Leveraging cues from person-generated health data for peer matching in online communities

Andrea L Hartzler[1,*], Megan N Taylor[2], Albert Park[3], Troy Griffiths[4], Uba Backonja[3], David W McDonald[2], Sam Wahbeh[4], Cory Brown[4], Wanda Pratt[4]

ABSTRACT

Objective Online health communities offer a diverse peer support base, yet users can struggle to identify suitable peer mentors as these communities grow. To facilitate mentoring connections, we designed a peer-matching system that automatically profiles and recommends peer mentors to mentees based on person-generated health data (PGHD). This study examined the profile characteristics that mentees value when choosing a peer mentor.

Materials and Methods Through a mixed-methods user study, in which cancer patients and caregivers evaluated peer mentor recommendations, we examined the relative importance of four possible profile elements: health interests, language style, demographics, and sample posts. Playing the role of mentees, the study participants ranked mentors, then rated both the likelihood that they would hypothetically contact each mentor and the helpfulness of each profile element in helping the make that decision. We analyzed the participants' ratings with linear regression and qualitatively analyzed participants' feedback for emerging themes about choosing mentors and improving profile design.

Results Of the four profile elements, only sample posts were a significant predictor for the likelihood of a mentee contacting a mentor. Communication cues embedded in posts were critical for helping the participants choose a compatible mentor. Qualitative themes offer insight into the interpersonal characteristics that mentees sought in peer mentors, including being knowledgeable, sociable, and articulate. Additionally, the participants emphasized the need for streamlined profiles that minimize the time required to choose a mentor.

Conclusion Peer-matching systems in online health communities offer a promising approach for leveraging PGHD to connect patients. Our findings point to interpersonal communication cues embedded in PGHD that could prove critical for building mentoring relationships among the growing membership of online health communities.

INTRODUCTION

Online health communities provide a rich source of peer support,[1,2] and there is a wealth of person-generated health data (PGHD) embedded within posts made in these communities.[3] Online posts are one type of data within a broad variety of PGHD[4,5] (eg, patient-reported outcomes,[6,7] observations of daily living,[8] quantified self,[9,10] and illness narratives[11]) that has enormous potential for healthcare.[12,13] Largely untapped, PGHD offers a foundation for connecting individuals with experienced "peer mentors" who share similar interests and health concerns.

Offline peer mentors, from "health coaches"[14] to "peer navigators,"[15] can improve patients' diabetes control,[16–19] treatment adherence,[20] and help patients cope with brain injury,[21] heart disease,[22] and cancer.[23] These programs generally match patients with mentors based on diagnosis and demographics,[16] but they have suffered from drop-outs,[22] limited uptake,[18] and complaints due to poor mentor matching.[16] Yet, "better patients," who model good self-management, do not necessarily make better health coaches.[24] Although patients also prioritize mentor qualities, such as having a non-judgmental or upbeat outlook,[15] when judging the merits of peer mentorship, other factors are rarely examined. Further research is needed to understand what makes a good mentor-mentee match for peer mentoring,[25] both offline and online.

Patients increasingly turn to online health communities for help from "someone like me,"[26] yet few communities leverage PGHD to cultivate mentoring relationships.[27–29] Mentoring through e-mail and discussion groups shows targeted successes in education,[30,31] professional development,[32,33] and support for patients with disabilities.[34–36] Although individual differences in expectations and communication style are major sources of failure and frustration,[37] careful mentor matching[31,34] and interpersonal similarity[32,38] appear vital to the success of mentorship programs.

Interpersonal similarity also promotes member attachment in online health communities.[39] Online referents to shared experiences,[40] personality,[41] and interests[33,42] facilitate homophily – a cornerstone of online connections between individuals.[43] The more homophilous users perceive discussion groups to be, characterized by an "optimal match"[44] in health experiences and needs,[45,46] the more likely it is that users will adopt advice they are given by members of these groups.[47] As online communities grow and diversify, forging connections becomes increasingly arduous, leaving the burden on users to locate compatible mentors. Gaining awareness of the peer support available in online communities often requires having multiple interactions to build relationships.[48] All too often, yielded advice does not fit the user's needs.[23] Thus, uncovering suitable peer mentors is a growing challenge.

We aim to address this challenge by leveraging PGHD embedded within online community posts to automatically profile and match peers for mentorships (ie, "peer matching"). Although user posts

*Correspondence to Andrea L Hartzler, PhD, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101-1448, USA; hartzler.a@ghc.org; Tel: 206-287-2030. For numbered affiliations see end of article.

provide rich insight into the diversity of the support available in online health communities,[49] traditional profiles that users enter manually do not capture this rich PGHD. PGHD can form the basis of automated health profiling for targeted prevention[50] and treatment strategies,[51,52] and shows promise for online health communities.[53,54] Determining what factors are important to mentees when choosing a peer mentor is critical for designing peer matching tools that extract and profile the PGHD that is of greatest value.

We extend our validated approach to automated profiling using PGHD,[55] by employing principles of "social matching systems"[56] to recommend online community members with similar interests.[57] Outside the health context, similar systems make social recommendations based on user profiling from digital.[58–63] Our peer-matching system is designed to forge mentoring connections by recommending peers with profiles that share similar health topics and other personal characteristics gleaned from PGHD.

OBJECTIVE

We designed a peer-matching system that automatically profiles and recommends peer mentors based on PGHD. Through a mixed-methods user study, we examined mentor profile characteristics that cancer patients and caregivers found most helpful when choosing a peer mentor. We addressed the research question: *What PGHD in an online user profile helps establish a good mentor-mentee match in online peer mentoring?*

APPROACH

Our peer-matching system is designed to facilitate connections between individual "mentees" and "peer mentors" who share similarities, which are gleaned from PGHD. The system recommends peer mentors by processing text to generate and match the profiles of individuals with shared characteristics. In prior work with CancerConnect.com, we demonstrated the validity and acceptability of extracting health-related terms from online community posts to summarize members' health interests in individualized profiles.[55] In this study, we extended this approach by extracting additional language style characteristics to generate profiles for mentors and mentees. We then matched profiles using a similarity metric that analyzed health interests, language style, and demographics to recommend matching mentor profiles (Figure 1).

Generating Profiles

We generated user profiles from PGHD using text extraction to summarize characteristics of peer mentors and mentees. Mentors were drawn from users of CancerConnect.com, our partnering online health community. This partnership enabled us to generate profiles for a large pool of hypothetical peer mentors by processing the community posts made by each member. Mentor profiles were generated through text processing to extract and display: (1) health interests, (2) personal language style, (3) demographics, and (4) sample posts. The mentees were study participants who submitted a two-page personal story about their cancer experience (see Materials and Methods section). Mentee profiles relied on the same text processing procedures to extract (1) health interests and (2) personal language style from mentees' personal stories. Figure 2 shows a sample mentor profile with the four key profile elements. Below, we briefly elaborate on the extraction methods and the use of the four extracted profile elements.

Health Interests

We applied pre- and post-processing to improve MetaMap's ability to extract health-related concepts from PGHD.[64] As in our prior work,[55]

we extracted health-related terms from mentor posts and grouped semantically similar terms into four categories: health problems, treatments, diagnostics and tests, and provider care.

Personal Language Style

We processed text with the Linguistic Inquiry and Word Count (LIWC) tool.[65] LIWC calculates the degree to which people use particular word categories in text, such as positive or negative emotional terms, references to one's self, and other cognitive and emotional characteristics.[66] We grouped the extracted terms into the 13 categories that LIWC associates with psychological processes and personal concerns: social, affective, cognitive, perceptual, biological, relativity, work, achievement, leisure, home, money, religion, and death.

