Abstract

As the Internet becomes an important source for health information, user-centric applications have arisen to empower end-users to self manage their illness. Users are able to share personal experiences in ways that have not been possible through the patient-clinician channel. Online health forums are a type of user-centric platform. In these forums it is difficult to evaluate the credibility of user-generated content because much of the communication is subjective, discursive, experiential and anonymous. Furthermore, users’ capacity to cognitively scrutinise incoming information can be effected by the high stakes or severe consequences of their illness. This work examines how users assess the health information generated by other users. We have used both qualitative and quantitative methods to explore how users develop their assessment criteria in judging the credibility of different types of information. Five criteria were discovered in the initial qualitative work. We then quantitatively tested the relationship between those criteria based on two types of information. Our analysis shows that different criteria are used by participants in online health forums for scientific information and experiential information. We used these novel findings to develop a model for how information credibility is assessed in online health forums.

Keywords: eHealth, Information Credibility, User-Generated Content, Online Community.
1 Introduction

Online health information exists in two forms: content published by institutions or organisations with established reputations (e.g., pharmaceutical companies, government agencies, research and educational institutes, media organisations, etc.), and content generated by end-users on user-centric platforms (e.g., forums, social networking sites, blogs, wikis and other social media). Both sources are commonly sought. For example, millions of Americans visited the websites of institutions and organisations during the swine flu pandemic to obtain information (Allen, 2009). Around 80% of Internet users particularly look for tailored information, and 41% of health information consumers have read someone else’s commentary or experience about health or medical issues on an online news groups, website, or blog (Fox and Jones, 2009).

While its convenience and anonymity makes the Internet an attractive medical information source, these same characteristics create risks. There is no uniform mechanism to ensure the accuracy, objectivity, truthfulness, and understandability of online health information (Pourmand and Sikka, 2011). Moreover, only a small proportion of the online health information has been reviewed or moderated by a medical professional (Fox and Rainie, 2000). Additionally, information can be jargon-laden and difficult to comprehend (Smart and Burling, 2001), and misusing online health information may lead to a worsening of the existing condition, delay of treatment, violation of privacy, or feelings of distress or trauma (Benigeri and Pluye, 2003).

Studies to date of how people evaluate online health information have tended to focus on information provided by institutions and organisations, and have therefore ignored online health content produced by other end-users (Metzger, Flanagin and Medders, 2010). Unlike end-users, these central authorities devote substantial resources to ensuring that the information they publish has merit and comes from a credible source (Callister Jr., 2000). Studies of these sites generally assess the attributes of information quality, such as source credibility, accuracy, currency, and completeness (Wang and Lui, 2007). However, many people rely heavily on websites where the content is produced by anonymous peers (Metzger et al., 2010), and user-generated health content exhibits other characteristics which challenge the adequacy of the existing evaluation criteria. In any event, ability to assess the source is negatively impacted by the fact that user-generated content is discursive, subjective, experiential, and “quality signals become more faint and diffused, and cognitive overload occurs frequently” (pp. 244) (Agarwal, Gupta and Kraut, 2008). Thus, the aim of this research is to obtain an in-depth understanding on how users assess user-generated health content online.

From amongst all user-centric applications, online health forums (OHFs) have been chosen for this research. Firstly, OHFs highlight text-based content and are also the most popular tool adopted (Lee, Vogel and Limayem, 2003). In contrast to blogs, OHFs centralise users in one place and better maintain users’ anonymity. As opposed to wikis, OHFs enable social interaction and influence between users. To maintain the communication dynamics between users and assess online health information produced by anonymous users, we exclude those OHFs that are explicitly involved with and moderated by medical professionals such as general medical practitioners, specialists, and registered nurses which tend to be more like Q and A and lack emotional exchange. By removing a central authority and medical professionals, we can better answer the research question: How do people assess health information generated by other anonymous end-users in OHFs?

The remainder of this paper is structured as follows: we start with a review of the key concepts and theories that guide our exploration, and then describe the methodology used to execute this research. We then present the emerging results from two studies, and finally discuss the insights and implications of our findings.

2 Theoretical Background

2.1 Online Health Forums and Users

As a repository of the wisdom of crowds, OHFs are equivalent to online support groups where people with similar medical conditions gather together and exchange information, knowledge, experience, feelings and support. First-hand experience with a medical condition determines community membership and the boundary of a group, and forms the source of knowledge in an online health community (Borkman, 1999). The prerequisite of the shared medical condition sets OHFs apart from other online communities.

