that were never recommended during the duration of the study. We visualize this set's ranks in Figure B.2 along with the resulting effect on the corresponding score assigned to each set by the recommendation algorithm. In Table B.3, we see that the set of recommended author/site pairs were significantly different from the group of sites that could have been recommended (pseudo-control). Values were calculated at the time a recommendation was clicked; in the case of non-clicked and pseudo-control sites, a click time was randomly chosen from the set of actual click times from the same recommendation batch.

In Tables B.4 and B.5, we extend this analysis to include the subset of clicked recommended author/site pairs. Again, we see many significant differences between the set of clicked recommended author/site pairs and the pseudo-control set. However, we find little significant differences between clicked and non-clicked recommended author/site pairs.

**Table B.3: Observed recommended pseudo-control USP site behavior from the 35 day window before the point they were clicked (or could have been clicked). Site tenure is in days, Total # of authors is since the sites creation, # of authors and # days visiting peers is for the entire window period, while all other rows are the number of weekly actions.**

| | Recommended ($n_{\mathrm{P}}$=4190) | | Pseudo-Control ($n_{\mathrm{C}}$=4190) | | | |
|---|---|---|---|---|---|---|
| | Med. | M (SD) | Med. | M (SD) | $\mathrm{M_P}$ - $\mathrm{M_C}$ | $U_{\mathrm{P}}/(n_{\mathrm{P}}n_{\mathrm{C}})$ |
| Site tenure (days) | 113 | 227.4 (519) | 152 | 262.7 (425) | -35.3* | 42.4%* |
| Journal updates | 1 | 1.9 (2.0) | 1 | 1.6 (2.3) | 0.3* | 40.4%* |
| # of authors | 1 | 1.3 (0.7) | 1 | 1.2 (0.5) | 0.2* | 44.9%* |
| Total # of authors | 2 | 1.8 (1.0) | 2 | 1.7 (0.8) | 0.2* | 46.1%* |
| Peer visits | 0 | 0.3 (1.2) | 0 | 0.1 (0.2) | 0.2* | 26.6%* |
| Repeat user visits | 0 | 2.1 (6.3) | 0 | 0.9 (1.5) | 1.2* | 49.5% |
| Peer initiations | 0 | 0.8 (1.7) | 0 | 0.4 (0.5) | 0.4* | 47.2%* |
| Peer interactions | 0 | 3.6 (8.0) | 0 | 1.8 (3.9) | 1.8* | 47.3%* |
| # days visiting peers | 7 | 11.6 (11.7) | 7 | 9.3 (9.0) | 2.3* | 47.4%* |
| Site author interactions | 0 | 0.0 (0.2) | 0 | 0.0 (0.2) | 0.0 | 50.0% |
| Site author initiations | 0 | 0.0 (0.0) | 0 | 0.0 (0.0) | 0.0 | 50.0% |
| Site author self ints. | 6 | 15.1 (24.9) | 3 | 8.4 (15.4) | 6.8* | 40.8%* |

**Table B.4: Observed clicked and non-clicked recommended site behavior from the 35 day window before the point they were clicked (or could have been clicked). Site tenure is in days, Total # of authors is since the site's creation, # of authors and # days visiting peers is for the entire window period, while all other rows are the number of weekly actions.**

| | Clicked ($n_{\mathrm{P}}$=220) | | Non-Clicked ($n_{\mathrm{C}}$=3970) | | | |
|---|---|---|---|---|---|---|
| | Med. | M (SD) | Med. | M (SD) | $\mathrm{M_P}$ - $\mathrm{M_C}$ | $U_{\mathrm{P}}/(n_{\mathrm{P}}n_{\mathrm{C}})$ |
| Site tenure (days) | 107 | 267.4 (609) | 113 | 225.2 (514) | 42.2 | 49.5% |
| Journal updates | 1 | 2.1 (2.2) | 1 | 1.9 (2.0) | 0.2 | 46.6% |
| # of authors | 1 | 1.4 (0.7) | 1 | 1.3 (0.7) | 0.1 | 45.7%* |
| Total # of authors | 2 | 1.9 (1.0) | 2 | 1.8 (1.0) | 0.1 | 47.4% |
| Peer visits | 0 | 0.5 (1.7) | 0 | 0.3 (1.1) | 0.2 | 43.9%* |
| Repeat user visits | 1 | 3.1 (8.1) | 0 | 2.0 (6.2) | 1.1 | 43.6%* |
| Peer initiations | 0 | 1.1 (2.1) | 0 | 0.8 (1.7) | 0.3 | 45.2% |
| Peer interactions | 0 | 4.6 (9.5) | 0 | 3.6 (7.9) | 1.0 | 45.7% |
| # days visiting peers | 11 | 13.0 (11.7) | 7 | 11.5 (11.7) | 1.5 | 45.0% |
| Site author interactions | 0 | 0.1 (0.4) | 0 | 0.0 (0.2) | 0.0 | 49.2% |
| Site author initiations | 0 | 0.0 (0.1) | 0 | 0.0 (0.0) | 0.0 | 49.2% |
| Site author self ints. | 5 | 14.0 (25.8) | 6 | 15.2 (24.9) | -1.2 | 48.0% |

