The Value of Social Media for Patients: Social Supports, Networking, and Learning in Online Healthcare Communities

Lu Yan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Yong Tan, Chair

Apurva Jain

Ming Fan

Program Authorized to Offer Degree:

Business Administration

University of Washington

Abstract

The Value of Social Media for Patients: Social Supports, Networking, and Learning in Online Healthcare Communities

Lu Yan

Chair of the Supervisory Committee:

Professor Yong Tan

Department of Information Systems and Operations Management

Social media are changing the way we learn about health and illness. Yet, there is limited information about whether online healthcare communities have a significant effect on patients' health conditions, as they do in healthcare communities in the offline world. To better understand what value social media bring to patients, research is needed to investigate this expanding phenomenon—patients joining OHCs and helping each other. In this dissertation, I conduct three studies to examine this increasingly important circumstance.

In the first study I proposed a nonhomogeneous Partially Observed Markov Decision Process model to examine patients' health outcome dynamics related to their online activities. The Internet provides easier and greater access to health- and medical-related information than ever before. OHCs enhance this access by providing user-generated content about firsthand experiences and various social supports that patients need. The results indicate that patients receive benefits from learning from others, and participation in the online community helps patients improve their health conditions and better engage in the disease self-management process.

Whereas the public accessibility of OHCs offers great promise for patients, it also carries the risk of information overload. In the second study, I propose a model that incorporates widely examined homophily features in offline social networking studies with endogenous online network measurements to examine network dynamics. I found that individuals' preference for communicating in the online social networking environment could be explained by network characteristics and their positions in the network.

Although the Internet provides the opportunity for patients to access "more people like me" in a very fast and highly convenient way, it remains unknown how effective these online social interactions are in improving patients' health conditions. In the third study, I investigate how patients use information gleaned from others' shared experiences in the online healthcare communities. The results provide evidence for the existence of social contagion in OHCs, particularly in the areas of health information seeking and knowledge creation.

In sum, these investigations provide evidence of the prominent role that social media are assuming in healthcare and raise important and promising questions for future research in this field.

TABLE OF COTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEGEMENTS

I would like to take this opportunity to acknowledge many of the people who have played a significant role in helping me complete my doctoral degree.

First of all, I would like to acknowledge my adviser Prof. Yong Tan, who has acted as my mentor during my doctoral career. Over the past few years, he has served as an excellent role model, helping me to become a productive researcher and educator. His guidance, encouragement, and support have helped me overcome many difficulties, both academic and personal, throughout my graduate study.

I also would like to acknowledge the members of my dissertation committee, Prof. Ming Fan, Prof. Apurva Jain, Prof. Theodore Klastorin, and Prof. Subohda Kumar, who, along with Prof. Dayong Gao, have provided me with invaluable input on this dissertation. In addition, I would like to thank Prof. Yong-pin Zhou, who provides me with insight into the academic world and encourages me in my academic pursuits.

The completion of my formal education would not have been possible without the love, encouragement, and support of my parents Da-sheng Yan and Yu-fen Ji and my brother Shen Yan. Through the years, they have never stopped supporting me. I truly appreciate and love them very much.

Finally, I would like to thank Jingxin Sun for always being there for me. For the past ten years, although he can’t stand by my side physically, he has done everything he can to keep me positive and optimistic. He has encouraged me when I was discouraged, made me happy when I was down, and challenged me to become a better researcher and scholar. Without his love and support, I probably never would have aspired to complete this level of education.

DEDICATION

To my family.

Chapter 1 Introduction

1.1 Background

Social media have brought tremendous changes to the healthcare system. Search engines help individuals locate a profusion of health information, and there are many healthcare-related websites today to provide services to meet our needs. Some of these websites help consumers take control of their health and healthcare, while others help whole communities to improve their health. According to a recent study by the Pew Research Center, a growing number of individuals turn to the Internet for health-related activities, as in "I don't know, but I can try to find out." These patients and their caregivers use the Internet to search for health information that is easy to understand as well as to learn about others' experiences and treatments that could potentially benefit their own health. Social media open a door for lay users—"the patients"—to add their contributions to help others, as in "I know, and I want to share my knowledge." This has been called the third dimension, "patient–patient," by some researchers (O'Grady 2008, O'Grady et al. 2008) in the healthcare decision framework. Given the opportunities presented by unprecedented openness among patients, who get together to share their knowledge to help each other improve their health conditions, I am interested in studying the social value generated by online healthcare communities, including how this social phenomenon reshapes patients' health knowledge and behaviors, what content providers can do to improve their services, and what roles firms should play in online healthcare communities.

1.2 Online Social Support and Health Condition Dynamics

Recent research indicates the number of people turning to the Internet for health-related knowledge is on the rise. According to the Pew Research Center, the number of Americans who sought health information online in 2008 stood at 61%, up from 25% in 2000 (Pew Internet & American Life Project 2009). In 2010, the 59% of American adults who used the Internet to research health problems constituted 80% of Internet users (Pew Internet & American Life Project 2011). In light of this heightened interest, health service providers are moving swiftly to address this behavior change. Already, the American Cancer Society (ACS) and the Centers for Disease Control and Prevention (CDC) have opened “offices” in Second Life, the online virtual world, to spread word about the importance of nutrition awareness, cancer screening, and infectious disease prevention (Landro 2006).

The Internet is changing the way people learn about health and illness (Ziebland et al. 2004). Features such as its reach to a vast audience in a cost-efficient way, 24/7 accessibility, and user anonymity make the Internet a place people can turn for social support at any time. Without the need to be spatially and temporally co-present, the Internet provides a safer environment for disease sufferers to engage in nonthreatening and supportive communication (Coulson2005). Patients who meet online develop friendly and comforting relations and provide each other with various forms of social support. The resulting trust generally increases an individual’s willingness to disclose details of her disease and experiences (Lamberg 2003).

Extensive studies over the past few decades indicate that the level of social support in people’s lives essentially predicts their physical and mental health outcomes (Clark 2006). However, people with chronic illnesses, especially mental problems, often cannot develop and maintain relationships offline (Leung 2011). Mental health is defined as an individual’s ability to respond to the many, varied experiences of life with flexibility and a sense of purpose (Oluwole et al. 2011). Mental illness, or mental disorder, is generally associated with distress or disability. People with serious mental problems have difficulty balancing themselves, other people, and the surrounding environment. Take depression as an example. One in five people in the United States have some form of depression. According to the National Institute of Mental Health, there were 33,000 suicides in the United States in 2006, and more than 90% of those individuals had been diagnosed with a mental disorder. Clinical depression is now the second-most-costly disorder among all medical diseases in the United States. While only a minority of those with mental disorders seek professional help (Christensen and Griffiths 2000), many find online social interactions attractive and effective to obtain much needed emotional support and companionship (Leung 2011).

1.3 Communication Network Dynamics

Information and communication technologies have made it increasingly easier for individuals to interact with others about health issues, unlike the largely a one-way flow of information from doctor to patient in the past. Encouraged by modern technical trends, patients are shifting their roles from passive recipients to active consumers of health information. Although health information is one of the most frequently sought-after topics on the Internet, patients do not see the Internet as a replacement for a health professional (McMullan 2005). Instead, patients add a new dimension to healthcare systems by utilizing their own values. Unlike health facts, which contain information that can only be provided by medical professionals like physicians, the actual feelings about a certain disease and what to expect next is unique information that only someone with personal experience can tell. As more interactive healthcare sites make the social networking revolution come to the healthcare, the online healthcare community (OHC) used for sharing interpersonal communications is of significant social, organizational, and economic importance. The advantages are obvious. Hundreds of thousands individuals exchange advice and experiences in these virtual communities (Preece and Ghozati 1999, Wang et al. 2008). Patients in these online spaces can raise questions, search disease information, share medical records, discuss their problems with other members, get second opinions from patients suffering from the same disease, and compare their self-outcomes with a larger population. It has already been shown that participation in online social networking activities helps patients gain better disease-management knowledge and make progress to better health conditions (Frost and Massagli 2008, Yan and Tan 2010).

In addition to these benefits, however, lurk disadvantages. For example, when there is an abundance of available information, it is often difficult to obtain useful and relevant information when it is needed (Yellowlees and Brooks 1999, Edmunds and Morris 2000). Not surprisingly, this phenomenon is not new. First mentioned by Bertram Gross (1964), the problem of “information overload” is widely recognized today in various contexts (Edmunds and Morris 2000). At the individual level, the problem is defined as receiving more information than one is capable of processing (O’Reilly 1980, Eppler and

Mengis 2004). When being applied to the organizational level, this term conveys the situation that when individuals' information overload is widespread in the organization, it reduces the overall effectiveness of the organization. In particular, increased information overload may make it difficult to accurately identify relevant cues, resulting in decreased performance. Discussions about the severity of the problem and its consequences are explored in several studies. For example, researchers across various disciplines have found the inverted U-curve to illustrate the correlation of an individual's performance and the amount of information she received (see Schroder et al. 1967). Once the amount of information exceeds a certain point, it will burden the individual's information load and detrimentally impact her decision making (Schick et al. 1990). In the field study "Dying for Information?" Lewis (1996) finds that information overload can lead to illness and other problems such as stress, loss of job satisfaction and ineffective actions (Edmunds and Morris 2000, Denning 2002). In health informatics studies, Lowe (2004) likewise identifies this problem and finds that thousands of people are confused by the plethora of available health information.

There is no doubt that the increased availability of information from digitalized resources contributes greatly to the accessibility of information. However, the problem of information overload is not simply a matter of more information available than people can readily acquire, process and learn from (Wilson 2001), because people need to find strategies and learn how to cope with information overload. By analyzing the causes of the problem, solutions are proposed based on three aspects: technological means, policy and management practices, and human factors. However, the existing literature on the information overload concentrates mainly on the first two areas. Various kinds of technologies are created to save users from sinking in the data sea. In particular, such advanced technologies created in the virtual space include search engines/algorithms, personalized channels, tagging, RSS feeds, and others. Meanwhile, training policies in terms of information demand and dissemination are also raised in order to limit the damaging effects of information overload, but the role of human factors is seriously ignored. In particular, how human behavior changes in face of information overload is not well studied.

As the modern technology becomes ubiquitous, the problem of information overload brings challenges to healthcare as it does to other areas. The unique context of healthcare makes the problem more complicated in the OHC. First, the special features of the Internet, such as convenient accessibility, make the issues of massive volume of information and quality more prominent (Christensen et al. 2004). The management training designed to overcome the problem of information overload is also hard to organize in the virtual communities. Second, the most up-to-date information is not necessarily what patients need. Sometimes, older messages may contain more important information for patients to understand their health conditions, making new content-oriented collecting technologies less effective than they are in other fields. Third, without proper professional education, patients may not know exactly what they want or what keywords to search, even if the information is available. Moreover, a paradox also exists, found by Edmunds and Morris (2000). The paradox is that it is often difficult to obtain useful and relevant information when it is needed, even though there is an abundance of information available. With health information in particular, patients experience anxiety about whether they have missed an important piece of information in the vast volume of materials available. Patients who suffer from life-threatening diseases are particularly concerned, because finding a critical piece of information might be a matter of life and death.

Furthermore, the healthcare industry is a combination of private and public sector activities. The credibility of the publicly-shared private information is largely dependent on "the kindness strangers," because it can't be independently verified. This also brings challenges to modern health informatics systems in terms of using patients' shared health data to improve health outcomes. Finally, there is no effective quality control for the user-generated content (UGC) in virtual space, and patients might obtain inaccurate or potentially dangerous information (Christensen et al. 2004). As the result, these factors make the proposed technical and managerial solutions less efficient than other areas. Consequently, human factors, such as individuals developing strategic behaviors to solve the problem of information overload, could take on a more important role in the context of healthcare.

1.4 Online WOM and Health Education

Social media is changing the way people learn about health and illness (Ziebland et al. 2004). A report from Manhattan Research suggests that more than 60 million Americans are consumers of "Health 2.0" services (Kane et al. 2009). Those who use these Web 2.0 applications to learn about health are changing their role from passive information receivers to active participants to build knowledge collaboratively (O'Grady et al. 2008). They use the Internet to find and share information, to look for insights into rare conditions and new treatments, and to exchange experiential and anecdotal knowledge online, all of which arises, broadly speaking, from the online word of mouth (WOM) that results from user-generated-content. The use of new tools that are designed to support collective knowledge sharing with interfaces increases the speed of healthcare community formation for collaborative web-based patient education and magnifies its impact and reach. These online healthcare communities are thus believed to have profound implications (O'Grady 2008, Fichman et al. 2011).

The phenomenon of online WOM is not new. The Internet has enabled individuals all over the world to make their personal experiences, thoughts, and opinions globally accessible (Dellarocas 2006), and online user reviews are believed to have greater potential impact than any other communication channel (Godes and Mayzline 2004). For this reason, online WOM plays a critical role in affecting consumers' perceptions of a product and in shaping firms' marketing strategies. Numerous analytical and empirical studies have been, and are, conducted to investigate the social influence of ex-post reviews. On the one hand, for instance, prior research has focused on establishing the casual impact of user ratings on sales. Chevalier and Mayzline (2006) find evidence to show the positive impact of average ratings on book sales, while Chen et al. (2004) find no evidence of contagion in a similar research setting. On the other hand, several studies have found evidence to suggest that online WOM can be used as a free "sales assistant" that provides firms with the opportunity to implement marketing strategies that may not otherwise be valid (Chen and Xie 2008, Fay and Xie 2008, Xie and Gestner 2007, Dellarocas 2006), and that promotional chat can be beneficial to both consumers and firms (Mayzline 2006). Although online conversation among patients is part of WOM in the virtual space, doubts exist as to whether or not the prior findings are valid in the online healthcare community setting. Even less is known about whether the typical methodologies that have been proposed in prior studies can be used to imply social influence and the consequent impact of patients' online information exchange behaviors, because people tend to be more emotional and exhibit greater risk-seeking behavior when faced with a life-death choice than they do with problems in other life domains, such as personal finances or public property (Druckman and McDermott 2008). If an individual is sick, she might disclose information in the hope that it might somehow improve her health, regardless of what the actual probability of getting better could be, and she may seek others' opinions (Anderson and Agarwal 2011). Given these significant differences, online participants with health issues tend to be irrational compared to participants we observed in non-healthcare environments. In addition, what patients communicate online is their first-hand experience based on their personal health situation and feelings, so the quality of the information may vary because the knowledge generated comes from patients with different backgrounds and health status, and the information provided does not follow standard guidance but rather comes from different aspects or angles, such as treatment effectiveness or the severity of side effects. More importantly, the ability to interpret various forms of health information and integrate them varies widely across populations (Murray et al. 2007). Taken together, collaborative health education is a complex process, and existing findings may or may not be applied to explain patients' experiential learning in online healthcare communities.

The increased participation in online communities strengthens the potential for patients to influence one another's decision-making, emphasizing a third decision-making dyad – patient-patient (O'Grady et al. 2008). In response to the flourishing online communication among patients, many studies have been conducted to investigate this phenomenon. Because the information seeking is a multifaceted and complex process, individual characteristics are important in affecting people's decisions regarding whether and how to seek information for a health problem (O'Grady et al. 2008). Patients are found to manifest different patterns, given the potentially stressful nature of their search. Some may be more likely than others to seek help as a copying strategy; for example, in the mental disease community, female patients are found to be more active in exchanging social support when they are experiencing extreme mood conditions (Yan and Tan 2010a). In addition, individuals may vary in their information seeking and behavior changes according to their different motivations and different information processing capabilities. Several theories have been proposed to explain these differences, including the Extended Parallel Process model, the Theory of Reasoned Action, and the Health Belief Model (for an overview, see O'Grady et al. 2008, Logan and Tse 2007). However, these aforementioned research works tend to explain people's information seeking motivations and identify their behavior patterns at the individual level, but they do not account for the collaborative activities in online communities. In spite of the fact that this experiential information seeking and knowledge sharing occurs with no boundary or limit, and that the consequent impact of the "wisdom of crowds" (Sarasohn-Kahn 2008) is described as having a big impact on health outcomes (Eysenbach et al. 2004), little research to date has examined the notion of collaborative behavior in relation to health information seeking and knowledge creation on the Internet.

1.5 Findings and Contributions

1.5.1 Online Social Support and Health Condition Dynamics

Despite the Internet's increasingly important role in the healthcare industry, how helpful online healthcare communities might be for patients is largely unknown (Lamberg 2003). To the best of our knowledge, there is little research that systematically studies the social influence brought about by patients' participation in online healthcare communities and the sharing of their disease information and knowledge. The objective of this work is to examine the impact of patients' activities in online social networks on their health outcomes. The challenge of studying this problem is the difficulty of measuring perceived utility through patients' online behaviors in online healthcare communities, especially when patients' health conditions are hidden most of the time. To overcome this obstacle, we propose a Partially Observed Markov Decision Process (POMDP) model, where a patient's health condition is partially observed and varies over time. The transition between different health conditions is determined by the extrinsic medical help and a set of covariates measuring the benefits patients receive from their activities in online healthcare community. The number of health condition states is determined by the complex characteristics of patients' online behavior and dynamics over time. To control for individual-specific characteristics, we include a set of random-effect coefficients to capture this unobserved heterogeneity. Finally, a maximum likelihood estimation procedure is conducted for this POMDP model.

There are countless reasons why people seek social support in online settings. For many people who participate in online healthcare communities, the platform is used to supplement traditional offline methods of support; for others, the online venues may be the only social support available. The POMDP model we propose identifies dynamic changes in health condition according to patients' online activities and provides evidence for the helpfulness and value of online healthcare communities. By incorporating partially observed patients' health conditions to examine (latent) dynamic changes, we find that patients receive various social supports from their online activities, and that communication with other similar patients has a positive impact on their health conditions. Although information is the major social support that patients seek/provide in online healthcare communities, emotional support has a higher magnitude of influence in helping patients move to a better health condition.

This work, to the best of our knowledge, is the first study to focus on online healthcare communities, where patients share their medical histories and health information to help one another. We find measurable evidence of how patients' social interactions affect their health conditions. Our work bridges the social networking and healthcare fields, offering the following contributions. First, we study patients with mental (chronic) problems and their online activities. We find quantitative evidence that online healthcare communities change patients' disease management behavior. Users who directly participate in social support exchanges receive support in various forms, such as information about their condition or the knowledge that others are experiencing similar stressful situations. These social supports help them to stop blaming themselves for their illness and present them with opportunities to actively engage in mutual aid and self-help. Second, we identify the effectiveness of online healthcare community as an online health university. Patients who are managing their disease and understand its progression are a tremendous resource for patients suffering similar problems. The access to a massive library of health data and visualized networking tools make it a true “health university” for the public (and provides the opportunity for post-trial analysis for the pharmaceutical industry). Third, the online platform helps patients to keep their health condition history and take control of their health and healthcare. By observing similar patients’ health condition dynamics, they can compare their outcomes with a large population and assess treatment effectiveness (Kane et al. 2009). Fourth, the proposed POMDP model can help to recover patients’ missing or unavailable information. It takes time and effort for patients to keep track of their health condition. And sometimes, patients may not find the opportunity to get their health condition assessed. Under such conditions, our work postulates a way to reveal the unobservable information effectively and accurately. Fifth, this work suggests a less costly diagnostic tool. Even if patients do have the chance to measure their health condition, the complex process reduces patients’ willingness to participate because it is time-consuming and there is a significant cost in effort. Our model thus provides a method to further simplify the procedure and generate a dynamic questionnaire covering patients’ health history and previous health condition. Hence, it can increase patients’ participation rate by requiring less time and effort and thus encouraging them to reveal more important and valuable information. Finally, our work is not limited to mental health problems, as it can be applied to other types of illnesses.

1.5.2 Communication Network Dynamics

We address challenges of how patients manage information in an OHC in the second essay. In other words, we study how people identify and select proper users with whom to communicate thereby reducing information overload in the context of an online subscription network. This strategic behavior can be explained by homophily (i.e., the preference for similarity), a concept which has been identified by social scientists as one of the important motivations for why people create, maintain, dissolve, and reconstitute communication networks (Monge and Contractor 2003). In this paper, we are particularly interested in studying the driving forces behind network formation and evolution. By tracking the network dynamic characteristics for patients’ subscription networks (or friendship networks) in the OHC, our model incorporates widely-examined homophily features in offline social network studies and online network measures, such as degree centrality, betweenness centrality, closeness centrality, clustering coefficient, structural equivalence, shortest path, and so forth.

We find that purely structural endogenous network features have positive impacts to increase the probability for individuals to find other people similar to them online and communicate with them. Although complex social network structures have been shown to contribute to the spread of information, norms, and social exchanges in many offline settings, Internet-based interpersonal communication is changing the nature of friendship and other interpersonal communications (Thelwall 2009). Our work therefore contributes to the understanding of social changes and the consequent impacts brought by information and communication digitalization. Our approach also contributes to expanding the homophily research when traditional homophily theories fail to explain the changes in individual preferences over time.

This research is also distinguished from other online network studies. Social structure and the individuals in it make social capital a complicated concept. As indicated by Reagans (2005), contextual factors play important roles in explaining network connections. In our context of an OHC, the OHC provides supplemental support for real-world health services. That is, only when offline acquaintances lack the expertise or empathy needed by an individual, do patients turn to online sources and hence form exceptional relationships with other individuals within the OHC. This unique characteristic highlights the OHC's differences from other online social networking. The perception of similarity could be even strengthened by patients' initial purpose of joining the online community in order to satisfy their needs. Under such circumstances, our work sheds light on practical applications for web service providers and has managerial implications for healthcare providers and practitioners. In general, although technologies contribute to the problem of information overload, as discussed earlier, technologies also provide the channels and mechanisms through which information can be distributed or accessed. Hence, healthcare providers have opportunities to offer better services and products through online communities, and patients have opportunities to be more involved their own disease management.

Our theoretical and practical contributions can be summarized as follows. First, on the theoretical side, our study explored the propensity that human factors are involved in the situation of information overload. Following the above discussions, our analysis and proposed solutions add to the existing literature on the effort of reducing the effect of information overload. Our empirical results show how patients gradually develop their information management skills and how they pull the information they need. As more and more researchers and practitioners become interested in studying online users' behaviors and how to integrate social capital into online applications, our work calls their attention to the initial step of understanding how such networks form and grow.

From the online service prospective, our work examines community members' behavior patterns and finds that patients use the OHC to share health information, exchange knowledge, and identify similar patients from whom to learn using their existing connections. These results suggest ways to design better patient websites and improve the service providers' recommendation systems. On the practical side, our work contributes in three ways. First, as shown by the existing literature, social capital is different from other forms of capital in that its resources are embedded in a social structure that is accessed and/or mobilized in purposive action (Lin 2001). Social capital is embedded in the relationship between individuals and in their connections with their communities (Putnam 1995). By studying patients' communication relationships, our work helps managers to take advantage and occupy the strategic position in the network so that they can lead the conversations and spread social influence. Second, given that the purpose of websites has a significant impact on the characteristics of their social networks, as shown by Golbeck (2009), managers should carefully design these online spaces, because it is the design of these online communities rather than users' networking that is most critical for information systems managers. Our work suggests what users need and what they care about. Third, considering why users share information online and the issues concerning of the credibility of the online content, website designers need to put more effort into encouraging users to provide useful and trustworthy information and share their knowledge with other members. It is especially important that the online space provides a trust-based recommendation system in which assumed trust is captured by similarity among users.

1.5.3 Online WOM and Health Education

To address these aforementioned challenges, we investigate how patients utilize and educate themselves by learning about others' shared experiences in the online healthcare communities, a process which can be described as "experiential learning" in the healthcare context in contrast to "observational learning" in other contexts. Our study combines patients' individual health condition data, the network data on their online connections of their followers and discussions, the treatment data, and online reviews for these treatments. In this way, we are able to investigate the social influence or social contagion effect in patients' collaborative health education, which is of great relevance to both practitioners and researchers (O'Grady et al. 2008, Fichman et al. 2011).

We find that, in general, there is no consistent WOM effect if we examine information attributes separately. Our empirical study suggests that the average rating of a product has no impact on the consumer's perception of product quality when considering the information attributes of side effects, persistency and hardship. Although this result conflicts with findings in some prior marketing literature, it confirms the finding from one study that an individual's opinion may not be useful in assessing another's own opinion because experiential products evoke many different emotional responses and render the possibility for multiple interpretations of product experiences (West and Broniarczyk 1998). Meanwhile, the variance of the ratings from the community dimension has been found to be of no significant influence, suggesting that although a high level of disagreement in product reviews can indicate a high level of risk for a product, it does not necessarily affect the perception of the product. The only exception is that the volume of WOM and the averaged proxy for the perceived treatment efficacy attribute is found to have significant impact. Surprisingly, WOM from an individual's close connections is found to be insignificant, which can be explained as follows. Consumers who frequently communicate know each other well, therefore the usefulness of divergence in their opinion in providing additional insights has been compromised in their communications. However, by integrating information attributes (formative model) and creating a constitutional construct, the perception of treatment quality, we find the online WOM has a significant impact on patients' health education. In addition, WOM sources, whether those of the public or of friends, are found to present different influential patterns, which suggests that WOM plays different roles in educating patients and helping them to improve their health knowledge.

Our work fits into the field of social media and contributes to the broad discourse of WOM from several aspects. From the theoretical perspective, this study constructs a solid framework to incorporate information attributes into consumers' perceptions of a product, identifies the evidence of different impacts of experience-based information attributes, and expands the effective use of information provided by social conversation. In particular, it complements the traditional marketing literature from a unique perspective. In most literature on WOM, a single indicator, i.e. rating, is often used to summarize the reviews of different aspects of a product (Archak et al. 2007). In this work, we decompose the effects of online reviews into multiple experiential aspects. By exploring the effects of online reviews from multiple aspects and the interactions among these different aspects, we are able to provide a theoretical model that comprehensively explains the relationship between WOM and the product quality perceived by customers. To the best of our knowledge, we also provide the first empirical demonstration of how the experiential aspects are significant determinants for customers' perceptions of product quality.

Chapter 2 Literature Review and Theoretical Framework

2.1 Online Social Support and Health Condition Dynamics

An online healthcare community is a special kind of virtual community where individuals meet online and exchange social support. Different from the widely used, email-based type of support group, an online healthcare community affords significant advantages, such as access to a voluminous dataset (the aggregated knowledge generated by members), and live online discussions. These modern social media-based communities are constructed on a commons-based peer production basis (Benkler 2002, Fichman et al. 2011) and are especially attractive to individuals with rare diseases and/or chronic health problems. An online healthcare community provides patients with new opportunities to gain support within a virtual network formed by individuals dealing with similar issues. Many of those who join online healthcare communities are patients who are active in their self-care process. Following Merton's (1976) description of the role of the "good doctor," Radley and Billig (1996) advocate that the "good patient" must be more than a patient to receive the entitlement. For this reason, aside from physical conditions, internal attitude plays an important role in "defining" patients' health conditions.

A positive relationship between people's health and social support has long been recognized. Social support is one of the most important factors in predicting individuals' overall physical health (Chernomas and Clarke, Clark 2006). Cobb (1976) explains that supportive interactions among people protect against the health consequences of stress. McCorkle et al. (2008) find that social support increases patients' adherence to treatments and enhances recovery. As a result, researchers across disciplines have been studying the social support of individuals in various scenarios. Today, a major body of sociology research categorizes social support in four forms: informational support, emotional support, companionship, and instrumental assistance (Berkman et al. 2000, Wortman and Conway 1985).

Informational support involves the transmission of information, including advice and referrals. The convenient access to the mass of information makes health-related topics one of the most popular searches on the Internet (McMullan 2006). The Internet is a source for mental health information for over

10% of the general population, and more than 20% of people who have a history of mental health problems (Powell and Clarke 2006). There is also a significant amount of research showing extensive use of online healthcare communities. Members of online healthcare communities create health profiles or blogs to share geographic and demographic information such as age and gender and compile the effects of various medical treatments. Rather than sorting through huge piles of test results and other paperwork, online health profiles help patients keep track of their treatment progress and medications easily and conveniently. This simplifies medical interpretation, enables patients to understand their condition better, and helps patients make treatment decisions (McMullan 2006).

The increased access to shared health information, medical experiences, and treatment history in online healthcare communities can produce more informed patients. The knowledge gained from informational support can help provide a greater understanding of problems and possible solutions. The more health information patients learn, the better they understand their condition and the better they can take steps to care for themselves (Kassirer 2000, McMullan et al. 2006, Wanless 2002). Learning experiential information from other patients' profiles provides an individual with a window on second opinions, information that is "difficult" to ask directly, and assistance in making sense of the stage of the disease (Ziebland 2004). Therefore, it makes the Internet and online healthcare communities an attractive resource for information about new treatments and discoveries.

Hypothesis I.1: *Informational support has a positive effect on patients' health conditions.*

Emotional support comes in the form of sharing happiness or sadness, or expressing caring and concern. It sends a signal that one is not alone, that one is taken care of and valued. This kind of support is especially important for patients with chronic mental problems. First, patients with mental health issues have difficulty in developing and maintaining relationships to receive meaningful help. At different stages, the disease can inhibit the ability of patients to cope with their illness. Family relationships can become strained and support withdrawn because of the various burdens that stem from the disease (Weinberg et al. 1995, Wright 2000). Second, due to the limits of time and resources, it may be difficult for offline relations to provide support when it is needed. However, with no geography boundary, online healthcare communities make it possible for patients to talk with other patients suffering from similar illnesses at any time (Bambina 2007, Lamberg 2003). Third, and most important, knowing that others have faced a similar problem, and even overcome it, can provide both relief from personal blame and renewed strength (Bambina 2007, Weiss 1974, Wills et al. 1985). It is found that online healthcare community members often develop intimate and trusting relationships; among other things, they provide referrals and encourage each other to stick with therapy (Lamberg 2003).