Demographics

Because we lacked demographic data for the mentors identified from CancerConnect.com, we assigned a fictitious sex and age to mentor profiles for the study's purpose of matching demographics with mentees.

Sample Posts

We added sample posts to the mentor profiles to provide additional context about the mentors. We chose snippets from each mentor's initial post on Cancerconnect.com, their most recent post, and an intermediate post.

Matching Profiles

To match profiles, we first established a pool of mentors whose profiles had a sufficient amount of text to allow for individualized matches with text from a mentee's profile. We excluded CancerConnect.com members whose profiles did not include enough terms with which to make a match or included so many terms that they matched with all of the mentees. We chose members who had made five or more community posts containing 20–100 unique health interests. Applying these criteria resulted in a pool of 415 potential mentors who, on average, contributed 13 posts covering 46 unique health interests.

We selected candidate mentors from this pool and presented them to the study participants, who played the role of mentees in a hypothetical scenario (see Materials and Methods section). For each mentee, we selected five candidate mentors who systematically varied in their level of health interest, language style, and demographic similarity, from high (ie, most similar match in the pool) to low (ie, least similar match in the pool) (Table 1). The selection of these five candidate mentors was based on a weighted similarity rank metric that equally weighted health interest similarity and language style similarity between mentor and mentee. For example, Profile B is the mentor with both the highest similarity of health interests and the lowest similarity of language style to the mentee. To make the demographic similarity of Profile B low, we assigned the mentor a contrasting age and the opposite sex of the mentee. A detailed description of our similarity matching approach is provided in the Supplementary Materials.

"Health interest similarity" was based on the cosine similarity of health-related terms extracted from the mentee's text and each potential mentor's text. We used term frequency-inverse document frequency (TF-IDF) to weigh the cosine similarity of health interest terms in a vector space model.[67]

"Language style similarity" was based on language style matching (LSM) of LIWC terms in the mentee's text and each potential mentor's text. LSM represents the average similarity of LIWC scores and has been used to predict personality[68] and the initiation of interpersonal relationships using text.[69] For each potential mentor, we calculated

Figure 1: Recommending peer mentors through peer matching.
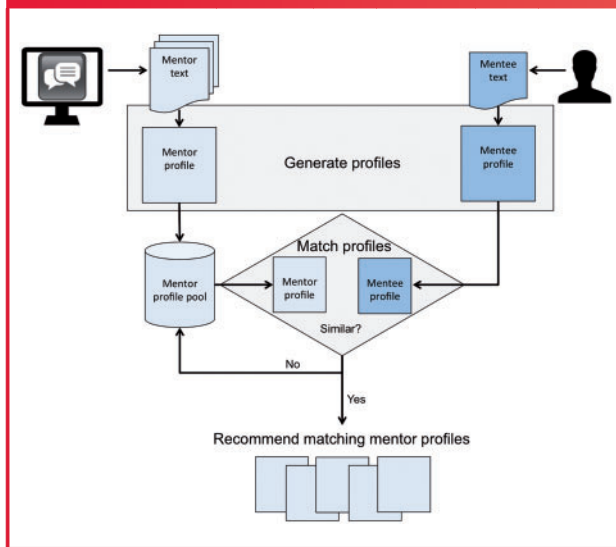
LIWC scores that represent the percentage of total words that fall into each of the 13 LIWC language style categories. We then compared the LIWC scores in each category for the mentee and each mentor to yield an average composite LSM similarity score.

"Demographic similarity" was based on the self-reported demographics of the participant (ie, the "mentee"). We systematically varied the similarity of the sex and age of the candidate mentors by assigning fictitious demographics to each of the mentor profiles.

MATERIALS AND METHODS

To examine what PGHD helps establish a good mentor-mentee match in online peer mentoring, we conducted a mixed-methods user study with cancer patients and caregivers. By asking the study participants to evaluate the profiles of candidate mentors recommended by our peer-matching system, we examined the relative importance of health interests, demographics, language style, and sample posts when choosing a mentor. The University of Washington institutional review board approved the study procedures.

Using flyers, we recruited people who had experience with cancer, as a patient or a caregiver, to take part in a two-part study comprised of a brief intake survey and a 1-hour audiotaped follow-up session. The 1-hour follow-up session was comprised of a scenario-based profile evaluation task followed by a semistructured interview with questions similar to those of Shami and colleagues.[70] Although the mentor profiles were generated from CancerConnect.com members' community posts, mentor recommendations were hypothetical – study participation carried no expectation for actual mentor contact. During the follow-up sessions, we elicited participants' perceptions about the relative importance of the four profile elements when choosing a mentor and their broader attitudes about our peer-matching approach. The study participants received $30 at the end of the session.

Data Collection Procedures

After providing informed consent, the study participants completed the intake survey by reporting their demographics, cancer experience, and technology experience. We then asked participants to compose and submit a two-page personal story about their cancer experience, which we processed to generate their "mentee" profile for mentor

matching. We scheduled a follow-up session with each study participant, during which the participants evaluated five recommended mentor profiles. To ground the session in the hypothetical context of online peer mentoring, we asked the participants to imagine themselves playing the role of a mentee in a peer mentoring scenario (Box 1).

Next, we asked participants to review and rank the five mentor profiles presented to them by preference. Participants provided a series of ratings about their choice of mentors. First, participants rated the likelihood that they would hypothetically contact each mentor on a scale from 1 (ie, not at all likely) to 9 (ie, extremely likely). Second, participants used the same 9-point scale to rate the helpfulness of each profile element for deciding whether to contact each mentor. We prompted participants to comment on their reasoning, so that we could qualitatively understand the characteristics they, as mentees, sought when choosing mentors. To close the follow-up session, we asked the study participants for feedback about the usefulness of peer matching and for feedback on how to improve the design of profiles.

Data Analysis

We summarized the study participants' characteristics and mentor profile ratings with descriptive statistics. We applied linear regression to examine the helpfulness ratings given to the profile elements as predictors for the likelihood of a mentee contacting a mentor. Because each study participant provided multiple ratings, we used a generalized estimating equation (GEE) to handle the correlation within each participant's data in our stepwise regression and identify the strongest predictors. We conducted regression analyses in R (v2.15.1),[71] using the "geeglm" function from the "geepack" package and utilizing the default parameters. We first conducted bivariable analyses to independently test the association between each predictor variable (ie, the helpfulness ratings for health interests, language style, demographics, and sample posts) and the outcome (ie, the rating for the likelihood of a mentee contacting a mentor). We then conducted a multivariable regression that included all the predictor variables in a single model to test the association between all the predictors and the outcome.

Based on observing a large proportion of study participants who had given extreme mentor contact ratings (ie, either a low rating of "1" or "2" or a high rating of "8" or "9"), we also conducted an exploratory analysis on this extreme rating subset to examine whether any profile element may have helped the mentees rule mentors in (ie, subgroups of highest ratings) or out (ie, subgroup of lowest ratings). Thus, among the subset with extreme ratings, we fit a binomial regression model with a log-as-link function to estimate the relative risk of mentees ruling mentors in (high rating) or out (low rating). Using this model, we conducted both bivariable and multivariable regression analyses.

We qualitatively reviewed the audio recordings of the study participant interviews to identify illustrative quotes the about participants' rationale for their mentor ratings and their feedback on our approach. We used affinity diagramming[72] to group quotes into qualitative themes about the characteristics mentees seek in peer mentors. Below, we report our results regarding participants' characteristics, mentor profile ratings, characteristics sought in peer mentors, and the perceived value of peer matching.

