Journal of the American Medical Informatics Association, 25(8), 2018, 931–944
doi: 10.1093/jamia/ocy012
Advance Access Publication Date: 18 April 2018
Research and Applications

Research and Applications

Can online social support be detrimental in stigmatized chronic diseases? A quadratic model of the effects of informational and emotional support on self-care behavior of HIV patients

Xunyi Wang,[1] Srikanth Parameswaran,[1] Darshan Mahendra Bagul,[2] and Rajiv Kishore[1]

[1]Department of Management Science and Systems, State University of New York at Buffalo, Buffalo, NY 14260-4000, USA and [2]Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260-4000, USA

Corresponding Author: Rajiv Kishore, Department of Management Science and Systems, State University of New York at Buffalo, 325C Jacobs Management Center, Buffalo, NY 14260-4000, USA; rkishore@buffalo.edu

Received 27 July 2017; Revised 17 January 2018; Editorial Decision 28 January 2018; Accepted 1 February 2018

ABSTRACT

Objective: We studied the impact of online social support on patient self-care behavior in an online health community for human immunodeficiency virus (HIV) patients. We conceptualized emotional and informational support provided by community members into nuanced sub-dimensions. We explored how the direct and interaction effects of these sub-dimensions impact the self-care behavior of a support seeker.

Methods: We used data from 330 255 posts in 30 050 threads from POZ, an online health community for HIV patients. Our key variables—self-care behavior$_i$, objective information$_j$, experiential support$_j$, and emotional tone$_j$—were operationalized using linguistic analysis with self-generated dictionaries and Python libraries. We tested our hypotheses using Tobit regression.

Results: Out of 6 null hypotheses, 5 were rejected. Objective information and emotional tone had an inverted-U relationship with self-care behavior. Experiential information and community involvement were positively related to self-care behavior. Community involvement amplified the inverted-U relationship between emotional tone and self-care behavior. No significant interaction effect was found between experiential support and objective information.

Conclusions: Beyond a threshold, both informational and emotional online social support had a deleterious impact on self-care behavior of HIV patients. Our results suggested that caution should be exercised in the use of online health community interventions for HIV patients, and perhaps patients with other stigmatized chronic diseases.

INTRODUCTION

Advances in information technology, particularly social media, has created a new avenue for providing and seeking online social support via the Internet. Patients who suffer from chronic diseases, disabilities, or cancers find social media particularly useful as it enables them to seek relevant information and support from peers or experts to help them with their self-management of such long-term diseases.[1] On the other side, patients also provide social support to their peers via social media by sharing health-related information and their personal experiences about coping with and self-care of their diseases. Recent studies have shown that 40% of those who sought health-related information on social media also shared their personal health experiences.[2] A recent meta-analysis showed an overall positive impact of social networking site interventions on health-related behavioral outcomes.[3] A study by Yan and Tan[4] shows a positive effect of online support on patients' mental health.

While the findings on beneficial health impacts of online social support looks promising, we identified 4 avenues for further advancing this stream of research. First, while the literature analyzes impact of online social support on mental and emotional states,[4] there is lack of research on the role of online social support in promoting self-care behavior of a chronic disease.[1] Although patients' mental states such as good mood and happiness are desirable outcomes, they do not guarantee that the patients are performing actions required to manage their disease. Self-care "involves the range of activities individuals undertake to enhance health, prevent disease, evaluate symptoms, and restore health."[5] Since most chronic diseases are treated in home and outpatient settings, active patient participation is pivotal in managing the disease.[6] Thus, it is important to analyze self-care behavior as a key outcome of online social support in chronic disease management.

Second, informational and emotional support, the 2 types of online social support analyzed in this literature, are analyzed in a very broad sense.[4] We argue that informational and emotional support could have important dimensions. For example, consider informational support. An online community member could provide informational support based on facts taken from the Internet or based on personal experience. This argument suggests that informational support could be conceptualized as objective information and experiential information. However, the literature is yet to disaggregate informational and emotional support into nuanced dimensions. The qualitatively different dimensions, taken separately and simultaneously (i.e., the interaction with one another) could have differential impacts on a support seeker's self-care behavior. Thus, it is important to conceptualize the fine-grained types of informational and emotional support.

Third, prior research indicates positive health outcomes due to online social support.[4,7,8] However, ample evidence across disease contexts in the offline social support literature shows that beyond a threshold social support can lead to negative health outcomes.[9,10] At higher levels, social support could manifest as over involvement from support providers, where the providers "become worrisome, overprotective, intrusive, and excessively indulgent and self-sacrificing in a way that burdens the patient and discourages autonomy and personal responsibility for self-care."[11] Moreover, discourse on the dark side of online health communities suggests these communities are a haven for overload of information and contacts, yielding a "paralyzing effect" on information seekers.[12,13] Thus, there is a need to reconcile these differing views, and hypothesize and test a valid relationship between online social support and self-care behavior.

Fourth, the literature is silent on the role of online social support in the management of stigmatized chronic diseases like human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS). Patients suffering from these diseases face discrimination and, hence, have trouble disclosing their identity and establishing relationships.[14] This could have an impact on the quantity and quality of online social support sought and provided, and the ensuing self-care behavior. So, it is important to analyze whether online social support will induce self-care behavior in the stigmatized chronic disease context. Having identified the gaps in the literature in this area, we asked the research question: does online social support work in improving self-care behavior of a stigmatized chronic disease?

We addressed this research question in the following ways and make four contributions to the literature. First, with self-care behavior as the outcome of interest, we theorized the impact of online social support. By doing so we added to the online social support literature, which has only studied emotional and mental health outcomes. Second, we conceptualized informational and emotional support into fine-grained distinct types and hypothesize their direct and interactive effects on self-care behavior. By doing so, we provided a nuanced understanding of online social support, and its various types. Third, we integrated contradicting views on the impact of online social support by hypothesizing and testing a curvilinear relationship with self-care behavior. Fourth, we tested our hypotheses using a large dataset we constructed using data from an online health community for HIV patients. As a chronic disease and as a source of discrimination, HIV has societal and economic impacts.[15] Studies have reported poor adherence on self-care behavior ranging from 33% to 88%.[16] By analyzing the HIV online community, we extended the research on the effects of online social support to stigmatized chronic disease context.

HYPOTHESES

We disaggregated online informational and emotional support into 2 sub-dimensions—objective information and experiential information as 2 dimensions of informational support, and emotional tone and community involvement as 2 dimensions of emotional support. Table 1 provides the definitions, rationale, and the key differences between these sub dimensions. It also describes how we conceptualized these dimensions in an online health community context. A glossary of key terms used in this study is also provided in this table.

We built an online social support model of self-care behavior having conceptualized the nuanced dimensions. In the following section, we develop our model's 6 hypotheses. Figure 1 illustrates our research model.

Objective information includes facts that explain the underlying scientific mechanisms[26] including information about the etiology of the disease, information about medication and treatment, and information about coping with the disease and self-care. By virtue of providing knowledge of health risks and benefits, objective information makes the patients more informed. It helps patients understand the problems and the possible solutions associated with their disease. So, more objective information to the patient would suggest better understanding of their condition, and better actions for health improvement.[4,27] However, beyond a threshold objective information manifests as information overload—a stressor for patients.[28] Due to information overload patients may perceive the support provider's reasoning behind self-care actions logically inconsistent.[29] They may tend to expend less cognitive effort[4,30] and over-simplify the support provided by the community.[29,31] So, patients may be less informed or even misinformed about the self-care. Overall, we argue that at low levels of objective information, patients will have no understanding of their disease. With increasing support, they become better informed and, hence, engage in self-care behavior. However, excessive information will manifest as overload and lead to superficial or even incorrect understanding of support provided, and; hence, patients will not engage in self-care behavior. Therefore, we hypothesize:

> *H1. There is an inverted U relationship between the objective information provided by online community members (j's) who respond to an individual member i and the engagement of self-care behavior by i.*

There is abundant research delving into the importance and the impact of emotional management on the treatment of chronic diseases.[32–34] Patient morale is highly dependent on the emotional tone of the stakeholders, such as family, friends, and similar others.[32] Emotional tone sends a signal to the support seeker that he/she is not alone and is taken care of.[4] Emotional tone sets the mood of a