Hypothesis I.2: *Emotional support has a positive effect on patients' health conditions.*

The informational and emotional support delivered through online healthcare communities may help patients cope better with their mental problems. Bambina (2007) investigated a cancer forum and finds that on average members receive more informational support than emotional support. However, it remains unclear, due to the lack of prior empirical evidence, which type of support is better for meeting patients' social needs. To help improve the effectiveness of online healthcare communities, it is important to contrast the respective effect of informational and emotional support.

Hypothesis I.3A: *For patients with mental problems, the emotional support they receive online plays a more important role in helping them to progress to healthier conditions than informational support.*

Hypothesis I.3B: *For patients with mental problems, the informational support they receive online plays a more important role in helping them to progress to healthier conditions than emotional support.*

While it is important to differentiate the type of social support, Jacobson (1986) points out that the "timing" or sequence of social support can affect its effectiveness. A medical problem may need different type of support as it moves through its stages (Pearlin 1985). This calls for social support to be examined as a dynamic process rather than just a resource or outcome (King et al. 2006). In the context of this study, the dynamics is characterized by a patient's fluctuating among health conditions. The helpfulness of social support may vary depending on patients' instant health conditions.

Hypothesis I.4: *The effect of social support is moderated by patients' health conditions.*

Companionship can consist of group meetings, chatting, and other social activities. It provides support by making individuals feel there are others who enjoy their presence and that they are a valuable part of something bigger than themselves (Wellman and Wortley 1990). In an online healthcare community, such support is usually exchanged by participating in a discussion forum. The various activities in online healthcare community act as "talk" therapy and can make each individual feel they are not isolated from the world and have social connections. Finally, instrumental or practical support refers to assistance in locating life-related resources. This kind of support is usually not available in online healthcare community settings, as it requires individuals to reveal their real-life identity. In this study, social support is classified to either informational, emotional, or companionship, so their respective effect cannot be empirically identified simultaneously. We have selected companionship as the base category.

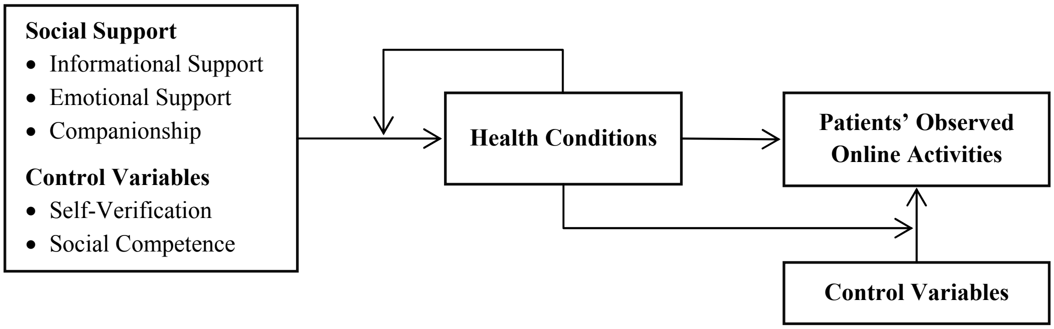


Figure 1: Conceptual Framework

Figure 1 shows the conceptual framework where social support influence patients' health conditions and there is a moderating effect. It is theorized that patients' health conditions consequently affect their online activities and also moderate other control variables.

2.2 Communication Network Dynamics

The propensity for people to request and to disseminate information is usually described as "information pull" and "information push" (Wilson 2001). In the OHC, information pull relates to patients seeking information if the information can help them better understand their health conditions. Patients' needs for information to mitigate uncertainty and improve understanding helps to explain the rich discussions and comments that researchers have observed in virtual communities. Information push, on the other hand, is more related to the “push technology” in how patients access the information they need. RSS, for example, is one of the popular tools used to send alerts and distribute new content. However, in the OHC, especially for patients who suffer from chronic diseases, historical health information may be more meaningful and may maximize the utility of health data for each contributor more than simply “the latest” content. Therefore, a subscription tool – which allows patients to subscribe to interesting topics, channels, or subgroups so that they can establish virtual friends or communication networks – is more suitable and useful for patients. Subscription tools are widely used to evaluate proposed commitment and to indicate the degree of importance or the relevance of a given piece information in online social websites. This tool is especially helpful as the total number of social relationships one can manage is around 150 (Gladwell 2000), among which 10 to 20 are close ones (Parks 2007). More importantly, subscription tools help to reduce information overload to certain extent. Among the various services provided in the virtual world, subscription services enable subscribers to be immediately notified of updated information and also to be linked with the direct access to such information (Souto et al. 2006). By subscribing to other members’ profiles or messages, users can precisely and quickly identify with whom to communicate and what changes have happened, and therefore they can reduce their search time. Moreover, the subscription service can also remind members of the topics that were discussed earlier and keep track of the progress of such conversations. Some websites even provide the option of categorizing or flagging within the subscription category to further classify the content and thereby reduce information overload. However, this communication network is not built once and for all. Instead, by establishing new connections and removing idle ones, patients are learning and adjusting their information management behaviors through their online communications, so that they can gradually develop their capabilities to identify and process useful information. This learning process is explained by examining the dynamic changes of the network.

2.2.1 Related Literature

Social networks are realizations of various patterns of relations among individuals, and these networks are not static but evolve over time (Snijders 2001, 2005). Individuals in the network are called actors, and their social connections are called ties. Specifically, in our study settings, patients join the online social networks to gain useful information and communicate with other members given the current network configuration. Community members make choices in the network evolution based on their evaluation of information level they receive at the present state of the network, and thereafter they try to obtain a "pleasant" structure of relationships. As stated by Snijders (2005), these changes could be caused by the mechanism related to individuals' characteristics and purely structural network endogenous mechanisms. The best-known driving force for relationship addition and removal is known as homophily.

A vast body of literature studying homophily has existed for the past 50 years. As a widely-used term, homophily describes the degree of similarity between people and the strength of such connections. Our study relates to a large body of literature on homophily across a wide variety of disciplines. In this section, we discuss the streams of particular relevance to our work.

The first stream of literature that relates to our work is on similarity and communication. Researchers agree that sharing an attribute would produce some baseline interpersonal attraction and homophily (Reagans 2005). The strength of such communication relations is also influenced by similarity on specific attributes of actors. Byrne (1961, 1971) proposes the similarity attraction hypothesis, that the more similar people are, the more they will be attracted to each other. Such attraction has powerful implications for people's social worlds, including the type of information they receive, the interaction they experience, etc. (McPherson et al. 2001). Several studies have shown evidence of communications selected based on similarity in age, gender, education, tenure, and so on. Zenger and Lawrence (1989) observe that there is a relationship between age and the frequency of technical communication. Brass (1985) reports that the communication network in an organization is clustered based on gender. Researchers in sociology and statistics have also extensively studied the influence of individual characteristics on communications. The importance of individual preference has been found to be of significant impact on personal and demographic similarities on network connections (e.g., Zenger and Lawrence 1989). For example, McPherson et al. (2001) find that in the U.S., race, ethnicity, age, religion, educational level, occupation, and gender are influential factors for friendship. Recently, a study examining whether online communication changes offline friendship patterns finds that the influence of gender similarity is eliminated but race, ethnicity, age, religion, marital status, and etc. still have strong influence on connections (Thelwall 2009). However, in light of the digital revolution and consequent blurring of distinctions between telecommunication and computing technologies, new communication technologies have induced some recent attempts to consider contextual factors and the changes brought to the social network (Reagans 2005, Monge and Contractor 2003). The emergence of Web 2.0 technologies and UGC calls for more investigations on Internet-based homophily and the subsequent changes brought to online social networks (Thelwall 2009). As a result of limited nonverbal cues in online environments, individuals may find it difficult to assess similarity (Wright 2000).

Next, consider the literature exploring the motivations and driving forces for individuals to participate in online communications. The theory of self-categorization (Turner et al. 1986) suggests that individuals define their social identity through a cognitive grouping process and classify others and themselves based on similarity characteristics such as age and gender. Identification therefore increases the probability of constructing the strong (direct) tie. Similarly, the social identity/deindividuation theory of computer-mediated communication (Lea and Spears 1992, Wang et al. 2008) states that it is social identity, or social similarity of online communicators with a salient group reference, that drives identification and relationship in online interactions. Another closely-related explanation for homophily is from social comparison theory (Festinger 1954), which demonstrates that self-evaluation is based on comparisons to other people. The relevant others are selected on the basis of homophily, and there is an upward drive toward achieving greater abilities.

However, there are several divergent findings in the recent studies on web-based communities. On one hand, a study on MySpace (Thelwall 2009) finds the existence of the widely-used geographic and demographic homophily such as ethnicity, religion, age, country, marital status, attitude towards children, sexual orientation, and reason for joining the community and networking together. Wang et al. (2008), in their study on health information and influence via the Internet, also find that people are increasingly likely to adopt advice from those people who are more homophilous. On the other hand, recent studies on web-based communities have found evidence for the diminishing explanatory power of baseline homophily, which has been widely studied previously in the offline settings. For example, Preece (1999, 2001) examines the contribution of empathy, also known as emotional support, to the online community and finds that people with similar backgrounds, such as similarity features in illness, addiction, disability, and other similar health experiences, are more likely to exhibit relationships. Thelwall (2009) also finds that, unlike offline world studies, the feature of gender in MySpace is found to be insignificant in the homophily selection process. However, these mixed findings suggest that baseline homophily could be undermined given the ease of Internet communication. Investigations focusing on Internet-based homophily are called for to understand the formation of online communities and the evolution of such networks (Thelwall 2009). All of this suggests the need to further examine the homophily phenomenon in the OHC environment.

While individual characteristics explain part of the decision why patients form a tie, the position where a patient is in the network will impact her receipt of social support and influence other activities of social events. As Snijders (2005) found, structure changes can be conducted by network endogenous mechanisms and patterns related to individual characteristics. Given that the structure of a network greatly affects information and knowledge seeking and diffusion dynamics (Singh et al. 2011, Burt 2004, Coleman 1988), the relationships in a subscription network are likely to influence the accessibility and validity of information that patients need. It is explicitly stated in Festinger's (1954) classic theory of social comparison that there are various ways for people to value others as similar references, including their structural position in the network. Patients influence each other and transfer information and experiences through their connections. The shared social connections and activities proposed in activity focus theory (Feld 1981), the cohesion mechanism and structurally equivalent positions (Krackhardt and Brass 1994), and the theories of physical proximity (Festinger et al. 1950, Korzenny and Bauer 1981, Monge et al. 1985) also address the importance of network position – that individuals meet and interact from different aspects – to offer the explanations for network connections.

2.2.2 Individual Characteristics

Without the limits of place and time, the virtual space allows patients to access information and communicate anytime and anywhere. However, because of the anonymity feature of the Internet, most of the information considered in traditional homophily studies is unavailable (for example, familial background). Typically, in the OHC, a patient's profile would contain her basic geographic information such as *state* and demographic information like *age* and *gender*, which are important characteristics for offline communication networks. But, these geographic and demographic data are not necessarily associated with the disease (here we refer to general chronic diseases excluding gender-specific cases). Therefore, we hypothesize that the effect of baseline homophily would be declined among patients' networks.

Hypothesis II.1A: *Geographic and demographic homophily is insignificant in OHC.*

As discussed earlier, the OHC is not a replacement for local healthcare service but a supplement to it when patients lack the social support they need. Under such circumstances, the purpose of patients joining and participating in an OHC is to exchange health information and other supports they cannot get elsewhere, and to learn from each other so as to improve their own disease management. Intuitively, this suggests that health information and disease experience is more valuable for OHC users to exchange. And these personal experiences (for example, similar symptoms or problems faced before) trigger the communication among patients (Wang et al. 2008) and follow-ups. The continuing relationship is not fixed and can be terminated based on patients' information needs, thus creating the online social network and its dynamic changes. Hence, we propose that in the context of an OHC, patients with similar illness conditions are more likely exhibit association.

Hypothesis II.1B: *Disease experience-based similarity increases the probability for patients to communicate.*

2.2.3 Network Characteristics

In this section, we provide some expositive details for the network measures used in our model to study the dynamics of social networks. Specifically, the social network in our study is patients' profile subscription network. By tracking the progress of homophilous patients' medical information and experience, patients can better engage in their own disease management and reduce the possibility of a worsening health condition. Thus, by identifying other similar patients, patients also develop their skills to process information and gain knowledge from communications.

2.2.3.1 Centrality

The most common use of graph theory in social network analysis is to identify the "important" actors in a social network (Wasserman and Faust 1994). An individual's location in a social network is usually identified and measured by prominence. Actors who are more prominent or visible are more likely to be found in the important location in a network, and thus are more likely to be connected with. The prominence, recognized by Knoke and Burt (1983), can be measured by centrality. In general, there are three prominent kinds of centrality, namely degree, closeness, and betweenness centralities.

Degree centrality is the measure of the number of connections an individual has. In a direct network, a prestigious actor is the one with high indegree centrality and is held in high regard by peers. This measure is simply calculated by counting the number of subscriber connections for each patient in the network and normalized to be between 0 and 1. These incoming connections make an individual particularly visible to others in a network, thereby creating a higher chance for her to be connected.

Closeness centrality focuses on how close an individual is to other actors in a network. Closeness is a function of geodesic distances, including direct and indirect connections. A high value on this measure implies the central position in a network and grants the individuals the ability to reach various resources. A patient who holds a more central position in the network will be involved more quickly in communication with all other members in the community, thus making the information exchange process more efficient. In our context, the closeness metric can be considered as the proxy for how similar the focal patient is to other patients in the network. Thus the closer they are, the more similar their health conditions.

Betweenness centrality is used to identify strategic positions in a network. Because communication between two non-adjacent patients depends on the other patients who lie on the path to connect them, this measure is calculated by normalizing the summation of probability that the patient is "involved" in the communication. According to Freeman (1979) and Friedkin (1991), the "actor in the middle" between the others has more "interpersonal influence" on others and therefore has the potential to control the influence of attention. In other words, betweenness measures an actor's position between two other members along the shortest path, or geodesic, that connects them, indicating a potential position of power and illustrating the amount of control and influence an actor can have over the interactions and the flow of information between those actors connected through her (Bambina 2007). In our study, individuals find similar patients to communicate with and share knowledge of their disease to improve their own health condition. Given a large pool of similar patients to choose from and thousands of files in the community to read, the prestigious patients are more visible to be identified and connected with. Similarly, patients with higher closeness centrality have a higher chance to be more similar. According to the theory of similarity attraction, closeness centrality has the advantage of speedy information exchange with other similar patients and patients can quickly initiate communication with those similar others. However, a higher betweenness centrality indicates a patient's strategic position of controlling the information flow to other members in the network. Although multiple paths ensure the transmission of the information, it also indicates the redundant connections she has and the amount of duplicated information she needs to process. Since she already has access to gain the necessary information from her connections, therefore, there is less incentive for her to connect with more patients and increase her information load.

Hypothesis II.2: *A higher indegree centrality will increase the probability for a patient to be identified and communicated with. A higher closeness centrality will increase the chance for a patient to initiate connections to other similar patients. A higher betweenness centrality will decrease a patient's willingness to make connections.*

2.2.3.2 Latent Space

In a virtual community, the social distance between two individuals is usually unobservable. Most studies on binary networks have the implicit assumption of equally weighted ties. However, there is a growing body of literature advocating that the probability of tie initiation should persist as a decreasing function with the latent distance. Furthermore, as noted by Hubbell (1965) and Friedkin (1991), prominence should be measured by looking not only at direct ties, but also at indirect paths involving intermediaries. These arguments suggest that individuals with similar characteristics will possess an enhanced relationship. Discussions about the social space mention various concepts (for example, Faust 1988). One of these discussions, (Hoff et al. 2002), considers a subset of individuals with close connections, and this connection pattern may indicate that group members have nearby positions in the social space of characteristics. An example is displayed in Figure 1. The closer they are, the less search effort is required. Therefore, we propose that the latent distance will have a negative relationship with the probability of communication.

Hypothesis II.3: *The number of ties for two patients to get to know each other (the shortest path) will be negatively related to the probability of constructing a strong (direct) tie between them.*

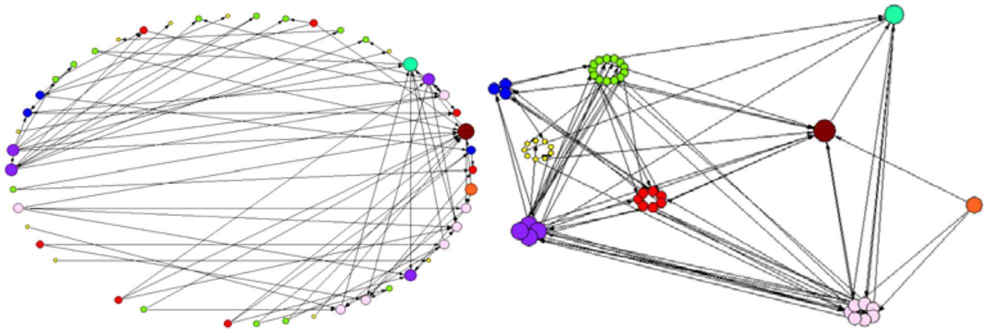


Figure 2. Patients grouped based on latent space

2.2.3.3 Structural Equivalence

Structural equivalence is calculated by considering both people and connections. Two actors are structurally equivalent if they have the same types of ties to the same (type of) people. Usually structural equivalence is calculated by the total number of common neighbors normalized by first and second largest degree. Patients who subscribe to similar patients' profiles will have a higher probability of identifying more similar patients from this transmittable subscription relationship. Therefore, this cohesion process helps to categorize relevant patients, reduces the amount of information that needs to be processed, and enables patients to build up information and knowledge exchange channels and exhibit communication connections. Furthermore, for a fixed level of structural equivalence, the increase in path length will decrease the probability of constructing such a strong connection.

Hypothesis II.4: *Structural equivalence indicates the similarity between pairs and will be positively related to the probability of constructing a strong (direct) tie between them.*

Hypothesis II.5: *The interaction of increased path length with structural equivalence will be negatively related to the probability of constructing a strong (direct) tie between them.*

Noteworthy in our study is that we are using the *dissimilarity* index to measure structural equivalence. We count and normalize the total number of uncommon neighbors for the pair of patients. Therefore, in our model, the structural similarity level decreases with this index. We also use this measure to control for the influence of non-similar patients. In our study, subscription initiation is interpreted as a connection, and neighbors are defined as such. It could be possible that patients share common friends outside the subscription network we are restricted to, and this relationship might increase the probability to create a tie in the "insider network." Hence, this index is constructed based on patient's acquaintances both inside and outside the subscription network.

2.2.3.4 Latent Trait

Clustering is used to describe the relatively strong internal relations shared by a group of people. Clustering states the probability that a patient's pair of subscribers also subscribe to each other, and is thus defined by the proportion of connections between the patients within their neighborhood. When a patient finds a similar patient and establishes a communication connection, it is easy for her to find out who else has subscribed to this patient. And she can also find the list of patients who have been followed up by her contactor. Since individuals tend to have very homogeneous beliefs within highly cohesive groups (Collins 1988), there is a higher chance for her to find other patients with similar concerns and common interests and establish communication relations through these existing connections. These paths suggest the possible information flow that she can construct so as to get what she needs to help her reduce her search effort and the amount of information to be processed. However, this cohesiveness will be decreased with the increased distance as the cost of identifying whom to contact and the amount of information to process grows.

Hypothesis II.6: *The cluster coefficient will be positively related to the probability of constructing a strong (direct) tie between them. This effect will be enhanced after controlling for the path length*.

2.3 Online WOM and Health Education

The search for health information through online health communities has been the focus of considerable attention as the number of patients relying on online health information has steadily increased (Nambisan 2011). Patients can more easily bypass traditional information intermediaries to learn about health topics, and communicate and share knowledge directly with one another in a timely manner. These behaviors, as shown in prior studies, are stimulated because patients are likely to seek others' opinions to reduce their cognitive effort or uncertainty when making decisions (West and Broniarczyk 1998, Dellarocas 2006). However, it must be noted that many of the information seeking patterns we now observe in healthcare communities are not new – they merely replicate the patterns and preferences for information seeking seen in other online environments. What is new is the increased ability for patients to access "more patients like me" and the ways in which information is used by patients and their families (Epstein and Street 2011). Patient-patient collaborations have been recognized as an important element in supporting those learning about health conditions (O'Grady 2008), and our effort in exploring the factors which acknowledge experiential learning are thus drawn from the following three aspects: the effectiveness of WOM in the online healthcare community, the homophily phenomenon, and the credence derived from the social role in communications. We address these facets in detail in the rest of this section.

2.3.1 The Effectiveness of WOM

When patients need health information to manage their personal health, they turn to other patients for their first hand experience and advice (Hartzler and Pratt 2011). Unlike the expertise offered by health professionals, patients offer other patients substantial personal health guidance based on the expertise they have gained from managing similar health situations. During interactions, patients discuss various topics including symptom interpretations, illness management, the nature and effective treatments, and so on (Coulson 2005, Christensen and Griffiths 2008). With access to the Internet, this informational support is not confined to a limited period of time but can be accessed at any time and anywhere. As a result, online healthcare communities become alternative sources of health information as well as connecting patients who have the same or similar disease. Significantly, the knowledge offered in these places is generated amongst a group of patients who share experiential or anecdotal information about health situations and offer feedback on treatments. Although online conversations can follow different formats, we believe that in general, WOM in healthcare communities has a significant impact on patients' healthcare education.

Hypothesis III.1 (HIII.1): *online WOM plays a significant role in patients' health education.*

Typically these online conversations in the healthcare communities are observed in numeric and text forms. Examples of the former include ratings for a particular drug and the distribution for these multitude ratings. The latter is typically found in forum discussions which contain text thread exchanges and comments for a rating that further reveals patients' experiences and opinions. In this study, we decompose the WOM into these two forms and examine each of them as follows.

2.3.1.1 The Volume and Dispersion of WOM

Whereas the overall ratings, as a proxy of perceived product quality which has been used widely in the existing literature, have been shown to be more influential than information content in affecting consumer interest (Wyatt and Badger 1990), recent studies have pointed out that the information embedded in product reviews cannot be captured simply by such a single scalar value (Archak et al. 2007). Other information dimensions are also valuable for developing a comprehensive understanding of customer online reviews; for example, the degree of disagreement among consumers' reviews is shown to be important when the average rating for a product is low (Sun 2012). In light of this evidence, we consider two distinct dimensions of numeric WOM, volume and dispersion, in our study. These measures are attractive in that they are provided by most online social networking sites as convenient tools for users to see the distribution of ratings.

Volume of WOM. The first and most obvious dimension of WOM is its volume (Godes and Mayzlin 2004). It has been examined as the essential measure in prior studies (e.g., Van de Bulte and Lilien 2001, Godes and Mayzlin 2004, Chevalier et al. 2006). The more conversations there are, the higher is the chance that more patients will be informed about the topic and consequently be aware of the focused treatment, or learn from other patients' experiences and knowledge. Because knowing and understanding a treatment is the necessary condition for patients to set up proper expectations for the treatment outcome, we expect that the higher volume of WOM will lead to higher ratings.

Hypothesis III.1.1 (HIII.1.1): *the volume of WOM is positively associated with future ratings*.

Dispersion of WOM. The overall rating, typically expressed as averaged WOM, has been widely used in the existing literatures as a proxy of perceived product quality. There have been numerous analytical and empirical attempts to study the impact of ex-post reviews on consumer valence and product sales, but the findings are mixed. For example, a positive impact of user reviews on book sales has been found by studying the WOM effect on online retail channels such as Amazon.com and BarnesandNoble.com (Chevalier and Mayzlin 2006). However, Chen and his colleagues find insignificant influence of user ratings from similar empirical resources. Among others, Duan et al. (2005) test the relation between online reviews for software downloads and find that the volume of reviews has a positive impact on product sales. By studying the online reviews for movies, Liu (2006) shows that online WOM has a significant effect on box office revenue. When it comes to the treatment outcome evaluations, the effect of the WOM becomes less obvious. On one hand, the higher average rating of a treatment indicates a satisfactory treatment outcome which, because it is generated by prior users, suggests a higher probability for a higher future rating. On the other hand, patients may set a high expectation for the treatment outcome based on what they learn from the shared information, which can cause a lower rating due to unfulfilled expectations. Therefore, we propose:

Hypothesis III.1.2A (HIII.1.2A): *the average rating is positively associated with future ratings.*

Hypothesis III.1.2B (HIII.1.2B): *the average rating is negatively associated with future ratings.*

In addition to the average rating of the numerical reviews, the variance of the rating offers the percentage of reviews that are associated with each numerical score (Sun 2012). The pie chart and bar chart make the distribution of ratings convenient for patients to compare their situation with other patients who also taking the treatment. The variance of patients' ratings thus suggests heterogeneity in patients' treatment outcomes. The higher measure indicates the reviews generated by patients with broader health conditions, and the effectiveness of the treatment is thus distributed according to specific, individual cases. This dispersion implies that patients who share their experience online confront various problems and difficulties during treatment trials; thus it is easier for a patient to find other patients who are involved in similar conditions and how they deal with these difficulties. Inconsistent opinions thus offer more information that is valuable for patients to learn from, thus we expect the variance of ratings to be positively associate with ratings.

Hypothesis III.1.3 (HIII.1.3): *a higher variance of prior ratings is positively associated with future ratings.*

2.3.1.2 The Opinion of WOM

Recent study has proposed that text messages are more detailed and are believed to more accurately reflect true sentiment (Archak et al. 2007). When patients provide textual comments in addition to ratings, they are offering additional information and evidence to support the decision they have shared – in the WOM context, the ratings they have given. This kind of review helps to reduce misinformation caused by unobserved factors, such as preference, values, and attitudes. Furthermore, text comment also improves communications and makes it easier to understand each other's subjective experiences through other pathways. Therefore, we hypothesize:

Hypothesis III.1.4 (HIII.1.4): *the volume and dispersion of text messages have a significant impact on future ratings.*

2.3.2 Homophily within Healthcare Communities

The potential role of homophily, as has been suggested in previous research, is a key ingredient for the effective function of online health communities (Nambisan 2011). Consumers are found to purposefully seek out information for evaluating experiential products such as advice from critics when selecting a movie and advice from friends when selecting a restaurant (West and Broniarczyk 1998). The Internet and user reviews allow consumers to overcome geographic boundaries and to communicate based on mutual interests (Mayzin 2006), and the increasing usage of online social media enables information about product quality to transmit through a large-scale user network. As an emerging source of product information, online WOM provides a substantial number of messages from the users' perspective; however, it becomes increasingly hard to determine whose comments should be taken into account. The chat channels, the recommendation sites, and customer reviews offer potential buyers a massive amount of information from individuals' from various backgrounds. It is intuitive that not all consumer generated reviews will be equally valuable. The online social network application allows consumers to leverage available information from strangers with low credibility to friends with high trustworthyness, and helps users to differentiate the importance of certain information and select trustworthy information sources. In a survey about consumer behavior, the majority of respondents report that they trust company or product information coming from "people like me" (the Edleman Trust Barometer report in 2008). Yan and Tan (2010b) find that patients are more likely to communicate with others who have similar health conditions or experiences. Although the summarized ratings of online consumers' reviews represent signals of product quality, the WOM from friends might provide exceptional value because they share similar preferences and tastes and their opinions can increase the chances of finding the best match product. For that reason, the online reviews generated from friends or the general public will impact users differently.

Hypothesis III.2 (HIII.2): *the WOM generated by friends has significant impact.*

2.3.3 Creditability of Patient Expertise

Similar to the essential role played by the concept of similarity, the credibility of the information can also play an important role as more lay users become recognized as experts and opinion leaders in the online environment. Broadly speaking, opinion leaders, or influentials, are individuals who are likely to influence others in the communication environment. Their words and opinions can reach more people and influence their decisions; however, unlike the offline world where various proxies can be used to measure people's social role, the online healthcare community is a self-reporting environment in which social status is rarely used as the proxy for the creditability of shared information. Instead, the feedback to the usefulness of a comment can be used as a proxy measure for assessing the impact of individuals within a collaborative learning environment (O'Grady et al. 2008). When patients exchange information and knowledge in healthcare communities, a common practice is that their inputs can also be evaluated by other lay users. This mechanism helps to leverage the shared information and also suggest the creditability of the opinion. A higher value indicates a more experienced patient, whose opinion is thus more persuasive. Therefore, we expect that creditability has a significant impact on ratings.