**Table B.5: Observed clicked and pseudo-control site behavior from the 35 day window before the point they were clicked (or could have been clicked). Site tenure is in days, Total # of authors is since the site's creation, # of authors and # days visiting peers is for the entire window period, while all other rows are the number of weekly actions.**

| | Clicked ($n_\mathrm{P}$=220) | | Pseudo-Control ($n_\mathrm{C}$=4190) | | | |
|---|---|---|---|---|---|---|
| | Med. | M (SD) | Med. | M (SD) | $\mathrm{M_P}$ - $\mathrm{M_C}$ | $U_\mathrm{P}/(n_\mathrm{P}n_\mathrm{C})$ |
| Site tenure (days) | 107 | 267.4 (609) | 152 | 262.7 (425) | 4.7 | 43.0%* |
| Journal updates | 1 | 2.1 (2.2) | 1 | 1.6 (2.3) | 0.5* | 37.3%* |
| # of authors | 1 | 1.4 (0.7) | 1 | 1.2 (0.5) | 0.2* | 40.7%* |
| Total # of authors | 2 | 1.9 (1.0) | 2 | 1.7 (0.8) | 0.2* | 43.4%* |
| Peer visits | 0 | 0.5 (1.7) | 0 | 0.1 (0.2) | 0.4* | 23.1%* |
| Repeat user visits | 1 | 3.1 (8.1) | 0 | 0.9 (1.5) | 2.3* | 43.4%* |
| Peer initiations | 0 | 1.1 (2.1) | 0 | 0.4 (0.5) | 0.6* | 47.2% |
| Peer interactions | 0 | 4.6 (9.5) | 0 | 1.8 (3.9) | 2.8* | 47.5% |
| # days visiting peers | 11 | 13.0 (11.7) | 7 | 9.3 (9.0) | 3.7* | 42.3%* |
| Site author interactions | 0 | 0.0 (0.2) | 0 | 0.0 (0.1) | 0.0 | 49.8% |
| Site author initiations | 0 | 0.0 (0.0) | 0 | 0.0 (0.0) | 0.0 | 49.8% |
| Site author self ints. | 5 | 14.0 (25.8) | 3 | 8.4 (15.4) | 5.6* | 42.5%* |

## B.6 Thematic analysis of Journal update previews

### B.6.1 Quantitative content analysis methods

Our thematic content analysis identified three high-level themes (discussed in sec. 5.5.2)—the full set of themes with descriptions and examples is shown in Table B.6. Based on these themes, we isolated four categories for quantitative analysis:

1. *Reporting Health Status, Symptoms, and Treatment.* Includes previews that update on specific health events, symptoms, or emotions of the patient. We include two sub-categories for *positive* and *negative* disclosures e.g. as used by Yang et al. [114]. Previews can contain both positive and negative disclosures if both are provided or if the author qualifies the news (*"in horrible pain, but finally a reason for hope"*).

2. *Managing Author/Audience Relationship.* The author is visible through descriptions of their role as a writer or their relationship with their blog's readers. See thematic description above. *"writing these posts is somewhat cleansing."*

3. *Expressions of Appreciation (EOA).* This theoretical construct was previously used by Smith et al., defined as explicit expressions of thanks, gratitude, blessing, or happiness that recognize the support received by the author or patient [8].