Table 1. Glossary of Terms Used in the Study

Term	Definition\conceptualization
Online health community	Websites that provide a means for patients and their families to learn about an illness, seek and offer support, and connect with others in similar circumstances[17] **Example: patientslikeme.com, dailystrength.org**
Thread	Online communities are subdivided into broader topics. Within each topic, a new discussion can be started by a user and responded to by many other users. Each new discussion that is started is called as a thread[18] We conceptualized each thread as a "virtual group" formed to provide support to the support seeker. The support seeker is a patient who started a thread by posting some text that seeks support
Social support	Social support refers to "information leading the subject to believe that he is cared for and loved, esteemed, and a member of a network of mutual obligations."[19] It is formed by the exchange of resources including verbal and nonverbal messages between 2 or more individuals.[20] Social support as a strategy to cope with stress, and improve behavioral, psychological, and physiological outcomes is well established in the public health, clinical psychology, and sociology literatures. By either directly integrating the receiver of support into a social network or by indirectly affording interpersonal resources to the receiver, social support affects health outcomes.[21,22]
Support Seeker	A user of the online health community who posted a question. We denoted the support seeker as (i)
Support Provider	Support providers are all online health community members who responded to a support seeker's question in a thread. We denoted support providers as (j).
Definition and conceptualization of variables in our model	
Self-care Behavior	The range of activities individuals undertake to enhance health, prevent disease, evaluate symptoms, and restore health[5] **Conceptualization:** Expression of self-care behavior by the support seeker in all his/her subsequent response after he/she initiated a thread
Objective Information	Objective information refers to receiving primarily factual information related to a particular disease condition and its management.[23] Objective information can be received by a patient suffering from a particular disease condition from a number of stakeholders in the healthcare ecosystem including healthcare professionals, family members, and friends (who may do some research on a particular disease and share information with their loved ones), and other patients also suffering from that same disease. **Conceptualization:** All the objective information a support seeker receives from community members in response to his/her question in a thread
Experiential Information	Experiential information refers to receiving information about actual experiences and insights of and the strategies used by patients suffering from the same disease as the focal patient.[8] Obviously, this type of information can only be provided by other patients suffering from the same disease condition and not by other stakeholders in the healthcare system. This type of support can focus on either the experiences pertaining to the disease and its symptoms or on experiences pertaining to treatment and management of the disease **Conceptualization:** All the experiential information a support seeker receives from community members in response to his/her question in a thread
Emotional Tone	Emotional tone is essentially psychosocial support received by a patient suffering from a chronic disease condition that is provided by the giver with the intent to provide care, cheer, comfort, and relief to the patient to enable him/her to cope with the symptoms of his/her chronic disease and to engage effectively in the self-care of the disease condition.[24,25] **Conceptualization:** Emotional tone from all the replies a support seeker receives from community members in response to his/her question in a thread
Community Involvement	Community involvement is the extent to which online community members participate in a thread. Community involvement is closely related to the companionship one would typically get with offline friends through chatting, group meetings, and other social activities.[4] We considered community involvement as another dimension of emotional support because it denotes the intimacy and the emotional attachment the community members have towards the online health community and its support seekers **Conceptualization:** Number of unique community members who reply to a support seeker in response to his question in a thread

focal post, and a positive tone may shift members' focus from anxiety about their condition to how to deal with it properly. Emotional tone in the online context is even more salient for our HIV context because of the stigma associated with the disease, difficulty in disclosing their disease, and building relationships in the offline context. Knowing that other patients are successfully coping with HIV will provide assurance, and reinforcement, and reduce the attribution of personal blame.[35–37] Support seekers develop intimate relationships with online community members and motivate each other to perform health enhancing self-care actions.[38] However, beyond a threshold,[9] support seekers may perceive excessive emotional tone as "forced optimism"—rated as one of the unsupportive interactions from support providers.[39,40] Evidence shows that such interactions lead HIV patients to use coping strategies such as denial and disengagement from self-care.[41] Patients may generate thoughts such as "I am dying and that's why people would say some good words to comfort me." Moreover, due to excessive emotional tone, support seekers might depend on others for satisfaction, thereby losing autonomy and control. Such a subversion of self-efficacy beliefs will negatively impact their self-care behavior. In sum, moderate

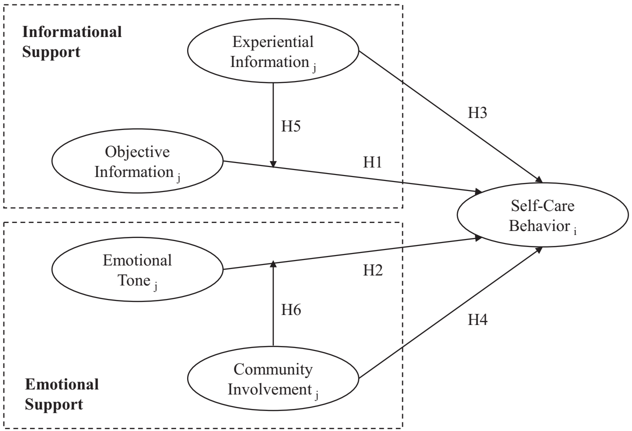

Figure 1. Research Model.

levels of emotional tone would motivate patients but excess optimism may cause hopelessness and loss of competence for self-care. At low levels of emotional tone, patients lack the emotional stability needed to get over the anxiety and stigma surrounding their condition. Thus, we hypothesize that:

> *H2. There is an inverted U relationship between the emotional tone in posts of online community members (j's) who respond to an individual member i and the engagement of self-care behavior by i.*

We posit that experiential information will be positively related to support a seeker's self-care behavior. Support seekers who receive experiential information will be better informed. However, as opposed to objective information where the seeker received scientific information about the disease, experiential information provides first-hand information from a patient who is suffering from or who has successfully overcome a similar health condition.[8] Therefore, in addition to making the patients more informed, experiential information provides a model for the support seeker to follow. The social-cognitive theory argues that when people observe a model performing a behavior and the effects of the behavior, they are better able to remember and replicate that behavior.[42] Accordingly, the higher the level of experiential information, the better will be the learning of the positive consequences of self-care from other patients' personal experiences, and better actions from the seeker to improve his/her health. Thus, we hypothesize that:

> *H3. The higher the experiential information provided by online community members (j's) who respond to an individual member i, the more likely i will engage in self-care behavior.*

We also argue that community involvement will be positively related to self-care behavior. Community involvement is closely related to the companionship one would typically get with offline friends through chatting, group meetings, and other social activities.[4] The companionship aspect of community involvement makes the support seeker believe that there are community members who enjoy the seeker's presence.[4] Consequently, support seekers believe that they are more than just an individual and get integrated into the threaded discussion.[4] Moreover, the support provided by the community members acts as a "talk therapy" to motivate them to perform self-care behaviors.[4] Community involvement is also closely related to the concept of social presence. In an online community context, social presence is the extent to which an online community projects or increases the awareness of community members to the support seeker.[43] Higher community involvement increases the social presence of online community members. By an enhanced sense of human contact, the community members act as a form of social capital providing the support needed to perform health enhancing actions.[44] When support seekers obtain a better sense of human contact, they feel a higher social presence and can trust the community members, which in turn alleviates the uncertainty associated with the support provided by them.[45,46] Thus, we hypothesize that:

> *H4: The higher the involvement of online community members (j's) who respond to an individual member i, the more likely i will engage in self-care behavior.*

Objective information is more influential if it is augmented with personal experience. In such situations, not only do the support seekers have information regarding the disease and its management through self-care behavior, they also have evidence to validate from personal experiences of community members. Experiential information adds validity and personal touch to the objective information. Yan and Tan[4] used a broader social support construct measuring both information and experiential support, and found significant health outcomes in online communities. So, we argue that objective information in a thread with high experiential support will influence a support seeker's self-care behavior better than a thread where there is low objective information. Thus, we hypothesize that:

> *H5. The experiential information provided by online community members (j's) who respond to an individual member i moderates the inverted U relationship between the objective information provided by (j's) to i and self-care behavior of i such that the positive impact of objective information on self-care behavior is stronger at higher levels of experiential information.*

Community involvement can also amplify the impact of emotional tone on self-care behavior. The social presence theory argues that media high on social presence is best suited for conveying affective information.[47] By this logic, high community involvement in a thread makes the community members more salient for the support seeker, and this should amplify the influence of emotional support provided by them. Research shows that perceived social presence increases the trust in the message source, usefulness of the message, and comfort in processing the affective cues.[48] So, perception of high social presence in threads makes support seekers more receptive to the emotional tone of community members. Users would view the thread as a warm, sociable, and personal virtual counseling team.[43] In sum, the emotional tone and the associated care, cheer, comfort, and relief provided by such threads will better motivate the support seeker to perform health-enhancing actions. Thus, we hypothesize that:

> *H6. The involvement of online community members (j's) who respond to an individual member i moderates the inverted U relationship between the emotional tone in posts of online community members (j's) to i and self-care behavior of i such that the positive impact of emotional tone on self-care behavior is stronger at higher levels of community involvement.*

METHODS

Data Sample

The dataset used in this research was constructed from an online health community—POZ, which was founded to facilitate information sharing that can help improve the lives of patients suffering from HIV. POZ serves its community members with daily news, treatment updates, personal profiles, videos, blogs, discussion forums, and an extensive online social network that includes 150 000 members. POZ's discussion forums (https://forums.poz.com) are one of the most popular and active online HIV/AIDS support forums on the internet with an outreach that includes 70% of the HIV positive people living in the United States. The forum supports more than 30 000 registered members and 20 forums organized by criteria such as HIV prevention and testing, disease stages ranging from newly diagnosed to long-term survivors, treatments and nutrition (e.g., antiretroviral therapy, Lipodystrophy and metabolic problems), and demographic characteristics (e.g., positive women). In these forums, members ask questions, share stories, and read posts from other community members.