What drives people to OHFs is the possession of a sense of alienation, feelings of isolation, anxiety about treatment and misconceptions and misinformation (McKenna, Wellisch and Fawzy, 1995). Emotional distress from a medical condition can be alleviated by satisfying informational needs (Hu, Bell, Kravitz and Orrange,
2.2 Information Quality vs. Information Credibility

Information quality is a subjective judgment of the goodness and usefulness of information to complete a particular task (Hilligoss and Rieh, 2008). Though many dimensions of information quality have been identified, a smaller core set of dimensions is used widely, including accuracy (intrinsic quality), completeness and currency (contextual quality), and format (representational quality) (Nelson, Todd and Wixom, 2005; Wang and Strong, 1996).

Credibility, the other major concept of this research, is the believability of information (Hilligoss and Rieh, 2008). It is closely associated with information quality to the extent that some researchers argue information credibility as an attribute of information quality (Flanagin and Metzger, 2000). Credibility is a complex factor, with research commonly partitioning the concept into multiple components such as information credibility, source credibility, media credibility and web credibility. Information credibility can be evaluated based on information itself, heuristic cues and interaction (Hilligoss and Rieh, 2008). As one heuristic cue, source credibility can be used to signal information credibility, because a credible source is more likely to produce credible information (Fragale and Heath, 2004). Cognitive authority, similar to source credibility, specifies a person as a credible source (Rieh, 2002).

Source credibility is operationalised using two dimensions: trustworthiness (an information provider’s intention to tell the truth or give unbiased information) and expertise (an information provider’s ability and/or qualification to know the truth of a topic) (Dutta-Bergman, 2004). Source credibility can be judged through presumed credibility (stereotyping), reputed credibility (source label such as doctor or professor), surface credibility (simple inspection of superficial features) and experienced credibility (direct experience with a source over time) (Tseng and Fogg, 1999). In our context, we are only interested in how users assess credibility of information rather than a contributor, so experienced credibility is not applicable. We exclude reputed credibility by removing the involvement of medical professionals. And presumed credibility is unlikely to be relevant due to anonymity. Thus, only surface credibility is appropriate.

2.3 Information Processing Theories

There are two main streams of information processing theories. The cognitive “miser” approach believes that people tend to utilise the least cognitive effort by taking shortcuts and processing a small amount of information for decision making (Fiske and Taylor, 1991). Prominence-interpretation theory, for example, argues that prominent aspects of a website dominate the evaluation process (Fogg, Kameda, Boyd, Marshall, Sethi, Sockol and Trowbridge, 2002). The limited capacity model of message processing (Lang, 2000) reinforces this view, proposing that limited cognitive capacity forces people to select only salient characteristics to encode, store and retrieve messages, particularly in situations where people feel overwhelmed. In these situations, people tend to seek mental shortcuts such as heuristic cues (stereotypes and categories) to minimise the amount of time and cognitive effort needed to process a message (Gigerenzer and Todd, 1999). Empirical studies show that in practice, verification methods that minimise time and cognitive effort tend to be used (Flanagin and Metzger, 2007).

However, the cognitive miser position has been challenged by dual-process theories, suggesting that people can be motivated to exert cognitive effort. The heuristic-systematic model (HSM) (Chen and Chaiken, 1999) and the Elaboration Likelihood Model (ELM) (Petty and Cacioppo, 1986) assert that when individuals’ motivation and ability to process are high, they are inclined to scrutinise all relevant information stemming from the source, message, context, and themselves (e.g. emotions) in an attempt to make an accurate judgment (referred to as the systematic or central route). When either motivation (e.g., personal relevance) or ability to process is low (e.g., insufficient knowledge), attitudes can be changed by relatively low-effort processes.
(referred to as the heuristic or peripheral route). In the health domain, the higher stakes of following the wrong advice and disclosing to the wrong people can motivate users towards effortful evaluation. Nonetheless, without years of systematic medical training, people may not even know when to be sceptical and be susceptible to misinformation (Hargittai, 2007). This challenges the predictability of the dual-process models.