4. *Expressive Writing and Reflection.* We adopt the definition used by Ma et al.: expressive writing is “completed using functionalities afforded by online communities to disclose users’ thoughts and feelings about personal experiences” [70]. We adapt this definition to the context of Journal previews by including personal health-related reflection as a component of expressive writing.

Using these categories, two researchers separately annotated 658 update previews in three rounds of coding. The researchers met between each round to discuss disagreements and update the codebook. We report pre-discussion IRR scores in Table B.2.

We used the resulting annotations to model the relationship between these categories and clicks. This problem is the “post-presentation reward prediction” problem [307]. Unfortunately, clicks are noisy and it is not possible to directly assess if the presence of a particular category is associated with a greater propensity to click. This impossibility results from confounding: the rank of the recommendation in question, the other recommendations in the same email, when the email was generated and opened, and the specific participant are all confounding factors. Instead, we build multiple models with varied sets of assumptions and look for convergence. Specifically, we assume that clicks are independent—a common assumption in click models [308]—and report analysis for three subsets of recommendations: (a) *Clicked Batches Only* includes only recommendations in emails for which some but not all of the recs were clicked, so aims to acquire the clearest view of *choice* among competing options. (b) *B1 Only* includes only recommendations sent in the first batch of emails, eliminating temporal confounding and reducing bias introduced by varying levels of participant activity. (c) *Clickers Only* includes only recommendations sent to participants who clicked at least once during the study, reducing bias introduced by never-seen recommendations. We fit logistic regression models to predict individual recommendation clicks. While we report only rec-level logistic regression models here, email-level models, multi-level models for participant and batch, models including subsets of the features, and feature transforms revealed similar patterns.

### B.6.2 Quantitative content analysis results

We summarize the relationship between all four categories and click behavior in Table B.7. None of the models with any subset of the variables outperforms the null model (F-test $p > 0.05$)—even the rank-only model.[5] Expressive writing was associated with a greater probability of being clicked among Clicked Batches (50% of recs with EW clicked vs 36% of recs without) and in B1 (47% of recs with EW clicked vs 29% of recs without), but not when considering all recs sent to clicking participants. Expressions of appreciation were associated with a decreased probability of being clicked among participants who clicked at least once (7% of recs with EOA clicked vs 12% of recs without) but not in the other subsets. Both of these results should be taken with a large grain of salt, and in general these results suggest that participants were making clicking decisions based on factors we did not annotate.

## B.7 Estimates of study impact on behavior for authors and sites

In sec. 5.5.3, we estimated the impact of recommendation on second-order behaviors. We provide additional method details and sensitivity analysis in this section.

In Figures 5.10 and 5.11, we showed post-study outcomes by comparing 5 weeks (35 days) of pre-study behavior to 12 weeks (91 days) of post-study behavior. We conducted a sensitivity analysis to determine the impact of two decisions: examining behavior during the post-study period (rather than the "during study" period) and choosing a time window of 12 weeks. Figure B.3 is a high-detail summary of those decisions, demonstrating that differences are small depending on the selected time window: smaller post-study time windows are generally associated with higher-variance effect size estimates. The estimators shown are as described in sec. 5.5.3: raw (solid, blue), OLS (dashed, orange), and DR (dotted, green). Figure 5.10 captures the vertical slice at post-study week 13. 95% confidence intervals are computed via bootstrapping (1000 iterations). In pre/post panel data analyses, it is common to use the same time window pre- and post-intervention, which helps to control for longer-term behavioral trends. We tried both approaches, and found no difference except increased variance, so we use data from the full pre-study time window (5 weeks) even if the post-study time window is less than 5 weeks.

Using a similar approach, we extend this analysis to investigate post-click outcomes on

[5] Recommendations listed first *were* most likely to be clicked, but this difference was statistically insignificant.