RESULTS

Participants

Table 2 provides the characteristics of the 13 study participants (Participant 1–Participant 13). The participants managed many types

Figure 2: Sample peer mentor profile. Profile elements include health interests, personal language style, demographics, and sample posts. The numbers in parentheses indicate term frequency across the mentor's posts. The font size of the personal style categories indicates the relative term frequency for each of the 13 categories.
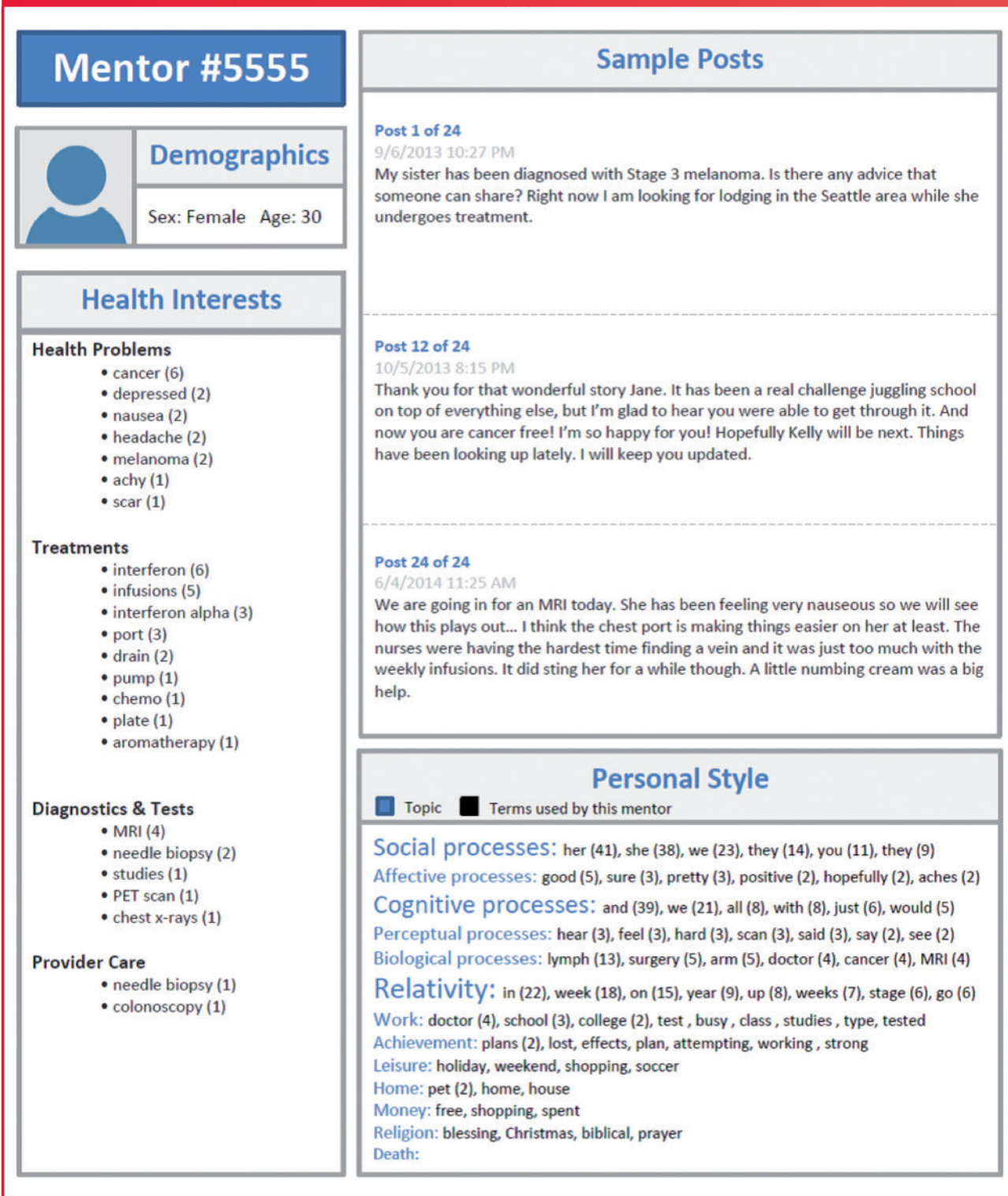
Mentor #5555

Demographics
Sex: Female   Age: 30

Health Interests

Health Problems
- cancer (6)
- depressed (2)
- nausea (2)
- headache (2)
- melanoma (2)
- achy (1)
- scar (1)

Treatments
- interferon (6)
- infusions (5)
- interferon alpha (3)
- port (3)
- drain (2)
- pump (1)
- chemo (1)
- plate (1)
- aromatherapy (1)

Diagnostics & Tests
- MRI (4)
- needle biopsy (2)
- studies (1)
- PET scan (1)
- chest x-rays (1)

Provider Care
- needle biopsy (1)
- colonoscopy (1)

Sample Posts

Post 1 of 24
9/6/2013 10:27 PM
My sister has been diagnosed with Stage 3 melanoma. Is there any advice that someone can share? Right now I am looking for lodging in the Seattle area while she undergoes treatment.

Post 12 of 24
10/5/2013 8:15 PM
Thank you for that wonderful story Jane. It has been a real challenge juggling school on top of everything else, but I'm glad to hear you were able to get through it. And now you are cancer free! I'm so happy for you! Hopefully Kelly will be next. Things have been looking up lately. I will keep you updated.

Post 24 of 24
6/4/2014 11:25 AM
We are going in for an MRI today. She has been feeling very nauseous so we will see how this plays out... I think the chest port is making things easier on her at least. The nurses were having the hardest time finding a vein and it was just too much with the weekly infusions. It did sting her for a while though. A little numbing cream was a big help.

Personal Style

■ Topic   ■ Terms used by this mentor

Social processes: her (41), she (38), we (23), they (14), you (11), they (9)
Affective processes: good (5), sure (3), pretty (3), positive (2), hopefully (2), aches (2)
Cognitive processes: and (39), we (21), all (8), with (8), just (6), would (5)
Perceptual processes: hear (3), feel (3), hard (3), scan (3), said (3), say (2), see (2)
Biological processes: lymph (13), surgery (5), arm (5), doctor (4), cancer (4), MRI (4)
Relativity: in (22), week (18), on (15), year (9), up (8), weeks (7), stage (6), go (6)
Work: doctor (4), school (3), college (2), test , busy , class , studies , type, tested
Achievement: plans (2), lost, effects, plan, attempting, working , strong
Leisure: holiday, weekend, shopping, soccer
Home: pet (2), home, house
Money: free, shopping, spent
Religion: blessing, Christmas, biblical, prayer
Death:

RESEARCH AND APPLICATIONS

Table 1: Candidate Mentors Matching Approach

Mentor Profile	Health interest similarity	Language style similarity	Demographic similarity
A	High	High	High
B	High	Low	Low
C	Low	High	Low
D	Low	Low	High
E	Low	Low	Low

Box 1: Peer Mentoring Scenario

Let's say you visit an online health community with an interest in connecting with a peer mentor. A peer mentor is an online community member, often a patient or caregiver, that you could turn to for help and advice about the cancer experience. For example, a peer mentor can provide advice from their own cancer experience, such as what to expect, help with treatment decisions, managing side effects, and other issues that impact everyday life. After visiting the online community, you receive recommendations for potential peer mentors. Your job is to review the recommendations to decide whom to contact for help.

of cancer, including breast, kidney, lung, ovarian, prostate, skin, and uterine cancer, as well as Hodgkin's lymphoma, leukemia, multiple myeloma, and myelodysplastic syndrome. The study participants contributed 320 mentor ratings across 65 mentor profiles. Data from one participant (Participant 9), who provided an incomplete set of mentor ratings, were excluded from the regression analysis.