The concept of Bounded Rationality provides a promising explanation. It states that, constrained by the limited capacity of the human brain and external resources, people adaptively seek a balance between cognitive effort and efficient outcomes of decision making (Gigerenzer and Todd, 1999). Thus, people tend to look for a satisfactory solution not an optimal one. For example, rather than understand the underlying mechanism, users tend to seek consistent pattern across various sources as an acceptable outcome (Nettleton, Burrows and O'Malley, 2005). When judging answers in a social question-and-answer site, the logic or plausibility of arguments as well as an answerer’s spelling/grammar, profile and credentials appear to be the most frequently used criteria (Kim, 2010). However, those studies fail to explain how information credibility is assessed when users seek satisfactory solutions in specific contexts. This research intends to answer this question.

3 Research Design and Methods

The research design contains two phases of investigation. In the first phase, a qualitative study was conducted to understand what criteria are used in evaluating user-generated health content and reveal the criteria applied by OHF users. In the second phase, a quantitative study was conducted to validate the findings from Study 1, and to explore the strategies used to assess different types of information. The research design highlights the purposes and strengths of the qualitative and quantitative studies, by first identifying a set of possible factors contributing to users’ judgments and then testing the strength of relationships (Hovorka and Lee, 2010).

3.1 Study 1: Qualitative Exploration

3.1.1 Data Collection and Analysis

The theoretical population is defined as every user of all computer-mediated asynchronous message boards, focusing on any physical condition caused by illness, injury, pain or discomfort. Simultaneously, we exclude any mental disorders and spiritual beliefs. Moreover, to eliminate potential confounding factors, we restrict the sampling population to users who have resided in Australia for at least two years. A purposive sampling approach was chosen to maximise the diversity in the sample (Miles and Huberman, 1994), with regard to medical conditions, user’s experience with forums, gender and age mix, education levels, purposes and roles (patient or care giver). The snowballing technique was employed to increase the chance to reach appropriate potential participants, by asking participants to recommend future interview candidates. Participants were recruited from an Australian university and a number of Australia-based OHFs. Recruiting from a university increased the chance of attracting lurkers (i.e., user who only consumer information without contributing), who are also important as part of the theoretical population.

Table 1 lists the medical conditions appearing among the interview participants. Each condition is decomposed based on the illness representation of the Common Sense Model (CSM) (Leventhal, Meyer and Nerenz, 1980). The classification of each medical condition is based on the perception of interview participants. It is appropriate as CSM aims to explain how people make sense of and respond to health threats and illness (Leventhal et al., 1980).

<table>
<thead>
<tr>
<th>Medical Condition</th>
<th>Chronic</th>
<th>Temporary</th>
<th>Manageable</th>
<th>Curable</th>
<th>Stigmatised</th>
<th>Not-Embarrassing</th>
<th>High Impact</th>
<th>Low Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>High Blood Pressure</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pregnancy</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eosinophilic Esophagitis</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinal Injury</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison’s Disease</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chronic Fatigue Syndrome</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degenerative Disc</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Mechanisms and research (e.g., medication, treatment, studies, explanation, etc.). It is usually shown in the scientific information refers to facts directly related to diseases and explains the underlying scientific (p7, A11). As personal experience varies, there is no clear standard to indicate what is credible. She actually passed away, she was on Tarceva [but] she lost all her hair. That only happens with chemotherapy out. At the very end they touch my glasses and

beneficial to some people but simple things like I have a problem that my eyelashes keep on growing

regularly attend the support group are on Tarceva. Everybody seems to be affected differently by it. It seems to be experiential information varies greatly, as stated by one participant: “You might find that most of the people who

1st hand experience with a condition or situation. For example, “The problem of using drug B is you got no way of gauging it. [There is] no way to accurately measure the dose for the steroid you are given. So a few years back somebody discovered it about drug B and they were writing to stop taking drug B. I reduced my own drug B intake from time to time and only use it as buffer” (p3, A12). Because everybody experiences the same illness differently (e.g., reacts to the same treatment or medication with a variety of symptoms and degrees), experiential information varies greatly, as stated by one participant: “You might find that most of the people who regularly attend the support group are on Tarceva. Everybody seems to be affected differently by it. It seems to be beneficial to some people but simple things like I have a problem that my eyelashes keep on growing. They don’t fall out. At the very end they touch my glasses and I can’t see. Nobody else has that problem. There is an old lady there, she actually passed away, she was on Tarceva [but] she lost all her hair. That only happens with chemotherapy” (p7, A11). As personal experience varies, there is no clear standard to indicate what is credible.