Hypothesis III.3 (HIII.3): *the creditability has significant impact to future ratings.*

Chapter 3 Feel Blue so Go Online: An Empirical Study of Online Supports among Patients

3. 1 Model

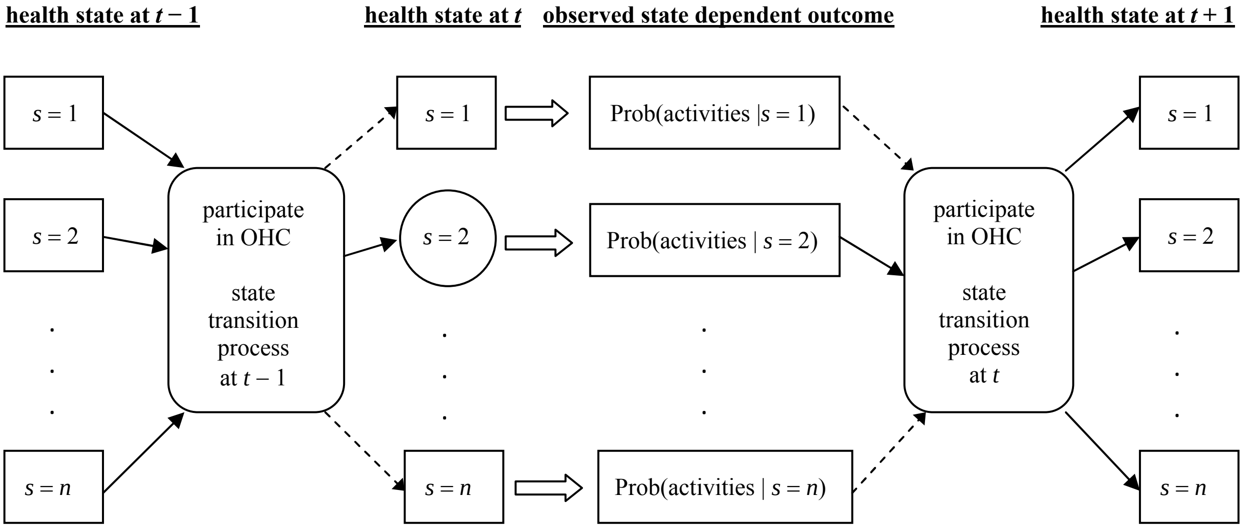
One of the challenges in this study is that the health conditions of a patient are partially observable and evolve over time. This makes it impossible to compare consecutive health conditions and draw the conclusion whether patients benefit from social support. To recover latent health conditions, we propose a model based on Partially Observed Markov Decision Process (POMDP) in which patients' health conditions can be inferred from other observables such as their online activities.

3.1.1 Partially Observed Markov Decision Process

The POMDP starts with the Hidden Markov Model (HMM), which is modified to account for the only partial observability of health conditions. HMM is a stochastic process that is not directly observable but can be inferred through another stochastic process that produces a sequence of observable outcomes (Rabiner 1989). It has been widely applied to different contexts. For example, Netzer et al. (2008) capture customers' dynamic relationships by modeling latent relationship states. Hauser et al. (2009) study customer's cognitive style in the context of website morphing. Singh et al. (2011) identify developer's learning dynamics from their past experiences and interactions with other peers in the context of open source software development. In other HMM models, Conati et al. (2002) examine students' understanding of a math concept. Zhang et al. (2010) study the vendor's learning curve from software security vulnerabilities.

In our case, we identify a patient's health condition as the latent state, whenever it is unobservable, and study whether a patient's online participation helps to change her latent state. Such a transition can be triggered by communication, the exchange of information and knowledge, or other activities with focal patients in the online healthcare community. The time variant for online activities is the observed outcome sequence for a patients' online behavior. The Markovian transitions account for the dependence on subsequent behaviors.

Figure 2 sketches the POMDP in this study. In contrast to HMM, a patient's (latent) health condition is not completely hidden. Therefore, by considering this information we are able to reduce the randomness of the HMM and redefine the latent states' distribution.



A rectangle indicates an unobserved state whereas a circle refers to an observed state. A solid arrow denotes a possible path whereas a dashed arrow indicates a forbidden path. OHC is the acronym for online health community.

Figure 3: POMDP Diagram

Following the notation of Rabiner et al. (1986, 1989), the proposed POMDP model is a combination of HMM and a probability adjustment process. It consists of three main components and an additional probability recalculation process for the partially observed health condition (state): (1) the initial state distribution, π; (2) the state transition probability distribution, A; (3) the observed outcome probability distribution, B; and (4) the recalculation of transition probability distribution, A', if the state is observed at the given period. For convenience, we use a compact notation for the overall model: $\lambda = (A/A', B, \pi)$. With this model specification, we proceed to find the values of parameters in $\lambda = (A/A', B, \pi)$ to best explain the observed outcome sequence, or to maximize the probability P (outcome sequence | λ).

3.1.2 State Transition Matrix

We assume that there are n states that discretize health conditions from the lowest health state 1 to the highest health state n. A patient takes medications, receives treatments, and other therapies. Even with such external controls, one still observes that very often a patient's health condition changes quite drastically. As the changes in health state can be very random, we relax the assumption of random walk in a typical POMDP model. The state transition probability is defined as $A = \{a_{it}(s,m)\}$, where $a_{it}(s,m) = P\left(S_{it+1} = m \mid S_{it} = s\right)$, $1 \le s,m \le n$; and S_{it} denotes the state of patient *i* at time *t*. For each state *s*, we have $\sum_{m=1}^{n} a_{it}(s,m) = 1$ and $a_{it}(s,m) \le 1$.

As discussed earlier, a patient's health condition changes according to her reception of social support through her online communications and other activities. A continuous measurement of this propensity needs to be modeled into the probability transition matrix. In other words, a patient can move to a higher health state if the benefit she gets from the online healthcare community is greater than a certain threshold, whereas she will transit to a lower state if the aggregate social impact is lower than a low threshold value. Hence, the matrix is defined as an ordered Logit model:

$$a_{it}(s,n) = 1 - \frac{\exp(\overline{\omega}_{s\to n} - \beta_s X_{it} - \xi_i)}{1 + \exp(\overline{\omega}_{s\to n} - \beta_s X_{it} - \xi_i)};$$

$$a_{it}(s,n-1) = \frac{\exp(\overline{\omega}_{s\to n} - \beta_s X_{it} - \xi_i)}{1 + \exp(\overline{\omega}_{s\to n} - \beta_s X_{it} - \xi_i)} - \frac{\exp(\overline{\omega}_{s\to n-1} - \beta_s X_{it} - \xi_i)}{1 + \exp(\overline{\omega}_{s\to n-1} - \beta_s X_{it} - \xi_i)};$$

$$\cdots$$

$$a_{it}(s,s) = \frac{\exp(\overline{\omega}_{s\to s+1} - \beta_s X_{it} - \xi_i)}{1 + \exp(\overline{\omega}_{s\to s+1} - \beta_s X_{it} - \xi_i)} - \frac{\exp(\underline{\omega}_{s\to s-1} - \beta_s X_{it} - \xi_i)}{1 + \exp(\underline{\omega}_{s\to s-1} - \beta_s X_{it} - \xi_i)};$$

$$\cdots$$

$$a_{it}(s,2) = \frac{\exp(\underline{\omega}_{s\to 2} - \beta_s X_{it} - \xi_i)}{1 + \exp(\underline{\omega}_{s\to 2} - \beta_s X_{it} - \xi_i)} - \frac{\exp(\underline{\omega}_{s\to 1} - \beta_s X_{it} - \xi_i)}{1 + \exp(\underline{\omega}_{s\to 1} - \beta_s X_{it} - \xi_i)};$$

$$a_{it}(s,1) = \frac{\exp(\underline{\omega}_{s\to 1} - \beta_s X_{it} - \xi_i)}{1 + \exp(\underline{\omega}_{s\to 1} - \beta_s X_{it} - \xi_i)}.$$

Here, s is the current state. $\underline{\omega}_{s\to k}$ is the threshold for the current state s to transit to a lower state k ($k < s$) and $\overline{\omega}_{s\to k}$ is the threshold for the current state s to transit to a higher state k ($k > s$). For a given s, we have $\overline{\omega}_{s\to n} \geq \overline{\omega}_{s\to n-1} \geq \cdots \geq \overline{\omega}_{s\to s+1} \geq \underline{\omega}_{s\to s-1} \geq \cdots \geq \underline{\omega}_{s\to 1}$. X_{it} is a vector containing variables that have impact for patients to switch between states. β_s is a set of state dependent parameters. ξ_i represents the patients' specific random effect and accounts for individual unobserved heterogeneity. As shown in Figure 2, patients in the lowest state can move to any one of $(n-1)$ higher states or stay idle while the highest-state patient can either stay unchanged or move down to any of $(n-1)$ lower states. Patient's health condition lying in any other states has the probability to move either up or down or stay unchanged in the same state.

3.1.3 State Dependent Outcome

As shown in the existing literature, patients use the Internet to search for health-related content and to access virtual support networks (e.g., Lamberg 2003, McMullan 2006). With the development and implementation of new technology, online healthcare communities are shifting from email-based support groups to multi-functional virtual communities. Patients in such online healthcare community environments can communicate through various channels, such as posts, to gain cumulative knowledge and aggregated support. Moreover, a patient's online behavior is closely related to her health condition. Therefore, the number of posts a patient initiates and answers determines the measurement of her observed online activity in the online healthcare community. Following Singh et al. (2011), we model that the number of posts, a count variable, follows a negative binominal (NB) distribution for a given health condition state:

$$P\left(O_{it} \mid S_{it}=s\right) = f_s\left(O_{it} \mid Y_{it};\gamma_s,\theta_s^2\right) = \frac{\Gamma(O_{it}+\theta_s^{-2})}{(O_{it}!)\Gamma(\theta_s^{-2})}\left(\frac{\theta_s^{-2}}{\theta_s^{-2}+\exp(Y_{it}\gamma_s+\eta_i)}\right)^{\theta_s^{-2}}\left(\frac{\exp(Y_{it}\gamma_s+\eta_i)}{\theta_s^{-2}+\exp(Y_{it}\gamma_s+\eta_i)}\right)^{O_{it}}$$

where O_{it} is the number of posts for patient *i* at time period *t;* and θ_s is the state-dependent parameter to capture the possible over-dispersion in O_{it}. Y_{it} is the vector composing variables that have direct impact on the outcome for patient *i* at period *t;* γ_s is the vector containing state-dependent parameters; and

$\exp(\gamma_s Y_{it} + \eta_i)$ specifies the expected value of O_{it}, according NB distribution. The symbol η_i is the patient-specific random effect that accounts for patient's unobserved heterogeneity.

3.1.4 Adjustment for State Transition Probability with Observed Patient Health Condition

The state transition matrix *A* for patient *i* at time period *t* must be modified if her health state is to be observed. Recall that $a_{it}(s,m) = P(S_{it+1} = m \mid S_{it} = s)$ is an element of *A*. If we observe that $S_{it+1} = m'$, that is, at time period (*t* + 1), with certainty, patient *i* enters a health state of m', then the corresponding state transition probability $a_{it}(s,m)$ is replaced by

$$a_{it}{}'(s,m) = \begin{cases} 1, & \text{if } m = m'; \\ 0, & \text{if } m \neq m'. \end{cases}$$

This adjustment process is in line with that of Kaelbling et al. (1998). For the time period that has the observed state information, all states from the previous time period will enter state m' with a probability of 1, and for the next time period, the possible routes will be initiated only from this state m'.

3.1.5 Likelihood of an Observed Sequence of Outcomes

Consider an observed sequence of outcomes $O(i) = O_{i1}O_{i2}\cdots O_{iT}$ for patient *i,* and a sequence of states $S(i) = S_{i1}S_{i2}\cdots S_{iT}$. The conditional likelihood, on two random effect control variables η and ξ, which account for unobserved patient heterogeneity, is the sum over all possible paths, explicitly,

$$L(O(i) \mid \eta, \xi) = \sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\cdots\sum_{s_T=1}^{n} P(S_{i1} = s_1)\prod_{t=2}^{T} P(S_{it} = s_t \mid S_{it-1} = s_{t-1})\prod_{t=1}^{T} P(O_{it} \mid S_{it} = s_t),$$

where $s_t \in \{1, \cdots, n\}$ is the state in which a patient can possibly reside in time period *t*. The likelihood of patient *i* can be obtained by integrating over η and ξ:

$$L(O(i)) = \int_{\eta}\int_{\xi} L(O(i) \mid \eta, \xi)\, dH(\xi \mid \eta)\, dG(\eta),$$

where the probabilities *H* and *G* are evaluated non-parametrically, that is, their supports and corresponding probability masses are considered as model parameters to be estimated.

3.2 Data

3.2.1 Research Context

We collect the data from the Health 2.0 website, where patients can create profiles and interact with others. This virtual place provides registration forms for patients to share their medical history and disease details, as well as communication platforms. To provide direct help and better service, this website is constructed and archived by health problem and patients are routed to their target communities based on the type of their disease. Members are required to disclose their health condition at the time of registration, and are thereafter directed to the targeted community at every login. Each community is a closed environment, because it is thought that patients suffering from similar diseases will better understand each other and thus provide social support more efficiently. Although members in different communities can view each other's profiles, information access is limited and patients from different communities cannot leave comments or initiate threads on any forum other than their own. Therefore, the boundaries of this online healthcare community are already defined by the website structure.

In the community that serves people with mental problems, patients, as usual, must first create their personal profiles. Similar to other online social networking sites, users need to provide basic information to introduce themselves. In the context of an online healthcare community, however, the "basic information" focuses more on their health, such as the type of their major problem (and perhaps a second or third health problem), the date of the first symptom, and the results of any diagnostic testing. There is also a medical application in the profile that allows patients to update recent health data and display these in chart form. Thus visualized, it is easy for patients to track their health history. To establish virtual relationships and communicate with other members in Health 2.0, patients can chat in the Forum and/or leave comments on profiles, thus establishing direct social ties.

Our data were sampled weekly for 4 months. The sample included patients' online activities on the website as well as their interactions with other members. A social network was constructed based on communications on patients' profiles. Since a comment could be in any content, and is a one-way communication representing a patient's willingness to interact with others, we can determine the degree of centrality of any given patient in the communication network. This denotes the extent to which individual patients are involved with others in the social network. This helps document how much benefit a patient receives from and contributes to the website. Table 1 reports social network statistics. Figure 3 shows a snapshot of a partial social network. It is a directed network in that an arrow indicates the direction of the communication. The size of the node represents the number of communications to or from that node; the bigger the node, the more connections. In other words, the size of a node can be viewed as a proxy for the visibility of the patient in the network.

Table 1: Social Network Statistics

Variable	Mean	St Dev	Min	Max
In-Degree Centrality	8.483	5.151	0	24
Degree Centrality	16.481	8.866	0	43

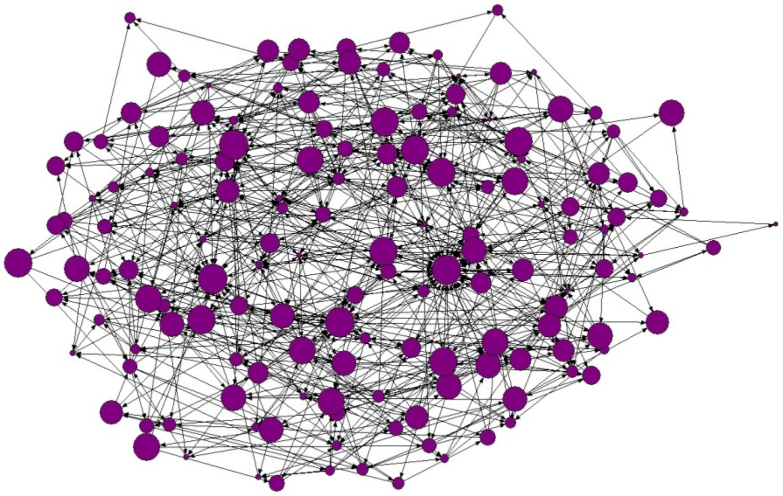


Figure 4: Patients' Social Network

3.2.2 Data Description

Like other social network websites, there are several ways for members to participate in the community. Each patient has her own profile, and each profile has an indicator on which the volume and quality of information on the website is identified and controlled. There are four levels of data quality, indicated by

0 to 3 stars. If there is only basic membership information with no health data, the profile receives no stars. One star is assigned to a patient who completes her profile with biographical and condition history information. Another star is added if the patient updates her treatments, symptoms, and mood maps for three months. Patients are also asked to provide names of prescription medications as well as significant supplements, equipment used, and other interventions. After completing four mood maps, patients receive a third star, indicating their profile is complete.

Each member in the mental online healthcare community can also use a search tool to easily find other patients who suffer from similar symptoms or experience similar treatments. The number of times a profile is displayed also indicates the visibility of the patient in the community. Once members find a valuable user—and think that person might have some information they need—the members can leave comments on the profile or send private messages. Hence, the number of communications reflects the quality of a patient's profile as well as her online interactions. As a reward for sharing outcome data, other patients can set a flag to express appreciation for the profile host's hospitality and generosity. Outside of the individual's profile level, the Forum is a social channel for every patient in the online healthcare community. There, members with general access can exchange general information, ask questions, seek help, provide useful information, or just chat. Each post is evaluated for usefulness by other patients. A utility score is added to the post if another reader finds it helpful.

Table 2: Profile Data Description

Variable	Mean	St Dev	Min	Max
Data Quality	1.481	1.117	0	3
Forum Activity Posts	27.028	14.506	1	70
Forum Activity Useful Marks	8.991	5.225	0	29
Number of Profile Views	14.739	7.922	0	43
Number of "Thank you"	4.255	2.724	0	15
Number of Comments	4.748	2.803	0	16

We collected both levels of patients' activities in the online healthcare community: patients' Forum activities statistics (the total number of posts is the aggregate number of conversation threads, including topic initiations and replies), helpfulness marks (other patients reward the post by marking it as helpful), as well as their profile activities. Data characteristics are shown in Table 2.

3.2.3 Partially Observed Health States

The health-related information that patients upload to their profiles and share with other patients in the mood community includes their instant mood, functionality level, overall distress level, and detailed distress components, treatments, symptoms, and counseling. Different from instant mood inputs, patients need to take a weekly multi-point survey to be diagnosed so as to obtain the assessments for their health condition measurements such as functionality level defined by this online healthcare community. For example, the online survey contains detailed questions about symptoms such as sleep quality, headache severity, problems concentrating, stomach pain, nervousness, hopelessness, and treatments such as drug dosage. As functionality level is a more comprehensive measure, we discretize it and use it to operationalize the variable for health condition (state) in our model. It can be overwhelming for patients to fill out such a detailed survey every week. The lack of effort from patients results in missing information. On average we observe in our dataset about 11% of patients' health conditions.

3.3 Variable Description

3.3.1 Social Support Measurements

As discussed in Section 2, there are four forms of social supports. Since we focus on online activities in this research, only three of them are considered: informational support, emotional support, and companionship. We followed the coding scheme proposed by Bambina (2007), and the details are provided in Table 3a.

The social supports are extracted from forum discussions. Different from the comments on the user profile, the Forum is a place commonly used for various kinds of social interactions, allowing richer insight into the experiences and needs of individuals affected by mental problems. Focusing on the patients who participated (and not those who just "lurked") in the Forum, we used LingPipe[1] to conduct a

---

[1] Alias-i. 2008. LingPipe 3.8.2. http://alias-i.com/lingpipe

semantic analysis on the forum threads. There were 5,192 topics initiated in the Forum and 371,562 posts made during our data collection period. For each post, a number was returned to indicate the probability that this post belonged in a certain category. Because patients tend to provide multiple pieces of information in each post, it would be improper to classify a post into one category only. Therefore, we assigned three probabilities (adding up to 1) to each post, corresponding to the topics addressed. With this classification scheme, we interpreted the probability as the amount of social support that a patient is offered or asks for. Table 3b gives the statistical details.

Table 3a: Social Support Coding Scheme

Support Categories	Support Subcategories
Informational Support	Advice Referral Teaching Information broadcasting/seeking Personal experience
Emotional Support	Understanding/empathy Encouragement Affirmation/validation Sympathy Caring/concern
Companionship	Chatting Humor/teasing Groupness

Table 3b: Social Support Statistics

Variable	Mean	Median	StDev	Min	Max
Informational support weekly	1.843	1.680	1.063	0.19	7.92
Emotional support weekly	0.712	0.560	0.578	0	3.68
Companionship weekly	0.542	0.420	0.442	0	2.8

3.3.2 Measurement of Variables That Impact Patients' Health Conditions

We also conducted analysis of variables that may affect patients' health condition dynamics, and these constitute vector X_{it}. For example, the total number of posts by patient *i* in time period *t*, called "posts," could indicate a change in the patient's health condition. Although the purpose of joining an online healthcare community is to get social support and other needs, a patient's health condition may limit her time online. For example, the most valuable information from individuals who suffer from mental problems—whose mental conditions and physical capabilities can change dramatically and periodically—may not appear in their newest posts. In that case it might only be possible to get the most salient information by digging into their historical messages. Another variable involves recognition and appreciation from other patients, which can help a patient to feel capable and valuable. This is measured by "views," which shows the total number of times that a patient's profile has been checked by other community members. If members find a profile particularly valuable, "thank you" is the simple way of showing appreciation for sharing information, while "comments" involve more detailed communications. Not all posts contribute equally; some might contain trivial information while others tend to be more useful. Therefore, "usefulness" is used to measure the value of posts, as assessed by other patients. Each patient can vote only once, for any post except her own.

3.3.3 Measurement of Variables That Directly Impact Patients' Online Behavior

The quality of a profile indicates a patient's level of concern about her disease, and therefore could directly impact her online activities. As four or more mood maps provide an accurate history of disease progress, it is important for not only sharing with the community but also working with the clinician. For this reason, we define "information quality" to measure a patient's data quality. There are other characteristics that could have a direct impact on patients' online behaviors. It is possible that female patients are more active in the online healthcare community, and this could lead to a different behavior pattern for their observed online behavior given a certain health state. For example, McPherson (2001) finds that gender is significant in predicting communication patterns. Hence, we include a "gender" variable in our model to control for this possibility. The date the patient joined the online healthcare community (membership), the frequency with which the patient updates his profile (update), and the willingness to communicate with other patients (out-degree) describe a patient's direct perception of the online healthcare community and her attitude toward it. Thus, these variables are proxies used to capture the patient's activity pattern.

Last, we use a patient's instant online activity, the number of new posts she initiates or replies in the period, to describe her state-dependent outcome. The change of this measure could result from a change in the patient's health condition. In other words, the social support she has is insufficient and she needs to communicate in the online healthcare community to fulfill her social demand. The more posts she contributes, the more likely her health condition has changed. Note, this does not imply whether a patient's health condition has deteriorated or improved. It suggests only that a patient's online activity relates to her current health state. The variable summary is presented in Table 4.

Table 4: Variable Description

	Variable Name	Description
Data Quality	*info quality*	The number of stars patient *i* receives for her profile
Activity Characteristics	*posts*	Cumulative number of posts on Forum by patient *i*
	new posts	The number of posts by patient *i* at time *t*
	usefulness	Total number of usefulness rated by other patients for Forum posts
	views	Number of views for patient profile
	thank you	Votes for shared information
	comments	Total number of comments left for this patient's profile
	update	1 if profile has been updated this period; 0 otherwise
	membership	Number of days staying in this forum
Social Support	*emo. support*	Total amount of Emotional Support by patient *i*
	info. support	Total amount of Informational Support by patient *i*
	companionship	Total amount of Companionship by patient *i*

3.4 Estimation

We started with a latent class model to estimate the initial distribution for the latent health state, and then used the maximum likelihood method to estimate the model parameters. To control for patients' heterogeneity, modeled by η and ξ, we followed the approach by Heckman and Singer (1984). The approximation process for the underlying unknown probability distribution was evaluated by finite sampled supporting points associated with probability mass distributions. After rescaling η and ξ by two parameters C_η and C_ξ, respectively, we set the boundary for each of the random effect variables to be between 0 and 1 (Kennedy 2003). We used the BFGS Netwon-Raphson algorithm (Whittaker and

Robinson 1967) to maximize the log-likelihood. The number of states *n* was chosen by the selection criteria derived by Greene and Hensher (2003), known as Bayesian Information Criterion (BIC):

$$BIC = \ln L - k \times \ln P / 2$$

where *P* is the sample size (the number of patients), *L* is the likelihood of the model, and *k* is the number of parameters to be estimated. The goal of the model selection process was to choose the model with a probability that approached one as the sample size got larger (Anderson et al. 1998). The results are shown in Table 5. Our simulation indicates that the three-state POMDP outperforms other models. It is noteworthy that although our model categorizes patients' health conditions into three health states, it does not necessarily mean that patients in State 1, the bad state, would prefer suicide. These health states describe patients' capability of balancing themselves with the environment they are in. And this ability is a combination of patients' physical and mental health.

Table 5: Selecting the Number of States

Number of States	Log-Likelihood	Variables	BIC
1	-14520.4	14	-14582.9
2	-14359.8	30	-14493.7
3	**-13872.1**	**48**	**-14086.3**
4	-14106.7	72	-14428.1

3.5 Results

In this section, we report the results from the POMDP model with three health states (bad, fair, and good). The initial state distribution probability, (0.7756, 0.15026, 0.07414), was obtained from a latent class model with three classes. The estimated parameters are presented in Table 6, where the corresponding standard errors are shown in parentheses. We first explain the various impacts of variables that would affect patients' directly observed activities in the online healthcare community, the results corresponding to discussions in Section 3.2. We then address various features that would impact patients' willingness to participate in an online community. In particular, as the key focus in our study, we use our empirical evidence to explain the discussions in Section 3.1.1 in detail and discuss the hypotheses we raised in Chapter 2 Section 2.1.

Table 6: Estimated Parameters for the Three-State POMDP [2]

Parameter	**State 1 (bad)**		**State 2 (fair)**		**State 3 (good)**	
θ Dispersion	0.8169^{***}	(0.1363)	0.4385^{***}	(0.0210)	0.5279^{**}	(0.2154)
Variables Impacting State Transition						
β_1 [views]	0.8430^{***}	(0.1096)	1.0405^{***}	(0.2108)	0.7635^{***}	(0.1567)
β_2 [thank you]	1.5249^{***}	(0.1785)	0.7661^{***}	(0.1199)	2.1804^{***}	(0.3254)
β_3 [comments]	3.0600^{***}	(0.1136)	0.7956^{***}	(0.2167)	1.5958^{**}	(0.7103)
β_4 [usefulness]	0.9838^{***}	(0.3457)	0.3796^{***}	(0.0982)	1.5081^{***}	(0.1556)
β_5 [in-degree]	-1.1192^{***}	(0.2064)	0.2119	(0.1784)	-0.7255^{***}	(0.2179)
β_6 [info. support]	0.7212^{***}	(0.1589)	0.6503^{***}	(0.1360)	0.6324^{***}	(0.1107)
β_7 [emo. support]	0.6605^{***}	(0.1665)	0.8445^{***}	(0.2069)	0.9798^{***}	(0.2429)
β_8 [posts]	2.3836^{***}	(0.1386)	1.6762^{***}	(0.1898)	4.5214^{***}	(0.4337)
Thresholds						
State 1 (bad)			0.9351^{***}	(0.1874)	2.8495^{***}	(0.6592)
State 2 (fair)	-2.5190^{***}	(0.4128)			2.4762^{***}	(0.3795)
State 3 (good)	-2.9527^{***}	(0.5882)	-1.0021^{***}	(0.2011)		
Variables Impacting State Dependent Outcome						
γ_0 [constant]	0.9639^{***}	(0.2765)	0.4837^{***}	(0.1273)	0.2152^{***}	(0.0817)
γ_1 [gender]	0.8302^{***}	(0.1724)	0.7594	(0.6781)	0.5737^{***}	(0.2162)
γ_2 [info quality]	0.2487^{***}	(0.0542)	0.4150	(0.3648)	0.1439^{**}	(0.0641)
γ_3 [membership]	0.9483^{***}	(0.0697)	0.5854	(0.7552)	1.0954^{*}	(0.6408)
γ_4 [out-degree]	-0.8281^{***}	(0.0564)	-0.6042^{*}	(0.3935)	-0.8942^{***}	(0.1198)
γ_5 [update]	-0.9178^{***}	(0.0440)	-0.9785^{**}	(0.2489)	0.2106^{***}	(0.0213)

Unobserved Heterogeneity (η , ξ)

$C_\eta = -0.209, C_\xi = -0.151$

	$\eta_1 = 0$	$\eta_2 = 0.3147$	$\eta_3 = 0.5289$	$\eta_4 = 1$
Probability $G(\eta)$	0.0612	0.3474	0.4011	0.1903
Conditional Distribution: $H(\xi \mid \eta)$				
$\xi_1 = 0$	0.9513	0.2411	0.1288	0.9331
$\xi_2 = 0.4135$	0.0277	0.5508	0.1875	0.0423
$\xi_3 = 1$	0.0210	0.2081	0.6837	0.0246

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

---

[2] Variables were checked for the problem of multicollinearity. Rescaling was performed to remove the potential correlation: views, thank you, comments, usefulness, in-degree centrality, and out-degree centrality were scaled down by a factor of 100; informational support was scaled down by a factor of 10; posts and membership were scaled down by a factor of 1000.

3.5.1 Hypothesis Test

As shown in Table 6, the estimated parameters for the effects of both informational and emotional supports are positive and significant across all three states. Hence, both Hypotheses I.1 and I.2 are supported. To test Hypothesis I.3, for a given state, we calculate the difference between the parameters of emotional and informational support, and the corresponding standard error. Note that since the covariate for informational support is scaled down by a factor of 10, the parameter β_6 measures the effect of 10 units of informational support. β_7 measures the effect of 1 unit of emotional support. For state 1 (bad), we compare β_7 with $\beta_6 / 2$ (or the effect of 5 units of informational support). The results are reported in the online supplement. They are all significant at 1% level. Hence, we conclude that Hypothesis I.3A is supported, and Hypothesis I.3B is rejected. That is, emotional support is significantly more effective in helping mental patients to progress to a better state. For Hypothesis I.4, we compare the estimated parameters across the states for informational or emotional support. We find that the effect of emotional support increases from bad to fair, and from fair to good state. Informational support is more effective for patients in a bad state. However, there is no significant difference between fair and good state. Therefore, Hypothesis 4 is partially supported.

3.5.2 State-Dependent Outcome

The parameters for state-dependent outcomes describe the variables that affect a patient's activities in online healthcare community at a given health state. It is interesting to note that—as indicated by the state-dependent constants that give the intrinsic propensity to contribute—the patients turn to create less posts as they progress to better health state. In the online supplement, we calculate the expected intrinsic number of new posts which are respectively 2.34 for bad state, 1.46 for fair state, and 1.12 for good state. These numbers are statistically different. The patients in a bad state want to learn more about their disease and hence have relatively more problems or questions to ask than those in fair or good state.