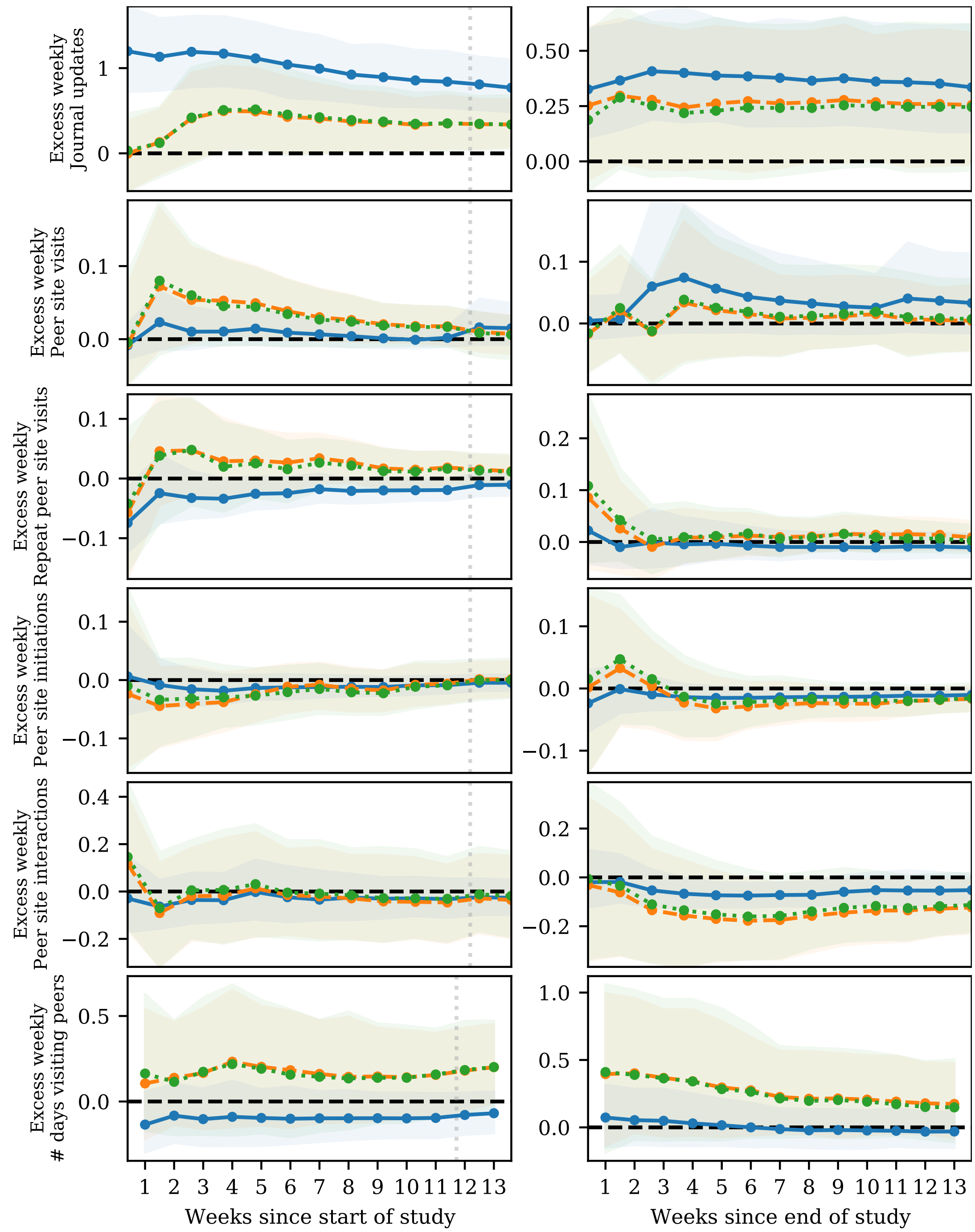



**Figure B.3: Estimated impacts of receiving Site Suggestion emails on author behavior, both during the study (left column, the vertical dotted line indicates the date of the last Site Suggestion email) and after the study (right column).**

recommended site behavior. In Figure 5.11, we showed post-study outcomes by comparing 5 weeks (35 days) of pre-click behavior to 12 weeks (91 days) of post-click behavior. Similar to Appendix B.5, values were calculated at the time a site was first clicked and in the case of non-clicked and pseudo-control sites, a click time was randomly chosen from the set of actual click times in the first batch that site was/could of be recommended. We conducted a sensitivity analysis to examine the impact a participant visiting a recommended site had on the sites behavior during the post-click period. Figure B.4 is a high-detail summary demonstrating we were unable to capture statistically significant effects from participant clicks on recommended site behavior. Moreover, we find that effect estimates based on comparison to the non-clicked and pseudo-control groups are similar.