Table 1. Medical Conditions Appeared among Interview Participants

Guided by the research questions and concepts derived from our literature review, participants were asked questions such as “What indicators tell you that the information is trustworthy” and “What factors help you determine the quality of the information in a posting”. Interview questions were pre-tested on two PhD students with expertise in health informatics and health education, and on two students who are regular OHF visitors. Accordingly, the wording of questions, illustrative examples and the order of questions were refined. In total, the primary investigator conducted one-to-one semi-structured interviews with 16 participants, who represent different medical conditions, length of using forums (from 3 months to over 3 years), genders, ages (from 18 to over 60), education levels (from high school to Master’s degree), roles (patient or care giver) and purposes. The length of interview ranged from 20 minutes to 1.5 hours. The interviews were conducted over Skype and recorded using MP3 Skype Recorder. Transcription and coding were conducted shortly after an interview was finished. Recruiting was stopped when theoretical saturation was reached and most of the categories and properties remained unchanged when a new transcript was introduced. It is recommended to have 12 to 20 interviews when intending to achieve maximum variation (Guest, Bunce and Johnson, 2006). During the open coding, both the number and the names of codes did not change significantly after coding the first 14 transcripts, which indicated data saturation. Data were coded using open, axial and selective coding (Strauss and Corbin, 1990). The iterative coding process is facilitated by computer software. The same set of data has been coded twice by the primary investigator using Excel and NVivo 8 with one month apart. The two coding processes generated nearly the same categories and similar properties. The codes were further affirmed by additional investigators, both individually and in team meetings to resolve differences of opinion regarding the meaning of each data point analysed. This multiple-assessor process ensured inter-rater agreement and sufficient depth of analysis.

3.1.2 Research Findings: Study 1

As discussed in the theoretical background, both emotional and informational support emerged from the data as important factors. Emotional support can help users cope with emotional distress caused by their medical related situation. A chronic fatigue syndrome patient mentioned the reason for using OHFs: “Loneliness basically. I started to use it mainly because I want to find someone who has the same illness. It’s isolating when you have a medical condition that is not well-understood by the outside. So isolation I suppose. Like to talk to someone who believes you, understands and recognises my illness (p1)...... outside world doesn’t recognise this illness and we get deflected a lot from doctors and from the general community” (p2, A13).

Furthermore, three types of informational support are uncovered. Experiential information stems from the first-hand experience with a condition or situation. For example, “The problem of using drug B is you got no way of gauging it. [There is] no way to accurately measure the dose for the steroid you are given. So a few years back somebody discovered it about drug B and they were writing to stop taking drug B. I reduced my own drug B intake from time to time and only use it as buffer” (p3, A12). Because everybody experiences the same illness differently (e.g., reacts to the same treatment or medication with a variety of symptoms and degrees), experiential information varies greatly, as stated by one participant: “You might find that most of the people who regularly attend the support group are on Tarceva. Everybody seems to be affected differently by it. It seems to be beneficial to some people but simple things like I have a problem that my eyelashes keep on growing. They don’t fall out. At the very end they touch my glasses and I can’t see. Nobody else has that problem. There is an old lady there, she actually passed away, she was on Tarceva [but] she lost all her hair. That only happens with chemotherapy” (p7, A11). As personal experience varies, there is no clear standard to indicate what is credible.

Scientific information refers to facts directly related to diseases and explains the underlying scientific mechanisms and research (e.g., medication, treatment, studies, explanation, etc.). It is usually shown in the

Table 1. Medical Conditions Appeared among Interview Participants

<table>
<thead>
<tr>
<th>Medical Condition</th>
<th>Fibromyalgia</th>
<th>Minor Ailments*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

*Here “minor ailments” is used to represent medical conditions that non-regular forum users have. Non-regular forum users usually locate a forum by googling and only visit OHFs when symptoms emerge or medical attention is needed. Their short-term visit will be terminated when they are cured. Some non-regular users in the sample did not have the intention to disclose their exact medical conditions and based on their reported experience with OHFs, it is reasonable to estimate that their ailments can be either stigmatised or not embarrassing.

1 Quotations from the interview data are listed as the page (P#) of transcript of participant (A#).
form of referrals to other websites or citing information from external media (e.g., the Internet, books, journal, etc.). The forum is utilised as an information filter, as one said “You often put in the forum that ‘I heard about this [new cancer drug]. Did anybody hear about it?’ You get a lot of information. If you look up on the Internet, you just get very critical. You don’t know all the information that if anyone ever tried it” (p3, A8).