As shown in Table 6, women participated more actively than men in the online healthcare community across all health states. Female patients tended to post more when they were in either the worst (3.05 more posts than the intrinsic number) or best condition (additional 0.86 posts). This may be due to the fact that women are more sensitive to changes in their emotional and physiological states (Hunt et al. 1981) and, hence, more willing to express themselves and their emotions when they are in extreme conditions (bad or good). This finding is consistent with prior studies (e.g., Hunt et al. 1981, Mechanic et al. 1978) that show a sex differential in admitting to certain problems.

A patient with good information quality or who updates her disease records regularly is one who keeps close track of her disease progression. She prefers to seek out social supports, find out underlying reasons, and try to improve her condition. This is confirmed by the significant and positive coefficients in our results. The duration of membership measures the attitude commitment to online healthcare communities. Our results show that a member with a longer tenure, when in good or bad condition, tended to contribute more to an online healthcare community than a newcomer. The significant and negative coefficients for out-degree centrality for all states suggest that a patient is more interested in talking with her own favored cluster of patients and less likely to participate in community-based communications.

3.5.3 State Transitions and Baseline Results

The thresholds provide the intrinsic propensity to transition from one state to another. As we allow patients to "jump" among different states, these thresholds ensure that moving involves some positive boundary requirements. The intrinsic probabilities[3] to transit among states are shown in Table 7.

Table 7: Intrinsic Transition Matrix

	Bad	Fair	Good
Bad	0.7279	0.2199	0.0523
Fair	0.0781	0.8477	0.0742
Good	0.0521	0.2264	0.7215

Although a patient's health state could change dramatically (even to the point of jumping to a nonadjacent state in our model), our results showed that patients are indeed relatively stable in their health

---

[3] They are evaluated with the estimated threshold values and random effects. The values of all variables were set to zero.

states. Different from mood changes, health status describes a patient's physical and mental ability and the variation is minimized by medication control. The stickiness in the current state could result from the effect of medical treatments that patients received for their mental disease. As for mood problems, medication is not always recommended for those with mild depression because the risks outweigh the benefits.[4] In our dataset, an average of 35% of patients were taking medical treatments. However, without help from external resources (e.g., various services provided in the online healthcare community), a patient has a lower probability of improving her health condition to a better level. Without participation in an online healthcare community, a patient is more likely to stay at her current health state or get worse.

3.5.4 Factors Influencing Patients' Health Transition Matrix

As our primary objective was to determine the helpfulness of online healthcare communities in improving patients' health condition, we provided detailed discussions about the variables that affect patients' health state and consequently influence their behavior in online healthcare communities. In what follows, we categorize these variables into three groups. The transition probabilities are evaluated with the average value of the focal variable and the values of the other variables set at zero, and compared with the intrinsic transition probabilities.

3.5.4.1 Impact of Social Supports on Health Condition

Table 8 shows the change of transition probabilities, due to informational support, from the intrinsic ones in Table 7 (also shown in the parentheses). By communicating with other members, a patient is more likely to get useful information and better understand her health conditions. Along with information about medical terms and symptom descriptions, personal advice and referrals make the communication more valuable. The firsthand experience available on the online healthcare community helps patients muster the strength to fight their disease; it is also a place to find guidance for self-management. All of this helps to increase the probability that patients will transit to a better health state. For example, compared with intrinsic propensity transition, the probability of a patient in a bad state moving to a fair state is increased

---

[4] For example, http://www.womens-healthcare.org/mental-health/depression.html

by 3.21%. With an increase of probability of 2.28%, a patient already at a good health status is more likely to stay in good condition when she receives information support. Our results also indicate that the possibility of moving down to a worse health condition decreases because of informational support.

Emotional support has a significant influence on patients in different states. As shown in Table 9, we observe the same pattern seen with informational support: The benefits of emotional support are significant and positive in all three states. In other words, it increases the probability that patients move to a better health condition. Many studies have found that emotional support plays a critical role in a patient's outcome. For example, in a study of heart failure, emotional support was found to have significant association with risk for heart disease (Krumhole et al. 1998). We also find evidence supporting the importance of emotional support. Patients in a bad state had a 8.94% higher possibility of moving to a fair state and a 10.22% lower possibility of staying in a bad state. A patient who was already in a good state was shown to be more likely to stay in good condition with a 11.73% higher probability.

Table 8: Change in Transition Probability – Informational Support

	Bad	Fair	Good
Bad	-0.0271 (0.7279)	0.0321 (0.2199)	-0.0050 (0.0523)
Fair	-0.0082 (0.0781)	-0.0005 (0.8477)	0.0087 (0.0742)
Good	-0.0061 (0.0521)	-0.0167 (0.2264)	0.0228 (0.7215)

Table 9: Change in Transition Probability – Emotional Support

	Bad	Fair	Good
Bad	-0.1022 (0.7279)	0.0894 (0.2199)	0.0127 (0.0523)
Fair	-0.0337 (0.0781)	-0.0196 (0.8477)	0.0534 (0.0742)
Good	-0.0258 (0.0521)	-0.0914 (0.2264)	0.1173 (0.7215)

Severe disease affects patients and changes their everyday activities. Researchers in psychosocial and social science have examined social supports in various contexts. Such work includes studying patients' need for emotional support (e.g., Slevin et al. 1996) and emotional and informational support for patients' relatives (Eriksson et al. 2000). The requirements for such social support change according to the magnitude and time in need. Hypothesis 3 indicates that the same unit of emotional support is 5 to 10 times more effective than that of informational support. Tables 8 and 9 show that emotional support is overall more influential in changing patients' conditions, although patients receive more units of informational support in this community.

3.5.4.2 Impact of Self-Verification and Identification on Health Condition

Multiple measurements can be used to evaluate how well patients communicate with other community members, and how personal images are built in such a virtual world. By searching for similar patients with certain criteria, a patient can learn more from those members by viewing their detailed profiles. Therefore, the number of times that her profile is viewed indicates how visible a patient is in this community. The results shown in Table 10 confirmed this theory. The coefficient for profile views was positively significant and therefore increased the probability that a patient would move to a better health state. Take a patient in a bad state as an example. The probability of moving to a fair state in the next time period increases by 3.07%, while there is a decrease of 2.53% in the probability of staying in the same state for the next period.

Table 10: Probability Change – views

	Bad	Fair	Good
Bad	-0.0253	0.0307	-0.0054
Fair	-0.0104	-0.0009	0.0113
Good	-0.0059	-0.0161	0.0220

A large number of "thank you" votes indicated the quality of a patient's data. It not only confirmed the patient's effort in disease self-management but also made her feel appreciation for helping others. This satisfaction can influence a patient's ability to move among different health states. As shown in Table 11, we found that patients in a bad state had a 2.1% increase in the probability of moving to a fair state and a 1.31% decrease in the probability of staying in a bad state. Patients in a good state also benefited from confirmation and encouragement, and therefore had a higher probability of staying well.

The measurement for "comments" was intended to signal patients' profile quality. As shown in Table 6, the number of comments on a patient's profile had a significant and positive impact in all health states. This may result from the profile owner being encouraged by recognition and care from other patients, thus increasing her probability of feeling better. Table 12 shows that even a patient in a bad health condition who received an average number of comments had a higher probability of moving to a better condition (a 3.41% increase to a fair state). If she was already in good condition, the possibility of staying well increased 1.49%, as compared to when this recognition was absent.

Table 11: Probability Change – thank you

	Bad	Fair	Good
Bad	-0.0131	0.0210	-0.0080
Fair	-0.0023	0.0000	0.0023
Good	-0.0050	-0.0132	0.0182

Table 12: Probability Change – comments

	Bad	Fair	Good
Bad	-0.0297	0.0341	-0.0044
Fair	-0.0027	0.0000	0.0026
Good	-0.0043	-0.0107	0.0149

3.5.4.3 Impact of Displaying Social Competence on Health Condition

The number of posts and the helpfulness of those posts helped to determine patients' social value in this online healthcare community. The number of "posts" that a patient created was an indicator of her attitude in facing the disease and her aggregated knowledge of the disease, which may contain valuable information and experience for others patients. The "usefulness" variable measured the effectiveness of her posts. Tables 13 and 14 depict the impact of such activities on patients' health conditions.

Table 13: Probability Change – Posts

	Bad	Fair	Good
Bad	-0.0129	0.0209	-0.0080
Fair	-0.0032	0.0000	0.0032
Good	-0.0063	-0.0175	0.0238

Table 14: Probability Change – Usefulness

	Bad	Fair	Good
Bad	-0.0179	0.0248	-0.0070
Fair	-0.0024	0.0000	0.0024
Good	-0.0069	-0.0194	0.0264

A patient in a bad health state could create posts in the online healthcare community to seek help. This could help her release unhappiness and pressure and receive advice about her next move, and hence prevent her from falling into a worse condition. As shown in Table 13, there was a 1.29% decrease in her probability of staying in a bad state, and there was a 1.75% smaller probability that a patient in a good state would move to a fair state. The recognition and reward for competence ("usefulness") also helped patients not to get worse. Table 14 shows that appreciation gave patients in a bad state a 2.48% greater chance of moving to a fair state. A patient already in good condition increased her probability of staying well by 2.64%.

3.5.5 Posterior Analysis

For this analysis, we applied the filtering approach proposed by Hamilton (1989) to recover patients' unobserved health conditions across time periods. This approach used the information until time *t* and calculated the posterior probability for a patient in a given state. This allowed a patient to be classified, in any given time period, into a health state according to the posterior probability calculation. As observed in Figure 5,[5] about 45% to 50% of patients were in a bad health condition, and 35% to 40% of patients were in a fair state over the time.

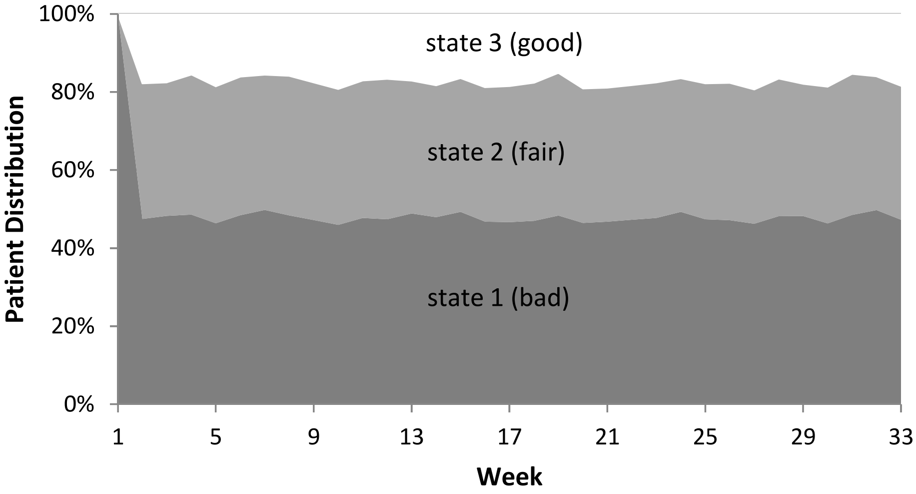


Figure 5: Posterior analysis for patient distribution

In Figure 6, we plotted two examples of individual patient behavior. There was no unique pattern: One remained at the same level and was more or less stable, while the other fluctuated a lot among states. Since the observed information on health states from Week 17 was not used to calibrate the model, the results of posterior analysis for Week 17 and beyond were purely predicted. Figure 5 shows that our POMDP model was very accurate in predicting patients' health conditions. For all the patients in our dataset, the overall accuracy was 91.3%. Therefore, our result shows this is a very effective way for patients and healthcare providers to recover missing or unavailable information.

---

[5] Figures 5 and 6 plot the results of posterior analysis on functionality state for 33 weeks. The parameters in our POMDP model are, however, estimated using the data of the first 16 weeks, due to computational complexity.

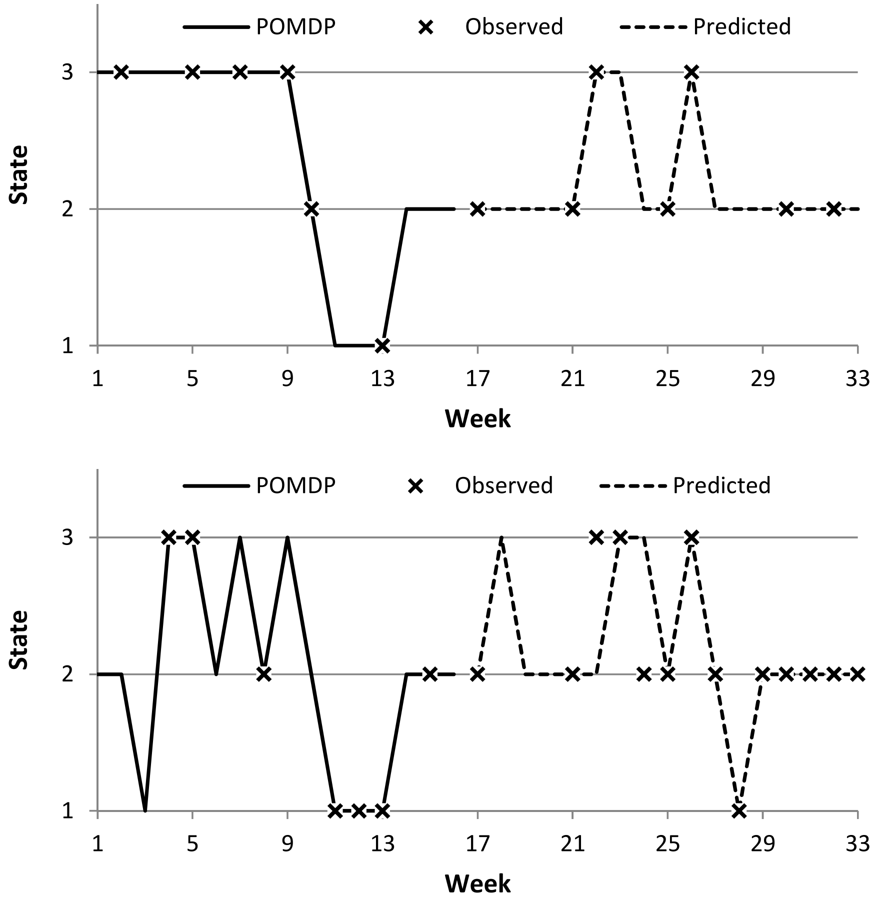


Figure 6: Posterior Analysis with Partially Observed Health State

3.5.6 Robustness Check

Various attempts were made to check the robustness of our results. First, we checked the robustness of observed health conditions. In our dataset, the observed functionality level was scaled from 0 to 100. We performed various classifiers to categorize (or discretize) functionality levels. The alternative trials did not produce qualitatively different results, and the likelihood does not exceed the result presented earlier. Next, the online healthcare community routinely asked that patients update their profiles weekly; hence, we chose a week as the length of time between observations. To account for potential concern about this timeframe, we used 2 weeks as the time period and re-conducted the analysis. The results showed no significant difference. Finally, we considered other controls for the analysis. For example, age is correlated with patients' online behavior, and therefore it could have affected state-dependent outcomes.

The results for controlling such effects did not produce any significant differences from the results presented earlier.

3.5.7 Limitations

There are several limitations in our study, and we do not want to overstate our findings. First, social supports were shown to have significant impact on patients' health condition changes, but our dataset did not allow us to distinguish between active and passive social supports. We were unable to separate "providing" from "seeking" social supports and, hence, cannot precisely measure the impact of each. Second, we used the number of posts as the measure for patients' online healthcare community outcomes. It is very possible, however, that patients possess different preferences in their online activities. For example, some patients may spend more time observing, rather than actively participating in others' communications. It would be helpful to incorporate more measures of patients' online behavior patterns. The third limitation is that we only considered direct communications among patients. Social supports can, however, also be transferred by word-of-mouth via common friends. Therefore, including other network measures could potentially shed more light on how online healthcare community can benefit patients.

3.6 Conclusion

In this paper, we developed a Partially Observed Markov Decision Process model to study patients' dynamic health outcomes. The POMDP model was estimated by a maximum likelihood procedure. Three health condition states were identified to best explain the data. Our results offered several insights into the driving forces behind patients' health condition changes and, hence, demonstrated the usefulness and value of online healthcare communities.

The main contributions of this work are (1) our proposed framework to measure how helpful an online social network can be and (2) new evidence of the efficiencies and benefits of such online services. Growing participation in online healthcare communities is well documented. While research on how these forms of social networking work and how well they serve patients' needs is underway, the important question of how social supports change patients' health outcomes remains unanswered (Lamberg 2003).

By investigating the online activities of patients suffering from mental disease, our results revealed the benefits and advantages of online healthcare communities in helping patients improve their health conditions.

Our procedure to identify a patient's unseen health condition distinguishes our model from other social networking studies on healthcare. Extending the sociological research on patient behavior, we used the POMDP model to explain patients' health condition changes with respect to social supports they received online. Patients are actively involved in disease self-management. Their participation in online discussions enabled them to learn from other patients, and enjoy a partial prevention effect that reduced the possibility of their condition deteriorating. These findings can be used to encourage users who are passively participating in online healthcare communities to reduce "lurking" behaviors. That would result in online healthcare communities becoming a place where social supports are contributed by an even more diverse membership. The investigation of transition distributions for various effects revealed that such communications were more effective for patients in good health conditions. We showed that a healthier patient gains more benefits from online healthcare community and has a higher probability of staying well.

We found, in our empirical analysis, that informational support was the most useful of all available online social supports. It was the main draw for patients and their family members to join an online healthcare community. However, its impact on changing patients' health condition was relatively lower than that of emotional support. Our results also indicated that recognition and positive feedback from other patients helped to improve an individual's health condition and encouraged patients to play their social roles competently. This effect was enhanced in the "sticky" dormant states.

Finally, our work is just a first step. It revealed the importance of studying the role of information systems in the context of healthcare. Fichman and his colleagues (2011) agree: The intersection of social media and healthcare is a promising direction for study. Our work combined theoretical modeling and data validation and exhibited the quantitative results. These findings signify a potential direction for healthcare reform, and suggest the effective and encouraging consequences of incorporating patients' self-help efforts into health management. These possibilities are tantalizing for both information systems and healthcare practices research.

Chapter 4: Network Dynamics: How Can We Find Patients like Us?

4.1 Research Methodology

4.1.1 Research Context

To incorporate the earlier discussions into our research context, we have collected data from a health 2.0 website. This website holds an OHC and provides a social platform for patients to share personal health information, exchange medical experiences and knowledge, and learn from each other to improve their own health management. Mainly, there are two channels for individuals to find similar patients or exchange knowledge. The first channel is patients' front pages. At the time of registration to the website, patients need to disclose their disease conditions and create their own health profiles. Based on the category of disease, patients are assigned to the corresponding disease-based communities, where they can meet other patients with similar health problems. Personal webpages contain a health profile for each member, including personal and demographic information, health condition history, outcomes, symptoms, counseling or therapy, and detailed records of treatments. Patients can update their current health conditions at any time by answering a few questions about themselves. A chart display gives a direct interpretation of each patient's current condition and also contains the links to her previous input, to show her health condition history and changes. Such profiles are publicly available and shared with focal members within the community, but only partially observable by others outside the focal community. Meanwhile, patients can also find and communicate with similar members in the forum, where they discuss various topics, concerns about the disease, and treatments or experiences. This channel possesses functions similar to those in chat rooms but without access limitation to community members. The website provides a subscription tool that helps patients keep track of the disease progress of members they select and to exchange information with them. The tool allows patients to receive updates about profile information, treatment changes, or replied topics. From the subscriber and subscription lists, a patient can construct her friend network and gradually embed herself into the community.

4.1.2 Model

Social network data comprises complex structures including a set of people, their characteristics, and their pairwise relationships (Goodreau et al. 2009). The structure of such social relations is defined by both the distribution of individual characteristics and the dynamics of interactions. The ERG (Exponential Random Graph) model is perfectly suited to for our study. It estimates the coefficients and determines the impact of statistics features for a specific set of data. That is, the method compares and specifies the probability of the current network configuration with all alternative structures for the given set of ties and a set of individuals and their attributes. However, the cumulative model requires significant computing complexity. Our goal is to identify the coefficient matrix that maximizes the model likelihood. Therefore, following the assumption for the ERG model that the number of actors and their attributes are fixed, we proposed the degenerate statistical model to specify the probability of a set of ties *Y* given a set of patients and their individual pairwise attributes as:

$$P\left(Y_t = y \mid N \text{ actors}\right) = \frac{\exp\left(\sum_k \eta_k Z_{k,t-1}(y)\right)}{k(y)} \text{, where } k(y) = 1 + \exp\left(\sum_k \eta_k Z_{k,t-1}(y)\right).$$

Specifically, we consider a binary relationship: presence ($y_{ij} = 1$) or absence ($y_{ij} = 0$). y_{ij} is defined as the variable for a direct tie established between patient *i* and patient *j*. Therefore, Y_t represents a network with patients as the nodes and ties representing their communication (friends) relationships. $Z_k(y)$ is a vector of model covariates and η is the vector containing coefficients which determine the impact of these covariates for a given network and its corresponding features. These are unknown parameters to be estimated.

Furthermore, the friend network among patients is not static. Over time, patients find new members to communicate with and thus establish new relationships and remove unnecessary ties to reduce the volume of information to be processed. Thus, the configuration of the network changes as patients gradually develop their information management skills. Therefore, over a given period of time, the friend network will contain different configurations as ties are added and removed. Thus, given an observed configuration of network Y_t at time *t*, the probability for observing the network distribution for *T* periods is thus given by:

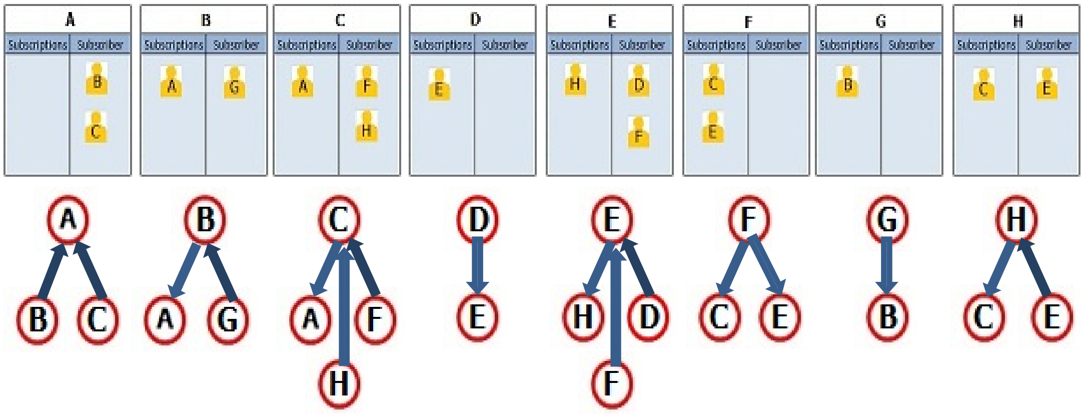
$$P\left(Y_1 Y_2 \cdots Y_T \mid N \text{ actors}\right) = \prod_t P\left(Y_t = y \mid N \text{ actors}\right)$$

In this actor-driven model, we assume that each patient has control over her outgoing ties y_{ij} and her characteristics Z_{ik} .

4.1.3 Data and Measures

The network considered in this study changes over time as patients update their decisions progressively. To subscribe to a patient's information is very easy. When a patient finds a piece of information that is helpful and wants to keep track of it, there is an icon next to the content, regardless of whether it is a patient's profile or forum discussion, thereby making it easy for community members to subscribe to it. To remove such subscription is also very straightforward, because there is a unsubscribe icon next to the subscribed content.

Subscriptions among patients and the network representations



The overall network

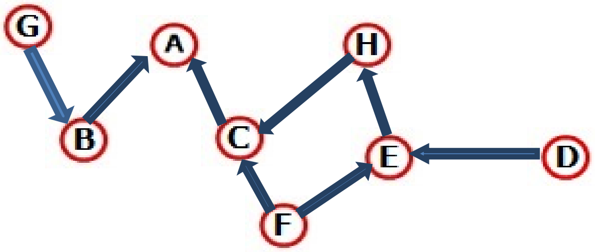


Figure 7. The construction of a patient subscription network

The subscription feature helps patients to establish a direct social network and manage relatively less information, thereby reducing information overload. On each patient's profile, there is a subscription tab and a subscriber tab for a patient to have an immediate view of the patients to whom she subscribes and who subscribe to her profile and other activities. On the subscription list, the patient receives notice of the latest information updates for those patients whom she cares about. She also has access to the list of other patients who have the same subscription decision she has made regarding other patients who are of interest. For example, if patient A subscribes to patient B's profile, on patient A's webpage, she has information about patient B's most recent activities, and who else subscribes to patient B. Others who have the same connections to patient B might potentially have common interests or similar concerns as patient A. Therefore, through the subscriber list of patient B, patient A can easily find similar patients or the information she needs. At the same time, patient A can also see who has subscribed to her, and review each of those patients' information to decide if there is a need for her to initiate a connection to them. Figure 7 gives an example of how the subscription network is constructed.

We have collected the data from a community where patients suffer from mental problems for four months starting in late 2009, for the reason that patients with mental diseases need more social support and are more active than the members in other communities. Note that the subscription network is the result of patients' selections of similar patients. This endorsement is a one-sided action, and patients can unsubscribe from others' profiles at any time they want. Under such circumstances, the social network is a directed graph, and the existence or absence of a tie indicates the connection relations among patients, whose connection relations change over time. Although the website provides a link on the friend's profile page for removing friends/subscriptions, the social implications of deleting friends can discourage users from doing so. Furthermore, little is gained by deleting friends. As a result, we observe the network changes with a much slower rate of relationship removal than relationship additions. Our dataset contains the subscription relationship records for 4908 patients and the connection changes that occurred in the four-month period. Due to computational complexity, we construct our major networks of 1322 patients whose dyadic records are first observed at the beginning of our data collection. Our study covers three time stamps (network layouts). The summary of the variable description and network statistics are shown in Tables 15 and 16.[6]

Table 15: Data Description and Statistics

Variable	Description	Mean	St. Dev	Min	Max
age	absolute value of age difference between (*i*, *j*)	14.3586	10.1597	0	42
gender	1 if (*i*, *j*) have the same gender	0.4991	0.5000	0	1
state	1 if (*i*, *j*) are from the same state	0.5005	0.5000	0	1
data quality	1 if patient's profile has a 3-star label	0.1581	0.3648	0	1
member days	the log transformed member days	162.5371	67.1107	28	302
same disease	1 if (*i*, *j*) are in the same category of disease	0.5003	0.5000	0	1
same symptom	number of shared symptoms	2.4984	1.7071	0	5
no. of drugs	number of drugs that patient is taking	11.9473	6.0766	2	29
no. of treatment	number of treatments (excluding drugs) that patient is taking	1.5149	1.1262	0	3
symptom_treatment	1 if (*i*, *j*) experience the same symptoms and take the same treatment	0.5121	0.4999	0	1
sympt_treat_s_c	number of same treatments patients share if they have the same symptom	4.6072	5.7009	0	17
symp_treat_d_c	number of different treatments patients share if they have the same symptom	3.0733	3.7578	0	11
symptom_drug	1 if (*i*, *j*) experience the same symptoms and take the same drug	0.4927	0.4999	0	1
sympt_drug_s_c	number of same drugs patients share if they have the same symptom	3.7902	4.8779	0	15
sympt_drug_d_c	number of different drugs patients share if they have the same symptom	2.9565	3.7323	0	11
post	number of posts	14.6601	8.0937	0	40
indegree	individual incoming degree centrality	0.0058	0.0046	0	0.0227
outdegree	individual outgoing degree centrality	0.0058	0.0045	0	0.0189
betweenness	individual between centrality	0.0020	0.0022	0	0.0208
closeness	individual closeness centrality	0.3044	0.0932	0	0.4424
structural dissimilarity	dissimilarity of (*i*, *j*)	0.3513	0.1423	0	1.0000
shortest path	shortest path between (*i*, *j*)	-	-	1	∞
clustering coefficient	CC considering only 1 neighborhood	0.0126	0.0185	0	0.3333
transitivity	number of transitive pairs	0.0021	0.0564	0	6
reciprocity	number of two-way subscriptions	0.5169	0.7088	0	2

Table 16: Net Change in Dynamics in Social Networks

	Profile Subscribers	Network	Edge	Removed	Added	Network Density
Network One	1322	1322	3041	-	-	0.0017
Network Two	2517	1322	8426	807	6192	0.0048
Network Three	3708	1322	19107	2082	12763	0.0109

---

[6] Shortest path, outdegree centralities, and closeness centralities are mean-centered. The average VIF is 3.39, and individual VIF is less than 9.49.

	Edge	Indegree				Outdegree			
		Mean	St.Dev	Min	Max	Mean	Std.	Min	Max
Network One	3041	2.3003	2.1269	0	13	2.3003	2.0750	0	11
Network Two	8426	6.3722	3.2801	0	18	6.3722	3.2568	0	19
Network Three	19107	14.4516	4.4833	2	30	14.4516	4.4962	2	29

4.1.3.1 Data on Patients' Profiles

Typically, a patient's profile contains her basic personal information, such as age, gender, location city and state. Hence, we keep track of these demographic and geographic and construct variables *age*, *gender*, and *state* by comparing the similarity between patient *i* and *j*. Meanwhile, health data contains information about a patient's primary condition, the time of diagnosis, detailed symptoms, medical and supplementary treatments history, and so forth. The patient's online activities are also recorded on the profile, including her member days in the community and the time of her last update. An overview of the patient's online social activities is also available. The overview gives not only the general picture of her past activities of communication and information exchanges but also access to these actions.