We collected data on posts, users and their profiles from May 2006 to March 2017 on 5 forums of POZ, namely "Living With HIV," "I Just Tested Poz," "Questions About Treatment and Side Effects," "Nutrition and HIV," and "Mental Health and HIV." Though we collected data from the forum "Research News and Studies," we did not include the threads from this forum in our analysis because these threads in that forum were focused on the discussion of HIV related news, and research studies on HIV medicines and cures. During the time period of our data collection, there were a total of 15 588 unique users who wrote 330 255 posts belonging to 30 050 discussion threads. In our analysis, we included every post during the data collection period. The median of total posts in a thread was 6 and the mean was 11. The median of the posts without support seekers' posts was 4 and the mean was 8. The median lifespan of a thread was 1 day and the mean was 45 days. Approximately 49% of the registered members did not actively participate in forum discussions by posting a message.

Measures

Prior literature used 2 broad automated methodologies for measuring variables from text data in online health communities. One methodology is based on linguistic features while the other is based on machine learning. Methods using linguistic features employ tools like Linguistic Inquiry and Word Count (LIWC) to extract the frequency count or percentage of words in a text that match the words from custom built and validated dictionaries representing specific linguistic dimensions (e.g., tense, pronouns), personal concerns (e.g., death, leisure), and psychological constructs (e.g., negative emotion).[49] Other operationalization methods using linguistic features extract features such as sentence count, words per sentence, parts of speech, question score that capture the complexity, style, and length of text messages. Tools such as OpinionFinder are used for these methods.[49,50] Methods using machine learning employ unsupervised and supervised learning techniques. In unsupervised learning, statistical models like the Latent Dirichlet Allocation are used to identify hidden topics and the words corresponding to these topics in text messages.[49] Experts are used to prune down the topics and identify meaningful ones. Subsequently, variables are operationalized based on the extent of occurrence of these topics using statistical prediction or a word frequency count. In supervised learning, human coders assign numerical values or value labels to denote the extent of a construct in text messages. Statistical techniques like Latent Semantic Analysis are used to extract features from the text messages and the extent to which a feature can predict a value or a label. These features and their weights are then applied to the larger sample to predict the extent to which a text message represents constructs. Tools like Python, R, and LingPipe are used for machine learning purposes.[4]

For this study, we used the LIWC program,[51] a popular linguistic analysis tool. We used LIWC 2015 for analyzing the texts posted by the patients on the discussion forums. LIWC 2015 also allows us to define a custom dictionary for analyzing text, a feature that we would be using for the calculation of self-care behavior scores. Prior research demonstrated a moderate correlation between ratings assigned by a human rater and LIWC Scores.[52] This tool has been used in the online health community context as well. Wang et al.[49] used selected dictionaries from LIWC dictionaries to investigate the effects of emotional and informational support on commitment of patients to online health support groups.

In this study, we conducted the analysis at the thread level. We concatenated all posts of community members (j) and support seeker (i) separately and use LIWC to analyze these texts. By focusing on the thread level, we were able to capture all social support a support seeker received from community members in response to his/her question. We were also able to model the self-care behavior as a function of social support by capturing the subsequent expression of self-care by the support seeker and regressing it on social support received from the community members. Using a thread-level ensures that the support seeker (i) indeed solicited and received social support from the community members (j).

Table 2 lists the dependent, independent, and control variables, and their operationalizations. For our 4 key variables, we validated our automated measurements by having expert raters manually rate posts from our HIV forums. We provided 105 threads containing 485 posts to an HIV physician, who helped us to validate the measurement of our dependent variable—self-care behavior. These threads were selected randomly from 3 strata reflecting no, low, and high self-care to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The 3 strata were based on an initial score that was assigned to self-care behavior using LIWC standard dictionaries for "achievement" and "reward" and a created custom dictionary based on the adherence questionnaire of AIDS Clinical Trials Group (ACTG). This coder coded the self-care behavior of the support seeker (i) in each post in a thread on a 3-level scale: 0 (no self-care behavior evident), 1 (low self-care behavior evident), and 2 (high self-care behavior evident). We then averaged the categorical values of self-care behavior for all posts in a thread to get an aggregate value of self-care behavior expressed by i in that thread.

Similarly, we provided 70 threads containing 585 posts to a senior HIV researcher, who helped us to validate the measurement of objective information and experiential information. These threads were selected randomly from 2 strata reflecting low and high objective information to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The 2 strata were based on an initial score that was assigned to objective information using LIWC standard dictionaries for "biological processes" and "cognitive processes." This coder coded objective information and experiential information provided by support provider j's in each post in a thread separately on a 2-level scale: 1 (low objective or experiential information) and 2 (high objective or experiential information). We then averaged the categorical values of objective and experiential information separately for all posts in a thread to get 2 separate aggregate values for objective information and experiential information provided by j's in that thread.

Finally, in a similar manner, we provided 70 threads containing 586 posts to an HIV nurse practitioner who is the director of patient experience at an HIV clinic and counsels all HIV patients upon their initial diagnosis of HIV. These threads were selected randomly from 2 strata

Table 2. Variable Operationalization

Variable	Operationalization
Dependent variable	
Self-care behavior$_i$	To measure self-care behavior of an "*i*," we created a self-made dictionary based on ratings by experts. We provided to an HIV physician 395 posts belonging to 105 threads that were selected randomly from 3 strata reflecting no, low, and high self-care to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The three strata were based on an initial score that was assigned to self-care behavior using LIWC standard dictionaries for "achievement" and "reward" and a custom dictionary created based on the adherence questionnaire of ACTG. Using Python's Natural Language Took Kit (NLTK) and after removing stop words, we selected top 300 words from the posts that the expert coded as reflecting high self-care behavior. From these 300 words, we chose the words that most appropriately represent self-care behavior. After several iterations, we finalized a concise, pithy dictionary that succinctly captures self-care behavior. The raw score for the extent of self-care behavior of an "*i*" expressed in a thread is calculated as the unique frequency count of words in our self-care dictionary that appear in the concatenated subsequent posts by this "*i*" in this thread after his/her initial post. Prior literature shows that the frequency of appearance is a good indicator of relevance theoretically and empirically.[53–55] We adjusted this raw self-care score with 2 factors. First, it is possible that an "*i*" expresses self-care words but is expressing his/her intention, and not actual behavior. To account for such futurity in expression of a support seeker, we multiplied the raw self-care behavior score with a factor (100 – future_focus_score)/100. By doing this we penalize a post for high future focus. The future_focus_score was calculated as the percentage of words in the text that denote futurity—this value is calculated using the future focus dimension of LIWC. Second, there is also a possibility that support seeker "*i*" used self-care behavior-oriented words but was complaining about it. To account for this, we calculated the net emotion score for *i*'s responses in the thread by subtracting the LIWC negative emotional tone score from the positive emotional tone score. When the net score was positive, we used the self-care score as it is. If the net emotion score was negative, self-care behavior was expressed in a negative emotional environment, indicating that the individual is probably not engaged in self-care behavior but complaining about it. In this case, we converted the adjusted score of self-care behavior as described above to zero
Independent variables	
Objective information$_j$	This variable is measured as the amount of factual information about the disease and treatment management provided by "*j*'s" in a thread. To measure objective information, we created a self-made dictionary based on ratings by experts. We provided to a senior HIV researcher 586 posts belonging to 70 threads that were selected randomly from 2 strata reflecting low and high objective information to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The 2 strata were based on an initial score that was assigned to objective information using LIWC standard dictionaries for "biological processes" and "cognitive processes." Using Python's NLTK, we removed the stop words and selected top 300 words from the posts that the expert coded as containing high objective information. From the top 300 words, we chose the words that most appropriately represent objective information. After several iterations, we finalized a concise, pithy dictionary that succinctly captures objective information. Using this dictionary as input, we obtained the LIWC percentage score to quantify objective information in the concatenated subsequent posts by "*j*'s" in a thread after a support seeker's question. This LIWC percentage score was our score for objective information
Experiential information$_j$	This variable is measured as the amount of experiential information provided by "*j*'s" in terms of their personal stories and anecdotes from the past. To measure experiential information, we created a self-made dictionary based on ratings by experts. We provided to a senior HIV researcher 586 posts belonging to 70 threads that were selected randomly from 2 strata reflecting low and high experiential information to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The 2 strata were based on an initial score that was assigned to experiential information using LIWC standard dictionaries for "past focus." Using Python's NLTK, we removed the stop words and selected the top 300 words from the posts that the expert coded as containing high experiential information. From the top 300 words, we chose the words that most appropriately represent experiential information. After several iterations, we finalized a concise, pithy dictionary that succinctly captures experiential information. Using this dictionary as input, we obtained a raw experiential information score as the LIWC percentage score for the concatenated posts by "*j*'s" in a thread after an "*i*'s" question. Further, we adjusted the experiential information score by 2 factors: (1) second-hand information and (2) first-person account score. We adjusted downward the raw experiential information score for second-hand information by a factor (100 – second-hand information score)/100 because there is a possibility that a "*j*" may have used words like doctor, telling, etc. that may indicate that the information contained in the thread was obtained from second hand sources like the Internet, a doctor, etc. rather than based on personal experience. We adjusted upward the raw experiential information score for first-person account by a factor (first-person account score/100) because we found that when "*j*'s" share their personal experiences, they typically use first person singular pronouns like "*i*," "me," "mine," etc. For example, a post "I took the medicine and I felt better" has high experiential content compared to the zero experiential information in the post "my friend took the medicine and he felt better." To measure second-hand information, we first created a custom dictionary consisting of most-frequently used words by support providers that reflect second-hand information. To do this, we performed a word frequency count analysis on the text obtained after removing stop words from the concatenated posts of support providers in a thread using Python's NLTK. This custom dictionary was used to generate second-hand information scores for each thread using LIWC. Next, we calculated a first-person account score using the "first person singular pronoun" dictionary of LIWC. We used these two scores for adjustment as above
Emotional tone$_j$	This variable is measured as the LIWC score on "emotional tone" obtained by analyzing the concatenated posts of "*j*'s" in the focal thread
Community involvement$_j$	This variable is measured as the number of unique "*j*'s" who reply to an "*i*" in response to his/her question in a thread