Non-medical factual information (in a narrow sense) pertains to facts that are indirectly related to the disease but help people cope with other aspects of daily life (including insurance policy, hospital policy, and government financial aid). For example, “I posted on the managing money section, and I asked suggestion about what other people think…… I do look for information a bit. A lot of reasons I’m on is payment, at the point of my life” (p4, A5).

Although both scientific information and non-medical factual information are objective and factual, scientific information may contain competing theories which lead to different explanations, as one pointed out: “Even those specialists and PhDs, they have various opinions on what the right answer is. So I cannot actually judge what is right. People don’t really have the same opinions really. Sometimes some studies say this and the others have different opinions” (p5, A10). Therefore, compared with non-medical factual information that has standard answers from authorities (e.g., governments, insurance companies, hospitals), it is more difficult but valuable to study the assessment criteria of credibility of experiential and scientific information, which are the focus of the rest of the paper.

Figure 1 depicts the factors used to assessment of the credibility of experiential and scientific information. Both types require argument quality, verification and contributor’s literacy competence. Reference credibility is related to scientific information and crowd consensus is used to assess experiential information.

![Figure 1. Assessment Criteria of Perceived Information Credibility](image)

**Argument quality** refers to a logical soundness based on common sense. Logical soundness is repeatedly mentioned by interview participants, using phrases like “it makes sense”, “reasonable”, “logical” and “they know what they are talking about”, to describe credible information. It is the result of cognitive effort in assessing argument strength, as one stated that “if there is only one person answered, I’ll be reading his post and based on my experience that he’s a rational person and what he said make sense, I’ll try” (p8, A13).

**Verification** is another cognitively demanding activity. It can be accomplished by confirming with external sources (e.g., websites, books, academic journals, etc.) or internal sources (e.g., self knowledge). A participant elaborated on external verification: “If the advice is like if you hurt your foot you should put a cold pad and someone else said if you hurt your foot you should use a hot pad. I’ll actually go get journal articles and search the nature of the injury and whether a hot pad is better or a cold one is better in control trials and based on that” (p7, A2). Self knowledge can be gained through studying and researching as well as experiencing the medical condition on a daily basis, as one said: “[To judge what others said,] I suppose it’s based on my own personal experience of the illness. I do research and get medical status and I belong to some medical association. I’m constantly reading up about my illness. So it’s based on that” (p2, A12).

**Contributor’s literacy competence** is the most frequently mentioned heuristic cue (13 out of 16 participants). It forms a first impression of the quality of a message. One participant explained: “[The information and advice in OHFs] are particularly trustworthy because some of the postings are extremely well-written, seen like medical literature and put into the words that the rest of us can understand” (p1, A13). Another stated: “If in the title and preview, the information is punctuated poorly or wrongly capitalised, then the quality is poor” (p2, A2). However, one pain sufferer does not respond to this cue: “It doesn’t matter if people can’t spell or their grammar is incorrect, they still suffer pain. If they are not articulate on what happened, if it’s very basic, I’m very thankful they
either business or people put information out there. You don’t know if they’re trustworthy. Business it’s government, university websites, if it’s research based, those are pretty trustworthy. You’ve got other websites
d as an inexpensive
Participants were recruited using Amazon’s Mechanical Turk (MTruk), which is recognise
3.2.1 Subjects and Procedures
contribute to perceived credibility of experiential information respectively.
Argument quality, verification, contributor’s literacy competence, and crowd consensus (CRC) positively
The one way you can get some sort of ideas is to get all your info, line them up and see how much they’re the same. The majority is the same and you get them from different sources, then you can pretty well say that they live with pretty much the same side-effects. So you get a pretty good idea how it’s going to react to you” (p7, A15). The social influence on adopting behaviour is obvious: “If the crowds really really like something, yes I would try. If 60 people say it’s good, I’ll give it a go. But if only one person mentions it, I’ll do some research. If I can’t find anyone else who thinks it’s good, I may be less likely to try” (p9, A8).