As noted by Hernán and Robins, successful causal inference predominantly rests on untestable assumptions [276]. The doubly robust estimates we produce depend on five assumptions:

- **Exchangeability** Exchangeability refers to the probability of being in the treatment group is independent of the causal outcome (depending only on $A$). Exchangeability is likely false: it is reasonable to assume that authors that opted to participate in a recommendation intervention are more likely to engage with a recommendation intervention than non-participating authors, independent of e.g. their tenure as an author or their prior peer interaction behavior. If participants are not very different from non-participating authors in terms of their responsiveness to treatment, then it may still be reasonable to assume exchangeability. The same reasoning holds for clicked and non-clicked sites.

- **Positivity** Positivity requires that the distributions of the covariates for the treated and pseudo-control groups to fully overlap. Positivity is approximately true in our data, i.e. all values of the covariates observed in the treated group are also observed in the pseudo-control group. The positivity assumption is why we omit causal estimates of the impact of deploying Site Suggestion emails to the whole author population in terms of e.g. click rates; we have no ability to estimate the behavior of the pseudo-control population when exposed to Site Suggestion emails.

- **Consistency** Consistency refers approximately to the assumption that the treatment varies only in ways unrelated to the outcomes. Recommendation varies, so participants will be exposed to different versions of the treatment; different participants are shown different recommendations, so clicked sites will be exposed to visits from different participants.