(continued)

Table 2. continued

Variable	Operationalization
Control variables—capturing support seeker's (*i*'s) characteristics	
Emotional tone$_i$	This variable is measured by the LIWC score on "emotional tone" obtained by analyzing the concatenated posts of "*i*" in the focal thread. This measure can act as a proxy for the support seeker's sickness
Self-disclosure$_i$	This variable measures the willingness of an "*i*" to share personal information. Based on the information collected by this online community, we accounted for disclosure of age, gender, and location, with each being a binary code (0—not shared, 1—shared) and contributing to the self-disclosure score for the "*i*." For each user, we aggregated the three disclosure scores to form the self-disclosure score for an "*i*" (an integer value between 0 and 3). This score captures the openness of "*i*" and the extent to which he/she trusts this community
Degree centrality$_i$	This is a social network measure gauging the centrality of an "*i*" in the reply network. It is measured by summing up the in-degree and out-degree of the "*i*"[56]
Creator involvement$_i$	This variable was measured as the number of posts by an "*i*" in a thread initiated by him/her. This variable reflects "*i*'s" importance to that thread, which in turn could be related to self-care behavior
Question score of posts$_i$	This variable measures the possibility that an "*i*" asked questions about self-care behavior rather than actually engaged in self-care behavior. We first split the concatenated posts by "*i*" in a thread into individual sentences using Python's NLTK tokenizers. We then classified each sentence in that thread into two categories: question or not a question. Finally, we obtained the question score as (number of question sentences/number of total sentences)
Word count of first post$_i$	This variable measures the number of words in the first post by an "*i*"
Community activity$_i$	This variable was measured by counting the total number of threads generated by an "*i*" in this online community. This variable reflects how active the "*i*" is in the online community
Control variables—capturing support provider's (*j*'s) characteristics	
Crowd consensus$_j$	This variable reflects the degree of agreement in the support provided by "*j*'s" in a thread. To calculate it, we considered all replies posted by all "*j*'s" in a given thread. We generated a vector representation for each reply using Term Frequency-Inverse Document Frequency weighting. We used cosine similarity to measure the similarity of content between a pair of replies by community members. For every thread, we calculated similarity scores for every pair of replies and averaged all the scores to define crowd consensus for a thread
Self-disclosure$_j$	This variable measures the willingness of "*j*'s" to share personal information. Based on information collected by this online community, we accounted for disclosure of age, gender, and location, with each being a binary code (0—not shared, 1—shared) contributing to the self-disclosure score. We aggregated the 3 disclosure scores for each "*j*" to form the self-disclosure score for *j* (an integer value between 0 and 3). For each thread, we averaged the self-disclosure scores of all "*j*'s" in the thread to generate a self-disclosure score$_j$ for the thread
Control variables—capturing thread characteristics	
Year dummy variable	Our dataset spanned 13 years, and there is a possibility that the online social activity might have changed over the years. We included 12 dummy variables in our model to account for the effect of each year from 2006 to 2017
Previous-interaction$_{i,j}$	This measure quantifies the familiarity between an "*i*" and "*j*'s" offering social support within this thread. Previous interaction is measured by the average number of interactions between an "*i*" and "*j*'s" prior to the creation of this thread. For example, we traced the number of interactions between an i and j_1, the i and j_2 . . . , the i and j_n prior to focal thread, and averaged these numbers to get the previous interaction$_{i,j}$ for the entire thread
Thread duration	This variable was calculated as the difference between the date of creation of the thread and the date of last post (either by *i* or *j*). This measure gives a sense of the level of activity in the focal thread

reflecting low and high emotional tone to ensure variability in expert coding. The coding sample provided to the expert did not contain strata information. The 2 strata were based on an initial score that was assigned to emotional tone using LIWC composite for emotional tone. This coder helped us to validate the measurement of emotional tone, and coded emotional tone of support provider *j*'s in each post in a thread on a 2-level scale: 1 (low emotional tone) and 2 (high emotional tone). We then averaged the values of emotional tone for all posts in a thread to get an aggregate value of emotional tone for that thread.

We should note that we used a 3-level scale (with values of 0, 1, and 2) for coding our dependent variable self-care behavior in order to be consistent with our automated scores for self-care behavior which also contain a value of 0 either due to no expression of self-care by a support seeker or due to net negative emotional tone of the support seeker in the thread, as explained in our operationalization in Table 2. For the other 3 variables, we only used a 2-level scale (with values of 1 and 2) as a score of 0 can be categorized in the low level category.

Since the values for our 4 variables at the thread level coded by human coders are fractional values (as they are averages of values for all posts in a thread), and the automated values are continuous ones that were generated using the LIWC program, we used Spearman's ρ to measure the correlation between human and machine coding of our 4 key variables. Spearman's ρ is the appropriate measure to use to understand the level of correlation between machine-generated and expert-coded scores for our key variables based on the nonparametric nature of our expert coding data. Prior studies in the health informatics area, suggested a Spearman's ρ value of greater than or equal to 0.60 as acceptable for human-machine agreement in validating measures from automated text analysis tools.[57–59] The Spearman's ρ values for correlation between the human coder and LIWC scores were as follows: 0.75 for self-care behavior, 0.70 for objective information, 0.69 for experiential information, and 0.64 for emotional tone. All these correlations are above the recommended value of 0.60 for acceptable human-machine agreement, and provide validity to our variable operationalizations using the LIWC program, as described earlier in Table 2.

We controlled for characteristics of different entities and different levels including the thread characteristics, the support givers' characteristics, and the support seeker's characteristics, as discussed in Table 2, as they may have an impact on a support seeker's self-care behavior.

Analysis

Our estimation method is based on the nature of our dependent variable. Since our dependent variable was left censored with potential negative values coded as 0 values in our dataset, as discussed below, we chose the Tobit model to test our hypotheses. A dependent variable Y is censored when we know the true value of Y only for a restricted range of observations but we observe all independent variables (X's) for all observations. For such observations, when values of Y are in a certain range, a single value is reported. Right censoring occurs when Y is above a certain value but we do not know by how much.[60] Left censoring occurs when Y is below a certain value but we do not know by how much.[60] In our dataset, we captured self-care behavior based on the support seeker's expression of the same in his/her replies in a thread. Non-expression of self-care behavior by a support seeker (which will be given a score of 0 in our operationalization) does not imply that the support seeker did not engage in self-care behavior. It is possible that the support seeker engaged in positive self-care behaviors and did not express it, receiving a value of 0 on his/her self-care score. It is also possible that the support seeker engaged in negative behaviors that are contrary to self-care behaviors, and did not express them in the thread. In this instance, the true value of self-care behavior will be a negative score but our program would still give this individual a score of 0 on self-care behavior in that thread. Finally, in some cases, individuals may use words associated with self-care but are overly negative in their emotional tone suggesting that they may actually be engaging in behaviors contrary to self-care. In such cases, we converted the self-care score to zero as discussed in our operationalization above. Therefore, there is an issue of left censoring in our dependent variable with many observations potentially having a negative value that were coded as 0 in our dataset. In the presence of such censoring, standard estimation methods such as ordinary least-squares regression will not yield consistent parameter estimates.[61] To address this issue, we performed Tobit regression to model our dependent variable. Tobit regression uses all observations, both those at the threshold value of 0 and those below it, to estimate a regression line. The Tobit model is preferred, in general, over alternative techniques that estimate a line only with the observations above or below the threshold and/or neglect information about censoring.[62]

A requirement of the Tobit model is that the observations are independent. Independence is violated if the regression residuals (errors) are correlated within a cluster or a group and uncorrelated across a group.[61] In our data, a support seeker i could create multiple threads and subsequently express self-care in those threads. Since we perform our analysis at the thread level, our data contains multiple observations from some of the support seekers. While our data contains independent units in threads, the errors of the threads across a support seeker i are correlated. So, if our Tobit model over predicts for one thread of a support seeker, it could also over predict for another thread of the same support seeker.[61] Consequently, we cannot make valid inferences based on the estimated standard errors. By clustering across i, we are telling our estimator the group (or the support seeker) to which each thread belongs. Accordingly, the Tobit estimator alters the calculation of standard errors that are robust to this clustering. We can now make valid inferences on our hypotheses because we have accounted for the fixed effects of i. In addition, to overcome the potential issue of multicollinearity among independent variables, we standardized all independent variables, i.e., transformed them to their respective z-scores. We used STATA 14 to estimate our model. Our model equation is provided below.