Study 1 enriches our understanding on assessing credibility of user-generated health content by revealing the criteria used to evaluate the credibility of scientific information and experiential information. Though cognitive authority was mentioned by experienced OHF users who have repeated interaction with a contributor, we excluded it because we intended to form a uniform model suitable for all OHF users and cognitive authority is missing among inexperienced user. Interestingly, other criteria mentioned in the literature review such as completeness and length is not prominent. Completeness is not necessary because users can learn one’s experience through discourse and the appropriateness of details enclosed in a message varies among situations. Also, the preferred length varies. When describing experience, users prefer some details without undue elaboration. When showing emotional support, a very short message can fail to convey compassion and sincerity without illustrating any shared experience. Moreover, though details are important for communication, some participants could not read long postings due to cognitive function decline or physical discomfort.

3.2 Study 2: Quantitative Validation

Study 1 generates a set of criteria used by OHF users to assess user-generated health information. However, it did not specify whether they work uniformly, or are weighed differently for different types of information. Moreover, as common limitations of qualitative studies, lack of generalisability and the nature of subjective interpretation are often mentioned. To address those concerns, a quantitative study was conducted using an online survey and structural equation modelling to validate two general hypotheses (as in Figure 1): 1) Argument quality (ARQ), verification (VER), contributor’s literacy competence (LIC), and reference credibility (REC) positively contribute to perceived credibility of scientific information respectively; 2) Argument quality, verification, contributor’s literacy competence, and crowd consensus (CRC) positively contribute to perceived credibility of experiential information respectively.

3.2.1 Subjects and Procedures

Participants were recruited using Amazon’s Mechanical Turk (MTruk), which is recognised as an inexpensive way to collect high quality representative data, compared with traditional methods (Berinsky, Huber and Lenz, 2012). 159 online forums users living in the United States were self-selected and completed a 10-minute online
survey on their perceptions of the posting they have the best memory of during their last visit to an OHF. The average age is 32 with a standard deviation of 11.8. 44.7% of participants are female and 54.1% have a Bachelor’s degree or a more advanced degree. The demographic pattern is consistent with previous studies on online health information seeking (Fox, 2011; Powell and Clarkem, 2006).

3.2.2 Measures

All measures were adapted from existing literature and all items were measured using a seven point Likert scale with anchors from strongly disagree to strongly agree. Following Podsakoff et al.’s (2003) remedies for common method bias, the measures of dependent variable and independent variables are drawn from different sources, and temporal separation is created between them. Perceived information credibility was measured using Hilligoss et al. (2008) and Rains et al.’s (2009) items, while measures of dependent variables were chosen from a number of studies (Cheung, Luo, Sia and Chen, 2009; Dutta-Bergman, 2004; Fadel, Durcikova and Cha, 2008; Lee, Strong, Kahn and Wang, 2002; Nelson and van Meter, 2007; Wagner, Puranik, Foorman, Foster, Wilson, Tschinkel and Kantor, 2011; Zhang and Watts, 2003). A pre-test on a sample of 104 OHF users using the same recruiting and analysing procedures were conducted to ensure the reliability and validity of the instrument. As a result, items of verification were reworded and adjusted to better fit the context based on the results of Study 1 and the pre-test. Each construct has three to five items and in total we have 23 items.

3.2.3 Measurement Model and Validity Analysis

The measurement model was analysed using Partial Least Squares (PLS) software package Smart PLS 2.0. Its tolerance for small sample sizes and purpose of prediction suits our needs. Due to the nature of the phenomenon, not every participant encountered both types of information in the last visit. Participants whose most memorable message contains only scientific information did not answer questions on crowd consensus, and same with experiential information on reference credibility. Hence, there are missing data values in the sample. Though PLS can handle missing data through case wise or mean replacement, the adequacy of its ability is seriously questioned recently (Parwoll and Wagner, 2012). Thus, we split the sample into two: Sample 1 for scientific information with 140 cases, and Sample 2 for experiential information with 128 cases.

Table 2. Correlation of Latent variables (Scientific information sample; n=140)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St.D</th>
<th>Cronbach’s Alpha</th>
<th>ARQ</th>
<th>VER</th>
<th>LIC</th>
<th>REC</th>
<th>INC</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARQ</td>
<td>4.7288</td>
<td>1.3633</td>
<td>0.8102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VER</td>
<td>4.8226</td>
<td>1.3237</td>
<td>0.7807</td>
<td>0.7226</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8836</td>
</tr>
<tr>
<td>LIC</td>
<td>4.8365</td>
<td>1.4379</td>
<td>0.8208</td>
<td>0.6839</td>
<td>0.6873</td>
<td></td>
<td></td>
<td></td>
<td>0.9099</td>
</tr>
<tr>
<td>REC</td>
<td>4.4239</td>
<td>1.3389</td>
<td>0.7978</td>
<td>0.6258</td>
<td>0.6547</td>
<td>0.5058</td>
<td></td>
<td></td>
<td>0.8932</td>
</tr>
<tr>
<td>INC</td>
<td>4.9638</td>
<td>1.4333</td>
<td>0.8459</td>
<td>0.7340</td>
<td>0.8021</td>
<td>0.6907</td>
<td>0.6910</td>
<td></td>
<td>0.9197</td>
</tr>
</tbody>
</table>