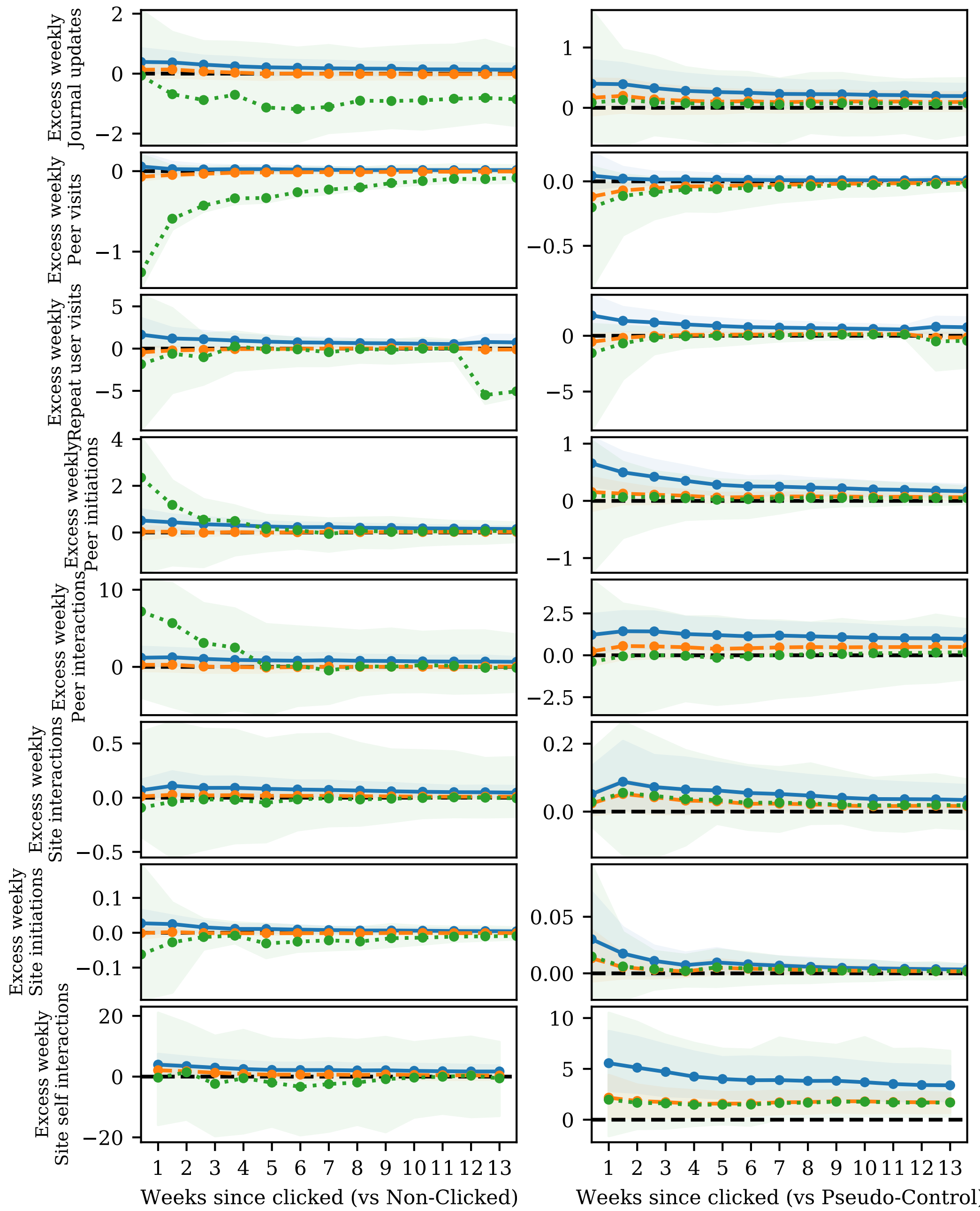



**Figure B.4: Estimated impacts of Site Suggestion emails on recommended site behavior post click, both using non-click sites (left column) and pseudo-control sites (right column). The estimators shown are as described in sec. 5.5.3: raw (solid, blue), OLS (dashed, orange), and DR (dotted, green). Figure 5.11 captures the vertical slice at week 13 in the right column. 95% confidence intervals are computed via bootstrapping (1000 iterations).**

Intuitively, these differences can be causally related to the outcomes: a participant with no prior interaction behavior, for example, may receive less relevant recommendations than a participant with a long history of peer interaction. A site clicked by a participant that visits and leaves a comment is different than a site clicked by a participant that only visits. Joachims et al. argue that recommender systems ought to be viewed as policies that select interventions in order to optimize a desired outcome [309]; it's a policy *designed* to select variable interventions. By assuming consistency, we assume that the recommendations are of similar quality and that participant visits have similar effects on the visited sites. The degree to which this assumption is violated will bias the resulting estimates.

- **No measurement error** Measurement error in the observed behaviors is likely non-existent, given the use of a complete database snapshot.
- **No model misspecification** To adjust for confounding, the models we use need to include all relevant covariates and assume the correct functional form. By fitting linear models, we make a parsimonious but possibly false assumption of functional relationship between the covariates and the outcome of interest. More serious is the assumption that all relevant covariates are available; as discussed, we believe that unmeasured confounders such as an interest in peer connection affect response to treatment in a way that is independent of the measured covariates.

Given this discussion, it seems reasonable to object that three of the assumptions are very likely false. In practice, we are assuming that these assumptions are *close enough* to true, such that we can still attain some insight from computing causal estimates based on these assumptions in order to compare to the observational estimates.

## B.8 Sample Size Calculations

In sec. 5.5.3, we described a power analysis for an uncontrolled peer recommendation study. In this section, we provide additional details and add estimates for an RCT. In Fig. 5.12, we showed sample size estimates for two variations of a replication of our feasibility study by computing the effect of two author behaviors associated with receiving recommendations. Here, we extend this analysis to include 5 behaviors: unique repeat visits to recommended sites, interactions with recommended sites, unique repeat visits to all sites, interactions with all sites, and recommended site updates. We consider the 35 days before and after the first exposure to recommendation or first stranger visit for the one-time email and 35 days before and the study period plus 35 days after for the 12 week recurring email intervention.

Table B.9 outlines the mean, variance, and sample size for each group in terms of weekly actions used in the calculations. The structure of our data let us consider two potential future interventions: a one-time recommendation email and a recurring, weekly recommendation email. We estimate the effects of a one-time recommendation email by including only participant exposures and visits from the first batch of recommendation emails (B1). We use the pseudo-control group (see sec. 5.4.1 and Table 5.6) and non-clicked recommendations (see Appendix B.5) to analyze the effect of receiving recommendations and stranger visits. For participants (recommended sites) behaviors, we report weekly actions from a 5 week pre-inflection period to a post-inflection period: 12 weeks + 5 weeks and 5 weeks. For participants (all sites) and recommended sites behaviors, we report the difference in weekly actions from similar pre and post-inflection periods. Effect size calculations were made using 17 decimal places.