$$
\begin{aligned}
\text{Self-Care Behavior}_i^* = {} & b_1\ \text{Year}_{2006} + b_2\ \text{Year}_{2007} + b_3\ \text{Year}_{2008} \\
& + b_4\ \text{Year}_{2009} + b_5\ \text{Year}_{2010} + b_6\ \text{Year}_{2011} \\
& + b_7\ \text{Year}_{2012} + b_8\ \text{Year}_{2013} + b_9\ \text{Year}_{2014} \\
& + b_{10}\ \text{Year}_{2015} + b_{11}\ \text{Year}_{2016} + b_{12}\ \text{Year}_{2017} \\
& + b_{13}\ \text{Thread Duration} \\
& + b_{14}\ \text{Creator Involvement}_i \\
& + b_{15}\ \text{Previous Interaction}_{i,j} \\
& + b_{16}\ \text{Emotional Tone}_i \\
& + b_{17}\ \text{Self-Disclosure}_i \\
& + b_{18}\ \text{Degree Centrality}_i \\
& + b_{19}\ \text{Word Counts of First Post}_i \\
& + b_{20}\ \text{Community Activity}_j \\
& + b_{21}\ \text{Crowd Consensus}_j \\
& + b_{22}\ \text{Self-Disclosure}_j \\
& + b_{23}\ \text{Question Score of posts}_i \\
& + b_{24}\ \text{Objective Information}_j \\
& + b_{25}\ \text{Emotional Tone}_j \\
& + b_{26}\ \text{Objective Information}_j \\
& \times \text{Experiential Information}_j \\
& + b_{27}\ \text{Emotional Tone}_j \\
& \times \text{Community Involvement}_j \\
& + b_{28}\ \text{Objective Information squared}_j \\
& + b_{29}\ \text{Emotional Tone squared}_j \\
& + b_{30}\ \text{Experiential Information}_j \\
& + b_{31}\ \text{Community Involvement}_j \\
& + b_{32}\ \text{Objective Information squared}_j \\
& \times \text{Experiential Information}_j \\
& + b_{33}\ \text{Emotional Tone squared}_j \\
& \times \text{Community Involvement}_j
\end{aligned}
$$

where

$$
\text{Self-care behavior}_i = \begin{cases} \text{Self-care behavior}_i^* & \text{if Self-care behavior}_i^* > 0 \\ 0 & \text{if Self-care behavior}_i^* \le 0 \end{cases}
$$

RESULTS

Table 3 reports the means, standard deviations, and correlations for all variables. The correlations provide a sense of relationship among the constructs. For instance, self-care behavior$_i$ was positively correlated to experiential information$_j$, emotional tone$_j$, and community involvement$_j$.

The results of the Tobit regression are presented in Table 4. Our Tobit analyses proceeds by first introducing our control variables (model 1), followed by the linear terms (model 2) and the quadratic terms (model 3), in a hierarchical way. A significant coefficient on the squared term indicates a curvilinear relationship. Positive sign for this coefficient would indicate a U-shaped pattern, whereas a negative sign would indicate an inverse U-shaped pattern.[63,64] Therefore, a negative sign on β_{28} and β_{29} in our results would be in line with H1 and H2, respectively. A significant interaction of a quadratic term with a linear moderator term would indicate linear moderation of a quadratic relationship. Accordingly, a negative sign on β_{32} and β_{33} in our results will be in line with H5 and H6.[64]

Table 3. Correlation Table

Variable	Mean	Min	Max	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0. Self-care behavior$_i$	3.2 (6.0)	0	79.88	1.00																											
1. Year-2006	0.1 (0.3)	0	1	0.00	1.00																										
2. Year-2007	0.1 (0.4)	0	1	0.00	−0.15	1.00																									
3. Year-2008	0.1 (0.3)	0	1	0.03	−0.11	−0.13	1.00																								
4. Year-2009	0.1 (0.3)	0	1	0.02	−0.11	−0.12	−0.09	1.00																							
5. Year-2010	0.1 (0.3)	0	1	0.03	−0.10	−0.12	−0.09	−0.08	1.00																						
6. Year-2011	0.1 (0.3)	0	1	0.02	−0.10	−0.12	−0.09	−0.09	−0.08	1.00																					
7. Year-2012	0.1 (0.3)	0	1	0.00	−0.11	−0.13	−0.10	−0.09	−0.09	−0.09	1.00																				
8. Year-2013	0.1 (0.3)	0	1	−0.02	−0.11	−0.13	−0.10	−0.10	−0.09	−0.09	−0.10	1.00																			
9. Year-2014	0.1 (0.3)	0	1	−0.03	−0.11	−0.13	−0.10	−0.10	−−0.09	−0.09	−0.10	−0.10	1.00																		
10. Year-2015	0.1 (0.3)	0	1	−0.04	−0.10	−0.12	−0.09	−0.09	−0.08	−0.08	−0.09	−0.09	−0.09	1.00																	
11. Year-2016	0.1 (0.2)	0	1	−0.02	−0.09	−0.11	−0.08	−0.08	−0.07	−0.08	−0.08	−0.08	−0.08	−0.08	1.00																
12. Year-2017	0.0 (0.1)	0	1	−0.03	−0.03	−0.04	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	1.00															
13. Thread duration	44.3 (195.0)	0	3840	0.13	0.07	0.01	0.02	0.01	0.01	0.00	−0.01	−0.03	−0.02	−0.03	−0.04	−0.02	1.00														
14. Creator involvement$_i$	2.8 (4.6)	0	112	0.54	0.05	0.02	0.03	0.01	0.03	0.03	0.00	−0.05	−0.04	−0.05	−0.05	−0.03	0.34	1.00													
15. Previous interactions$_{i,j}$	1.1 (2.9)	0	53	0.00	−0.03	0.08	0.03	−0.01	0.05	0.03	0.01	−0.03	−0.02	−0.05	−0.07	−0.03	−0.06	−0.04	1.00												
16. Emotional tone$_i$	36.5 (30.7)	0	99	0.03	0.06	0.05	0.03	0.03	0.02	0.00	−0.01	−0.04	−0.05	−0.05	−0.06	−0.03	−0.03	−0.08	0.12	1.00											
17. Self-disclosure$_i$	0.9 (1.3)	0	3	0.07	0.09	0.12	0.10	0.03	0.06	0.02	−0.02	−0.08	−0.10	−0.12	−0.13	−0.06	−0.06	−0.08	0.31	0.21	1.00										
18. Degree centrality$_i$	141.3 (374.8)	0	6696	0.02	0.05	0.07	0.03	0.01	0.07	0.02	−0.02	−0.05	−0.06	−0.07	−0.08	−0.03	−0.04	−0.02	0.49	0.10	0.32	1.00									
19. Word counts of first post$_i$	216.8 (219.2)	0	3801	0.16	0.00	0.00	0.00	−0.02	−0.02	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.08	−0.03	−0.12	−0.03	0.02	1.00								
20. Community activity$_i$	650.1 (2373.2)	0	30665	0.03	0.02	0.03	0.03	0.01	0.07	0.03	−0.01	−0.03	−0.04	−0.05	−0.06	−0.02	−0.03	0.01	0.51	0.07	0.28	0.85	0.04	1.00							
21. Crowd consensus$_j$	0.1 (0.1)	0	1	−0.02	−0.08	−0.07	−0.03	−0.02	0.00	0.00	0.01	0.05	0.06	0.05	0.06	0.01	0.01	0.00	−0.06	−0.07	−0.12	−0.08	0.03	−0.07	1.00						
22. Self-disclosure$_j$	2.0 (0.8)	0	3	−0.08	−0.11	−0.06	0.01	−0.07	−0.04	−0.04	0.00	0.13	0.16	0.07	−0.03	−0.01	−0.02	−0.02	−0.01	−0.08	−0.09	−0.02	0.01	−0.01	0.12	1.00					
23. Question score of posts$_i$	0.4 (0.4)	0	1	−0.04	−0.04	−0.03	−0.03	−0.01	−0.01	−0.01	0.01	0.02	0.04	0.03	0.04	0.02	0.01	0.01	−0.14	−0.17	−0.18	−0.13	0.02	−0.11	0.05	0.03	1.00				
24. Objective information$_j$	11.5 (7.4)	0	66.67	−0.14	−0.12	−0.17	−0.11	−0.06	−0.10	−0.03	0.01	0.09	0.21	0.18	0.15	0.06	0.04	0.02	−0.29	−0.30	−0.49	−0.28	0.05	−0.21	0.21	0.30	0.20	1.00			
25. Emotional tone$_j$	47.2 (29.2)	0	99	0.07	0.15	0.08	0.05	0.02	−0.02	−0.03	−0.05	−0.06	−0.04	−0.07	−0.07	−0.06	−0.07	−0.12	0.13	0.35	0.26	0.12	0.01	0.08	−0.08	−0.08	−0.20	−0.32	1.00		
26. Experiential information$_j$	0.1 (0.1)	0	3.333	0.27	−0.03	0.01	0.03	0.04	0.05	0.02	0.02	−0.01	−0.04	−0.04	−0.03	−0.03	−0.01	0.11	0.02	0.05	0.18	0.02	−0.04	0.00	0.00	−0.11	−0.02	−0.29	0.04	1.00	
27. Community involvement$_j$	5.1 (5.7)	0	94	0.28	0.13	0.15	0.06	−0.01	0.05	0.02	−0.04	−0.09	−0.10	−0.11	−0.12	−0.06	0.09	0.20	0.23	0.14	0.31	0.28	0.03	0.24	−0.08	−0.03	−0.13	−0.40	0.17	0.18	1.00