Note: The diagonal is the square root of the average variance extracted (AVE).

Table 3. Correlation of Latent variables (Experiential information sample; n=128)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St.D</th>
<th>Cronbach’s Alpha</th>
<th>ARQ</th>
<th>VER</th>
<th>LIC</th>
<th>CRC</th>
<th>INC</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARQ</td>
<td>4.8706</td>
<td>1.2903</td>
<td>0.8133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VER</td>
<td>4.9914</td>
<td>1.2256</td>
<td>0.7507</td>
<td>0.6475</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8664</td>
</tr>
<tr>
<td>LIC</td>
<td>4.9402</td>
<td>1.4026</td>
<td>0.8355</td>
<td>0.5013</td>
<td>0.5594</td>
<td></td>
<td></td>
<td></td>
<td>0.9141</td>
</tr>
<tr>
<td>CRC</td>
<td>4.8404</td>
<td>1.2951</td>
<td>0.8744</td>
<td>0.5421</td>
<td>0.7323</td>
<td>0.4926</td>
<td></td>
<td></td>
<td>0.9351</td>
</tr>
<tr>
<td>INC</td>
<td>5.0753</td>
<td>1.3064</td>
<td>0.8132</td>
<td>0.6526</td>
<td>0.7733</td>
<td>0.6238</td>
<td>0.7110</td>
<td></td>
<td>0.9018</td>
</tr>
</tbody>
</table>

Note: The diagonal is the square root of the average variance extracted (AVE).

PLS analysis was performed on both samples. Table 2 and 3 demonstrate strong reliability and validity of the measures. All Cronbach’s Alpha scores fall between 0.75 and 0.85, which satisfies Nunnally’s (1978)
recommendation of 0.7. Moreover, all items have loadings over 0.8, except one item of verification, which is 0.7814 in Sample 1 and 0.6967 in Sample 2, with 0.2 differences from all cross-loadings in both samples. Overall, the instrument satisfies the general agreed cut-off point of 0.7 for loading on a factor (Doll, Raghunathan, Lim and Gupta, 1995). To analyse the factorial validity, all items load with a significant t-value on its latent construct. The minimum t-value of loadings is 11.842 in Sample 1 and 8.897 in Sample 2. A further support of convergent validity is that the average variance extracted (AVE) of constructs exceeds 0.5 (Chin, 1998). Moreover, discriminant validity is demonstrated with low cross-loadings (0.2 differences from the loading on a factor) and reinforced by meeting the Fornell-Larcker Criterion (the square root of the AVE should be higher than the correlations between the latent variable and all other latent variables). The diagonal in both tables is the square root of AVE, which is higher than any values in its row and column. Based on the above tests, we conclude that the measurement model is reliable and valid. Next, we test the structural model.

### 3.2.4 Structural Model Analysis and Results

Both the PLS Algorithm and Bootstrapping procedure were performed. Since PLS does not provide fitness indices like LISREL or AMOS, results were evaluated based on path coefficients and the R squares that indicate the explained variance by the models.

<table>
<thead>
<tr>
<th>Sample 1 (Scientific Information)</th>
<th>Sample 2 (Experiential Information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2 = 0.733$</td>
<td>$R^2 = 0.702$</td>
</tr>
<tr>
<td>ARQ - INC</td>
<td>VER - INC</td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.183**</td>
</tr>
<tr>
<td>t-value</td>
<td>2.002</td>
</tr>
</tbody>
</table>

*p ≤ 0.1, **p ≤ 0.05, ***p ≤ 0.01, ****p ≤ 0.005

### Table 4. Path Analysis Results

As shown in Table 4, all paths in both samples are significant and both models have explained more than 70% of the variance in perceived information credibility. Among all criteria, verification has shown the strongest impact on perceived credibility of both scientific information and experiential information, while argument quality has shown less impact across both samples. To assess scientific information, reference credibility is the second most important criterion, and a contributor’s literacy competence is the least useful and salient factor. On the other hand, to evaluate experiential information, crowd consensus and a contributor’s literacy competence is the second and third most influential criteria.