Besides baseline homophily variables, we further construct eight disease experience related variables to describe the information and knowledge exchange among patients based on their communication ties. It is worth noting that the first six measures are medical characteristics calculated at the dyadic level, while the last two measures focus on the communication connection initiator's health records. Considering a subscription tie (directed), the pair of patients who established a connection covariates *same disease*, *symptom_drug*, *symptom_treatment*, *sympt_drug_s_c*. *sympt_treat_s_c* and *same symptoms* are calculated by comparing patients' health condition similarities. The remaining two measures, *no. of treatments* and *no. of drugs,* are constructed based on the tie initiator's disease conditions.

4.1.3.2 Data on Patients' Subscription Friend Networks

On each patient's profile, there is a "Feeds" tab that helps the patient construct her narrowed-down friend network. By using this tool, a patient can subscribe to other patients' profiles in which she is interested and feels there is a need to keep contact with. Such interests include various situations such as the disease history, (which contains useful information for others to learn from), disease treatments, disease progress, (which gives the firsthand experience of individual outcomes), and the discussion topics that deliver meaningful information. Meanwhile, the profile owner can also see the list of patients who subscribe to her. Therefore, based on this subscription technique, we construct the patients' subscription friend network and keep track of each patient's network characteristics. The network is mapped by three timestamps for the configurations in the four-month period, and Table 16 explains the dynamics of the network.

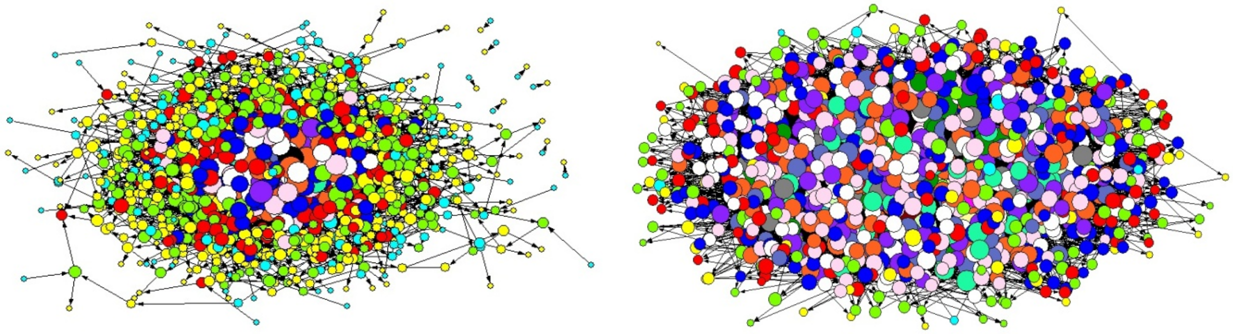
The network position covariates are constructed for each patient for a given network configuration. First, *indegree* and *outdegree* centralities count the number of incoming and outgoing ties, and the degree of *reciprocity* focuses on relations and is calculated as the number of ties that are involved in fully connected relations. The *shortest path* between patients is constructed as the number of connections for a pair of patients in the minimum distance. If a pair of nodes is not connected, then the shortest path between those two nodes is set to be infinite. The variable *being contacted*$_i$ indicates if patient *i* has received contact from patient *j* in the previous time. We further calculate the *betweenness* centrality based on Freeman's definition and also construct *closeness* centrality. The first and second largest degree (incoming ties and outgoing ties) in the network are used to calculate the variable *structure dissimilarity*. For the *clustering coefficient*, we only consider patients within 1- neighborhood scope in the network. Finally, *transitivity* is constructed as the ratio of the number of transitive triads in the existing network over the number of possible transitive triads.

4.1.3.3 Dependent Variable

As discussed earlier, individuals in the online healthcare community are searching for similar patients and sharing information and experiences. They will construct a subscription tie to others only if those health profiles are informative and contain some similarities to their own situation, from which they could learn. Under such circumstances, the initiation of a subscription tie from patient *i* to patient *j* indicates patient *i*'s belief of being similar. In other words, the subscription network is an indication for homophily. This endorsement is a one-sided action and patients can unsubscribe from others' profiles at any time they want. Hence, the social network we considered here is a directed graph, and the probability of a relational tie may increase as the characteristics of two individuals become more similar. In our model, the absence or presence of a tie is our dependent variable.

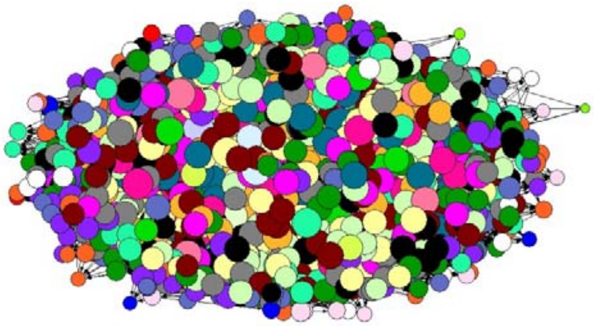
4.2 Results

There are 5,239,086 observations of subscription ties in our study of patients' network dynamics in the three observation periods. The network configuration change is depicted in Figure 8. The size of a node indicates its number of connections. The color of a node displays structural similarity: patients who are more structurally similar are represented by the same color. We use maximum likelihood procedure to estimate the coefficients to best fit our model in describing the network and its changes.



Network One

Network Two



Network Three

Figure 8. The evolution of subscription networks over time

Following the related literature, we adopt the assumption of homogenous classes of dyads and take a parsimonious approach. We assume that each instance is equiprobable (Goodreau et al. 2009). That is, a covariate's effect is weighted the same for all observations. To account for the potential heterogeneities associated with patient attributes, we control for the unobserved individual attribute-specific configurations in our estimation procedure. A variable for the average number of ties made by each patient is inserted as the instrument variable into our model. Furthermore, the concerns related to patients' positions and their communications may also arise since people outside the target group of members could impact on patients' connecting probability. For example, suppose that two patients have the same number of friends, and one of them has more friends outside the study group. If these outsiders have members in the group, the patient with more outside connections occupies a more important position in the network. To account for this possibility, dissimilarity is calculated based on the connections from the full subscription network and thus captures this indirect impact. We explicitly involve the lags of covariates in our model to ensure the causal relationship. We apply the random effect model (RE) because our short panel dataset has an extremely large number of entities (Torres-Reyna 2007), and the lack of the null case of similarity in the fixed effect (FE) model suggests that that RE should be used (Drury 2008). Meanwhile, RE also returns a higher likelihood. Serial correlation and heteroskedasticity are explicitly controlled in the estimation. Table 17 presents the estimation results.

4.2.1 Personal, Demographic, and Health Problem Similarities

Unlike the study on MySpace (Thelwall 2009), our result shows significant evidence of *age* homophily (at 10% significance level). This result is consistent with the finding by Suitor et al. (1995), suggesting that patients find it much easier to communicate with others of similar age and that the information exchanges are useful However, we find no evidence of *state* and *gender* homophily, confirming the diminishing explanation power of basic homophily. We observe that experiential similarity (i.e., similarities in disease or symptoms) is more important than structural similarity (i.e., similarity in age, gender, state). When patients are taking prescription drugs and receiving some additional treatments, they tend to participate more actively online. The negative significance of *no. of drugs* indicates the less tracking control for patients. Our data set contains a case where a patient takes 29 different types of drugs at the same time (six different drugs on average). This exacerbates a patient's burden of disease management and reduces her ability to identify the prominent information and process it effectively.

Intuitively, if patients are experiencing the same symptoms and use the same drugs, there is a higher incentive for them to keep track the other's progress and compare the effectiveness of such treatments. On the other hand, if they are facing similar problems but taking different drugs, it is difficult to make direct comparisons and learn from each other's experience. Our result supports this intuition. The negative correlation between symptoms and drugs (*sympt_drug_d_c*) indicates less incentive for patients to establish direct contacts when they are taking different approaches. Besides prescription drugs, the information about other treatments (*no. of treatment*) can also be useful for patients, as these treatments, such as supplements and individual therapy, are typically easy to deploy and less costly. Other control variables for patients' health records are found to have no significant contributions to the formation of subscription network. Our results explore evidence explaining personal, demographic, and health experience homophily to a certain extent, and therefore Hypothesis II.1A and B are partially supported.

4.2.2 Network Prominence Measures

Patients' visibility and their network positions have been found to have significant impact on the decision of building up connections. Our results show strong support for Hypothesis II.2. First, in our directed network, prestigious patients with large indegree centrality are particularly visible to the other patients in the network. Everything else being equal, these patients are much easier to be identified in the community and thus contacted. The positive and significant coefficient for this measure (*indegree*) thus confirms that a more prestigious patient is more likely to be sought. Next, because betweenness and closeness centralities are not uniquely decided by the patient's preference but also depend on others' connections configurations in the network, we therefore discuss implications without considering the marginal effects. Patients who have large betweenness centrality are more likely to be on the paths of other patients' communication. Therefore, from their position in the current network configuration, the cost for them to have extra ties to get additional useful information is outweighed by the efforts they need to expend in order to process the duplicated information. As expected, we obtain a negative and significant coefficient for the *betweenness* centrality. On the other hand, a patient with large closeness centrality is closer to all other patients in the network and can get information faster and, therefore, there is incentive for them to establish such connections and find similar patients (a positive significant coefficient for *closeness centrality*).

Table 17: Results of Random Effects Model

Variable Type	Variable Name	Coefficient	St. error
Activities in Different Channels	$\textit{post}_{i,t-1}$ – learning from forum	0.0129^{***}	(0.0004)
	$\textit{outdegree}_{i,t-1}$ – learning from profile	81.6413^{***}	(1.7786)
Patient's Visibility	$\textit{post}_{j,t-1}$	0.0008.	(0.0003)
Network Prominence Position	$\textit{betweenness}_{i,t-1}$ – "I am the middle man"	-44.3303^{***}	(0.9637)
	$\textit{closeness}_{i,t-1}$ – learning at faster rate	5.8484^{***}	(0.0503)
	$\textit{indgree}_{j,t-1}$ – visibility in the network	42.2618^{***}	(0.8833)
Network Structural Characteristics	$\textit{inversed shortest path}_{ij,t-1}$	0.5946^{***}	(0.0354)
	$\textit{inversed shortest path}_{ij,t-1}^{2}$	0.4019^{***}	(0.0418)
	$\textit{structural dissimilarity}_{ij,t-1}$	-0.5823^{***}	(0.0235)
	$\textit{structural dissimilarity}_{ij,t-1} \times \textit{inversed shortest path}_{ij,t-1}$	-2.6546^{***}	(0.0840)
	$\textit{clustering coefficient}_{ij,t-1}$	-0.0958^{***}	(0.0400)
	$\textit{clustering coefficient}_{ij,t-1} \times \textit{inversed shortest path}_{ij,t-1}$	-0.1591^{**}	(0.0581)
	$\textit{transitivity}_{ij,t-1}$	0.1797^{***}	(0.0241)
	$\textit{reciprocity}_{ij,t-1}$	-0.0505^{***}	(0.0018)
Individual Similarity	*age*	-0.0003.	(0.0001)
	gender	-0.0012	(0.0026)
	state	0.0036	(0.0026)
Health Condition Similarity	*same disease*	0.0031	(0.0026)
	same symptom	0.0003	(0.0007)
	$\textit{no. of drugs}_{i}$	-0.0061^{***}	(0.0008)
	$\textit{no. of treatment}_{i}$	0.0053^{***}	(0.0008)
	$\textit{symptom\_treatment}_{ij}$	-0.0040	(0.0027)
	$\textit{sympt\_treat\_d\_c}_{ij}$	-0.0015	(0.0021)
	$\textit{symptom\_drug}_{ij}$	-0.0030	(0.0027)
	$\textit{sympt\_drug\_d\_c}_{ij}$	-0.0055^{*}	(0.0020)
Profile Characteristics	$\textit{data quality}_{i}$	-0.0694^{***}	(0.0036)
	$\textit{member days}_{i}$	0.0450^{***}	(0.0034)
	$\textit{commenter}_{i,t-1}$ – leaving comments on other's profile	0.0327^{***}	(0.0039)
	$\textit{commenter}_{j,t-1}$ – received comments	-0.0039	(0.0037)
	$\textit{being contacted}_{i,t-1}$	0.0361^{***}	(0.0031)
Base Case of Similarity	*constant*	0.9883^{***}	(0.0208)

i = initiator, *j* = contactor

Significance code: 0 '***' 0.001'**' 0.01'*' 0.05'.' 0.1'_' 1

The positive and significant coefficient for *inversed shortest path* provides confirmative support for Hypothesis II.3. That is, the probability of creating a strong relationship (tie) from an initiator to the contactor decreases with the distance between them. The explanation could be very intuitive. If two patients are connected indirectly, it suggests there is a chance for them to be similar and thus creates incentive to do so. However, if they are apart by several degrees, i.e., they are linked through several other intermediates, the amount of information that must be processed in order to initiate the final connection could trigger the problem of information overload. This potential overload reduces the probability of such direct contact when the distance between the two patients grows. Furthermore, this effect of proximity on communication is found to be a nonlinear function (also see Monge and Contractor 2003). In other words, if the distance between two patients is increased twofold, the likelihood of communication is reduced by more than half. In our estimation, the results present strong evidence to support Hypotheses II.4 and II.5. Among network structural similarity characteristics, the variable *structural dissimilarity* is found to be negatively correlated with the probability of having a connection between initiator and contactor. Therefore, the more dissimilar two patients are, the less likely they will be to communicate and exchange medical knowledge. Furthermore, this heterophily effect is found to be enhanced by the increased path length (*structural dissimilarity* × *inversed shortest path*), as indicated by the negative and significant coefficient. However, Hypothesis II.6 is not supported. The clustering coefficient index (*clustering coefficient*) is significant but negative in our estimation result. It indicates that the subgroup is not growing to be more cohesive. The interaction between *clustering coefficient* and *inversed shortest path* does not increase the probability for two patients to communicate. This interesting result suggests a role similar to that of betweenness centrality in that patients who are close together have shared information and hence have reduced incentive to enhance such connections. At the same time, patients who are not nearby might provide useful and valuable information into this subgroup.

4.2.3 Other Interesting Findings

Our estimation yields some other interesting results. For example, patients who have multiple posts in a discussion forum indicate familiarity with a notable body of medical information and experience with their diseases. This social competence makes them particularly visible and attractive for other patients to learn from, and thus increases the probability of receiving a tie (a positive significant coefficient for *post*). Patients who “talk” (*post*) more in the community are relatively more active than others and therefore there is a higher chance for them to “go out” (*outdegree*) and make friends.

4.3 Implications and Discussions

In addition to the advantage of using the subscription feature to reduce information overload, there are also other reasons for the configurations that patients construct in such close-knit communication networks. In this section, we discuss issues related to privacy and content credibility for such online services and the implications of patients’ attitudes toward these online health informatics platforms based on our findings.

4.3.1 Trust

Trust is a social concept that has been identified as the action based on the belief of contingent actions of others (Sztompka 1999). Although trust in web-based networks is usually hard to measure due to limited information, namely social relations and user profiles, trust has been shown as an important component for the success of online communications (Piccoli and Ives 2003). Typically, for a system with a component of trust, members in the system will make some direct statements about others they trust (Golbeck 2009). These statements, such as an initiation of a direct communication relation to whom they trust, shapes their behaviors and thus forms a social network. Even though most of the key information that sociology studies use to measure trust is not available in online settings, we are able to infer several dimensions of trust operating on multiple levels from our empirical results based on patients’ communication ties.

In healthcare-related services, trust is a very important component. Trust may be more important in the OHC than in any other fields, because patients in the community need to reveal their medical problems and disclose their detailed health information. Confronted with the threat of being identified offline and encountering various subsequent difficulties, patients weigh their trust on two levels: to other patients and to the system. First, trust is shown to have a strong correlation with the similarity between two people (Zieler and Golbeck 2007). The more similar two people are, the more trust exists between them. The similarity of individuals' characteristics has been widely studied, and interpersonal attraction or baseline homophily, i.e., *state*, *age* and *gender*, has been shown to lead to the formation of friendship (Burgess and Wallin 1943, Newcomb 1961, Byren 1961, 1971). However, in our empirical results, we cannot find support for baseline homophily and therefore find no evidence suggesting that trust among patients is abstracted from this category. We further observe the lack of trust among patients, as supported by the negative and significant coefficient for *clustering coefficient*. Intuitively, patients in a cluster are connecting more intensely. Such dense relationships can foster trust among individuals. However, we find conscientious patients tend to depend on self-management and are less active in sharing their sensitive private information, indicating less trust to share their information with others. Being contacted or subscribed to shows trust to some extent, as the variable *being contacted* defines another dimension which can initiate trust. The fact that a patient in the OHC is contacted is an indication of the similarity between the initiator and contactor as well as a signal of trust. However, this positive effect (*being contacted*) could not overcome the trust problem. Meanwhile, the limited support for baseline homophily does not provide a sufficient basis to form initial trust. The variable *reciprocity* is found to have a negative and significant effect, indicating less likelihood of constructing two-way communications. This result, again, implies the lack of trust in the OHC.

Second, there is community-based trust in the OHC. The network under study posits that patients will communicate mood problems, share information, and exchange knowledge; and our results empirically confirm this community goal. As a component that defines trust, *commitment* describes the occurrence of trust when belief in others' behavior is used to make a commitment to a particular action (Sztompka 1999). The variable *member days* measures such commitment and identifies the signal that a person is linked to the organization (Sheldon 1971).

The above discussion suggests the low level of trust among patients in the OHC while confirming the community goals. Service providers need to take this issue into consideration when designing websites, since their goal is to link similar patients and provide a platform for them to work together and solve problems. If patients in this community do no trust each other, then it is very hard for them to achieve this goal. As a result, online social networking will perform more like a repository for health data rather than helping patients develop their medical knowledge and disease management skills on other patients' shoulders. Even worse, the system might not be stable because a network that has a predominance of null or reciprocated ties over asymmetric connections may be a more "equal" or "stable" than one with a predominance of asymmetric connections (Hanneman 2005).

4.3.2 Attitude Commitment and Privacy

How much information to disclose is closely related to an individual's privacy concerns and is of importance to the success of online interactions (Coppola et al. 2004, Jarvenpaa and Leidner 1998, Meyerson et al. 1996, Piccoli and Ives 2003). In our study, the variable *data quality* can be viewed as a measure indicating attitude commitment and privacy concerns. Attitude commitment exists when the goal of the organization is increasingly integrated or congruent with the goals of the individuals in it (Hall et al. 1970). Patients spend time and effort to upload and share their private health records and information, suggesting their commitment to the community. Their actions are rewarded by profile stars, a label showing the quality of the shared information. The more detailed health information that individuals upload, the more stars they will receive. This process not only helps patients to keep track of their disease progress, but it also provides an integrated picture of disease history from which other patients can learn and use to show to their physicians to better understand their conditions. Although the benefits are easy to see, there are also drawbacks. The biggest problem is privacy, while the benefit is the gain. The significant but negative effect for *data quality* indicates the inefficient mechanisms provided by the website to encourage patients to contribute more. Although the chart-based interpretations for medical information and history give patients an easy understanding of their disease progress, the incentive or gain for filling out questionnaires and receiving graph-based diagnose results does not appear to compensate patients' effort and time. Furthermore, the evidence from *clustering coefficient* and *betweenness centrality* (both negatively significant) suggest that patients do not tend to form a cohesive group. Despite easier access to shared information about treatment, patients are concerned about being identified outside the OHC. Self-presentation and detailed profiling, which would enhance perceived identity verification (Ma and Agarwal 2007), would also increase the risk of being recognized in the real life. This risk prevents patients from constructing a more close network. This factor has been found in many studies as the critical determinant of sharing information and developing relationships (Fukuyama 1995, Lewis and Weigert 1985). Although trust in the electronic medium generally increases an individual's willingness to disclose, privacy-based protection mechanisms are especially necessary for health-related networks compared to other online social networks.

4.3.3 Knowledge Sharing

Knowledge sharing is consistently found to be positively related to demographic similarity (Pelled 1996), status similarity (Cohen and Zhou 1991), and network position (Wasko and Faraj 2005). Two people who are not directly connected do not have a foundation for knowing each other and sharing knowledge. However, Brown and Duguid (2000) point out that when individuals have a common practice, knowledge readily flows across that practice, enabling individuals to create social networks to support knowledge exchange. People contribute their knowledge when they have experience to share and when they are structurally embedded in the network (Wasko and Faraj 2005), thereby reducing the probability of removing the connection. Bouty (2000) states that reciprocal knowledge exchange between individuals is an important aspect of being a member of the community.

In this study, there are various measures explaining knowledge sharing. Referring to geographic and demographic similarities, we find that the impact of demographic homophily on initiating knowledge exchange is absent. This agrees with the findings in an earlier study by Lewis et al. (2008) who explain that Internet-based communication technologies have reshaped user behaviors. When considering the information and knowledge exchange triggered by health condition similarity, the measures constructed from health similarity, namely *same disease* and (number of) *same symptoms* provides insignificant support for conducting such activities. Finally, the network position results return some significant evidence for knowledge sharing in the OHC. We find that position in the network has significant explanation power on patients' knowledge sharing. Although a cohesive or strongly-connected subgroup ensures the consistency of information, surprisingly, we find that patients tend to reduce the volume of information to be processed. Patients do not tend to form a cohesive subgroup. This effect is particularly significant if an individual is in the middle of some other connections (*betweenness centrality*) and is also enhanced by *clustering coefficient*. Since a cohesive subgroup indicates the existence of redundant links, meaning receiving the same social support multiple times, patients are more attracted to different experiences and knowledge which are typically brought by outliers. This is similar to the difference between strong and weak ties where a strong tie ensures the speed and effectiveness of communications, while a weak tie always brings in new characteristics and ideas. This observation supports the conjecture that the purpose of using subscription tools is to reduce the problem of information overload at the cost of credibility of the exchanged information.

All of the discussion above suggests that patients focus more on the diversity of information regardless of the concern for reliability, suggesting the need for the website to enhance the service and to improve information quality. Various mechanisms should be deployed to enhance patients' desire to give and receive information. Although the online medium makes it very hard to control for information quality and knowledge exchange, a framework emphasizing trust could be a good initial step to launch, because trust is strongly related to information disclosure (Ridings et al. 2002, Metzger 2004).

4.4 Conclusion

In this paper, we examine the homophily in a dynamic online healthcare social network for patients to communicate and exchange knowledge and experiences. The distinct characteristics of the Internet (for example, the lack of geographical boundary) call for extending current theories and analyses on homophily so as to better describe online social relationships. We propose an econometric framework at the dyadic level to analyze longitudinal data. This approach incorporates relevant network level measures to investigate individual preferences for social contacts and then posits to empirically explain online social network evolution. Our results show evidence of homophily for various network measures. The findings of our work are of especial interest to the healthcare industry to engage in online social networking websites and improve their services.

An OHC is a novel way to make personal health records accessible. This technology is believed to contribute to the transformation of healthcare delivery systems. Future research should further investigate customized search engines to deliver the most relevant and timely information. New recommendation mechanisms should be relationship based and designed by taking into consideration network characteristics.

Chapter 5: Shared Minds: How Patients Use Collaborative Information Sharing

5.1 Data and Variables

We collected the data from a popular healthcare social networking website where patients can communicate various health related topics and share their experience and knowledge online. Because the Internet plays a significant part in mental health information-seeking (Powell and Clarke 2006), we focus on the online social interactions and virtual relationship among patients who have mental problems. Our data set consists of patients' individual levels of health conditions, their online relations (hereafter, referred to as 'friends'), treatment measurement reports within this community, discussion of treatments and forum thread exchanges.

For the purpose of our study, we explore discussions about 15 treatments widely used for patients who are diagnosed with mental problems; these treatments cover 93% of users who provide their treatment information and experiences in the community. Community level interactions are collected from the treatment evaluations that are available to every member in the community. For each patient appearing in treatment discussions, we construct her friend network, if any, by looking at the subscription list on her profile. Like Twitter or other popular blogs, where users can "follow" or "subscribe" to an account, patients in the focused healthcare community can "become a friend" of another patient and learn about her health condition dynamics and treatment changes by simply subscribing to her profile. This virtual relation is a one way initiation so that two patients will become mutual friends only if they are "friended" to each other.

5.1.1 Attribute Information and WOM Measures

Distinguished from prior studies, our data set contains patients' WOM on treatment outcomes from various perspectives, such as from its effective prospect, the severity of side effects, to what extent they adhere to treatment, and how difficult it is being on the treatment, all of which we refer to here as the attribute information of a treatment, following a similar notion by West and Broniarczyk (1998).

Another difference in this work is that prior studies focus on one feature, one information attribute such as ‘effectiveness’, for the treatment outcome to study patients’ perceptions about the treatment quality, while they are ignorant of the potential impact of other features. For instance, confronted with a severe side effect from a treatment, Efficacy may not play the critical role in patients’ adoption decisions. Taking Clenbuterol as an example, although it is effective in helping patients with chronic breathing disorders, it has been banned by many countries for its severe side effects and danger. When a patient reads the online reviews that discuss a product from multiple angles, the opinions embedded in the reviews that discuss the different attributes of a treatment are amalgamated to affect the patient’s perception of the treatment quality. Thus, it is important to understand how the reviews of various aspects of a treatment integratively affect a user to form her perception of treatment quality. This challenge is compounded by the fact that when the online product evaluations involve multiple experiential aspects of a product, the evaluations from different customers can significantly differ from one another depending on the individual’s experience with a product and their preferences for different experiential aspects of the product. We will re-examine these difficulties in Sections 5.2 and 5.3.

5.1.2 Ratings Data

In the focused healthcare community, online reviews of a treatment are often provided in two formats. The first and most common practice of WOM is the expression of numerical ratings that a patient chooses to report for a treatment from various aspects, such as treatment effectiveness and side effects. Detailed discussion of a treatment is sometimes also reported in text. To capture a patient’s assessment for a treatment, we use numerical ratings $rating_i$ as the dependent variable in our study, where i is the indicator for the experiential attributes. The complementary text reviews are used as covariates to measure social contagion. Table 18 reports the overall statistics of reviews for 15 treatments.

Table 18: Review Statistics for 15 Treatments

	Mean	St.Dev	Min	Max
No. Rating Reviews	2226.8	1319.33	163	4480
No. Text Reviews	1862.33	1101.99	150	3778

We collect ratings of four experiential aspects for treatments: the effectiveness of the treatment, the side effects that patients experienced when taking the treatment, the persistency of taking the treatment, and lastly the hardship of taking the treatment. The perceived treatment efficacy is a “must provide” assessment for patients to write their reviews. It expresses patients’ beliefs of the potentially positive consequences of taking the treatment. The measure of side effects indicates patients’ perceived adverse effects from taking the treatment. These two measures are not simply judged or controlled by patients’ self-disease management, because they are also related to patients’ health condition. However, the other two measurements, the adherence and burden, are completely controlled by patients, as each patient decides how to manage the process of adopting treatments. All four experiential aspects together reflect patients’ beliefs in the quality of the product. There are, as a result, 20023 effectiveness ratings, 15968 side effect ratings, 11474 persistency ratings, and 10392 hardship ratings for 15 treatments across 15 months in our panel data set.

Each of these aspects is measured in an ordinal score which ranges from the lowest score 1 to the highest score 4 and indicates a patient’s perception of the treatment for the corresponding aspect. The vertical information about consumers’ perception of product quality (e.g., low ratings vs. high ratings) combined with the horizontal aspects of a product, including the efficacy level, the side effects, adherence and burden, describes a complete picture of patients’ opinions of the treatment.

5.1.3 WOM Data

WOM are collected from two sources: the community and the friend. The former is considered as the “buzz effect” because it is drawn from the treatment evaluations shared within the entire community, and the latter can be described as the “similarity effect” because it captures the opinions from patients “more like me.” One advantage for patients who communicate in online healthcare communities is finding similar patients to learn from. Once a patient identifies someone similar to her, she can establish a friendship style relationship with the other party and remain closely connected. Under such circumstances, we expect patients to react differently to the opinions of friends compared to general community opinions.

To gain a deeper understanding of social contagion in the healthcare community, we differentiate the source of WOM in our study, and construct WOM measures in terms of dimension and form for each source. For simplicity, we eliminate the source in the variable explained in this section.

5.1.3.1 The Volume and Dispersion of WOM Measures

We adopt Godes and Mayzlin's (2004) method to measure the volume and dispersion of WOM as follows. Take "buzz effect" reviews as example. Let $j = 1, ..., J$ index the patients. We define $NumRating\,\mathrm{Re}view_{t,j,k}$ as the total number of numerical reviews patient j write for the treatment k till time t, and the volume of patients' reviews for treatment k at time t is thus defined as:

$$NumRating\,\mathrm{Re}view_{t,k} = \sum_{j=1}^{J} NumRating\,\mathrm{Re}view_{t,j,k} \; .$$

Since we can not only directly observe the number of the reviews, but also keep track of who posts which review at what time, we can average these numerical review ratings by attribute. $AvgEfficacy_{t,k}$ is the average rating of the effectiveness assessment for treatment k till time t, $AvgSideEffect_{t,k}$ captures the average side effect patients reported for treatment k till time t, and $AvgPersit_{t,k}$, $AvgHardship_{t,k}$ are covariates controlling the averaged persistency and hardship that patients reported in their reviews respectively. We further construct the variance of these values $VarEfficacy$, $VarSideEffect$, $VarPersistency$, and $VarHardship$ correspondingly, all of which summarize other patients' experiential opinions for the treatment till time t. Tables 19, 20 and 21 give review distributions for a sample treatment. For reasons of anonymity, the treatment identity is not revealed.