Mentor Profile Ratings

Table 3 shows the distribution of the study participants' likelihood ratings for contacting mentors. On average, the participants rated the likelihood of contacting any mentor a 5.2 out of 9 (standard deviation [SD] = 2.8), with a rating of 7.9 (SD = 1.6) given to the highest-ranked profile and a rating of 2.5 (SD = 1.7) given to the lowest-ranked profile. Over half of the participants' mentor contact ratings were at the extremes – either "low," with a rating of "1" or "2" (n = 17), or "high," with a rating of rating of "8" or "9" (n = 17).

The study participants' helpfulness ratings of the profile elements varied significantly (*Friedman x^2* = 56.7, P < .001). Table 4 shows the results of our regression analyses, including the association between the mentor contact ratings and the ratings for each profile element independently (ie, bivariable analyses) as well as among all four profile elements together (ie, multivariable analysis). The study participants rated the sample posts as the most helpful profile element for determining whether to contact a mentor, followed by demographics, health interests, and language style (Table 4). However, no profile element significantly predicted the likelihood of a mentee contacting a mentor, either in the bivariable or multivariable analyses. Although choosing a mentor was a nuanced process that varied among the study participants, language style was the strongest predictor of a mentee's

likelihood to contact a mentor, but was rated as the least helpful profile element relative to the other profile elements.

Given the large proportion of mentor contact ratings at the extremes – either "low" (n = 17) or "high" (n = 17) – we examined, through an exploratory analysis, whether any profile element may have particularly helped mentees rule mentors in or out. For each subset of extreme ratings (ie, "low" and "high" likelihood of a mentee contacting a mentor), Table 5 shows the relative risk of a mentee ruling a mentor in and out. In both the "low" and "high" subsets, the helpfulness of the mentors' sample posts significantly predicted the mentees' mentor contact ratings. Thus, the study participants appear to have found sample posts to be quite helpful for deciding whether to rule mentors in or out.

Characteristics Sought in Mentors

The study participants' comments about choosing mentors offer insight into both the helpfulness of the individual profile elements and ways to improve the design of the mentor profiles (Box 2). Qualitative themes highlight desirable mentor characteristics, such as being knowledgeable, sociable, and articulate based on cues from PGHD. The participants also emphasized the need for streamlined profiles that minimize the amount of time required to chose a mentor. For example, the participants suggested shortening lists of terms to enable scanning for key topics and layering content to enable details on demand. Other suggestions included adding private messaging functionality and an "*e-hug*" (Participant 2) button for users to show support.

Health Interests

Although some study participants found it challenging to scan the mentors' health interests for key terms, they quickly learned about the mentors' specific medical experiences from this profile element. Health interests helped the participants assess the applicability of the mentor's knowledge to their own health situation. Profiles that listed the mentors' cancer type helped participants understand a mentor's story quickly. The study participants suggested adding cancer type, stage, and dates of diagnoses and treatments to the mentor profiles. They also found listed treatments especially helpful for assessing how far along a mentor was in their cancer experience. Some participants noted that, even if a mentor profile listed a different type of cancer than their own, they were interested in connecting with that mentor if treatments were similar.

Language Style

Language style helped the study participants learn about each mentor's personal life and character. The participants used the mentors' language style to assess the interpersonal compatibility between themselves and the mentors. However, the majority of the participants found this profile element confusing and overwhelming, because the terms to describe language style could not be categorized in a relatable way. LIWC's categories reflect its design as an analytic tool for personality analysis. Participants who did not shy away from this element tended to prefer concrete categories (eg, work, leisure, money, religion), because they were easy to understand and provided personal life cues that might strengthen interpersonal affinity with the mentor. The participants suggested listing only concrete categories and shortening the term lists for easy scanning. One study participant suggested color-coding categories to highlight the mentor's key language style.

Table 2: Study Participant Characteristics	
Demographics	
Age, mean (SD)	48 (12.8)
Sex	
Female	11/13
Male	1/13
Transgender	1/13
Race	
Asian	1/13
Black	2/13
White	10/13
Ethnicity	
Hispanic/Latino	1/13
Education	
High school degree	1/13
College degree	8/13
Post-graduate degree	4/13
Cancer experience	
Role	
Patient	7/13
Caregiver	4/13
Both patient and caregiver	2/13
Years managed	
<1 year	4/13
1–3 years	6/13
≥3 years	3/13
Technology experience	
Computers	
Moderate experience	6/13
Extensive experience	7/13
Online communities	
No experience	2/13
Moderate experience	5/13
Extensive experience	6/13

SD, standard deviation.

Table 3: Distribution of Mentor Contact Ratings	
Rating	Frequency
1	11
2	6
3	5
4	3
5	7
6	5
7	11
8	8
9	9

mentor. Other participants placed less importance on demographics than on other profile elements. Several participants suggested adding more details to the demographic information available in the profile, such as marital status, whether or not the individual has children, nationality, and location, because treatment protocols can vary by region or country, and some mentees may wish to meet mentors in person. The study participants also noted that adding the mentor's role, both with respect to cancer (eg, patient, survivor, caregiver) and everyday life (eg, mother, sister), was important.

Sample Posts

Consistent with our regression analysis, the study participants found that the mentors' sample posts were the most universally helpful profile element, apparently due largely to communication cues embedded in their text. Through these cues, the participants discovered contextual insights about the mentors that were not captured elsewhere, such as tone, attitude, and perspective (eg, optimism); empathy and social courtesies (eg, greetings, apologies); personality; and the mentor's overall "story." The study participants varied in the characteristics to which they were drawn. For instance, Participant 3 liked one mentor because their posts exhibited "*more fight and attitude and don't take it lying down*," whereas Participant 2 told us, "*I don't want to be preached to.*" The mentors' posts were also used as indicators of their online community involvement and availability. For example, some participants viewed short posts as a sign that a mentor could be difficult to connect with regularly. The participants suggested adding features, such as links to a mentor's additional posts, that could be used to gain additional context about potential mentors.

Overall, the study participants considered a broad number of characteristics when choosing a mentor. They were drawn to people who "*know what they are talking about*" (Participant 11). Providing links to websites, book titles, descriptions of treatments, and advice from personal experience were factors that made mentors stand out as knowledgeable. Participants sought articulate mentors who communicated clearly in easy-to-understand terms. Other informative communication cues included frank and specific posts; an informal and positive tone; transparency; having a similar life mantra (eg, "*attitude is everything*" – Participant1); having similar religious or spiritual beliefs; and sociability, such as greeting others, apologizing for mistakes, and not talking in platitudes or 'yelling' through the use of all capital letters. Many of these characteristics, which are critical to the formation of

Demographics

The study participants varied with respect to how helpful they perceived the demographic information to be when choosing a mentor. Some participants preferred that mentors have a similar age and sex or gender to their own. For instance, one participant preferred female mentors with whom she could discuss fertility concerns. A transgender participant expressed the importance of representing diverse gender categories. Two participants shared that similar demographics might be more important than sharing the same diagnosis with a

Table 4 Regression Analysis (N = 60)

Profile element	Rating	Bivariable analyses	Multvariable analysis
	Mean (SD)	Coefficient (SE), P-value	Coefficient (SE), P-value
Health interests	5.7 (2.5)	−0.05 (0.23), P = .80	−0.09 (0.18), P = .62
Language style	4.4 (2.3)	0.18 (0.16), P = .24	0.22 (0.16), P = .18
Demographics	6.4 (3.2)	−0.03 (0.17), P = .86	−0.03 (1.76), P = .76
Sample posts	7.9 (1.8)	0.09 (0.19), P = .61	0.11 (0.17), P = 0.50

N, number of profiles rated; SD, standard deviation; SE, standard error.