Table 4. Tobit Regression Result

Variables	Model 1	Model 2	Model 3
Year-2007	0.46	0.52*	0.25
	(0.275)	(0.251)	(0.234)
Year-2008	1.34***	1.23***	1.18***
	(0.327)	(0.296)	(0.279)
Year-2009	1.76***	1.72***	1.65***
	(0.315)	(0.293)	(0.282)
Year-2010	1.60***	1.63***	1.48***
	(0.344)	(0.309)	(0.294)
Year-2011	1.34***	1.60***	1.46***
	(0.348)	(0.314)	(0.298)
Year-2012	1.82***	2.24***	2.03***
	(0.328)	(0.292)	(0.279)
Year-2013	1.84***	2.29***	2.10***
	(0.319)	(0.288)	(0.277)
Year-2014	1.76***	2.09***	2.15***
	(0.324)	(0.294)	(0.285)
Year-2015	1.27***	1.70***	2.06***
	(0.332)	(0.301)	(0.296)
Year-2016	1.30***	1.96***	1.78***
	(0.349)	(0.318)	(0.315)
Year-2017	0.16	1.60**	1.76**
	(0.592)	(0.568)	(0.550)
Thread Duration	−0.58***	−0.40***	−0.43***
	(0.096)	(0.089)	(0.083)
Creator Involvement$_i$	4.96***	4.95***	4.35***
	(0.139)	(0.132)	(0.124)
Previous Interactions$_{i,j}$	−0.23	−0.40*	−0.20
	(0.206)	(0.174)	(0.163)
Emotional Tone$_i$	0.83***	0.27***	0.36***
	(0.071)	(0.066)	(0.065)
Self-Disclosure$_i$	1.25***	0.39***	0.28**
	(0.104)	(0.100)	(0.100)
Degree Centrality$_i$	0.28	−0.02	−0.10
	(0.251)	(0.224)	(0.231)
Word Counts of First Post$_i$	1.36***	1.37***	1.35***
	(0.099)	(0.093)	(0.088)
Community Activity$_i$	−0.42	−0.09	−0.11
	(0.246)	(0.251)	(0.269)
Crowd Consensus$_j$	0.51***	0.69***	0.55***
	(0.060)	(0.058)	(0.055)
Self-Disclosure$_j$	−0.70***	−0.49***	−0.38***
	(0.060)	(0.056)	(0.058)
Question Score of posts$_i$	−0.24***	0.01	−0.20***
	(0.063)	(0.060)	(0.057)
Objective Information$_j$		−2.14***	0.28*
		(0.102)	(0.127)
Emotional Tone$_j$		1.28***	1.92***
		(0.068)	(0.079)
Objective Information$_j$ × Experiential Information$_j$		−3.02***	−0.44**
		(0.176)	(0.144)
Emotional Tone$_j$ × Community Involvement$_j$		0.42***	0.90***
		(0.108)	(0.115)
Objective Information squared$_j$			−0.52***
			(0.123)
Emotional Tone squared$_j$			−1.47***
			(0.068)
Experiential Information$_j$			2.35***
			(0.141)
Community Involvement$_j$			1.67***
			(0.108)
Objective Information squared$_j$ × Experiential Information$_j$			−0.02
			(0.193)

(continued)

Table 4. continued

Variables	Model 1	Model 2	Model 3
Emotional Tone squared$_j$ × Community Involvement$_j$			−0.59***
			(0.111)
Constant	−2.53***	−3.68***	−0.99***
	(0.240)	(0.221)	(0.218)
Pseudo R^2	0.0644	0.0828	0.1016
Observations	30 035	30 035	30 035

Robust standard errors in parentheses ***$P < .001$, **$P < .01$, *$P < .05$

Our results in Model 4, Table 4 show that the coefficients of Objective Information squared$_j$ ($\beta_{28} = -0.52$, $P < .001$) and Emotional Tone squared$_j$ ($\beta_{29} = -1.47$, $P < .001$) were negative and significant. Thus, the null hypotheses of H1 and H2 were not supported, and we can conclude that objective information and emotional tone of support provider j's have an inverted-U effect on self-care behavior of i. In H3 and H4, we hypothesized that experiential information$_j$ and community involvement$_j$ were positively related to self-care behavior$_i$. Our results in Table 4 showed that Experiential Information$_j$ ($\beta_{30} = 2.35$, $P < .001$) and Community Involvement$_j$ ($\beta_{31} = 1.67$, $P < .001$) were positively and significantly related to self-care behavior$_i$. Therefore, the null hypotheses of H3 and H4 were not supported, and we can conclude that experiential information provided by and community involvement of support provider j's have a positive impact on self-care behavior of i. The coefficient of "Objective Information squared$_j$ × Experiential Information$_j$" ($\beta_{32} = -0.02$, $P > .05$) was statistically not significant. However, the coefficient for "Emotional Tone squared$_j$ × Community involvement$_j$" ($\beta_{33} = -2.22$, $P < .001$) was negative and significant. Therefore, we failed to reject the null hypotheses of H5 but the null hypothesis of H6 was not supported. Accordingly, we can conclude that community involvement amplifies the inverted U relationship between emotional tone$_j$ and self-care behavior$_i$.

The nature of the above interactions is illustrated in Figures 2 and 3. We plotted these graphs by examining the parameters of our regression equation at mean levels for all variables except the focal variables and at two different levels of the moderator terms, one at one standard deviation above the mean value and the other one standard deviation below the mean value of the moderator variable. The values of all the variables used in these graphs were calculated as described in Table 2 pertaining to the operationalization of variables. Figure 2 plots the impact of objective information (mean = 11.46, SD = 7.41) on self-care behavior (mean = 3.25, SD = 6.05). The range of objective information in this plot is [0,25]. When objective information approximately reaches 13, the self-care behavior is at the peak. Figure 3 plots the impact of emotional tone (Mean = 47.2, SD = 29.17) on self-care behavior at different levels of community involvement (Mean = 5.1, SD = 5.74). Based on ±1 SD, the values of high and low community involvement are set to 10.84 and 0 (the negative value of −0.64 is set to the minimum value of 0 for the community involvement variable). The range of emotional tone in this plot is [0,100] and when it approximately reaches 70–80, the self-care behavior reaches its peak. The coefficient for Emotional Tone squared$_j$ ($\beta_{29} = -1.47$) is approximately three times the coefficient for Objective Information squared$_j$ ($\beta_{28} = -0.52$), indicating that emotional tone has 3 times the amount of impact on self-care behavior than objective information. This is also supported by the range of Self-care Behavior$_i$ and the maximum values of self-care behavior achieved at different levels of Objective Information squared$_j$ and Emotional Tone squared$_j$ as shown in Figures 2 and 3. While self-care behavior does not have a negative value in our operationalization, the negative values for self-care behavior in Figures 2 and 3 are achieved because we use the mean values for all variables except the focal variables (which will not always be true in reality) in the regression equation to calculate the impact only for the

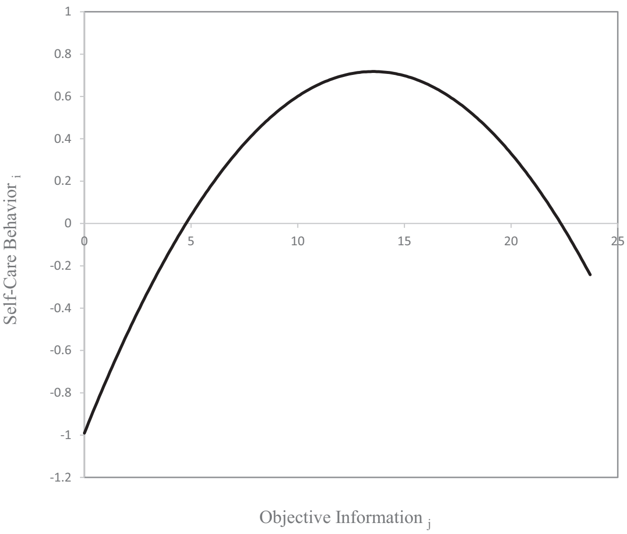

Figure 2. The inverse-U relationship between Objective Information$j$ and Self-Care Behavior$_i$.

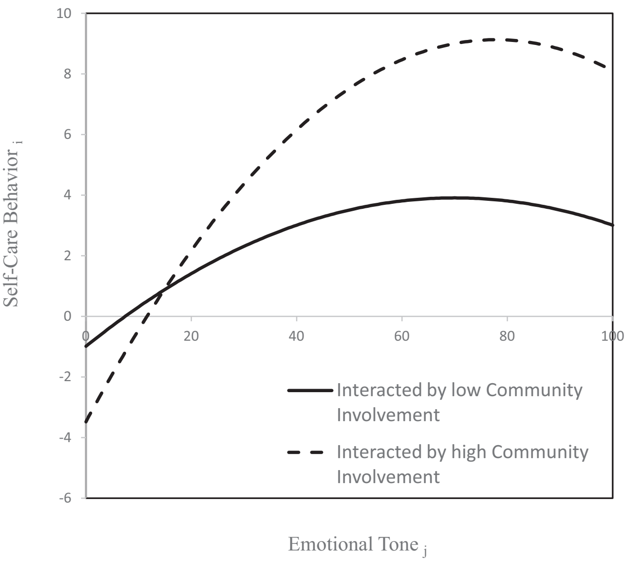

Figure 3. Community Involvement$_j$ moderates the inverse-U relationship between Emotional Tone$_j$ and Self-Care Behavior$_i$.

focal variables on self-care behavior. Figures 2 and 3 thereby provide further support for H1, H2, and H6.