5.1.3.2 The Opinion of WOM Measures

For text reviews, we count the number of comments for the treatment to calculate the volume of textual WOM $NumComt_{t,k}$ for the treatment k till time t, and entropy of the text reviews as the dispersion control

$Entropy$ (refer to appendix for the definition). In addition, we conduct sentiment analysis[7] to extract the opinion embedded in the text. Instead of assigning +/- sign to a certain text comment, we use the probability to describe the likelihood that the patient is in favor of the treatment. Thus $PosComment_{t,k}$ is calculated by summarizing the probability of positive opinions embedded in the context of all text reviews for treatment k till time t. It measures how positive patients think about the treatment. Table 22 shows the overall distribution for opinions extracted from "buzz" and "similarity."

5.1.4 Opinion Leadership

Among various measures, such as indegree centrality which can be used to measure of status or prestige in a network, we look at the feature in which users rate each other's content, a tool offered by the healthcare community. In other words, once a patient posts something to the online community, other members can review the content and assess the quality and usefulness of this particular piece of information. The higher the score of the usefulness or helpfulness a patient receives for her shared online information, the higher the perceived quality of information; a high score also suggests a knowledgeable status in the community. This feedback mechanism provides direct observation of the creditability of the information a user has shared. We construct $Creditablity$ to assess the impact of individuals within the collaborative community.

5.1.5 Other Controls

In addition to WOM in different dimensions, we are able to take controls for the treatment characteristics and patients' individualized characteristics to differentiate the potential WOM effect from alternative possibilities. The detailed variables are summarized in Table 23.

5.1.5.1 Treatment Characteristics

We capture treatment related features from their visibility to its specific tangible usage data. Specifically, we count the total number of patients who report taking the treatment in the community $NumPatient$, the number of subsets of patients who take the treatment and also post reviews $NumPatientRating$, and the

---

[7] LingPipe: Alias-i. 2008. LingPipe 4.1.0. http://alias-i.com/lingpipe (accessed October 1, 2008)

number of discussions related to the treatment $NumTopics$. These three controls indicate the popularity of the treatment and can also be used as a proxy for treatment quality. In addition, to avoid potential bias brought about by the use of the treatment, we construct $AvgDosage$, $VarDosage$, $AvgDuration$, $VarDuration$, $AvgStopTakingTime$, $VarStopTakingTime$, $AvgCost$, and $VarCost$ as the treatment specific usage information to demonstrate the treatment's general adoption distribution.

5.1.5.2 Individual Characteristics

Although the data are collected from patients mainly suffering from mood problems, there are significant differences among these patients. First, patients are characterized by their demographic information such as gender and age. $Female$ is the dummy variable for patient's gender, and Age captures a patient's demographic feature. To control for the patients' differences in disease severity and their knowledge of the health problem, we construct variables to indicate their individualized disease features, namely the number of symptoms that they are currently suffering $Numsymptoms$, how severe the health problem is $AvgSeverity$, and how long they have been diagnosed with this problem $\mathrm{D}iseasediagnoseTime$. We also construct variables to capture treatment usage difference such as how long they have been on the treatment $TimeonDrug$ and the drug dosage $CurrentDosage$. Lastly, their online activities are also recorded to eliminate potential unobserved individual specific personality effects. $DataQuality$ describes how much health information a patient has shared in the community, while $Membership$ indicates the number of days a patient has been a member of this health community. $Ins\tan tMood$ is a proxy that captures the current mental condition of a patient at the time she writes the treatment evaluation.

5.2 Baseline Models and Analysis

5.2.1 Model Specification

To study the effect of WOM in the online healthcare environment, we first test separately how different information attributes, i.e., *efficacy*, *side effect*, *persistency*, and *hardship*, shared in online reviews affect consumers' beliefs through a latent response model. After controls for the treatment difference and individual level factors, we differentiate the WOM generated by community users from WOM generated by friends, and investigate the social influence of both on patients' perceived treatment outcomes using the following model:

$$rating_{i,t,j,k} = \alpha Treatment_{t,k} + \beta Patient_{t,j,k} + \gamma ComOpinion_{t,k} + \theta FrOpinion_{t,j,k} + \varepsilon_{i,t,j,k} \qquad (5.2.1)$$

where t refers to the occasion or time when a patient writes a review, j and k are the indicators for the patient and treatment respectively. It is worth noting that patients rate their treatment outcomes from different aspects, thus item i represents the information aspect that a patient shares in her opinion, namely $i = 1$ refers to the perceived treatment efficacy outcome, $i = 2$ refers the perceived treatment side effect, $i = 3$ indicates the patient's perception of her adherence level, and $i = 4$ is the perceived treatment burden. For simplicity, we use rating as an indicator in the model for the perceived treatment outcome and conceal the subscript thereafter, but we conduct separate analysis for each experiential attribute. $Treatment_{t,k}$ controls the information available for treatment k's characteristic till time t, and $Patient_{t,j,k}$ summarises patient j's individual-level related information that can be traced till time t. $ComOpinion_{t,k}$ and $FrOpinion_{t,j,k}$ represent the WOM factors extracted from general users and friends respectively.

Ideally, a patient's evaluation in the review is a continuous variable $rating^*_{t,j,k}$ reflecting her true perception of the quality of the treatment. However, it is hard to implement these measures in the online environment so that users' inputs are mostly transformed by the threshold model (note that all three thresholds κ_s, $s = 1,2,3$ are identified):

$$rating_{t,j,k} = \begin{cases} 1 & if \quad rating^*_{t,j,k} \le \kappa_1 \\ 2 & if \quad \kappa_1 < rating^*_{t,j,k} \le \kappa_2 \\ 3 & if \quad \kappa_2 < rating^*_{t,j,k} \le \kappa_3 \\ 4 & if \quad \kappa_3 < rating^*_{t,j,k} \end{cases}$$

As stated by Yabe-Hesketh et al. (2003), the cumulative models are equivalent to the latent response model and the proportional odds model equivalent to the above model is thus defined as:

$$\log\left(\frac{P(rating^{*}_{t,j,k}\le s)}{1-P(rating^{*}_{t,j,k}\le s)}\right)=\log\left(\frac{P(rating_{t,j,k}\le s)}{1-P(rating_{t,j,k}\le s)}\right)$$

$$=\kappa_{s}-\alpha Treatment_{t,k}-\beta Patient_{t,j,k}-\gamma ComOpinion_{t,k}-\theta FrOpinion_{t,j,k}$$

5.2.2 Unobserved Heterogeneity

In the model (5.2.1), the reviews belonging to the same cluster share the same cluster-specific influence, which leads to the problem of unobserved heterogeneity. In other words, patients who take the same treatment will behave similarly, as if they are reporting homogenous evaluations. The reviews of those who make multiple evaluations will be identical to one another; however, these assumptions are unlikely to hold because of the unobserved heterogeneity. It is intuitive that patients are different, and there are different patterns among treatments reviews. To control for the unobserved dependence between responses for the reviews in the same cluster, we explicitly add random intercepts $\zeta^{(2)}_{0j,k}$ and $\zeta^{(3)}_{0k}$ into the model where $\zeta^{(2)}_{0j,k}\sim N(0,\psi_0{}^{(2)})$ and $\zeta^{(3)}_{0k}\sim N(0,\psi_0{}^{(3)})$ to control for patients and treatment related variations. Thereby, we have the following model:

$$rating_{t,j,k}=\zeta^{(2)}_{0j,k}+\zeta^{(3)}_{0k}+\alpha Treatment_{t,k}+\beta Patient_{t,j,k}+\gamma ComOpinion_{t,k}+\theta FrOpinion_{t,j,k}+\varepsilon_{t,j,k} \qquad (5.2.2)$$

5.2.3 Hierarchical Structure

It is noteworthy that model (5.2.2) is in a hierarchical structure that responses $rating_{t,j,k}$ are nested by individuals because multiple treatment reports can be generated over time, and individuals are clustered by treatment k. In such a scenario, our dataset is organized in a three level hierarchical structure, and the treatment is the highest level in the model, followed by patients, and the response at the given time is the lowest level. To test the significance of variances of effects at each level, we conduct an ANOVA test with nested random factors using item $i=1$ perceived treatment efficacy data for models with single level, two-level, and three-level respectively. The returned Log-likelihood indicates that the two level model outweighs other models (LL_1 = -76937.238, LL_2 = -76936.921, and LL_3 = -76939.228). In other words, unobserved heterogeneity between patients is significant while the treatments are similar to one another.

Hence, our initial model is dropped from a three-level to a two-level structure by considering only two levels of item responses and patients.

5.2.4 Findings

Table 24 reports the estimation results for the models. As discussed earlier, we use each experiential attribute measure as the dependent variable to run the model (5.2.2), and the estimation results are listed in Model 1 to Model 4 in Table 24 respectively. Overall, we find that WOM from different social dimensions has different influences patients and no consistent pattern is recognized. Surprisingly, the WOM from community users has a higher impact than the friends' opinions in general. Depending on the aspect on which the experiential experience shared among patients is focused, the influential power of such WOM to patients' perception is found to differ in effect. We discuss these results below.

When considering the discussions about perceived treatment efficacy, i.e. Model 1, we find that community opinions exhibit significant influence. As expected, the more patients provide the rating for the treatment efficacy ($NumRatingReviews$), the higher score a patient tends to rate the treatment, as supported by positive and significant estimate γ_1 . However, the higher other patients rate their perceived efficacy outcome ($AvgEfficacy$), the lower a future patient tends to rate her own. When referring to the opinions embedded in the text reviews, the greater the volume of this detailed experience sharing ($NumComment$)or the more positive experience shared by prior users with the treatment (*PosComment*), the lower a patient would like to rate the score. These negative and significant results suggest the difference between what patients expect and what they truly feel. When various discussions and opinions are shared among patients, a patient can collect more information about the treatment she is taking; she observes other patients' health conditions and the effectiveness of the treatment, and uses this knowledge to assess her own condition. From others' experiences, she can set her expectation for the efficacy outcome of the treatment, and compare it with her true feeling. The volume of WOM is closely related to the visibility of a treatment. The more reviews about a treatment receives, the more patients know about it, and thus the more information will be available regarding the experience of taking the treatment.

Knowing how well others are doing and what is the best outcome others have achieved, the patient expects that she can receive similar results regardless of her own health condition. The gap between what a patient expects and how she really feels after taking the treatment leads to the negative impact of community opinions.

The WOM from community members is found to have different influence patterns in three other perception aspects. In the side effect discussions, the volume and dispersion of WOM is found to have no significant impact, regardless of whether the WOM is from the "buzz" or from "friend." However, the averaged ratings are found to have significant impact in the other two information attributes, i.e., treatment persistency and hardship in Models 3 and 4. These two measures are assessments from individuals' self-disease management perspective, so the focus of these two aspects leans toward how patients evoke sympathy from conversations with others. For this reason, the averaged perception of persistency and hardship are estimated to have a correspondingly positive and significant impact on patients; however, the opinions embedded in the text comments, volume or sentiment, are not influential.

It is shown in our estimation results that friends' opinions have less influential power across all four experiential aspects. Overall, the results are mixed. The number of friends' ratings (*NumFrRatingReviews*) is negative and significant in the discussions of patients' adherence to the treatment, i.e. θ_3, while insignificant in influencing future ratings for the other three aspects. Referring to the average ratings measures, none of the patients' ratings are significantly influenced by either these averaging measures or the distributions of the corresponding ratings (for example, *AvgFrEfficacy VarFrEfficacy AvgFrSideEffect AvgFrAdherence AvgFrBurden*) in Model 1 and 4, while the averaged efficacy rating θ_4 in Model 2 and variance of the rating θ_5 are found to have significant impact.

As discussed above, our initial models identify limited evidence showing the influential power of WOM to patients' perceptions for their treatment outcome from different angles. Although there is also a learning effect from patients who provide multiple evaluations for their treatment (positive and significant result for PreviousEval*Time), it remains unclear how these attributes affect patients' perceived treatment quality as a whole, as we do not find consistent patterns in the results returned by the above models. In addition, the WOM from the friend network is found to have less impact than the WOM from the buzz. As a result, we do not find support for Hypotheses III.1 to 3. On the contrary, we observe increasing participation in such information sharing behavior in the community, as shown in Figure 9. If there is no social impact, why would patients willing to take time to do participate? We address this challenge in detail in the next section.

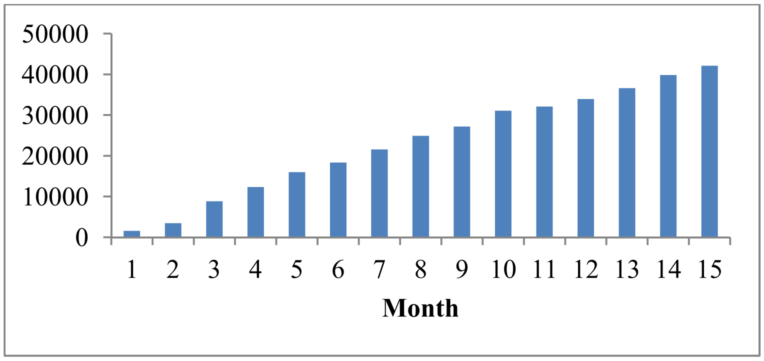


Figure 9: Numerical Reviews over Months

5.3 Constitute Construct and Analysis

From the discussions in Section 5.2, we find that different information aspects have an impact on the rating of a treatment, although this does not lead to a comprehensive understanding of how these information attributes represent the integrated realization of patients' perceived treatment quality. Nonetheless, it is possible to partially explain the effects of WOM by considering each experiential attribute separately and testing the effect of prior WOM on patients' perceived treatment quality from different aspects. However, we still need a framework that is capable of suggesting the overall patients' perceived treatment quality through integrating different information attributes posted online by consumers. Therefore, in this section, we develop a constitute construct "perception of treatment quality" measured by four experiential attributes.

5.3.1 Constitute Construct Model

From earlier discussion, we find that a different information aspect impacts the ratings in the corresponding experiential attribute. The outcome of interest, i.e., patients' perceived treatment quality, can be measured by the four experiential attributes. Looking at these aspects separately cannot explain the process by which patients integrate these aspects, and it is also unclear whether these four attributes are equally weighed. To gain a better understanding of a treatment, it is necessary to combine the information of these four aspects, each of which indicates a different attribute of the treatment, and then to form a conceptual perception of treatment outcome based on the four aspects. The latent response $rating_{i,t,j,k}$ for the i^{th} item at time t for patient j treatment k is therefore modeled as:

$$rating^{*}_{i,t,j,k} = \delta_i + \lambda_i \eta^{(2)}_{t,j,k} + \varepsilon_{i,t,j,k},\ \lambda_1 = 1, \delta_1 = 0 \tag{5.3.1}$$

Note that, from the discussion in Section 4, we have found no significant heterogeneity among treatments, hence we lower our multilevel data from three levels to two levels and omit the subscript k for treatments. By combining the measurement model with a structural model for the treatment outcome, $\eta^{(2)}_{t,j,k}$ is denoted as $\eta^{(2)}_{t,j}$ and modeled as follows:

$$\eta^{(2)}_{t,j} = \alpha Treatment_t + \beta Patient_{t,j} + \gamma ComOpinion_t + \theta FrOpinion_{t,j} + \eta^{(3)}_{j} + \zeta^{(2)}_{t,k} \tag{5.3.2}$$

where $\eta^{(3)}_{j}$ is a random intercept at patient level, and $\zeta^{(2)}_{t,k}$ an time specific random intercept.

5.3.2 Thresholds Models

As patients' perceived values are not directly observable, we model the latent response by a thresholds model and integrate the responses from the four different aspects. This approach presents some challenges. Since there are experiential aspects encoded in the framework, it remains unclear whether the thresholds κ_s, $s = 1,2,3$ are constant across four items i, or four different specifications of thresholds κ_{is} and intercepts δ_i should be modeled. Therefore, we consider three different specifications:

Case 1: Different thresholds for each item i and no intercepts:

$$rating_{i,t,j,k} = \begin{cases} 1 & if \quad rating^*_{i,t,j,k} \le \kappa_{i1} \\ 2 & if \quad \kappa_{i1} < rating^*_{i,t,j,k} \le \kappa_{i2} \\ 3 & if \quad \kappa_{i2} < rating^*_{i,t,j,k} \le \kappa_{i3} \\ 4 & if \quad \kappa_{i3} < rating^*_{i,t,j,k} \end{cases}$$

Case 2: One set of thresholds for all four items and no intercepts:

$$rating_{i,t,j,k} = \begin{cases} 1 & if \quad rating^*_{i,t,j,k} \le \kappa_{1} \\ 2 & if \quad \kappa_{1} < rating^*_{i,t,j,k} \le \kappa_{2} \\ 3 & if \quad \kappa_{2} < rating^*_{i,t,j,k} \le \kappa_{3} \\ 4 & if \quad \kappa_{3} < rating^*_{i,t,j,k} \end{cases}$$

Case 3: One set of thresholds for all four items and intercepts δ_2 , δ_3 and δ_4 for items 2, 3 and 4.

These models return log-likelihood as -76932.137, -76936.921, and -76938.05. According to BIC, reported in Table 25, we adopt the case 2 model to conduct further analysis.

Table 25: Model Selection

Model	Log-Likihood	Parameters	BIC
1	-76932.137	12	-76997.93
2	**-76936.921**	**3**	**-76953.37**
3	-76938.05	5	-76965.46

5.3.3 Findings

The parameter estimates for the constitute construct model are given in Table 26. In this analysis, we find evidence for the social contagion of online WOM in the healthcare communities. First, we find that the volume of the community WOM (*NumRatingReviews*) and its dispersion (i.e., the averaged ratings *AvgRating* and *VarRating*) have a significant impact on future ratings indicating a patient's perception of the treatment. Specifically, the positive and significant estimate γ_1 provides the supporting evidence for Hypothesis HIII.1.1. The more information patients acquire from others' opinions, the better knowledge they will gain, and the greater chance they will have to obtain more complete feedback from the experience of being on the treatment. This more comprehensive information helps patients to set proper goals, given their health condition, for the expectation of taking the treatment. The perception of the treatment quality is adjusted as the difference between what patients' expect and what they receive diminishes in the experiential learning process. However, the negative and significant impact of the averaged rating indicates that Hypothesis HIII.1.2B is supported while HIII.1.2A is rejected. When a patient observes a high averaged rating for a certain treatment, the overall effective feedback causes her to set a high expectation for the treatment, which adversely leads to a lower rating at the time she provides the rating, even if she achieves the health condition she wants. In addition, we find the variance of such review ratings has a positive significant influence on the future rating (*VarRating*). This positive influence can be explained by the fact that a large variance indicates the large difference in health conditions and treatment outcomes achieved. As such, it is easier with a large variance for a patient to find someone with similar health conditions and to measure their treatment experience more realistically. Hence, it is expected that there will be a positive effect on the consumer ratings. Similar to the previous findings, the impact of positive comments (*PosComment*) to patients' reviews is negatively significant, which again confirms the difference between the patients' expectations and their real experiences. Hypothesis III.1 is thus supported.

However, in contrast to the discussions in Section 5.2, we find increasingly influential power from friends' WOM, and these findings support Hypothesis III.2. That is, a patient who has more friends who rate a treatment might choose to report a higher score for her experience with the treatment (positive and significant estimate for *NumFrRatingReviews*). On the other hand, a large variance identified by a patient from her friends' ratings can cause the patient to doubt why her friends in a similar health condition report quite different treatment reviews. As such a doubt increases with the variance of friends' ratings, a negative impact of *VarFrRating* on future reviews is expected. Our estimation result supports this argument. However, we do not find support for Hypothesis III.3.

Although side effects, persistency, and hardship appear "harder" to discern than measuring perceived treatment efficacy (negative estimates for the intercepts variables in the measurement model), there is no significant evidence showing that patients are quite different from each other and from the time they provide their evaluations, as indicated by the insignificant results for δ_2 and δ_3. The exception is the intercept for the hardship δ_4, indicating that patients find this attribute hard to measure, and they are quite different both between patients and between the times of providing the measures within patients. Lastly, we find that the factor loading is very close to 1. The low standard errors, significant at 99% level, suggest that these items should not be constrained to 1. That is, these different experiential attributes should be weighed differently when integrating patients' perceived treatment quality. This finding provides additional support for our assertion that it is necessary to look at attribute information integratively when investigating the WOM effect in the online environment, especially when information comes from a variety of aspects to describe the product and its attributes.

5.4 Managerial Implications and Conclusion

Although online WOM has presented a unique channel for consumers to share their experiences of a product from multiple aspects, little attention has been devoted to examining how a consumer understands the user-generated content and to what extent WOM impacts consumer behaviors when the information shared online involves different experiential aspects about a product and the evaluations differ significantly from one to another depending on the unique backgrounds and personal preferences of individuals. In the last few years, the number of online health communities has increased rapidly as more patients seek to access alternative sources of health information as well as to connect with other patients with the same or similar disease (Nambisan 2011). In addition, although the Internet is changing the way people learn about health and illness (Ziebland et al. 2004) and a multitude of online health information exists, there is little consistency in terms of how many people actually use this information, their expectations of it, and the implications of this usage (Cotton and Gupta 2004). To fill this void, this paper investigates how to understand online reviews in the presence of substantive experiential information shared from various aspects and the impacts of experiential information sharing on a customer's perception of product quality. With the data collected from a healthcare social community website, we have empirically compared two analytical models – one model which examines the influence of WOM from each experiential information aspect separately and the other which integrates the different experiential information aspects by using a constitute construct. We have found evidence that the model using the constitute construct better explains the relationship between WOM and the product quality perceived by customers.

There are two unique aspects of online WOM in a collaborative healthcare environment that fundamentally distinguish our work from prior studies. First, the experiential attributes of a treatment are an important aspect for any patient who is considering adopting the treatment, and the most effective channel for obtaining the information of experiential attributes is often by WOM from those who have experienced the treatment. Doctors can provide the facts of a treatment, which are often also available in the product description provided by the pharmaceutical company, but other patients with personal experience can tell what a treatment truly feels like and what real challenges to expect when receiving the treatment. Besides the effectiveness of the outcome, other experiential aspects about treatment can also be shared with other patients, such as the severity of the treatment, the cost of the treatment, and so on. These information aspects may be critical for a patient who can choose other alternatives. Thus, in the healthcare context, effective sharing of treatment information, particularly experiential information, is critical for those patients who are seeking a second source of communication about a medical problem in addition to talking to the doctor.

Second, it is worth noting that the evaluations of treatments are dependent on patients' perceptions of the experiential aspects of treatments. Because patients may have different physical conditions and be at a different stage of severity of a disease, the same treatment may show varying degrees of effectiveness in different patients, leading to differing perceptions and ratings of the same treatment. Furthermore, an online review can evaluate a treatment from multiple experiential aspects and rate the attributes of the treatment differently. Depending on the patient's weighting function for the various attributes, the overall evaluation of a treatment may vary significantly from one patient to another. Given such heterogeneity, an overall rating of a treatment that integrates individuals' opinion may not be so useful for a patient seeking to learn from the online community. Thus, a critical challenge to be addressed even for the healthcare domain per se is to understand how a patient integrates and interprets the information gathered from the online community when facing such an extensive heterogeneity of evaluations.

Three directions of future research are promising. First, to assure the generalizability of our study, we could test our models on other data sets in addition to the data sets from the healthcare domain and study whether consistent results regarding experiential information sharing can be obtained from other domains. Second, it could be very interesting to examine how a customer integrates the opinions from the community with those of friends and how WOM influences customer behaviors when the WOM from the community conflicts with the WOM from friends. Third, as a step further to extend this study, the findings from this study could be used to guide the design of the social community website so that a user could more easily identify the reviews relevant to her.

Table 19: Numerical Review Distribution

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. Numerical Reviews	163	456	144	450	73	320	184	171	560	436	234	386	150	358	606

Table 20: Text Review Distribution

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. Text Reviews	150	361	125	380	68	279	107	170	436	382	80	382	139	177	542

Table 21: Community Review Distribution

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. Text	150	511	636	1016	1084	1363	1470	1640	2076	2458	2538	2920	3059	3236	3778
Mean	10	34.07	42.4	67.73	72.27	90.87	98	109.33	138.4	163.87	169.2	194.67	203.93	215.73	251.87
St.Dev	6.47	39.09	46.25	64.80	65.50	84.67	92.56	102.21	129.07	141.45	139.71	161.56	163.99	182.44	205.98
Min	1	4	4	9	12	12	12	15	30	24	31	38	44	42	46
Max	20	137	170	194	199	228	262	265	391	416	435	453	473	497	555

Table 22: Distribution of Evaluations of Text Reviews

	Number	%	Mean	St.Dev	Min	Max
Positive Reviews from Community Opinion	16126	58.2	1075.067	624.198	86	2148
Positive Reviews from Friend Opinion	15962	49.5	0.276	0.574	0	3

Table 23: Data Description and Statistics

Dimension	Variable	Description	Mean	STD	Min	Max
	item	the responses: (efficacy -1, side effect -2, adherence -3, burden -4)				
	time	indicator for month t				
	patient	indicator for individual id				
	treatment	indicator for treatment(drug) id				
Individual Review	Efficacy	patient's perceived outcome effectiveness of a treatment	2.17	1.05	1	4
	SideEffect	patient's (perceived) experienced side effect when taking a treatment	0.64	1.14	0	4
	Persistency	patient's self-revealed adherence to a treatment	0.16	0.64	0	4
	Hardship	patient's (perceived) experienced difficulty when taking a treatment	0.04	0.33	0	4
Community Opinion	NumRatingReviews	the number of rating reviews in a community	324.97	303.33	5	1197
	AvgEfficacy	the average efficacy rating in the community for a treatment	1.55	1.15	0	3
	VarEfficacy	the variance of efficacy rating in the community for a treatment	1.07	0.24	0.12	2
	AvgSideEffect	the average side effect rating in the community for a treatment	1.69	1.15	0	3
	Avg Persistency	the average adherence rating in the community for a treatment	1.61	1.04	0	3
	AvgHardship	the average difficulty rating in the community for a treatment	1.73	1.06	0	3
	NumComment	the number of text reviews in the community for a treatment	204.43	159.80	4	555
	PosComment	the aggregated positive probabilities of the text reviews in the community	80.33	64.29	1.89	244.12
	Credence	the total number of usefulness flags users has get for their online participation in the community	34317.79	8056.92	20005	47819
	AvgCredence	the average number of usefulness flags users has get for their online participation in the community	1.95	0.46	1.14	2.72
	ComEntropy	the aggregated comments entropy in the community	0.25	0.15	0	0.51
Friend Opinion	NumFr	the number of friend the patient has	3.51	2.30	0	7
	NumFrTakingTreatment	the number of friend who is taking the same treatment	2.32	1.81	0	7
	NumFrRatingReviews	the number of ratings reviews in the friend network	1.14	1.08	0	4
	AvgFrEfficacy	the average efficacy rating in the friend network for a treatment	1.34	1.26	0	4
	VarFrEfficacy	the variance of efficacy rating in the friend network for a treatment	0.38	0.61	0	2.25
	AvgFrSideEffect	the average side effect rating in the friend network for a treatment	0.94	1.14	0	3
	AvgFrPersistency	the average adherence rating in the friend network for a treatment	0.95	1.15	0	3
	AvgFrHardship	the average difficulty rating in the friend network for a treatment	0.95	1.15	0	3
	NumFrComment	the number of text reviews in the friend network for a treatment	0.57	0.83	0	3
	PosFrComment	the aggregated positive probabilities of the text reviews in the network	0.20	0.41	0	2.25
	Fr_c	the total number of usefulness flags users has get for their online participation in the network	10.95	8.16	0	25

Table 23: Data Description and Statistics Cont'd

Dimension	Variable	Description	Mean	STD	Min	Max
Friend Opinion	AvgFr_c	the average number of usefulness flags users has get for their online participation in the network	4.03	4.85	0	25
	FrEntropy	the aggregated comments entropy in the friend network	0.17	0.42	0	1.58
	AvgFrDosage	the average dosage in the friend network	4.25	3.40	0	9.9
	AvgFrDuration	the average duration in the friend network	4.03	3.33	0	10
	AvgFrStopTakingTime	the average time of stop taking the drug in the friend network	4.03	3.33	0	10
Treatment Characteristics	NumPatient	the number of patient who is taking the treatment in the community	958.46	678.09	23	2296
	NumPatientRating	the number of patient who is taking the treatment and also write rating in the community	242.82	184.57	4	641
	AvgDosage	the average dosage in the community	5.80	2.66	1	10
	VarDosage	the variance of dosage in the community	0.35	0.09	0.00	6.12
	AvgDuration	the average duration in the community	3.48	1.65	1	6
	VarDuration	the variance of duration in the community	1.07	0.23	0.00	3.03
	AvgStopTakingTime	the average time of stop taking the drug in the community	3.34	1.49	1	6
	VarStopTakingTime	the variance of time of stop taking the drug in the community	1.02	0.21	0.04	4.76
	AvgCost	the average cost in the community	1.47	0.69	1	6
	VarCost	the variance of cost in the community	0.02	0.00	0	0.19
	NumTopics	the number of forum topics for the treatment	57.45	27.11	17	101
Patient Characteristics	Female	is 1 if the patient is female	0.50	0.50	0	1
	Age	patient's reported age	41.35	14.58	17	66
	DataQuality	the data quality for patient's medical profile	1.52	1.12	0	3
	InstantMood	patient's reported current mood	2.99	1.41	1	5
	Membership	the number of days since the patient joined the community	235.13	167.43	1	895
	NumSymptoms	the number of symptoms the patient is suffering	5.99	2.60	2	10
	AvgSeverity	the average severity level of the disease for the patient	1.50	1.12	0	3
	DiseaseDiagnoseTime	the number of days since the patient be diagnosed	270.40	146.95	30	510
	TimeonDrug	the number of days the patient is on the drug	282.85	155.94	30	540
	CurrentDosage	the current dosage for the drug	256.71	377.56	0.5	1200

Table 24: Latent Response Model Estimation Results[8]

		Model 1 *dep = Efficacy*				Model 2 *dep = Side Effect*			
		Coef.	**STD**	**Coef.**	**STD**	**Coef.**	**STD**	**Coef.**	**STD**
Community Opinion	γ_1 NumRatingReviews	0.3324^{***}	(0.0926)	0.3440^{***}	(0.0939)	-0.1097	(0.1053)	-0.0977	(0.1068)
	γ_2 AvgEfficacy	-23.4858^{*}	(12.7987)	-21.7975^{*}	(12.7551)	21.3495	(14.5567)	21.2571	(14.5083)
	γ_3 VarEfficacy	6.3340	(6.1846)	6.0887	(6.1870)	-2.4141	(7.0832)	-2.5287	(7.0857)
	γ_4 AvgSideEffect	-1.3048	(1.2623)	-1.1309	(1.2675)	0.3621	(1.4345)	0.4443	(1.4404)
	γ_5 AvgPersistency	0.5099	(1.3657)	0.8071	(1.3623)	0.7203	(1.5599)	0.7658	(1.5557)
	γ_6 AvgHardship	-2.0366	(1.3990)	-1.8925	(1.3984)	-1.8866	(1.5880)	-1.8650	(1.5873)
	γ_7 NumComment	-0.1124^{***}	(0.0186)			-0.0225	(0.0210)		
	γ_8 PosComment			-0.2788^{***}	(0.0470)			-0.0665	(0.0530)
	γ_9 Credence	-0.0004^{**}	(0.0002)	-0.0004^{**}	(0.0002)	0.0000	(0.0002)	0.0000	(0.0002)
	γ_{10} CommentEntropy	-5.8458	(9.0607)	-5.8926	(9.0606)	-13.3915	(10.3031)	-13.4036	(10.3029)
Friend Opinion	θ_1 NumFr	-1.1820	(8.8640)	-1.0552	(10.2975)	15.1270	(11.5802)	15.1660	(11.5799)
	θ_2 NumFrTakingTreatment	2.3411	(12.2929)	2.1459	(12.2740)	8.3715	(13.7586)	8.3223	(13.7583)
	θ_3 NumFrRatingReviews	36.1800	(25.6197)	35.9048	(25.8035)	-0.0138	(29.5429)	0.0276	(29.5421)
	θ_4 AvgFrEfficacy	0.6633	(1.6259)	0.6431	(1.6329)	-7.6904^{***}	(1.8930)	-7.6944^{***}	(1.8930)
	θ_5 VarFrEfficacy	-0.8293	(2.9098)	-0.8402	(2.9188)	2.5325	(3.3801)	2.5304	(3.3801)
	θ_6 AvgFrSideEffect	1.0314	(1.4772)	1.0569	(1.4794)	-0.0151	(1.7091)	-0.0059	(1.7092)
	θ_7 AvgFrPersistency	-0.1981	(1.4585)	-0.1953	(1.4596)	-0.5117	(1.6859)	-0.5131	(1.6858)
	θ_8 AvgFrHardship	0.1988	(1.4574)	0.1983	(1.4603)	0.1521	(1.6830)	0.1546	(1.6829)
	θ_9 NumFrComment	0.8470	(4.1022)	0.9789	(4.1009)	-4.0757	(4.7382)	-4.0475	(4.7382)
	θ_{10} PosFrComment	-0.8027	(4.6750)	-0.9458	(4.6666)	0.6179	(5.4354)	0.5790	(5.4355)
	θ_{11} FrCommentEntropy	-9.4515	(6.8522)	-9.5709	(6.8572)	-2.3521	(7.9408)	-2.3839	(7.9409)

---

[8] Variables are checked for the problem of multicollinearity. NumComment, PosComment, NumPatient, and NumPatientRating are log-transformed and mean centered. The mean VIF is less than 2.25. Rescaling is performed to remove the potneital correlation: NumRatingReviews, NumRatingReviews^2, NumFr, NumFrTakingTreatment, InstantMood, Membership, NumSymptoms, DiseasediagnoseTime, TimeonDrug, and CurrentDosgae are scaled down by factor of 1000; and other variables are scaled down by a factor of 100.