Table 5 Exploratory Analysis

Profile element	Bivariable analyses	Multivariable analysis
	RR (95% CI), P-value	RR (95% CI), P-value
Low likelihood of mentee contacting mentor (rating = 1 or 2) (N = 17)		
Health interests	1.02 (0.78-1.34), P = .88	1.11 (0.94-1.20), P = .63
Language style	0.86 (0.72-1.03), P = .09	0.88 (0.71-1.11), P = .29
Demographics	1.04 (0.83-1.30), P = .74	1.03 (0.88-1.22), P = .70
Sample posts	1.11 (0.74-1.65), P = .62	**2.06 (1.44-2.96), P < .001**
High likelihood of mentee contacting mentor (rating = 8 or 9) (N = 17)		
Health interests	0.93 (0.79-1.11), P = .43	0.93 (0.83-1.04), P = .22
Language style	0.99 (0.84-1.16), P = .87	1.10 (0.98-1.24), P = .11
Demographics	0.95 (0.85-1.06), P = .35	0.96 (0.87-1.05), P = .35
Sample posts	**4.02 (1.59-10.2), P = .003**	**4.04 (1.96-8.31), P < .001**

CI, confidence interval; N, number of profiles rated; RR, relative risk of "low" or "high" mentor contact rating. Cells are bolded that show significance at 0.05 level or better.

interpersonal relationships, were evident from communication cues embedded within the text of the mentors' posts.

Perceived Value of Peer Matching Based on PGHD

When asked about the usefulness of online peer matching, the study participants noted the advantage of saving valuable time and energy during the process of finding mentors who had similar circumstances to their own. For instance, Participant 2 told us that, "*You choose your spouse, it's this special person and when you lose them and watch them die it's really awful. It would be really nice to have somebody else who's gone through that say 'Keep breathing, you're going to be ok'.*" The participants thought that recommending peer mentors through profiles "*helps me make judgments faster because you spend a lot of time on those forums*" (Participants 1). Overall, the participants found that the peer-matching system was a valuable tool for helping them quickly making connections with peer mentors: "*I think it's a fabulous way of finding support*" (Participant 3).

The study participants also noted the potential disadvantages of peer matching. The cancer experience itself can be emotionally and physically taxing: "*You're so exhausted and everything is emergent and you're trying to have a normal life when nothing is normal*" (Participant

2). Peer matching helped participants find mentors who offer both knowledge and emotional support: "*These three* [mentors] *seem like people I could talk to either on an intellectual level or on a feelings level*" (Participant 3). Although participants valued a "give and take" relationship between mentor and mentee, some expressed concerns about balancing the expectations of reciprocity while dealing with cancer: "*I wouldn't have the energy or the strength to help 'me' when I needed to because I'd be wanting to help him. I didn't do group therapy for that reason. It would be too hard. I would want to help other people*" (Participant 2). Thus, forming reciprocal relationships with others through online communities can feel taxing, due to the perceived social obligations involved in such relationships: "*When you're diagnosed you get a lot of Facebook messages and you get a lot of emails and you feel obligated to answer them . . . it becomes another obligation*" (Participant 1). Privacy concerns also surfaced: "*there's anonymity issues*" (Participant 1). Some participants suggested chatting and secure messaging with mentors as solutions to these concerns.

Generally, the study participants felt that peer matching could add value to online health communities. Existing online communities can make searching for key information about peers burdensome. A successful peer-matching system can address these barriers with PGHD, but it must be sensitive to the context of the cancer experience and prioritize ease of use. The users of online health communities are likely to be emotionally and physically exhausted. For instance, Participant 2 told us that, "*I don't want to have to do research to reach out to this person to find out what they're talking about.*" Thus, to build relationships with peers online, mentor profiles should leverage PGHD to provide a succinct snapshot of a mentor that allows mentees to make judgments about potential mentors without asking them probing questions or having in-depth interactions that require valuable time and energy.

DISCUSSION

Our peer-matching system recommended mentors to mentees by profiling individuals using PGHD and matching profiles with shared health interests, language style, and demographics. The study participants found that the mentors' sample posts were the most helpful profile element when they were choosing mentor, particularly because of the interpersonal communication cues embedded therein. This finding is important, because the weakest predictors in our linear regression – namely, health interests and demographics – are key dimensions upon which peer mentors are traditionally matched in offline interventions.[16–18,24] In our exploratory analysis, these traditional profile elements were also associated with a lower relative risk for mentees ruling mentors in or out compared to the sample posts. Because we lacked demographic data for the potential mentors, however, our systematic assignment of sex and age to the sample mentor profiles could have lost important nuances in the ways people speak based on their age or gender. Our qualitative findings reflected differences among the study participants in how they judged the importance of age, gender, and other demographics commonly used in concordance matching, such as race, culture, or ethnicity. Even though, on average, the participants rated demographics as the second most helpful element of a mentor profile, our small participant sample limited the power to detect significant predictors. Thus, further investigation is warranted.

Online settings offer PGHD that points to additional dimensions of a good mentor-mentee match. In addition to diagnoses and demographics, interpersonal communication cues about an individual's knowledge, sociability, communication skills, and availability provide invaluable context when judging mentor compatibility. Our findings

Box 2: Qualitative Comments About Choosing Mentors	
Health interests	"*Health interests were probably the most helpful, because there were a lot of very specific things in there.*" (Participant 9, patient)
	"*[It was] hard to get a good read on what is his cancer, but then I looked for the health interests and saw prostate cancer.*" (Participant 4, caregiver)
	"*Because I have such a rare cancer, I go for the people that have my specific type of cancer . . . naturally I would gravitate towards someone who has survived for 7 years with stage 4, because it gives me hope.*" (Participant 5, patient)
	"*I'm looking for someone who has gone through the treatments [that I have] or I know I could be in the future, and he has less talk about the treatments than [other mentors].*" (Participant 4, caregiver)
	"*There's tons of people that have breast cancer . . . but at the same time, we don't exchange a whole lot of information with each other because the treatment is so different.*" (Participant 5, patient)
Language style	"*There's almost a little bit more of an outside life kind of bit . . . where he talks about books and he talks about work and he talks about songs and all these other things besides his cancer.*" (Participant 4, caregiver)
	"*I probably can relate to the cognitive processes, the uncertainty.*" (Participant 2, caregiver)
	"*Religion stands out to me . . . money is [also] a big issue because it's so expensive.*" (Participant 12, patient)
	"*I'd like to scan over some stuff . . . it's just a lot to read . . . how often did they say this . . . I don't really care!*" (Participant 2, caregiver)
Demographics	"*Extremely helpful, 9, based on demographics. Female, age . . . first thing I looked at.*" (Participant 2, caregiver)
	"*I would want a female mentor because I like mentors that have gone through the same thing as me. Anything that would be more similar to me.*" (Participant 6, patient)
	"*Since I have prostate cancer, I probably would like to talk to a man. You know, because that's specific to men. But I would also like to [talk with a woman], as I have learned, there is a very powerful relationship between the characteristics and treatments of breast cancer and prostate cancer.*" (Participant 11, patient and caregiver)
	"*When I look at this person with similar things, demographics and symptoms, it encourages me to want to reach out even though it might not be my same cancer.*" (Participant 5, patient)
	"*I was more interested in what they had to say, and less interested in their gender or age.*" (Participant 4, caregiver)
Sample posts	"*[The post] tells a lot about the mentor, what you can expect if you start a discussion.*" (Participant 13, caregiver)
	"*I feel like I can get a better view of how they communicate, if they are easy to follow or not . . . She was very thorough, she writes a lot. She's very specific.*" (Participant 4, caregiver)
	"*Personal style became less important as I read the posts . . . it's something about the tone of the posts – straightforward and not judgmental . . . I'm very familiar with the rah-rah type. Even though it's the same diagnosis, it's a different mentality.*" (Participant 9, patient)
	"*Those [posts] are really helpful because, again, I get sort of an idea of how empathetic he is, how he uses sort of down to earth language.*" (Participant 1, patient)
	"*The person seems to be very authentic . . . with the description and his vulnerability and openness about the treatment.*" (Participant 5, patient)
	"*I like how this person reaches out to this other person and says 'if you need a friend or whatever else I can help you with.'*" (Participant 10, patient)
	"*I just don't like the capital letters. I feel like he's yelling at someone. Too opinionated or something.*" (Participant 8, patient and caregiver)