Using these statistics, Table B.10 estimates the sample size required for a replication (same as Fig. 5.12 and for an RCT, with 50% not receiving recommendation emails. For author behaviors targeted at recommended sites, we compute the effect as simply as $d = M_P/SD$. For all other behaviors, we control for prior behavior by computing the effect as $d = (M_C^{before} - M_C^{after}) - (M_P^{before} - M_P^{after})/SD_{pooled}$, where P is the group of participants or clicked recommended sites and C is the control group of pseudo-control participants or non-clicked recommended sites. We present $M^{before} - M^{after}$ because the expected churn of users dictates over time they will become less active. We find this is true in our sample; all users that published at least one journal update in July 2021 subsequently published, on average, 1.32 (5.44) less updates in the following month. Here, a positive effect size means that being a participant results in less of a difference from before and after when compared to the control group. Using G*Power 3.1.9.7, the estimated required sample sizes for an appropriately powered RCT were calculated based on observed effect sizes using a one tailed point biserial model at 80% power with $\alpha = 0.5$ [278].

Unsurprisingly, we see large differences in participant behavior towards recommended sites when compared to the pre-study period. More interesting is the fact that participants, on average, sustained an increase in weekly actions including non-recommended sites after receiving recommendations. While this translates to a seemingly large effect size during calculation, we attribute these differences primarily to the significant differences pre-study between both groups outlined in Table 5.4.1. Here, we believe any differences that exist between groups can largely be attributed to confounding variables (i.e. author/site tenure).

**Table B.10: Observed effect sizes and required sample size needed for an appropriately powered RCT for a one-time recommendation email and a recurring, weekly recommendation email. Site descriptions in parenthesis for participant groups denote the target of the intended action.**

| | Recurring 12 Week | | One-Time | |
|---|---|---|---|---|
| Behavior | Effect Size | Sample Size | Effect Size | Sample Size |
| **Participants (Rec)** | | | | |
| Second visits | 0.28 | 75 | 0.26 | 81 |
| Interactions | 0.12 | 444 | 0.14 | 314 |
| **Participants (All)** | | | | |
| Second visits | 0.85 | 5 | 0.67 | 10 |
| Interactions | 0.43 | 29 | 0.18 | 189 |
| **Recommended Sites** | | | | |
| Updates | -0.14 | 299 | 0.23 | 110 |

**Table B.6: Thematic analysis: summary of themes in recommended Journal update previews**

| Theme | Description | Example(s) |
|---|---|---|
| **Reporting patient/caregiver status** | | |
| Past vs Future News | An axis that differentiates past occurrence and future plans. | This last week was oh so busy with all of my tests and appointments.<br>Friday is a huge milestone for me. |
| Events | Those describing events, specifically related to a singular point in time. | Bryan met with his neurosurgeon yesterday. |
| Patient Symptoms, Status, and Health Processes | Those detailing how the patient is doing, and health processes related to a prolonged interval of time. | Well, Sid is getting stronger by the day. |
| Beginnings, Endings, and Transitions | Those describing singular health events that transition the patient into or out of a health process. | Kristen is finally off the ventilator. |
| Status sentiment | An axis that defines the sentiment of reported news (as “good” or “bad”). | Today Ashley had her second day of PT and she did amazing in every way.<br>We waited all night for and unfortunately he’ll have to go through another round of chemo. |
| Emotion-laden Reporting | Those explicitly stating the author’s attitudes towards an update using emotive language. | Tears of joy as I write this. We are sooo happy to announce Andrea is out of the ICU. |
| Reflection | Those reflecting on the author’s or patient’s experiences. | It’s hard to believe it was only a year ago today that Jen started chemo. |
| **Managing author/audience relationship** | | |
| Expressions of Gratitude | Those that positively acknowledge a specific type of support. | We are so thankful for your prayers. |
| Update Context | Those that explicitly provide contextual information about the post or its contents for the audience. | Today is going to be a little different. I don’t have any medical updates, I just need to let some stuff out. |
| Comments on Update Frequency | Those acknowledging the time between multiple CaringBridge updates. | First, I need to apologize for how long it’s been since our last update. |
| Reflection on Writing Process | Those reflecting on the author’s experiences as a Caring Bridge author. | I’ve really struggled to put my feelings into words in these posts. |