	θ_{12} AvgFrDosage	-0.4217	(0.5185)	-0.4098	(0.5185)	0.0694	(0.5861)	0.0740	(0.5861)
	θ_{13} AvgFrDuration	0.0170	(0.5256)	0.0285	(0.5257)	0.1857	(0.5930)	0.1889	(0.5930)
	θ_{14} AvgFrStopTakingTime	0.5142	(0.5258)	0.5215	(0.5258)	1.1705**	(0.5953)	1.1739**	(0.5953)
	θ_{15} FrCredence	-0.2405	(0.2767)	-0.2396	(0.2766)	-0.0787	(0.3137)	-0.0795	(0.3137)
	θ_{16} AvgFrCredence	-0.2476	(0.4711)	-0.2493	(0.4710)	0.3518	(0.5336)	0.3519	(0.5336)
Treatment Characteristics	α_1 NumPatient	16.9817***	(3.2032)	17.7295***	(3.2345)	2.9512	(3.6382)	3.2654	(3.6726)
	α_3 NumPatientRating	-26.3235***	(3.3367)	-27.3606***	(3.3799)	0.8630	(3.7856)	0.4743	(3.8335)
	α_4 AvgDosage	0.5114	(0.5455)	0.4950	(0.5454)	0.1124	(0.6204)	0.1255	(0.6202)
	α_5 VarDosage	9.1413	(16.3544)	8.2777	(16.3629)	99.8144***	(19.9283)	99.3389***	(19.9273)
	α_6 AvgDuration	1.4650*	(0.8641)	1.5668*	(0.8668)	1.4697	(0.9835)	1.5189	(0.9867)
	α_7 VarDuration	0.6751	(5.9279)	0.6691	(5.9277)	1.5120	(6.8347)	1.4446	(6.8358)
	α_8 AvgStopTakingTime	-0.6291	(0.9683)	-0.9658	(0.9661)	0.7205	(1.1019)	0.6547	(1.0992)
	α_9 VarStopTakingTime	2.7649	(6.5253)	3.3984	(6.5211)	16.2325**	(7.5214)	16.3201**	(7.5179)
	α_{10} AvgCost	-1.2434	(2.1457)	-1.4994	(2.1502)	-1.1407	(2.4129)	-1.2409	(2.4179)
	α_{11} VarCost	-44.6392	(324.3064)	-30.9241	(324.2443)	-934.3092**	(382.9748)	-933.6721**	(382.8667)
	α_{12} NumTopics	0.0802	(0.0544)	0.0767	(0.0545)	0.0176	(0.0618)	0.0144	(0.0620)
Patient Characteristics	β_1 Female	0.5166	(2.7234)	0.5371	(2.7234)	-0.6838	(3.1288)	-0.6817	(3.1286)
	β_2 Age	-0.1434	(0.0936)	-0.1451	(0.0936)	-0.0476	(0.1074)	-0.0477	(0.1074)
	β_3 DataQuality	1.4814	(1.2162)	1.4702	(1.2162)	-0.8398	(1.3993)	-0.8381	(1.3993)
	β_4 InstantMood	-0.9332	(9.6382)	-0.9170	(9.6382)	13.1746	(10.9798)	13.1637	(10.9795)
	β_5 Membership	-0.0091	(0.0901)	-0.0123	(0.0901)	0.0556	(0.1030)	0.0600	(0.1029)
	β_6 NumSymptoms	0.3944	(5.2319)	0.3703	(5.2318)	10.8044*	(5.9502)	10.8103*	(5.9500)
	β_7 AvgSeverity	-2.0116*	(1.2172)	-2.0182*	(1.2172)	0.6939	(1.3816)	0.6938	(1.3815)
	β_8 DiseaseDiagnoseTime	-0.0365	(0.0945)	-0.0368	(0.0945)	0.0179	(0.1075)	0.0170	(0.1075)
	β_9 TimeonDrug	-0.1556*	(0.0874)	-0.1552*	(0.0874)	-0.0939	(0.0995)	-0.0938	(0.0995)
	β_{10} CurrentDosage	0.0574	(0.0361)	0.0581	(0.0361)	0.0508	(0.0407)	0.0510	(0.0407)
	β_{11} PatientCred	-0.1430	(0.1024)	-0.1432	(0.1024)	-0.0142	(0.1165)	-0.0142	(0.1165)
	β_{12} PreviousEval * Time	45.9255***	(0.5594)	45.9198***	(0.5593)	4.0597***	(0.4463)	4.0624***	(0.4462)

Significance code: 0 ‘***’ 0.01‘**’ 0.05‘*’

Table 24: Latent Response Model Estimation Results Cont'd

		Model 3 *dep = Persistency*				Model 4 *dep = Hardship*			
		Coef.	**STD**	**Coef.**	**STD**	**Coef.**	**STD**	**Coef.**	**STD**
Community Opinion	γ_1 NumRatingReviews	-0.0259	(0.1854)	-0.0195	(0.1877)	0.0898	(0.3493)	0.1665	(0.3547)
	γ_2 AvgEfficacy	-6.0220	(25.5031)	-6.0466	(25.4250)	17.9159	(48.2210)	16.0473	(48.1348)
	γ_3 VarEfficacy	-4.1736	(12.3387)	-4.2283	(12.3416)	10.6554	(24.4810)	10.1139	(24.5210)
	γ_4 AvgSideEffect	-6.4578**	(2.4962)	-6.4038**	(2.5084)	4.5241	(4.7687)	4.9578	(4.7789)
	γ_5 AvgPersistency	6.0623**	(2.7441)	6.1020**	(2.7390)	8.6914	(5.2868)	8.8065*	(5.2950)
	γ_6 AvgHardship	2.7906	(2.8226)	2.8077	(2.8210)	11.7889**	(5.4102)	11.8377**	(5.4126)
	γ_7 NumComment	-0.0161	(0.0370)			-0.0556	(0.0678)		
	γ_8 PosComment			-0.0449	(0.0931)			-0.2092	(0.1712)
	γ_9 Credence	-0.0002	(0.0004)	-0.0002	(0.0004)	0.0003	(0.0007)	0.0002	(0.0007)
	γ_{10} CommentEntropy	-4.1723	(18.0733)	-4.1741	(18.0733)	37.4787	(34.2194)	37.5793	(34.2200)
Friend Opinion	θ_1 NumFr	31.9800	(19.8795)	31.9940	(19.8791)	-36.4181	(39.6349)	-36.3795	(39.6316)
	θ_2 NumFrTakingTreatment	-5.8939	(23.8070)	-5.9074	(23.8068)	28.8408	(44.3679)	28.8051	(44.3693)
	θ_3 NumFrRatingReviews	-95.5116*	(53.6764)	-95.5039*	(53.6767)	13.4055	(95.4106)	13.7798	(95.4133)
	θ_4 AvgFrEfficacy	-4.6668	(3,4218)	-4.6696	(3.4218)	-0.4049	(6.0839)	-0.4186	(6.0849)
	θ_5 VarFrEfficacy	14.2615**	(5.9833)	14.2603**	(5.9835)	-4.2413	(11.0424)	-4.2637	(11.0453
	θ_6 AvgFrSideEffect	-1.1702	(3.1034)	-1.1658	(3.1035)	5.5557	(5.4785)	5.5971	(5.4792)
	θ_7 AvgFrPersistency	-3.9809	(3.0733)	-3.9817	(3.0734)	-2.7361	(5.4456)	-2.7543	(5.4456)
	θ_8 AvgFrHarship	0.3467	(3.0576)	0.3489	(3.0576)	3.3806	(5.4162)	3.3955	(5.4162)
	θ_9 NumFrComment	-9.8694	(8.6097)	-9.8496	(8.6100)	3.2302	(15.3571)	3.3687	(15.3494)
	θ_{10} PosFrComment	4.1996	(10.1004)	4.1717	(10.1006)	26.9226	(17.8701)	26.7445	(17.8707)
	θ_{11} FrCommentEntropy	7.1671	(14.5400)	7.1446	(14.5405)	-34.8013	(25.1876)	-34.9546	(25.1908)
	θ_{12} AvgFrDosage	-0.3168	(1.0346)	-0.3143	(1.0346)	-1.6383	(1.9784)	-1.6249	(1.9785)
	θ_{13} AvgFrDuration	-0.2575	(1.0478)	-0.2549	(1.0478)	-0.1833	(1.9958)	-0.1601	(1.9961)
	θ_{14} AvgFrStopTakingTime	0.5672	(1.0495)	0.5688	(1.0496)	-0.2131	(2.0015)	-0.1892	(2.0016)
	θ_{15} FrCredence	-0.8204	(0.5542)	-0.8209	(0.5542)	0.6750	(1.0480)	0.6695	(1.0481)
	θ_{16} AvgFrCredence	1.2906	(0.9188)	1.2907	(0.9188)	-1.8382	(1.8524)	-1.8376	(1.8525)
Treatment	α_1 NumPatient	1.4952	(6.4051)	1.6816	(6.4596)	6.0692	(11.7204)	7.7492	(11.8075)

Characteristics	α_3 NumPatientRating	-2.3732	(6.5980)	-2.6016	(6.6716)	8.0323	(12.5988)	6.1944	(12.0730)
	α_4 AvgDosage	0.4002	(1.0898)	0.4091	(1.0902)	2.4326	(2.0700)	2.5667	(2.0700)
	α_5 VarDosage	-91.3305**	(37.0934)	-91.6621**	(37.1339)	-78.0490	(54.8243)	-80.1488	(55.8129)
	α_6 AvgDuration	1.0594	(1.7242)	1.0895	(1.7300)	0.3623	(3.2470)	0.6746	(3.2626)
	α_7 VarDuration	52.9963***	(11.7471)	52.9815***	(11.7483)	17.8749	(23.6234)	17.5144	(23.5406)
	α_8 AvgStopTakingTime	-1.4761	(1.9316)	-1.5238	(1.9278)	-1.5590	(3.6253)	-1.7590	(3.6221)
	α_9 VarStopTakingTime	18.3066	(13.0599)	18.3763	(13.0524)	49.0535*	(25.6234)	49.4601*	(25.6221)
	α_{10} AvgCost	2.9834	(4.2705)	2.9198	(4.2785)	-0.2295	(7.7868)	-0.6708	(7.8006)
	α_{11} VarCost	-1051.7200	(666.0816)	-1051.3800	(666.0817)	2123.9320*	(1239.4260)	2148.7350*	(1247.0700)
	α_{12} NumTopics	-0.1256	(0.1092)	-0.1273	(0.1094)	0.0456	(0.2078)	0.0201	(0.2086)
Patient Characteristics	β_1 Female	6.3873	(5.4365)	6.3871	(5.4364)	10.7592	(10.2815)	10.7142	(10.2813)
	β_2 Age	0.1338	(0.1866)	0.1337	(0.1866)	-0.2141	(0.3532)	-0.2123	(0.3531)
	β_3 DataQuality	1.6682	(2.4331)	1.6686	(2.4331)	-0.7935	(4.5850)	-0.7881	(4.5847)
	β_4 InstantMood	-28.2576	(19.2924)	-28.2613	(19.2924)	2.2907	(36.3835)	2.0597	(36.3848)
	β_5 Membership	0.0957	(0.1794)	0.0978	(0.1793)	-0.3518	(0.3445)	-0.3167	(0.3446)
	β_6 NumSymptoms	5.7861	(10.4677)	5.7885	(10.4677)	-46.7668**	(19.8863)	-46.7251**	(19.8872)
	β_7 AvgSeverity	-0.1870	(2.4237)	-0.1876	(2.4236)	-1.8390	(4.5879)	-1.8419	(4.5880)
	β_8 DiseaseDiagnoseTime	-0.0005	(0.1887)	-0.0010	(0.1888)	-0.3734	(0.3579)	-0.3809	(0.3579)
	β_9 TimeonDrug	0.3485**	(0.1743)	0.3486**	(0.1743)	0.6541**	(0.3300)	0.6547**	(0.3300)
	β_{10} CurrentDosage	0.1067	(0.0702)	0.1068	(0.0702)	0.1097	(0.1327)	0.1100	(0.1327)
	β_{11} PatientCred	0.0913	(0.2043)	0.0914	(0.2043)	0.0116	(0.3861)	0.0143	(0.3861)
	β_{12} PreviousEval * Time	0.5681	(0.7415)	0.5699	(0.7315)	-1.0494	(1.4291)	-1.0226	(1.4283)

Significance code: 0 ‘***’ 0.01‘**’ 0.05‘*’

Table 26: Estimates for Latent Growth Model[9]

	Est	**(SE)**
Structural Model		
Regression Coefficients		
γ_1 NumRatingReviews	0.3185^{***}	(0.0507)
γ_2 AvgRating	-17.3764^{**}	(0.7403)
γ_3 VarRating	22.5734^{***}	(3.4689)
γ_4 PosComment	-0.3158^{***}	(0.0566)
γ_5 CommentEntropy	-3.8617	(5.2909)
γ_6 Credence	0.0060	(5.2909)
θ_1 NumFr	-0.1464	(0.1067)
θ_2 NumFrTakingTreatment	2.6606	(7.0709)
θ_3 NumFrRatingReviews	31.6482^{*}	(12.3691)
θ_4 AvgFrRating	0.7590	(0.7876)
θ_5 VarFrRating	-2.6132^{*}	(1.5286)
θ_6 PosFrComment	-0.6284	(2.5351)
θ_7 FrCommentEntropy	-5.3509^{**}	(2.6204)
θ_8 FrCredence	-0.0054	(0.1591)
θ_9 AvgFrDosage	-0. 4166	(0.3003)
θ_{10} AvgFrDuration	0.2326	(0.3039)
θ_{11} AvgFrStopTakingTime	0.0336	(0.3035)
α_1 NumPatient	15.5919^{***}	(1.8732)
α_2 NumPatientRating	-26.2171^{***}	(1.9200)
α_3 AvgDosage	0.4252	(0.3178)
α_4 VarDosage	0.4142	(8.7980)
α_5 AvgDuration	1.5645^{***}	(0.4960)
α_6 VarDuration	4.9871	(3.7111)
α_7 AvgStopTakingTime	-1.1528^{**}	(0.5515)
α_8 VarStopTakingTime	2.0530	(3.9404)
α_9 AvgCost	-2.2597	(1.1970)
α_{10} VarCost	-291.215	(200.227)
α_{11} NumTopics	0.0588^{*}	(0.0309)
β_1 Female	-0.8385	(1.5841)
β_2 Age	-0.0684	(0.0543)
β_3 DataQuality	0.5372	(0.7071)
β_4 InstantMood	0.8265	(5.6191)
β_5 PatientCred	0.0617	(0.0583)
β_6 NumSymptoms	-1.6113	(3.0458)
β_7 AvgSeverity	-0.0532	(0.7082)

---

[9] Variables are checked for the problem of multicollinearity. NumComment, PosComment, NumPatient, and NumPatientRating are log-transformed and mean centered. The mean VIF is less than 2.25. Rescaling is performed to remove the potneital correlation: NumRatingReviews, NumRatingReviews^2, NumFr, NumFrTakingTreatment, InstantMood, Membership, NumSymptoms, DiseasediagnoseTime, TimeonDrug, and CurrentDosgae are scaled down by factor of 1000; and other variables are scaled down by a factor of 100.

β_8 DiseaseDiagnoseTime	-0.0192	(0.1539)
β_9 TimeonDrug	0.0056	(0.0506)
β_{10} CurrentDosage	0.0617	(0.0210)
β_{11} PreviousEval * Time	45.9827^{***}	(0.6311)
Measurement Model		
Intercepts		
δ_1 [Efficacy]	0	
δ_2 [Side Effect]	-0.0167	(0.0362)
δ_3 [Persistency]	-0.0638	(0.0365)
δ_4 [Hardship]	-0.0972*	(0.0364)
Factor Loading		
λ_1 [Efficacy]	1	
λ_2 [Side Effect]	0.9983^{***}	(0.0176)
λ_3 [Persistency]	1.0214^{***}	(0.0180)
λ_4 [Hardship]	1.0318^{***}	(0.0181)
Thresholds		
κ_1	0.1783	(0.1253)
κ_2	2.1171^{***}	(0.1257)
κ_3	3.7658^{***}	(0.1268)

Significance code: 0 ‘***’ 0.01‘**’ 0.05‘*’

Chapter 6 Conclusion and Directions for Future Research

Getting connected in virtual space has been recognized for decades as leading to health benefits. In the late 1990s, scholars advocated the advantages of sharing ideas and experiences with others through online health support groups (Eng et al. 1999). Today, the online healthcare community is believed to be the single most important aspect of the web with the biggest impact on health outcomes (Eysenbach et al. 2004, O'Grady et al. 2008). Social media communities have recently been particularly active in the healthcare domain (Kane et al. 2009), and this intersection of healthcare and social media represents a promising space for future IS research and has attracted many researchers' attention (Fichman et al. 2011). The primary driver of value in online communities—commons-based peer production (Benkler 2002)—appears especially well suited to healthcare (Fichman et al. 2011).

In this dissertation, I present what I believe to be the first measurable evidence that patients benefit by participating in online healthcare communities, which help them improve their health conditions and better engage in their disease self-management process. I then study how patients in an online health social network identify and connect to other people to obtain useful and timely information. In the third study, I investigate online word of mouth, especially online treatment reviews in the presence of substantive experiential information shared from various sources, and the impact of experiential information sharing on patient's perception of treatment quality.

As my early studies of the patient–patient dimension in online healthcare communities suggest, there are quite a few promising directions for future research. The first could be additional intensive study of the social value generated from the dimension of patients helping patients. For example, to what extent does this platform help patients in their disease management process? Furthermore, as social needs differ at various disease stages, research is needed to identify the most important needs for patients with respect to their current health conditions.

Another valuable direction of study would be how to use online healthcare communities to improve physician–patient relations. This relationship has long been strained, partly because of the high information asymmetry in the healthcare field, allowing physicians to act autonomously. However, with increasingly easy access to health information and the ability to assess the quality of health service, patients now have the potential to change this predicament. Another potentially fruitful area of research in this realm would be how to encourage doctors to participate in online healthcare communities to learn more about their patients.

A third direction for future study could be the health service delivery system. Important questions, such as how to use these interactive platforms to improve health service delivery, are worth rigorous investigation.

In sum, the interface of social media and healthcare exemplifies a promising space for IS research. My studies are a starting point. But much additional work is needed so that OHCs can not only enhance patients' experiences and outcomes, but also contribute to improving the healthcare system itself.

Reference

Reference of Essay One:

Alexander, G.C., C.W. Tseng. 2004. Six strategies to identify and assist patients burdened by out of pocket prescription costs. *Cleveland Clinic Journal of Medicine* 71(5) 433-437.

Anderson, C., R. Agarwal. 2009. Privacy risks emotions and consumers' willingness to share personal health information. Research Briefing, University of Maryland 4 (1A).

Anderson, D.R., K.P. Burnham, G.C. White. 1998. Comparison of AIC and CAIC for model selection and statistical inference from capture-recapture studies. *Journal of Applied Statistics* **25** 263 – 282.

Bambina, A.D. 2007. Online Social Support: The interplay of Social Networks and Computer-Mediated Communication.

Benkler, Y. 2002. Coase's penguin, or, Linux and the nature of the firm. Yale Law J. 112(3) 369–446.

Berger, P.L., T. Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Anchor, First Thus edition, July 11, 1967.

Berkman, L.A., L. Breslow. 1984. Health and Ways of Living: The Alameda County Studies. *The Journal of Ambulatory Care Management* 7(1) 80.

Berkman, L.F., T. Glass, I. Brissette, T.E. Seeman. 2000. From social integration to health: Durkheim in the new millennium. *Social Science and Medicine* 51 843-857.

Chernomas, W.M, Clarke, D.E. Social Support and Women Living with serious Mental Illness. *Project #23 of the Prairie Women's Health Centre of Excellence* **ISBN 0-9689692-4-9**.

Christensen H, Griffiths K. 2000. The Internet and mental health literacy. *Aust N Z J Psychiatry* **34**(6) 975-9.

Clark, C.M. Relations Between Social Support and Physical Health. 2006. Working paper, Rochester Institute of Technology.

Cline, R.J.W., K.M. Haynes. 2001. Consumer health information seeking on the Internet: the state of the art. *Health Education Research* 16(6) 671-692.

Cobb, S. 1976. Social Support as a Moderator of Life Stress. *Psychosomatic Medicine* 38(5).

Conati, C., A. Gertner, K. Vanlehn. 2002. Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Modeling and User-Adapted Interaction* 12(4) 371-417.

Eriksson, E., S. Lauri. 2000. Informational and emotional support for cancer patients' relatives. *European Journal of Cancer Care*. 9: 8–15. doi: 10.1046/j.1365-2354.2000.00183.x

Fichman R. G., Kohli, R., Krishnan, R. and Kane, G.C. 2011. The Role of Information Systems in Healthcare: Current Research and Future Trends, lead article by the senior editors in *Information Systems Research,* special issue on Healthcare IT.

Forman, C., A. Ghose, B. Wiesenfed. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research* 19(3).

G. Reeves, Greco, P., J.S. Pendley, K. McDonell. 2001. A peer group intervention for adolescents with type 1 diabetes and their best friends. *Journal of Pediatric Psychology* 26(8) 485-490.

Hamilton, J.D. 1989. A New Approach to the Economic-Analysis of Nonstationary Time-Series and the Business-Cycle. *Econometrica* 57(2) 357-384.

Hauser, J.R., G.L. Urban, G. Liberali, M. Braun. 2009. Website Morphing. *Marketing Science* 28 202 - 223.

Heckman, J., Singer, B., 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* **52**(2) 271-320

Hellerstein, M. 1999. Measurement of cell proliferation: Special problems for T lymphocytes and recent methodological advances. *Immunol. Today* 20 438–441.

Hensher, D.A., W.H. Greene. 2003. The Mixed Logit model: The state of practice. *Transportation* 30(2) 133-176.

Hunt, S.M., S.P. Mckenna, J. Mcewen, J. Williams, E. Papp. 1981. The Nottingham Health Profile – Subjective Health-Status and Medical Consultations. *Social Science and Medicine Part A-Medical Sociology* 15(3) 221-229.

Jacoboson, D.E., 1986. Types and Timing of Social Support. *Journal of health and Social Behavior* **27**(September) 250-264.

Janakiraman, R., S. Dutta, C. Sismeiro, P. Stern. 2008. Physicians' Persistence and Its Implications for Their Response to Promotion of Prescription Drugs. *Management Science* 54(6) 1080-1093.

Kaelbling, L.P., M.L. Littman, A.R. Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101 99-134.

Kane, G.C., Fichman, R.G., Gallaugher, J. and Glaser, J. 2009. Community Relations 2.0 *Harvard Business Review*.

Kassirer, J.P. 2000. Patients, Physicians, and the Internet. *Health Affairs* 19(6) 115-123.

Kennedy, P. 2003. A Guide to Econometrics. MIT Press.

King, G., Wiloughby, C., Specht, J.A., Brown, E. 2006. Social Support Processes and the Adaptation of Individuals With Chronic Disabilities. *Qualitative Health Research* **16**(7) 902-925.

Lamberg, L. 2003. Online Empathy for Mood Disorders-Patients Turn to Internet Support Groups. *American Medical Association* 289(233) 3073.

Landro, L. 2006. Social Networking Comes to Health Care. *Wall Street Journal* 1227.

Lazarus, R.S., Folkman, S. 1984. Stress, Appraisal and Coping. New York: Springer.

Leung, L. 2011. Loneliness, social support, and preference for online social interaction: the mediating effects of identity experimentation online among children and adolescents. *Chinese Journal of Communication* **4**(4) 381-399.

Ma, M., R. Agarwal. 2007. Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. *Information Systems Research* 18(1) 42-67.

McCorkle, B.H., Rogers, E. S., Dunn, E.C., Lyass, A., Wan, Y.M. 2008. Increasing Social Support for Individuals with Serious Mental Illness: Evaluating the Compeer Model of Intentional Friendship. *Community Mental Health Journal* **44** 359-366.

McLuhan, M., W.T. Gordon. 2002. *Understanding media*. Gingko Press.

Mcmullan, M. 2006. Patients using the Internet to obtain health information: How this affects the patient-health professional relationship. *Patient Education and Counseling* 63 24-28.

McPherson, M., L. Smith-Lovin, J.M. Cook. 2000. Birds of a Feather: Homophily in Social Networks? *Annu. Rev. Sociol.* 27 415-44.

Mead, G.H. 1967. *Mind, self and society*. University Of Chicago Press.

Merton, R.K. 1976. *Sociology Ambivalence and Other Essays*. Free Press, New York.

Monge, P.R., N. Contractor. 2003. *Theories of Communication Networks*. Oxford University Press, USA.

Netzer, O., J.M. Lattin, V. Srinivasan. 2008. A Hidden Markov Model of Customer Relationship Dynamics. *Marketing Science* 27(2) 185-204.

Oluwole, D.A., Hammed, A.T., Awaebe, J. 2011. Patterns of Stress, Social Spoort, and Mental Health among Nigerian Women. *Advancing Women in Leadership* AdvancingWomen.com.

Pearlin, L.I. 1985. Social Structure and Processes of Social Support. *Social Support and Health* 43-60.

Porter, C.E., N. Donthu. 2008. Cultivating Trust and Harvesting Value in Virtual Communities. *Management Science* 54(1) 113-128.

Powell, J., Clarke, A. 2006. Internet information-seeking in mental health: Population survey. *British Journal of Psychiatry* **189** 273-277

Rabiner, L.R. 1989. A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77(2) 257-286.

Rabiner, L.R., B.H. Juang, S.E. Levinson, M.M. Sondhi. 1986. Recognition of Isolated Digits Using Hidden Markov-Models with Continuous Mixture Densities. *AT&T Technical Journal* 64(6) 1211-1222.