We conducted robustness analyses with alternate model specification and variable operationalizations (see Supplementary Appendix A1 for details). Our results did not change in these analyses.

DISCUSSION

Theoretical and Practical Implications

Our study and its findings make four contributions to the health informatics and online social support literature. Our model and findings on the influence of online social support on self-care behavior addresses the gap in this literature which has primarily addressed mental and psychological health outcomes,[4,65] and seldom addressed self-care behavior. Industry reports of low self-care behaviors in HIV patients attest the notion: medical answers to the self-care puzzle are inadequate.[16] We need behavioral solutions as well. In light of these alarming numbers, our study makes a valuable contribution. Moreover, though there is ample evidence about the beneficial self-care impacts of social support in an offline context, there is a paucity of literature about the health behavior impacts of social support received from other patients through social media in the management of chronic diseases.[66] Our study fills this gap. Second, our conceptualization of informational and emotional support lends a deeper view on online social support, its dimensions, and sub-dimensions. These sub-dimensions, taken separately and together (interactions) produced different impacts on self-care behavior. We filled the gap in the literature where social support is examined very broadly. Our conceptualization provides online community managers a nuanced way to classify the ever-growing textual data in online communities and predict health outcomes. Third, contrary to existing findings we showed an inverse-U relationship between both objective information and emotional tone on self-care behavior. We also showed the moderators of these effects. It is crucial to know if and how online social support has an impact on patient self-care given the proliferation of health-related social media sites, and growing participation in these sites. More importantly, are there any dark sides to online social support? By integrating the arguments from the offline social support literature, our results reveal the merits and demerits of online communities, especially of online social support to patients. Fourth, we studied online health community for HIV patients, as a case of stigmatized chronic disease context. Our results suggest that caution should be exercised in the use of social media interventions for HIV patients, and patients with other stigmatized chronic diseases. It also points to a "social support paradox" where patients who are stigmatized have no support, live in isolation, and seek support from the online health community. However, too much support can have adverse impacts on their health outcomes.

Implication for Health Informatics

We state 4 ways in which our research is novel from an informatics perspective. First, by studying experiential information we enhanced the online social support typology and provide informatics professionals an additional lens to view, store, and present text data posted by support providers in online health communities. We conceptualized experiential information as an additional dimension of online social support in addition to the objective information and emotional tone dimensions studied in the literature.[4,67] Two, by studying community involvement, we contribute a novel social support metric that captures the user activity, and that complements the 3 text-based metrics of social support. Online communities are presented in a thread format allowing many support providers to contribute. Text based metrics such as objective information, emotional tone, and experiential information do not capture the number of support providers and the associated companionship aspect crucial to a support seeker. Our study fills this gap and directs the health informatics arena to leverage metrics that capture the more dynamic user involvement aspects of an online health community. Such metrics could inform us more about a threaded conversation, and about the involvement and collaboration of the various stakeholders in terms of the support they provide and its impact on the patient health as hypothesized and tested in this study. Third, ours is the first study to use a tool like LIWC that uses linguistic dimensions (and corresponding dictionaries) to operationalize the variables. Fourth, we build custom dictionaries based on ratings by experts to quantify online social support and self-care behavior for an HIV forum. The tool use case and our custom dictionaries can help informatics professionals to easily quantify massive, unstructured textual data in online HIV forums in automated ways, for better presentation, decision making, and impact analysis. Coupled with our findings on the quadratic effects of online social support, meaningful interventions can be devised to limit data in online health communities and to enhance self-care behavior of HIV patients.

Limitations

We acknowledged the limitations of this study and identify opportunities for future research. First, though we considered experiential information and objective information as separate dimensions of informational support, there is a potential overlap between these 2 variables. For example, a community member can provide objective information based on his/her experience thereby creating an overlap between objective and experiential information. Future work may apply supervised machine learning techniques to better differentiate objective information and experiential information. Second, as described in the previous section, we were only able to capture the self-care behavior of those patients who expressed this in their threads. We assigned a value of zero to both those patients who may have performed self-care behavior but did not report in their postings, as well as to those who did not actually perform self-care behavior or engaged in negative behaviors but did not write about this in their posts on the online forum. This implies a possible measurement issue with our dependent variable. We would have also liked to use a Heckman 2-stage model with instrumental variables that predict a patient's likelihood of disclosing their self-care behavior to correct for this bias. However, given that HIV is a highly-stigmatized disease, there is practically no personal information available about the patients on the online forum that could be used as instrumental variables (that meet the inclusion/exclusion criteria) for the first stage model. Accordingly, we used a Tobit model to account for the left censoring of data with negative values assigned a value of zero in our study. Nonetheless, future studies may explore the possibility of obtaining good instrumental variables and using them to perform a 2-stage regression model to correct for the self-selection bias.

CONCLUSION

We studied the impact of online social support on patient self-care behavior in an online health community for patients with HIV, a stigmatized chronic disease. We disaggregated informational and emotional support into nuanced types and hypothesized their direct and moderating effects on self-care behavior. We used data from the POZ online health community, and operationalized our variables using LIWC, self-developed dictionaries, and Python's NLTK and Gensim libraries. We tested our model using a Tobit regression model. Based on our results, 5 of the 6 null hypotheses were not supported providing strong support for our research model.

COMPETING INTERESTS

None.

FUNDING

This research was not supported by any grant from a funding agency in the public, commercial, or not-for-profit sectors.

CONTRIBUTORS

RK conceptualized the original idea for the paper that was further developed by XW, SP, and DB. All the 4 authors were involved in the design of the study. XW and SP developed the hypotheses under the guidance of RK. DB collected the data and performed text mining under the guidance of RK and with input from XW and SP. DB, SP, and XW, performed the reconciliation of the expert coding with the computer operationalization of variables under the guidance of RK. XW and SP performed the regression analyses under the guidance of RK. All four authors were involved in interpreting the findings of the study. XW, SP, and DB prepared the initial draft of paper that was refined by RK prior to submission and subsequent revisions.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors are grateful to Dr Alyssa Shon, MD, Dr Supriya Mahajan, PhD, and Ms. Lori Anthony for serving as experts and coding data on user posts in threaded conversations for our variable operationalization validation procedure as described in the paper. We are also thankful to Dr Alyssa Shon, MD and Dr Supriya Mahajan, PhD for their comments and suggestions on this paper. This study is based on an analysis of the content of the posts of the users of www.poz.com in the online forums of this website and their user profiles. Users posted messages and provided their profile information in these online forums of their own volition and with no expectation of privacy pursuant to the various notices in this regard provided by www.poz.com website on its various forum pages. Further, the contents of the posts provided by www.poz.com users and their profile information itself was not reprinted in any portion in this paper. Accordingly, the study does not violate the terms and conditions of the www.poz.com website outlined at https://www.poz.com/page/terms-use, including those with respect to paragraph #13 pertaining to the proprietary rights of the website owners over the contents of the website's forums. The research contained in this paper did not involve any interaction or intervention with any human subject. The various numerical variables that were used in this study to test the proposed hypotheses were derived from the data provided by www.poz.com users to the website's forums of their own volition, including their user profile data and the original textual content provided by them as posts in the 5 online forums, as outlined above, on the www.poz.com website. Users of www.poz.com provided all this data with the understanding that their data will be shared with any and all users of the online forums of www.poz.com website, and that they had no expectation of privacy while providing their profile information or posting any message on this website. Accordingly, this paper complies with the guidance provided by National Institutes of Health regarding use of social media tools for data collection, particularly item #1, available on the https://www.nih.gov/health-information/nih-clinical-research-trials-you/guidance-regarding-social-media-tools website. This study was reviewed by the Institutional Review Board at the University at Buffalo, State University of New York. It was determined to not constitute human research and, accordingly, does not require IRB review and approval.