support prior research on online peer mentoring that has suggested that interpersonal similarity of goals, attitudes, and beliefs may be as important as demographic similarity in building effective mentoring relationships.[32] This work also contributes a deeper understanding of the breadth of PGHD that can help establish a good match between mentees and mentors and also suggests design considerations for peer-matching systems (Figure 3).

Because we engaged only a small number of cancer patients and caregivers who were mostly female and white, our findings may have limited generalizability to other groups or contexts. However, our mixed-methods approach led to rich qualitative insights into the value of PGHD

for technologies that connect people with shared interests and health concerns for peer support. Our study participants found value in profiling peers with a succinct snapshot that enables quick and confident judgments about compatibility. Thus, our future research directions include extracting and representing a greater breadth of profile characteristics from PHGD and exploring additional matching strategies with a larger number of individuals from diverse backgrounds. Whereas some profile characteristics can be manually entered, other characteristics can be extracted automatically by processing PGHD. Future research that examines communication cues in text structure (eg, emoticons, text typed in all capital letters)[73,74] and language style,[69] such as sentiment,[75–77]

affect,[78] or personality,[68,79,80] could show promise for peer-matching systems. Although homophily engenders attraction (ie, "birds of a feather flock together"),[43] there is merit in exploring alternative mentor-matching strategies, such as "difference matching" (eg, "opposites attract"), "optimal matching" based on support needs,[44,45] or "reciprocal matching" that considers the preferences of both the mentor and the mentee.[81] Extensions of our work could investigate the design and implementation of peer-matching systems in settings other than online cancer communities.

Profiling users based on PGHD extracted from online community posts also presents challenges, requiring advances in natural language processing.[82] Just because a mentor's profile lacks a term like "chemotherapy" does not necessarily mean that that individual lacks experience with chemotherapy, and vice versa. Advances in machine learning, unstructured information management architectures, and other natural language processing techniques[3,76,83] can reveal "digital phenotypes" from symptoms, diagnoses, treatments, and other features extracted from PGHD.[84] Because online health communities have unique nomenclatures and language structures,[85,86] more research is required to fully leverage available PGHD from these communities.

Our study participants valued peer matching because it could save them time and energy when seeking support from others in online communities. With improvements to the design and content of mentor profiles, the study participants agreed that peer matching could help them efficiently forge connections with others in online health communities. A key design recommendation includes a layered approach to overview a mentor's distinguishing characteristics, paired with links to deeper details.[87,88] Flexibility also appears to be important for design. Different participants valued different matching characteristics, rendering a one-size-fits-all solution disadvantageous. Instead, peer-matching systems should provide multiple individualized matches based on those characteristics that a given user values most. Although most participants viewed the resulting mentor match as a give-and-take relationship with demonstrated reciprocal benefits,[18] future research is needed to reduce burden and promote supportive mentoring dynamics.[15] For example, patients and caregivers could have different needs and preferences for mentors. With a larger sample of participants, we could stratify our results to compare preferences among the subgroups to inform the design of tailored approaches for different mentoring dynamics.

Finally, privacy was found to be an important consideration for peer-matching systems, one which extends to digital health innovations in the era of precision medicine.[89,90] Many people take care in disclosing and selectively sharing personal information in online health communities.[91] The study participants suggested that private messaging and chatting features in peer-matching systems would be valuable. We should explore other solutions that guard users' anonymity[91] and facilitate user agency in the age of Big Data,[92] such as granular sharing controls[93] and user ability to opt in and edit their profiles. Likewise, we, as researchers, designers, and other stakeholder groups, must take care to ground the extraction and use of sensitive PGHD in ethical principles.[91,94,95]

CONCLUSION

Looking for a compatible peer mentor among the growing diversity of candidates in online health communities is a major challenge. Peer-matching systems address this problem by leveraging PGHD to convey the interests, health concerns, and communication cues of available peer mentors to mentees. Our findings suggest that there is more to a good match than a shared diagnosis and demographics. Sample posts on user profiles, which often embed communication cues to judge interpersonal similarity, help establish a good mentor-mentee match in online peer mentoring. Leveraging various types of PGHD may prove critical for facilitating mentoring relationships among the growing membership of online health communities.

Figure 3: Design considerations for peer-matching systems.

• Leverage the breadth of available PGHD containing health-related, demographic, and communication cues to peer knowledge, sociability, articulate expression, and availability.

• Profile peers with a succinct snapshot that enables quick and confident judgments about compatibility.

• Employ ethical principles in the extraction, use, and display of PGHD to endorse agency and preserve privacy.

• Provide sufficient flexibility for users to efficiently examine matching characteristics of greatest personal interest, whether viewing, searching or filtering extracted text.