Radley, A., M. Billig. 1996. Accounts of health and illness: Dilemmas and representations. *Sociology of Health & Illness* 18(2) 220-240.

Rideout, V. 2001. Generation Rx.com: How young people use the Internet for health information. Menlo Park, CA: Kaiser Family Foundation.

Singh, P.V., Y. Tan, N. Youn. 2011. A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. *Information Systems Research*. Forthcoming.

Slevin, M.L., S.E. Nichols, S.M. Downer, P. Wilson, T.A. Lister, S. Arnott, J. Maher, R.L. Souhami, J.S. Tobias, A.H. Goldstone, M. Cody. 1996. Emotional support for cancer patients: what do patients really want? *British Journal of Cancer* 74 1275-1279.

Tessler, R., D. Mechanic. 1978. Psychological Distress and Perceived Health Status. *J. Health and Social Behavior* 19(3) 254-262.

Umberson, D., Crosnoe, R., Reczek, C. 2010. Social Relationships and Health Behaviors across the Life Course. *Annual Review of Sociology* **36**139–57.

Wanless, D. 2002. Securing our future health: taking a long-term view. London: HM Treasury.

Wasserman, S., K. Faust. 1993. *Social network analysis: Methods and applications*. Cambridge University Press, Cambridge.

Weinberg, N., Schmale, J.D., Uken, J., Wessel, K. 1995. Computer-mediated support groups. *Social Work with Groups* **17** 43–54.

Weissman, M.M., R.C. Bland, G.J. Canino, S. Greenwald S, H.G. Hwu, P.R. Joyce, E.G. Karam, C.K. Lee, J. Lellouch, J.P. Lepine, S.C. Newman, M. Rubio-Stipec, J.E. Wells, P.J. Wickramaratne, H.U. Wittchen, E.K. Yeh. 1999. Prevalence of suicide ideation and suicide attempts in nine countries. *Psychological Medicine* 29(1) 9-17.

Williams, A.W., Ware, J.E., Donald, C.A. 1981. A Model of mental Health, Life Events, and Social Supports Applicable to General Populations. *Journal of Health and Social Behavior* **22**(4) 325-336.

Wittgenstein, L. 1953. *Philosophical investigations.* Oxford: Basil Blackwell.

Whittaker, E.T., Robinson, G. 1976. The Ruffini-Horner Method. *Calculus of Observations: A Treatise on Numerical Mathematics*, New York: Dover.

Wright, K. 2000. Computer-Mediated Social Support, Older Adults, and Coping. *Journal of Communication* **50** (3) 100-118.

Wortman C. & Conway T. 1985. The role of social support in adaptation and recovery in physical illness. *Social Support and Health* Wortman C. & Conway T. eds. Academic Press, New York.

Yabrra, M.L., Eaton, W.W. 2005. Internet-Based Mental Health Interventions. *Mental Health Services Research* **7**(2).

Zhang, G., Y. Tan. 2010. Learning from Past Failures: an Empirical Study of Security Vulnerabilities. Working paper, University of Washington.

Ziebland, S., A. Chapple, C. Dumelow, J. Evans, S. Prinjha, L. Rozmovits. 2004. How the internet affects patients' experience of cancer: a qualitative study. *British Medical Journal* 328 564-569.

Reference of Essay Two:

Bambina, A.D. 2007. *Online Social Support: The interplay of Social Networks and Computer-Mediated Communication*. Cambria Press.

Brass, D.J. 1985. Men's and women's networks: A study of interaction patterns and influence in an organization. *Academy of Management Journal*, 28: 327-343.

Brown, J.S., P. Duguid. 2000. *Mysteries of the region*. Stanford University Press.

Bouty, I. 2000. Interpersonal and interaction influences on informal resource exchanges between R&D researches across organizational boundaries. *Academy of Management Journal*. **43** 50-65.

Burgess, E., P. Wallin. 1943. Homogamy in Social Charastics. *American J. Sociology*. **49**(2) 117-124.

Butler, B.S. 2001. Membership size, community activity, and sustain ability: A resource-based model of online social structures. *Information Systems Research*. **12**(4) 346–362.

Burt, R.S. 2004. Structural holes and good ideas. *The American Journal of Sociology*. **110**(2) 349.

Byrne, D. 1971. *The attraction paradigm.* Orlando, FL: Academic Press.

Byrne, D. 1961. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*. **62**(3) 713-715.

Cohen, B.P., X. Zhou. 1991. Status Processes in Enduring Work Groups. *The American Sociological Review*. **56**(2) 179-188.

Coleman, J.S. 1988. Social capital in the creation of human capital. *The American Journal of Sociology*. **94** S95-S120.

Collins, F. 1988. Change-related behaviour and information systems. *Omega*. **16**(5) 369-381.

Coppola, N., S.R. Hiltz, N. Rotter. 2004. Building Trust in Virtual Teams. *IEEE Transactions on Professional Communication*. **47**(2) 95-104.

Denning, P. 2002. Internet time out. *Communications of the ACM* **45**(33).

Edmunds, A., A. Morris. 2000. The problem of information overload in business organizations: a review of the literature. *International Journal of Information Management* **20**(1) 17-28.

Eppler, M.J., J. Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *Information Society*.

Feld, S.L. 1981. The focused organization of social ties. *American Journal of Sociology* **86** 1015-35.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations, 7,*117-140.

Faust, K. 1988. Comparison of Methods for Positional Analysis: Structural and general Equivalence. *Social Networks*. **10** 313-41.

Faust, K., S. Wasserman. 1994. *Social network analysis: methods and applications*. Cambridge University Press.

Flew, T. 2008. *New Media: an introduction*. Third Edition, Oxford University Press: Australia.

Freeman, L.C. 1979.Centrality in social networks conceptual clarification. *Social Networks*. 1(3) 215-239.

Friedkin, N.E. 1991. Theoretical foundations for centrality measures. The *American Journal of Sociology*. **96**(6) 1478-1504.

Fukuyama, F. 1995. *Trust*. New York: Free Press.

Gladwell, M. 2000. The tipping point: How little things made a big difference. New York: Brown, Little , & Co.

Golbeck, J. 2009. Trust and Nuanced Profile Similarity in Online Social Networks. *ACM Transactions on the Web* **3**(4).

Golbeck, J. 2007. The dynamics of web-based social networks: Membership, relationships, and change. *First Monday* **12**(11).

Goodreau, S.M., J.A. Kitts, M. Morris. 2009. Birds of a Feature, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography* **46**(1) 103-125.

Gross, B.M. 1964. *The Managing of Organizations: The Administrative Struggle*. pp. 856.

Hagel, J., A.G. Armstrong. 1997. Net gain: expanding markets through virtual communities, *Mckinsey Quarterly* (Winter), 140–146.

Hall, D., Schneider, B., & Nygren, H. 1970. Personal factors in organizational identification. *Administrative Science Quarterly*. **15** 176-89.

Hoff, P.D., A.E. Raftery, M. Handcock. 2002. Latent Space Approaches to Social Network Analysis. *Journal of American Statistical Association*. **97** 1090.

Hubbell, C.H. 1965. An Input-Output Approach to Clique Identification. *Sociolmetry*. **28** 377- 399.

Jarvenpaa, S., D. Leidner. 1998. Communication and Trust in Global Virtual Teams. *Organization Science*. **10**(6) 791-815.

Jones, Q., G. Ravid, S. Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* **15**(2) 194-210.

Knoke, D., Burt, R.S. 1983. Prominence. In: R.S. Burt and M. Minor, Editors, *Applied Network Analysis*, Sage, Beverly Hills, CA. pp. 195–222.

Krackhardt, D., D.J. Brass. 1994. Intraorganizational networks: The microside. *Advances in social network analysis*. London: Sage.

Korzenny, F., C. Bauer. 1981. Testing the Theory of Electronic Propinquity. *Communication Research*. **8**(4), 479-498.

Lea, M., R. Spears. 1992. Social Influence and the influence of the "social' in computer-mediated communication. In M. Lea (Ed.), *Contexts of computer-mediated communication*. 30-65.

Lewis, J.D., A. Weigert. 1985. Trust as a Social Reality. *Social Forces*. 63(4) 967-985.

Lewis, D. 1996. Dying for Information? *London: Reuters Business Information*.

Lewis, K., J. Kaufman, M. Gonzalez, A. Wimmer, N. Christakis. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* **30**(4) 330-342.

Lin, N. *Social Capital: A Theory of Structure and Action*. London and New York: Cambridge University Press.

Lowe, J. 2004. *Health Information Overload*: *Health Topics*. University of Iowa.

Ma, M., R. Agarwal. 2007. Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. *Information Systems Research* 18(1) 42-67.

March, J.C., H.A. Simon. 1958. *Organizations*. New York: Wiley.

McMullan, M. 2006. Patients using the Internet to obtain health information: How this affects the patient-health professional relationship. *Patient Education and Counseling*. **63** 24-28.

McPherson, M., L. Smith-Lovin, J.M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27** 415-444.

Metzger, M.J. 2004. Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication*. **9**(4) 00.

Meyerson, D., K. Weick, R.M. Kramer. 1996. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research*. Sage Publications.

Monge, P.R., N. Contractor. 2003. *Theories of Communication Networks*. Oxford University Press.

Monge, P.R., L.W. Rothman, E.M. Eisenberg, K.I. Miller, K.K. Kirste. 1985. The dynamics of organizational proximity. *Management Science*. **31**(9) 1129-1141.

Newcomb, T. 1961. *The Acquaintance Process*. New York: Holt, Rinehart & Winston.

O'Reilly, C.A. 1980. Individuals and Information Overload in Organizations - Is More Necessarily Better? *Academy of Management Journal* **23**(4) 684-696.

Parks, M.R. 2007. Personal networks and personal relationsips. Mahwah, NJ:Lawrence Erlbaum Associates.

Pelled, L.S. 1996. Demographic Diversity, Conflict, and Work Group Outcomes: An Intervening Process Theory. *Organization Science*.**7**(6) 615-631.

Piccoli, G., B. Ives. 2003. Trust and the unintended effects of behavior control in virtual teams. *MIS Quarterly*. **27**(3) 365-395.

Preece, J. 1999. Empathic communities: balancing emotional and factual communication. *Interacting with Computers* **12**(1) 63-77.

Preece, J., K. Ghozati. 1999. Empathy online: A review of 100 online communities. In Proc. *Association for Information Systems*. Americas Conference (Baltimore, MD, 1998).

Preece, J., K. Ghozati. 2001. Observations and Explorations of Empathy Online. In. R. R. Rice and J. E. Katz, *The Internet and Health Communication: Experience and Expectations*. Sage Publications, Thousand Oaks. 237-260.

Putnam, R.D. 1993. The prosperous community: Social capital and public life. *American Prospect.* **13** 35-42.

Reagans, R. 2005. Preferences, identity, and competition: Predicting tie strength from demographic data. *Management Science* **51**(9) 1374-1383.

Rheingold, H. 1993. *The Virtual Community Homesteading on the Electronic Frontier*. The MIT Press.

Ridings, C.M., D. Gefen, B. Arinze. 2002. Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems. **11** 271-295.*

Roloff, M.E. 1981. *Interpersonal communication: The social exchange approach.* Sage Publication.

Royce, J. 1997. Surviving information overload: lessons from the reading reluctance research. *School Libraries Worldwide.* **3** 39–45.

Schick, A.G., L.A. Gorden, S. Haka. 1990. Information overload: A temporal approach. *Accounting Organizations and Society*. **15**:199–220.

Schroder, H.M., M.J. Driver, S. Streufert. 1967. Human information processing - Individuals and groups functioning in complex social situations. *New York: Holt, Rinehart, & Winston*.

Sheldon, M.E. 1971. Investments and Involvements as Mechanisms Producing Commitment to the Organization. *Administrative Science Quarterly*. **16**(2) 143-150.

Singh, P.V., Y. Tan, V.S. Mookerjee. 2011. Network Effects: The Influence of Structural Social Capital on Open Source Project Success. *Management Information Systems Quarterly*. Forthcoming.

Snijders, T.A.B. 2001. The statistical evaluation of social network dynamics. *Sociological methodology* 361–395. Boston and London: Basil Blackwell.

Snijders, T.A.B. 2005. Models for Longitudinal Network Data. Chapter 11 in P. Carrington.

Suitor, J.J., K. Pillemer, S. Keeton. 1995. When Experience Counts - the Effects of Experiential and Structural Similarity of Patterns of Support and Interpersonal Stress. *Social Forces* **73**(4) 1573-1588.

Souto, E., G. Guimaraes, G. Vasconcelos, M. Vieira, N. Rosa, C. Ferraz, J. Kelner. 2006. Mires: a publish/subscribe middleware for sensor networks. *Personal and Ubiquitous Computing*. **10**(1) 37-44.

Sztompka, P. 1999. *Trust: a sociological theory*. Cambridge University Press.

Thelwall, M. 2009. Homophily in MySpace. *Journal of the American Society for Information Science and Technology* 60(2) 219-231.

Turner, J.C., P.J. Oakes, S.A. Haslam, C. Mcgarty. 1986. Self and Collective-Cognition and Social-Context. *Personality and Social Psychology Bulletin* **20**(5) 454-463.

Van den Bulte, C., G.L. Lilien. 2001. Medical Innovation Revisited: Social Contagion versus Marketing Effort. *The American Journal of Sociology* **106**(5) 1409-1435.

Wang, Z.M., J.B. Walther, S. Pingree, R.P. Hawkins. 2008. Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health Communication* **23**(4) 368-368.

Wasko, M., S. Faraj. 2005. Why Should I Share? Examine Social Capital and Knowledge contribution in Electronic Networks of Practice. *MIS Quarterly* **29**(1) 35-57.

Watts, D.J., S.H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*. **393**:440-42.

Wilson, T.D. 2001. Information overload: implications for healthcare services. *Health Informatics Journal* **7** 112–7.

Wright, K. B. 2000. Perceptions of on-line support providers: An examination of perceived homophily, source credibility, communication and social support within on-line support groups. *Communication Quarterly*. **48** (1) 44–59.

Yan, L., Tan, Y. 2010. Feel Blue so Go Online: An Empirical Study of Online Supports among Patents. Working paper, Foster School of Business, University of Washington.

Yellowlees, P.M., P.M. Brooks. 1999. Health online: the future isn't what it used to be. *The Medical Journal of Australia*. **191** 522-525.

Zenger, T.R., B.S. Lawrence. 1989. Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal*, 32, 353-376.

Reference of Essay Three:

Aggarwal, R., Gopal, R., Sankaranarayanan, R., Param, V.S. Blog, Blogger, and the Firm: Can Negative Employee Posts Lead to Positive Outcomes? *Information Systems Research* Forthcoming.

Archak, N., Ghose, A., Iperiotis, P.G. 2011. Deriving the Pricing Power of Product Feature by Mining Consumer Reviews. *Marketing Science* forthcoming.

Ben-Tovim D.I, Dougherty M.L.G., Stapleton A.M.F., Pinnock C.B. 2002. Coping with prostate cancer: a quantitative analysis using a new instrument, the centre for clinical excellence in urological research coping with cancer instrument. *Urology*. **59** 383-8.

Boulding, W., Kalra, A., Staelin, R. 1999. The Quality Double Whammy. *Marketing Science* **18**(4) 463-484.

Cao, Q., Duan, W., Gan, Q. 2011. Exploring determinants of voting for the "helpfulness" of online user views: a text mining approach. *Decision Support Systems* **50**(2011) 511-521.

Chen, Y., Xie, J. 2008. Online Consumer Review: A New Element of Marketing Communications Mix. *Management Science* **54**(3) 477–491.

Chen, P.Y., Wu, S.Y., Yoon, J. 2004. The Impact of Online Recommendations and Consumer Feedback on Sales. *Proceedings of the 25th International Conference on Information Systems* 711-724.

Chevalier, Judith A., Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345–354.

Cotten S.R., Gupta S.S. 2004. Characteristics of online and offline health information seekers and factors that discriminate between them. *Soc Sci Med*. **59**(9) 1795-806.

Duan, W., Gu, B., Whinston, A.B. 2009. Information Cascades and Software Adoption on the Internet: An Empirical Investigation. *MIS Quarterly* **33**(1) 23-48.

Epstein, R.M., Street, R.L. 2011. Shared Mind: Communication, Decision Making, and Autonomy in Serious Illness. *Annals of Family Medicine*. **9**(5) September/October 454-461.

Eysenbach G. 2007. From Intermediation to Disintermediation and Apomediation: New Models for Consumers to Access and Assess the Credibility of Health Information in the Age of Web2.0. *In Stud Health Technol Inform.* **129** 162-166.

Fay, S., J. Xie. 2008. Probabilistic goods: A creative way of selling products and service. *Marketing Science* **27**(4) 674-690.

Fichman R. G., Kohli, R., Krishnan, R. and Kane, G.C. 2011. The Role of Information Systems in Healthcare: Current Research and Future Trends, lead article by the senior editors in *Information Systems Research,* special issue on Healthcare IT.

Godes, D., Mayzlin, D. Using Online Coversations to Study Word-of-Mouth Communication. *Marketing Science* **23**(4) 545-560.

Hartzler1, A., Pratt, W. 2011. Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians. *J Med Internet Res*. **13**(3) e62.

Hauser, J.R., Urban G.L., Liberali, G., Braun, M. 2009. Website Morphing. *Marketing Science* **28**(2) 202-223.

Herr, P.M., Kardes, F. R., Kim, John. 1991. Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research* **17**(4) 454-62.

Iyengar, I., Van den Bulte, C., Valente, T.W. 2011. Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science* March/April **30**(2) 195-212.

Kane, G.C., Fichman, R.G., Gallaugher, J. and Glaser, J. 2009. Community Relations 2.0 *Harvard Business Review*.

Kvalem I., Traeen B. 2000. Self-efficacy, scripts of love and intention to use condoms among Norwegian adolescents. Journal Youth and Adolescence **29***(*3) 37–53.

Lee, Thomas, Eric Bradlow. 2007. Automatic construction of conjoint attributes and levels from online customer reviews. Working Paper, The Wharton School, University of Pennsylvania.

Lee, Y.J., Tan, Y. Hosanagar,K. Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Rating. Working paper.

Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. Journal of Marketing 70(July) 74-89.

Logan RA, Tse T. 2007. A multidiscipline conceptual framework for consumer health informatics. *Stud Health Technol Inform*. **129** 1169-73.

Mayzlin, Dina. 2006. Promotional Chat on the Internet. *Marketing Science* **25** (2) 155-163.

Meyer, Robert J. 1981. A Model of Multiattribute Judgments under Attribute Uncertainty andInformational Constraint. *Journal of Marketing Research* **18**(4) 428–441.

Martin, Jolie, Gregory Barron and Michael I. Norton. 2007. Choosing to Be Uncertain: Preferences for High Variance Experiences. Working Paper, Harvard Business School.

Miller S.M. 1987. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. *J Pers Soc Psychol*. **52** 345-353.

Miller S.M. 1995. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease: implications for cancer screening and management. *Cancer*. **76** 167-177.

Mudambi S.M., Schuff, D. 2010. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly* **34**(1) 185-200.

Nambisan P. 2011. Information seeking and social support in online health communities: impact on patients' perceived empathy. *J Am Med Inform Assoc*. **18**(3):298-304.

O'Grady, L.A., Witteman, H., Wathen, C.N. 2008. The experiential health information processing model: supporting collaborative web-based patient education. BMC Medical Informatics and Decision Making. **8**(58).

O'Grady L.A. 2008. Web 2.0: Collaboration in web-based patient health education. *Health Education Research Trends*. Hauppauge, NY: Nova Science Publishers Inc.

Pavlou, P.A., Dimoka, A. 2006. The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller differentiation. *Information Systems Research* **17**(4) 392-414.

Polonsky, W.H., Skinner, T.C. 2010. Perceived Treatment Efficacy: An Overlooked Opportunity in Diabetes Care. *Clinical Diabetes* **28**(2) 89-92.

Sarasohn-Khan J. 2008. The wisdom of patients: Health care meets online social media. Oakland, CA: California Healthcare Foundation.

West, Patricia M. and Susan M. Broniarczyk. 1998. Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus. *Journal of Consumer Research* **25**(1) 38–51.

Williams-Piehota P, Pizarro J, Schneider TR, Mowad L, Salovey P. 2005. Matching health messages to monitor-blunter coping styles to motivate screening mammography. *Health Psychol.* **24** 58-67.

Wilson, E.J. D.L. Sherrell. 1993. Sources effects in communication and persuation research: A meta-analysis of effect size. *Journal of Academic Marketing Science* **21**(Spring) 101-112.

Sun, M. How Does Variance of Product Ratings Matter? *Management Science*, forthcoming.

Tong, S. T., Van Der Heide, B., Langwell, L., & Walther, J. B. (2008). Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *Journal of Computer-Mediated Communication* **13**(3) 531-549.

Xie, J., E. Gestner. 2007. Service escape: Profiting from customer cancellations. *Marketing Science* **26**(1) 18-30.

Yan, L., Tan, Y. 2010a. Feeling Blue so Going Online: An Empirical Study on Effectiveness of Virtual Social Networking. *Workshop on Health IT and Economics (WHITE), Washington DC, October*.

Yan, L., Tan, Y. 2010b. Homophily in Online Social Networking Website: The Way We Find Patients Like Us. *Proceedings of Workshop on Information Technology and Systems (WITS),* St. Louis, December.

Ziebland, S., A. Chapple, C. Dumelow, J. Evans, S. Prinjha, L. Rozmovits. 2004. How the internet affects patients' experience of cancer: a qualitative study. *British Medical Journal* **328** 564-569.

Appendix A to

Feel Blue so Go Online: An Empirical Study of Online Supports among Patients

A1. Statistics for Testing Hypothesis 3

To test Hypothesis 3, we compare the effect of emotional support ($\beta_{7,s}$) and informational support ($\beta_{6,s}$) for a given state *s*. Since we have rescaled the covariant for informational support by dividing a factor of 10, $\beta_{6,s}$ measures the effect of 10 units of informational support while $\beta_{7,s}$ measures 1 unit of emotional support. For state 1, we compare $\beta_{7,s}$ with $\beta_{6,s}$ (or, the effect of 5 units of informational support).

Parameter Difference between Emotional and Information Support		
State 1 (bad) $\beta_{7,1} - (\beta_{6,1}/2)$	State 2 (fair) $\beta_{7,2} - \beta_{6,2}$	State 3 (good) $\beta_{7,3} - \beta_{6,3}$
0.2999^{***} (0.0874)	0.1942^{***} (0.0719)	0.3474^{***} (0.1323)

Note: Standard errors in parentheses. *** Significant at 1%.

A2. Statistics for Testing Hypothesis 4

To test Hypothesis 4, we check whether the estimated coefficient for informational ($\beta_{6,s}$) or emotional ($\beta_{7,s}$) support varies across states. For example, we test whether the effect of informational support in state 1 is significantly different from that in state 2, that is, $\beta_{6,1} - \beta_{6,2}$.

	Parameter Difference between State		
	1 (bad) and 2 (fair)	1 (bad) and 3 (good)	2 (fair) and 3 (good)
info. support β_6	0.0709^{***} (0.0269)	0.0888^{*} (0.0482)	0.0179 (0.0255)
emo. support β_7	-0.1840^{***} (0.0538)	-0.3193^{***} (0.1067)	-0.1353^{***} (0.0197)

Note: Standard errors in parentheses. Significance level: *** 1%; * 10%.

A3. Comparison of Intrinsic Propensity to Contribute across States

Here, we test the constant coefficient $\gamma_{0,s}$ is different across state *s*. The following table shows the differences of these estimated parameters and their standard errors.

Parameter Difference between State		
1 (bad) and 2 (fair)	1 (bad) and 3 (good)	2 (fair) and 3 (good)
0.4802^{**} (0.2265)	0.7487^{***} (0.2182)	0.2685^{***} (0.0467)

Note: Standard errors in parentheses. Significance level: *** 1%; ** 5%.

A4. Calculation of State Transition Probabilities

The support points and probability masses for individual heterogeneity in the state transition are:

Support Point ξ_k	0	0.4135	1
Probability $P(\xi_k)$	0.4704	0.3454	0.1842

The transition thresholds are:

$\overline{\omega}_{1\to2}$	$\overline{\omega}_{1\to3}$	$\overline{\omega}_{2\to3}$	$\underline{\omega}_{2\to1}$	$\underline{\omega}_{3\to1}$	$\underline{\omega}_{3\to2}$
0.9351	2.8495	2.4762	−2.519	−2.9527	−1.0021

The intrinsic propensity to transition from i to j (Table 7) is calculated as follows with the rescaling parameter $C_\xi = -0.151$. Specifically, for $i = 1, j = 2$:

$$P_{1,2} = \sum_{k=1}^{3}\left(\frac{\exp\left(\overline{\omega}_{1\to3} - C_\xi \xi_k\right)}{1+\exp\left(\overline{\omega}_{1\to3} - C_\xi \xi_k\right)} - \frac{\exp\left(\overline{\omega}_{1\to2} - C_\xi \xi_k\right)}{1+\exp\left(\overline{\omega}_{1\to2} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.2199;$$

for $i = 1, j = 3$:

$$P_{1,3} = \sum_{k=1}^{3}\left(1 - \frac{\exp\left(\overline{\omega}_{1\to3} - C_\xi \xi_k\right)}{1+\exp\left(\overline{\omega}_{1\to3} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.0523;$$

for $i = 2, j = 1$:

$$P_{2,1} = \sum_{k=1}^{3}\left(\frac{\exp\left(\underline{\omega}_{2\to1} - C_\xi \xi_k\right)}{1+\exp\left(\underline{\omega}_{2\to1} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.0781;$$

for $i = 2, j = 3$:

$$P_{2,3} = \sum_{k=1}^{3}\left(1 - \frac{\exp\left(\overline{\omega}_{2\to3} - C_\xi \xi_k\right)}{1+\exp\left(\overline{\omega}_{2\to3} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.0741;$$

for $i = 3, j = 1$:

$$P_{3,1} = \sum_{k=1}^{3}\left(\frac{\exp\left(\underline{\omega}_{3\to1} - C_\xi \xi_k\right)}{1+\exp\left(\underline{\omega}_{3\to1} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.0521;$$

and for $i = 3, j = 2$:

$$P_{3,2} = \sum_{k=1}^{3}\left(\frac{\exp\left(\underline{\omega}_{3\to2} - C_\xi \xi_k\right)}{1+\exp\left(\underline{\omega}_{3\to2} - C_\xi \xi_k\right)} - \frac{\exp\left(\underline{\omega}_{3\to1} - C_\xi \xi_k\right)}{1+\exp\left(\underline{\omega}_{3\to1} - C_\xi \xi_k\right)}\right)P\left(\xi_k\right) = 0.2264.$$

A5. Expected Intrinsic Number of New Posts

The support points and probability masses for individual heterogeneity in the outcome model are:

Support Point η_k	0	0.3147	0.5289	1
Probability $G(\eta_k)$	0.0612	0.3474	0.4011	0.1903

The rescaling parameter is $C_\eta = -0.209$. The expected number of new posts for state 1:

$$\sum_{k=1}^{4} \exp\left(\gamma_{0,1} + C_\eta \eta_k\right) G(\eta_k) = 2.34;$$

for state 2:

$$\sum_{k=1}^{4} \exp(\gamma_{0,2} + C_\eta \eta_k) \Pr(\eta_k) = 1.46;$$

for state 3:

$$\sum_{k=1}^{4} \exp(\gamma_{0,3} + C_\eta \eta_k) \Pr(\eta_k) = 1.12.$$

Appendix B to

Shared Minds: How Patients Use Collaborative-Based Information Sharing

The entropy of conversation is operationalized at both community and friend levels. Let $reviews_{t,j,k}$ be the number of text reviews patient j write for the treatment k between time t-1 to t . Different from *NumComment* which is an aggregated measure for text reviews, the volume of patients' reviews in this period is thus defined as $reveiws_{tk} = \sum_{j=1}^{J} reviews_{t,j,k}$. Following the common practice in information theory literature, we define the entropy for community level and friend level respectively as follows:

$$ComEntropy_{t,k} = \begin{cases} -\sum_{j=1}^{J} \frac{reviews_{jtk}}{reviews_{tk}} \log\left(\frac{reviews_{jtk}}{reviews_{tk}}\right), & \text{if } reviews_{tk} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

and

$$FrEntropy_{t,k} = \begin{cases} -\sum_{j=1}^{J} \frac{frReviews_{jtk}}{frReviews_{tk}} \log\left(\frac{frReviews_{jtk}}{frReviews_{tk}}\right), & \text{if } frReviews_{tk} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Compared with the friend reviews, community based treatment evaluations are gathered from patients suffering from the similar disease but with various health conditions. In the contrast, a patient and her friends are more similar in terms of sharing some symptoms and treatments, and they contact frequently to exchange disease progress information and other experiential information. Therefore, we expect that a patient is more likely to be influenced by the community based opinion than by the opinions contributed by her close contact because the treatment reviews do not provide extra value in addition to their social interactions. As such, the entropy generated from the friend network is less influential than the entropy generated from the community.

Vitae

Lu Yan was born in China. She received her Bachelor’s degrees in Computer Science from Nanjing University of Aeronautics and Astronautics in 2001, and Master degree in Computer Science from San Diego State University in 2005. Before joined the Ph.D. program at the University Of Washington Foster School Of Business, she worked as an assistant software engineer at Air China, China. She is broadly interested in the topics that identify social capital and social influence in the online settings, with a specific focus on online healthcare systems. Her current research studies how online healthcare community changes patients' role in the disease management process and the consequent impact to the healthcare industry. Her work has appeared in the proceedings of conferences such as ICIS Workshop on Information Technology and Systems (WITS), INFORMS Conference on Information Systems and Technology (CIST), and Workshop on Health IT and Economics (WHITE).