### 4 Discussion

The goal of the paper is to provide an in-depth understanding on how people assess health information generated by other anonymous users in OHFs. We highlighted the difference between OHFs and other user-centric applications, emphasised the distinct features of user-generated content, and questioned the adequacy of using existing credibility assessment criteria on user-generated content. Employing sequential mixed-method, we conducted a qualitative enquiry to understand what criteria are used by users of OHFs for evaluation, followed by a quantitative validation of the criteria and an exploration on how those criteria are employed in judging different information. The findings show that users apply different strategies to assess scientific information and experiential information. Across all five criteria, verification is the most influential and salient factor. Reference credibility and crowd consensus are the second most valuable factors to evaluate scientific information and experiential information respectively. The low coefficients of argument quality suggest that users are not full confident in their logical reasoning ability to confirm or refute arguments. A contributor’s literacy competence is less effective when a message is based on reference to external scientific content.

The findings are consistent with previous studies. Toulmin’s model of completeness (1958) suggests that a logically complete argument contains three elements: claim assertions, evidence (grounds), and authority (warrants and backing). In our context, due to anonymity, authority is difficult to establish without repeated interaction. Thus, claim assertions (argument quality) and evidence (verification, reference credibility and crowd consensus) are highly relied on. Particularly, extreme weight has been placed on evidence by OHF users, which echoes the advocacy of evidence-based public health practice and self-management (Burns and Grove, 2011). Moreover, reference credibility and crowd consensus dovetail with informational influence and
the normative influence of Deutsch and Gerrard’s (1955) dual process theory. People are more likely to depend on objective standards (e.g., scientific mechanisms and explanation) if they are available; when objective standards are absent or obscured (e.g., assessing experiential information), people are inclined to apply social reality testing by detecting shared patterns among subjective and conflicting information (Festinger, 1954). In addition, to contrast with previous findings on user-centric applications where heuristic cues are largely adopted, users affected by health issues are more motivated to adopt an effortful approach and are less likely to depend on heuristic cues. Also, the low but significant coefficient of the ARQ – INC path indicates the high knowledge barrier and cognitive consumption faced by OHF users. Hence, when motivation is high but ability is low, users tend to balance the cognitive effort and seek a satisfactory outcome. Finally, a contributor’s literacy competence has been less helpful when assessing scientific information than experiential information. It can be explained by the nature of scientific information, which is cited from external sources and perceived information credibility is more likely to originate from the external sources.

5 Conclusion

The research contributes to the literature on information credibility by providing insights into how people assess credibility of information that is subjective, experiential, discursive, and anonymous. It also adds value to duel-process theories and user-generated content research by explaining how people strategise the evaluation when the stakes are high but their ability to discern credibility is low. The results show that a different mix of assessment criteria and preferences on each criterion are employed to evaluate credibility of scientific information and experiential information. These findings may have implications for non-medical forums such as those sharing investment advices where the stakes can be high and trust is important.

The findings have immediate implications for facilitating the self-management of medical conditions by patients. Governments that subsidise public healthcare encourage self empowerment through self education so that patients, especially those with chronic conditions, are able to monitor and manage the signs and symptoms by themselves. These findings can help governments to evaluate the outcome of self-education (e.g., evidence-based practice) and identify the underachieving areas. Finally, the findings can guide self-management services to enhance the usability and sociability of interface design.

One direction of future research could be examining potential moderators. The high R squares but low coefficients imply a possible moderating effect of medical condition. Different medical conditions may be perceived to have different degrees of impact on an individuals’ life. With high impact conditions, the risks of following wrong advice are higher, and individuals may be more motivated to conduct a comprehensive evaluation and less likely to rely on heuristic cues.

Finally, the boundary of the research is constrained by the context: user-generated content in online health forums. The criteria may be applicable, with caution, to other non-medical discussion-driven medium in addressing perceived information credibility in scenarios where the attributes of a contributor is easy to access and evaluate (e.g., blog), social interaction and influence is minimum (e.g., wiki), and stakes are not high. We also did not take the characteristics of users into account.

References