• Design profiles in layers to summarize an individual's most distinguishing characteristics with links to detailed background on their experiences and interests.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Dennis C. Peer support within a health care context: a concept analysis. *Int J Nurs Stud.* 2003;40(3): 321–332.
2. Doull M, O'Connor AM, Welch V, *et al.* Peer support strategies for improving the health and well-being of individuals with chronic diseases. *Cochrane Database Syst Rev.* 2005;3:CD005352.
3. Wang YC, Kraut RE, Levine JM. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. *J Med Internet Res.* 2015;17(4):e99.
4. Shapiro M, Johnston D, Wald J, *et al.* Patient-generated health data: white paper prepared for the Office of the National Coordinator for Health IT by RTI International. April 2012. www.rti.org/pubs/patientgeneratedhealthdata.pdf Accessed 28 May 2015.
5. Deering MJ. *ONC Issue Brief: Patient-generated Health Data and Health IT.* Office of the National Coordinator for Health Information Technology; 2013. www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf. Accessed 28 May 2015.
6. Wu AW. *Advances in the Use of Patient Reported Outcome Measures in Electronic Health Records.* 2013. www.pcori.org/assets/2013/11/PCORI-PRO-Workshop-EHR-Landscape-Review-111913.pdf. Accessed 28 May 2015.
7. Cella D, Riley W, Stone A, *et al.* The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179–1194.
8. Backonja U, Kim K, Casper GR, *et al.* Observations of daily living: putting the "personal" in personal health records. *Nurs Inform.* 2012;2012:6.
9. Appelboom G, LoPresti M, Reginster J-Y, *et al.* The quantified patient: a patient participatory culture. *Curr Med Res Opin.* 2014;30:2585–2587.
10. Choe EK, Lee NB, Lee B, *et al.* Understanding quantified-selfers' practices in collecting and exploring personal data. *Proc SIGCHI Conf Hum Factor Comput Syst.* 2014:1143–1152.
11. Sharf BF, Vanderford ML. Illness narratives and the social construction of health. In: Thompson TL, Dorsey A, Parrott R, Mille K, eds. *Handbook of Health Communication.* Mahwah, New Jersey, Lawrence Erlbaum Associates, 2003:9–34.
12. Chung AE, Basch EM. Potential and challenges of patient-generated health data for high-quality cancer care. *J Oncol Pract.* 2015;11(3):195–197.
13. Hansen MM, Miron-Shatz T, Lau AYS, *et al.* Big Data in science and healthcare: a review of recent literature and perspectives. Contribution of the IMIA Social Media Working Group. *Yearb Med Inform.* 2014;9:21–26.
14. Heisler M. *Building Peer Support Programs to Manage Chronic Disease: Seven Models for Success.* California Health Care Foundation; 2006. http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/B/PDF%20BuildingPeerSupportPrograms.pdf Accessed 28 May 2015.
15. Giese-Davis J, Bliss-Isberg C, Carson K, *et al.* The effect of peer counseling on quality of life following diagnosis of breast cancer: an observational study. *Psychooncology.* 2006;15(11):1014–1022.
16. Long JA, Jahnle EC, Richardson DM, *et al.* Peer mentoring and financial incentives to improve glucose control in African American veterans. *Ann Intern Med.* 2012;156(6):416–424.
17. Thom DH, Ghorob A, Hessler D, *et al.* Impact of peer health coaching on glycemic control in low-income patients with diabetes: a randomized controlled trial. *Ann Fam Med.* 2013;11(2):137–144.
18. Heisler M, Vijan S, Makki F, *et al.* Diabetes control with reciprocal peer support versus nurse care management: a randomized trial. *Ann Intern Med.* 2010;153(8):507–515.
19. Tang TS, Funnell M, Sinco B, *et al.* Comparative effectiveness of peer leaders and community health workers in diabetes self-management support: results of a randomized controlled trial. *Diabetes Care.* 2014;37(6): 1525–1534.
20. Jerson B, D'Urso C, Arnon R, *et al.* Adolescent transplant recipients as peer mentors: a program to improve self-management and health-related quality of life. *Pediatr Transplant.* 2013;17(7):612–620.
21. Hanks RA, Rapport LJ, Wertheimer J, *et al.* Randomized controlled trial of peer mentoring for individuals with traumatic brain injury and their significant others. *Arch Phys Med Rehabil.* 2012;93(8):1297–1304.
22. Turner BJ, Hollenbeak CS, Liang Y, *et al.* A randomized trial of peer coach and office staff support to reduce coronary heart disease risk in African-Americans with uncontrolled hypertension. *J Gen Intern Med.* 2012;27(10): 1258–1264.
23. Rini C, Lawsin C, Austin J, *et al.* Peer mentoring and survivors' stories for cancer patients: positive effects and some cautionary notes. *J Clin Oncol.* 2007;25(1):163–166.
24. Rogers EA, Hessler DM, Bodenheimer TS, *et al.* Diabetes peer coaching: do "better patients" make better coaches? *Diabetes Educ.* 2014;40(1):107–115.
25. Barg FK, Weiner MG, Joseph S, *et al.* Qualitative analysis of peer coaches' experiences with counseling African Americans about reducing heart disease risk. *J Gen Intern Med.* 2012;27(2):167–172.
26. Fox S, Jones S. The social life of health information. Pew Internet & American Life Project; May 2011. http://www.pewinternet.org/files/old-media//Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf. Accessed 28 May 2015.
27. Zheng K, Newman MW, Veinot TC, *et al.* Using online peer-mentoring to empower young adults with end-stage renal disease: a feasibility study. *AMIA Annu Symp Proc.* 2010;2010:942–946.
28. Nepal S, Paris C, Bista SK. SRec: a social behaviour based recommender for online communities. *3rd Workshop on Social Recommender Systems* at UMAP. 2012;872:SRS paper 7.
29. Nepal S, Paris C, Pour PA, Freyne J, Bista SK. Interaction-based recommendations for online communities. *ACM Trans Internet Technol.* 2015;15(2):Article 6.
30. Bull S, Greer JE, McCalla GI, Kettel L, Bowes J. User modelling in i-help: What, why, when and how. *User modeling 2001;2109 of the Lecture Notes in Computer Science*:117–126.
31. Taylor C, Zeng H. Developing online peer-to-peer mentoring programs for distance degree programs. *Proceedings of TCC Worldwide Online Conference.* 2008:81–105.
32. Ensher EA, Heun C, Blanchard A. Online mentoring and computer-mediated communication: new directions in research. *J Vocat Behav.* 2003;63(2): 264–288.
33. Liu H, Macintyre R, Ferguson R. Exploring qualitative analytics for e-mentoring relationships building in an online social learning environment. *Proceedings of the ACM 2nd International Conference on Learning Analytics and Knowledge.* 2012:179–183.
34. Cohen K, Light J. Use of electronic communication to develop mentor-protégé relationships between adolescent and adult AAC users: Pilot study. *AAC.* 2000;16(4):227–238.
35. Burgstahler S, Cronheim D. Supporting peer-peer and mentor-protégé relationships on the Internet. *J Res Technol Educ.* 2001;34(1):59–74.

RESEARCH AND APPLICATIONS

RESEARCH AND APPLICATIONS

36. Bowker NI, Tuffin K. Understanding positive subjectivities made possible on-line for disabled people. *New Zealand J Psychol.* 2007;36(2):63.

37. Shpigelman C, Weiss PL (Tamar), Reiter S. E-mentoring for all. *Comput Hum Behav.* 2009;25(4):919–928.

38. Allen TD, Eby LT. Relationship effectiveness for mentors: factors associated with learning and quality. *J Manag.* 2003;29(4):469–486.

39. Ren Y, Harper FM, Drenner S, *et al*. Building member attachment in online communities: applying theories of group identity and interpersonal bonds. *Mis Quarterly.* 2012;36(3):841–864.

40. Lampe C, Ellison N, Steinfeld C. A familiar Face(book): profile elements as signals in an online social network. *Proc SIGCHI Conf Hum Factor Comput Sys.* 2007:435–44.

41. Gosling SD, Gaddis S, Vazi, S. Personality impressions based on Facebook profiles. *Proceedings of the International Conference on Weblogs and Social Media.* 2007:435–444.

42. Crandall D, Cosley D, Huttenlocher D, *et al*. Feedback effects between similarity and social influence in online communities. *Proceedings of 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2008:160–168.

43. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annu Rev Sociol.* 2001;27:414–444.

44. Cutrona CE, Russell DW. The type of social support and specific stress: toward a theory of optimal matching. In: Sarason BR, Sarason IG, Pierce, GR, eds. *Social Support: An Interactional View. Wiley Series on Personality Processes.* Oxford, England: John Wiley & Sons; 1990:319–366.

45. Turner JW, Grube JA, Meyers J. Developing an optimal match within online communities: an exploration of CMC support communities and traditional support. *J Commun.* 2001;51(2):231–251.

46. Vlahovic TA, Wang YC, Kraut RE, Levine JM. Support matching and satisfaction in an online breast cancer support community. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems.* 2014:1625–1634.

47. Wang Z, Walther JB, Pingree S, *et al*. Health information, credibility, homophily, and influence via the internet: Web sites versus discussion groups. *Health Commun.* 2008;23:258–368.