**Table B.7: Logistic regression models predicting recommendation clicks from preview content for the rec subsets defined in sec. 5.4.2. Preview contents are poor predictors of clicks, as evidenced by near-chance ROC AUC scores (estimated using leave-one-out cross-validation). Note: $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001**

| | Clicked Batches Only | B1 Only | Clickers Only |
|---|---|---|---|
| Intercept | -0.467 | -1.943$^{***}$ | -1.714$^{***}$ |
| | (0.325) | (0.488) | (0.206) |
| Rank within email | 0.096 | -0.023 | 0.034 |
| | (0.085) | (0.115) | (0.052) |
| EOA | -0.581 | 0.156 | -0.546$^{*}$ |
| 5.8% of B1 recs | (0.396) | (0.621) | (0.277) |
| Expressive Writing | 0.638$^{*}$ | 0.762$^{*}$ | 0.050 |
| 31.2% of B1 recs | (0.250) | (0.353) | (0.152) |
| Managing Audience Relationship | -0.293 | 0.239 | -0.358 |
| 17.5% of B1 recs | (0.328) | (0.409) | (0.192) |
| Reporting health status (categorical) | | | |
| Base level: None (14.8% of B1 recs) | | | |
| Neutral disclosure | -0.171 | -0.235 | -0.013 |
| 31.5% of B1 recs | (0.338) | (0.498) | (0.211) |
| Positive disclosure only | -0.615 | 0.125 | -0.078 |
| 25.8% of B1 recs | (0.340) | (0.494) | (0.205) |
| Negative disclosure only | -0.395 | -0.540 | -0.091 |
| 22.2% of B1 recs | (0.396) | (0.546) | (0.247) |
| Pos & neg disclosures | -1.070 | -0.541 | -0.706 |
| 5.8% of B1 recs | (0.551) | (0.768) | (0.368) |
| Observations | 310 | 365 | 1,590 |
| Clicks | 120 | 51 | 220 |
| Log-Likelihood | -199.69 | -142.94 | -632.92 |
| ROC AUC | 0.554 | 0.469 | 0.488 |

**Table B.8: Activity variables included in the participant author model and the visited site model. We removed some variables to avoid collinearity, but otherwise opted to incorporate as many potentially relevant behaviors as possible [2].**

| Author model variables | Site model variables |
|---|---|
| # Journal updates | # Journal updates |
| # first visits to other authors' sites | # unique author visits (recent) |
| # repeat visits to other authors' sites | # unique authors (all time) |
| # unique days visiting other authors' sites | # first visits to site from peers |
| # interactions on other authors' sites | # repeat author visitors to site |
| # interactions on their own sites | # unique days other authors visited |
| # self-authored sites interacted with | # interactions from other authors |
| Author tenure (log) | # initiations to site |
| | # peer interactions by site authors |
| | # self-site interactions by site authors |
| | # initiations by site authors |
| | Site tenure (log) |

**Table B.9: Mean and standard deviation estimates for a one-time and recurring recommendation email intervention. Site descriptions in parenthesis for participant groups denote the target of the intended action.**

| | Recurring 12 Week | | One-Time | |
|---|---|---|---|---|
| Behavior | $M_P$ ($\sigma^2$) | $M_C$ ($\sigma^2$) | $M_P$ ($\sigma^2$) | $M_C$ ($\sigma^2$) |
| **Participants** | **Study** | | **Study** | |
| **(Recommended Sites)** | ($n_P$=79) | | ($n_P$=73) | |
| Second visits | 0.06 (0.06) | | 0.04 (0.02) | |
| Interactions | 0.47 (16.36) | | 0.07 (0.22) | |
| **Participants** | **Participants** | **Pseudo-control** | **Participants** | **Pseudo-control** |
| (All Sites) | ($n_P$=79) | ($n_C$=1759) | ($n_P$=73) | ($n_C$=1759) |
| Second visits | -0.07 (0.07) | 0.02 (0.01) | -0.05 (0.03) | 0.02 (0.01) |
| Interactions | -0.34 (14.81) | 0.11 (0.49) | -0.04 (0.72) | 0.09 (0.49) |
| **Recommended Sites** | **Clicked** | **Non-clicked** | **Clicked** | **Non-clicked** |
| | ($n_P$=158) | ($n_C$=368) | ($n_P$=51) | ($n_C$=55) |
| Updates | 0.024 (2.11) | 0.023 (1.80) | 0.02 (1.33) | 0.04 (3.45) |