REFERENCES

1. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *J Biomed Inform* 2013;46(6):957–969.
2. Fox S, Duggan M. Health online 2013. Washington, DC: Pew Research Center. http://www.pewinternet.org/2013/01/15/health-online-2013/. Accessed July 1, 2017.
3. Laranjo L, Arguel A, Neves AL, *et al*. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2014;22(1):243–256.
4. Yan L, Tan Y. Feeling blue? go online: an empirical study of social support among patients. *Inform Syst Res* 2014;25(4):690–709.
5. Cline RR, Worley MM. Osteoporosis health beliefs and self-care behaviors: an exploratory investigation. *J Am Pharm Assoc* 2006;46(3):356–363.
6. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. *Effect Clin Pract* 2000;4(6):256–262.
7. Erfani SS, Blount Y, Abedin B. The influence of health-specific social network site use on the psychological well-being of cancer-affected people. *J Am Med Inform Assoc* 2016;23(3):467–476.
8. Snyder KA, Pearse W. Crisis, social support, and the family response: exploring the narratives of young breast cancer survivors. *J Psychosoc Oncol* 2010;28(4):413–431.
9. Mallinckrodt B, Armer JM, Heppner PP. A threshold model of social support, adjustment, and distress after breast cancer treatment. *J Counsel Psychol* 2012;59(1):150.
10. Krause N. Assessing stress-buffering effects: a cautionary note. *Psychol Aging* 1995;10(4):518.
11. Coyne JC, DeLongis A. Going beyond social support: The role of social relationships in adaptation. *J Consult Clin Psychol* 1986;54(4):454.
12. Ellison NB, Steinfield C, Lampe C. Connection strategies: social capital implications of Facebook-enabled communication practices. *New Media Soc* 2011;13(6):873–892.
13. Longo DR, Schubert SL, Wright BA, LeMaster J, Williams CD, Clore JN. Health information seeking, receipt, and use in diabetes self-management. *Ann Fam Med* 2010;8(4):334–340.
14. Parker R, Aggleton P. HIV and AIDS-related stigma and discrimination: a conceptual framework and implications for action. *Soc Sci Med* 2003;57(1):13–24.
15. Piot P, Bartos M, Ghys PD, Walker N, Schwartländer B. The global impact of HIV/AIDS. *Nature* 2001;410(6831):968–973.
16. Mills EJ, Nachega JB, Bangsberg DR, *et al*. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. *PLoS Med* 2006;3(11):e438.
17. Montero B, Watts F, García-Carbonell A. Discussion forum interactions: text and context. *System* 2007;35(4):566–582.
18. Maloney-Krichmar D, Preece J. The meaning of an online health community in the lives of its members: roles, relationships and group dynamics. Technology and Society, 2002. (ISTAS'02). 2002 International Symposium on *IEEE*; June 6–June 8, 2002; Raleigh, NC.
19. Cobb S. Social support as a moderator of life stress. *Psychosomatic Med* 1976;38(5):300–314.
20. Pfeil U. Online support communities. In: Zaphiris P, Ang CS, eds. *Social Computing and Virtual Communities*. Boca Raton, FL: Chapman & Hall; 2009:121–150.
21. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. *Psychol Bull* 1985;98(2):310.
22. Kawachi I, Berkman LF. Social ties and mental health. *J Urban Health* 2001;78(3):458–467.

23. Coulson NS. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *CyberPsychol Behav* 2005;8(6):580–584.
24. Bunde M, Suls J, Martin R, Barnett K. Online hysterectomy support: Characteristics of website experiences. *Cyberpsychol Behav* 2007;10(1):80–85.
25. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. *J Med Internet Res* 2012;14(3):e54.
26. Fan H, Lederman R, Smith SP, Chang S. *How Online Health Forum Users Assess User-Generated Content: Mixed-Method Research*. Utrecht, Netherlands: ECIS; 2013.
27. McMullan M. Patients using the Internet to obtain health information: how this affects the patient–health professional relationship. *Patient Educ Counsel* 2006;63(1):24–28.
28. Josefsson U. Coping with illness online: The case of patients' online communities. *Inform Soc* 2005;21(2):133–141.
29. Epstein RM, Peters E. Beyond information: exploring patients' preferences. *JAMA* 2009;302(2):195–197.
30. Jones Q, Ravid G, Rafaeli S. Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration. *Inform Syst Res* 2004;15(2):194–210.
31. Arguello J, Butler BS, Joyce E, *et al*. Talk to me: foundations for successful individual-group interactions in online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Montréal, Québec, Canada: ACM; 2006:959–968.
32. Chesla CA, Fisher L, Mullan JT, *et al*. Family and disease management in African-American patients with type 2 diabetes. *Diabetes Care* 2004; 27(12):2850–2855.
33. Chesla CA, Fisher L, Skaff MM, Mullan JT, Gilliss CL, Kanter R. Family predictors of disease management over one year in Latino and European American patients with type 2 diabetes. *Fam Process* 2003;42(3):375–390.
34. Fisher L, Chesla CA, Skaff MM, *et al*. The family and disease management in Hispanic and European-American patients with type 2 diabetes. *Diabetes Care* 2000;23(3):267–272.
35. Bambina A. *Online Social Support: the Interplay of Social Networks and Computer-mediated Communication*. Youngstown, NY: Cambria Press; 2007.
36. Weiss RS. Online support communities. In: Rubin Z. *Doing Unto Others*. Englewood Cliffs, NJ: Prentice-Hall;1974:17–26.
37. Wills TA. Supportive functions of interpersonal relationships. In: Cohen S, Syme SL, eds. *Social support and health*. San Diego, CA: Academic Press;1985:61–82.
38. Lamberg L. Online empathy for mood disorders. *JAMA* 2003;289(23): 3073–3077.
39. Lehman DR, Ellard JH, Wortman CB. Social support for the bereaved: Recipients' and providers' perspectives on what is helpful. *J Consult Clin Psychol* 1986;54(4):438.
40. Wortman CB, Lehman DR. Reactions to victims of life crises: support attempts that fail. In: Sarason IG, Sarason BR, eds. *Social Support: Theory, Research and Applications*. Dordrecht: Springer Netherlands; 1985:463–489.
41. Song YS, Ingram KM. Unsupportive social interactions, availability of social support, and coping: their relationship to mood disturbance among African Americans living with HIV. *J Soc Personal Relationships* 2002;19(1):67–85.
42. Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004;31(2):143–164.
43. Short J, Williams E, Christie B. *The Social Psychology of Telecommunications*. London: John Wiley and Sons Ltd; 1976.
44. Tseng FC, Huang HC, Teng CI. How do online game communities retain gamers? Social presence and social capital perspectives. *J Comp Mediated Commun* 2015;20(6):601–614.
45. Beck R, Pahlke I, Seebach C. Knowledge exchange and symbolic action in social media-enabled electronic networks of practice: a multilevel perspective on knowledge seekers and contributors. *MIS Quart* 2014;38(4):1245–1270.
46. Lu B, Fan W, Zhou M. Social presence, trust, and social commerce purchase intention: An empirical research. *Comput Hum Behav* 2016;56:225–237.
47. Kreijns K, Kirschner PA, Jochems W. Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Comput Hum Behav* 2003;19(3):335–353.
48. Cyr D, Hassanein K, Head M, Ivanov A. The role of social presence in establishing loyalty in e-service environments. *Interacting Comput* 2007;19(1):43–56.
49. Wang Y-C, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. Seattle, Washington: ACM; 2012:833–842.
50. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Vancouver, British Columbia, Canada: Association for Computational Linguistics; 2005:347–354.
51. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC *2015*. Austin, TX: University of Texas at Austin; 2015.
52. Alpers GW, Winzelberg AJ, Classen C, *et al*. Evaluation of computerized text analysis in an Internet breast cancer support group. *Comput Hum Behav* 2005;21(2):361–376.
53. Chau M, Xu J. Business intelligence in blogs: Understanding consumer interactions and communities. *MIS Quart* 2012;36(4):1189–1216.
54. Arazy O, Woo C. Enhancing information retrieval through statistical natural language processign: a study of collocation indexing. *MIS Quart* 2007;31(3):525–546.
55. Maarek YS, Berry DM, Kaiser GE. An information retrieval approach for automatically constructing software libraries. *IEEE Transactions on Software Engineering* 1991;17(8):800–813.
56. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks* 1978;1(3):215–239.
57. Oksanen A, Garcia D, Sirola A, *et al*. Pro-anorexia and anti-pro-anorexia videos on YouTube: sentiment analysis of user responses. *J Med Internet Res* 2015;17(11):e256.
58. Collier N, Son NT, Nguyen NM. OMG U got flu? Analysis of shared health messages for bio-surveillance. *J Biomed Semant* 2011;2(5):S9.
59. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. *J Med Internet Res* 2013;15(11):e239.
60. Wikipedia contributors. (2017, May 31). Censoring (statistics). In: *Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Censoring_(statistics)&oldid=783218590. Accessed July 4, 2017.
61. Cameron AC, Trivedi PK. *Microeconometrics Using Stata*. College Station, TX: Stata Press; 2010.
62. McDonald JF, Moffitt RA. The uses of Tobit analysis. *Rev Econ Stats* 1980;62(2):318–321.
63. Seery MD, Holman EA, Silver RC. Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience. *J Personality Social Psychol* 2010;99(6):1025.
64. Schilke O. On the contingent value of dynamic capabilities for competitive advantage: the nonlinear moderating effect of environmental dynamism. *Strat Manage J* 2014;35(2):179–203.
65. Chen L, Straub D. *The Impact of Virtually Crowdsourced Social Support on Individual Health: Analyzing Big Datasets for Underlying Causalities*. In: Proceedings of the Americas Conference on Information Systems (AMCIS); August 13–15, 2015; Puerto Rico, USA.
66. Gielen A, McDonnell K, Wu A, O'campo P, Faden R. Quality of life among women living with HIV: the importance violence, social support, and self care behaviors. *Soc Sci Med* 2001;52(2):315–22.
67. Deetjen U, Powell JA. Informational and emotional elements in online support groups: a Bayesian approach to large-scale content analysis. *J Am Med Inform Assoc* 2016;23(3):508–13.