48. Hoey LM, Ieropoli SC, White VM, *et al*. Systematic review of peer-support programs for people with cancer. *Patient Educ Couns.* 2008;70(3):315–337.

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

50. Swan M. Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. *J Pers Med.* 2012;2(3):93–118.

51. Wicks P, Vaughan TE, Massagli MP, *et al*. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol.* 2011;29(5):411–414.

52. Nakamura C, Bromberg M, Bhargava S, *et al*. Mining online social network data for biomedical research: a comparison of clinicians' and patients' perceptions about amyotrophic lateral sclerosis treatments. *J Med Internet Res.* 2012;14(3):e90.

53. Wen M, Rose CP. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. *GROUP ACM SIGCHI Int Conf Support Group Work.* 2012:179–188.

54. Jha M, Elhadad N. Cancer stage prediction based on patient online discourse. *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing.* 2010:64–71.

55. Hartzler AL, McDonald DW, Park A, *et al*. Evaluating health interest profiles extracted from patient-generated data. *AMIA Annu Symp Proc.* 2014;2014:626–635.

56. Terveen L, McDonald DW. Social matching: a framework and research agenda. *ACM Trans Comput Hum Interact.* 2005;12(3):401–434.

57. Civan-Hartzler A, McDonald DW, Powell C, *et al*. Bringing the field into focus: user-centered design of a patient expertise locator. *Proc SIGCHI Conf Hum Factor Comput Sys.* 2010;2010:1753577.

58. McDonald DW, Ackerman MS. Just talk to me: a field study of expertise location. *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work.* 1998:315–324.

59. Ehrlich K, Shami NS. Searching for expertise. *Proc SIGCHI Conf Hum Factor Comput Sys.* 2008:1093–1096.

60. Reichling T, Wulf V. Expert recommender systems in practice: evaluating semi-automatic profile generation. *Proc SIGCHI Conf Hum Factor Comput Sys.* 2009:59–68.

61. Liu H, Maes P. Interestmap: harvesting social network profiles for recommendations. *Beyond Personalization Workshop, International Conference on Intelligent User Interfaces.* 2005:56.

62. Carullo G, Castiglione A, De Santis A. Friendship recommendations in online social networks. *Proceedings of IEEE International Conference on Intelligent Networking and Collaborative Systems (INCoS).* 2014:42–48.

63. Chen J, Geyer W, Dugan C, Muller M, Guy I. Make new friends, but keep the old: recommending people on social networking sites. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems.* 2009:201–210.

64. Park A, Hartzler A, Huh J, McDonald DW, Pratt W. Automatically detecting failures in natural language processing tools for online community text. *J Med Internet Res.* 2015;17(8):e212.

65. Pennebaker JW, Booth RJ, Francis ME. *LIWC2007: Linguistic Inquiry and Word Count.* Austin, Texas. http://www.liwc.net. Accessed 28 May 2015.

66. Pennebaker W, Francis ME. Cognitive, emotional, and language processes in disclosure. *Cogn Emot.* 1996;10(6):601–606.

67. Salton G, Wong A, Yang CS. A vector space model for automatic indexing. *Communications of the ACM.* 1975;18(11):613–620.

68. Golbeck J, Robles C, Turner K. Predicting personality with social media. *Proc SIGCHI Conf Hum Factor Comput Sys Extended Abstracts.* 2011:253–262.

69. Ireland ME, Slatcher RB, Eastwick PW, *et al*. Language style matching predicts relationship initiation and stability. *Psychol Sci.* 2011;22(1):39–44.

70. Shami NS, Ehrlich K, Gay G, *et al*. Making sense of strangers' expertise from signals in digital artifacts. *Proc SIGCHI Conf Hum Factor Comput Sys.* 2009:69–78.

71. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria; 2013. ISBN 3-900051-07-0. http://www.R-project.org. Accessed 28 May 2015.

72. Beyer H, Holtzblatt K. *Contextual design: defining customer-centered systems.* San Francisco (CA): Morgan Kaufmann Publishers; 1998.

73. Preece J, Ghozati K. Observations and explorations of empathy online. In: Rice RR, Katz JE, eds. *The Internet and Health Communication: Experience and Expectations.* Thousand Oaks, CA: Sage Publications Inc.; 2001:237–260.

74. Preece J. Etiquette online: from nice to necessary. *Commun ACM.* 2004;47(4):56–61.

75. Qiu B, Baojun, Kang Zhao K, *et al*. Get online support, feel better – sentiment analysis and dynamics in an online cancer survivor community. *Proceedings of Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom).* 2011:274-281.

76. Wang Y, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. *Proceedings the 2012 ACM Conference on Computer Supported Cooperative Work.* 2012:833–842.

77. Zhang S, Bantum E, Owen J, *et al*. Does sustained participation in an online health community affect sentiment? *AMIA Annu Symp Proc.* 2014;2014:1970–1979.

78. Brooks M, Kuksenok K, Torkildson MK, *et al*. Statistical affect detection in collaborative chat. *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work.* 2013:317–328.

79. Argamon S, Koppel M, Pennebaker JW, *et al*. Automatically profiling the author of an anonymous text. *Commun ACM.* 2009;52(2):119–123.

80. Back MD, Stopfer JM, Vazire S, *et al*. Facebook profiles reflect actual personality, not self-idealization. *Psychol Sci.* 2010;21(3):372–374.

81. Pizzato L, Rej T, Akehurst J, Koprinska I, Yacef K, Kay J. Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Model User-Adapt Interact.* 2013;23(5):447–488.

82. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *JAMIA.* 2011;18(5):544–551.

83. MacLean DL, Heer J. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. *JAMIA.* 2013;20(6):1120–1127.

84. Jain SH, Powers BW, Hawkins JB, *et al.* The digital phenotype. *Nat Biotechnol.* 2015;33(5):462–463.

85. Elhadad N, Zhang S, Driscoll P, *et al.* Characterizing the sublanguage of on-line breast cancer forums for medications, symptoms, and emotions. *AMIA Annu Symp Proc.* 2014;2014:516–525.

86. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. *J Med Internet Res.* 2011;13(2):e37.

87. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. *Proc IEEE Symp Visual Lang.* 1996:336–343.

88. Shneiderman B, Plaisant C, Hesse, B. Improving health and healthcare with interactive visualization tools. *IEEE Computer.* 2013;46(5):58–66.

89. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? *JAMA.* 2013;310(22):2395–2396.

90. Ashley EA. The precision medicine initiative: a new national effort. *JAMA.* 2015; Published online April 30, 2015. doi:10.1001/jama.2015.3595

91. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. *J Med Internet Res.* 2014;16(5):e126.

92. Peacock SE. How web tracking changes user agency in the age of Big Data: the used user. *Big Data Soc.* 2014;1(2):2053951714564228.

93. Hartzler A, Skeels MM, Mukai M, Powell C, Klasnja P, Pratt W. Sharing is caring, but not error free: transparency of granular controls for sharing personal health information in social networks. *AMIA Annu Symp Proc.* 2011:559–68.

94. Li J. Privacy policies for health social networking sites. *JAMIA.* 2013;20(4):704–707.

95. Bruckman A. Teaching students to study online communities ethically. *J Inform Ethics.* 2006;15(2):82–98.

AUTHOR AFFILIATIONS

[1]Group Health Research Institute, Seattle, USA

[2]Human Centered Design and Engineering, University of Washington, Seattle, USA

[3]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, USA

[4]The Information School, University of Washington, Seattle, USA

RESEARCH AND APPLICATIONS