

Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search

RYEN W. WHITE and ERIC HORVITZ
Microsoft Research

The World Wide Web provides an abundant source of medical information. This information can assist people who are not healthcare professionals to better understand health and illness, and to provide them with feasible explanations for symptoms. However, the Web has the potential to increase the anxieties of people who have little or no medical training, especially when Web search is employed as a diagnostic procedure. We use the term *cyberchondria* to refer to the unfounded *escalation* of concerns about common symptomatology, based on the review of search results and literature on the Web. We performed a large-scale, longitudinal, log-based study of how people search for medical information online, supported by a survey of 515 individuals' health-related search experiences. We focused on the extent to which common, likely innocuous symptoms can escalate into the review of content on serious, rare conditions that are linked to the common symptoms. Our results show that Web search engines have the potential to escalate medical concerns. We show that escalation is associated with the amount and distribution of medical content viewed by users, the presence of escalatory terminology in pages visited, and a user's predisposition to escalate versus to seek more reasonable explanations for ailments. We also demonstrate the persistence of post-session anxiety following escalations and the effect that such anxieties can have on interrupting user's activities across multiple sessions. Our findings underscore the potential costs and challenges of cyberchondria and suggest actionable design implications that hold opportunity for improving the search and navigation experience for people turning to the Web to interpret common symptoms.

Categories and Subject Descriptors: H3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Search process, Query formulation*
General Terms: Human Factors, Experimentation
Additional Key Words and Phrases: Cyberchondria

Authors' addresses: Microsoft Research, One Microsoft Way, Redmond, WA 98052; email: {ryenw, horvitz}@microsoft.com

1. INTRODUCTION

The World Wide Web has the potential to provide valuable medical information to people, where Web sites such as WebMD (<http://www.webmd.com>) and MSN Health and Fitness (<http://health.msn.com>) provide answers to such questions as whether concerning symptoms might indicate the onset of a serious, acute or chronic condition, or whether such fears are unfounded. However, the use of Web search as a diagnostic methodology—where queries describing symptoms are input and the rank and information of results are interpreted as diagnostic conclusions—can lead users to believe that common symptoms are likely the result of serious illnesses. Such escalations from common symptoms to serious concerns may lead to unnecessary anxiety, investment of time, and expensive engagements with healthcare professionals. We use the term *cyberchondria* to refer to the unfounded escalation of concerns about common symptomatology, based on the review of search results and literature on the Web.

The large volumes of medical information on the Web, some of which is erroneous, may mislead users with health concerns. Much has been written in the medical community about the unreliability of Web content in general [Eysenbach 1998; Jadad and Gagliardi 1998; Eysenbach et al. 2002] or content about specific conditions such as cancer [Biermann et al. 1999]. Indeed, studies have shown that, although 8 in 10 American

adults have searched for healthcare information online, 75% refrain from checking key quality indicators such as the validity of the source and the creation date of medical information [Pew Internet and American Life Project 2007]. Berland and colleagues [2001] suggest that medical information present on Web sites is generally valid, although they also find that it is likely to be incomplete. Eysenbach and colleagues [2002] systematically reviewed health Website evaluations and found that the most frequently used quality criteria included accuracy, completeness, and design (e.g., visual appeal, layout, readability). In their review, the authors noted that 70% of the studies they had examined concluded that the quality of health-related Web content is low. In addition, Benigeri and Pluye [2003] show that exposing people with no medical training to complex terminology and descriptions of medical conditions may put them at risk of harm from self-diagnosis and self-treatment. These factors combine to make the Web a potentially dangerous and expensive place for health seekers.

The information obtained from healthcare-related searches can affect peoples' decisions about when to engage a physician for assistance with diagnosis or therapy, how to treat an acute illness or cope with a chronic condition, as well as their overall approach to maintaining their health or the health of someone in their care. Beyond considerations of illness, information drawn from the Web can influence how people reflect and make decisions about their health and wellbeing, including the attention they seek from healthcare professionals, and behaviors with regard to diet, exercise, and preventative, proactive health activities.

In this article, we present the findings of a log-based study of anonymized data about online searches for medical information drawn from a large set of data on Web search behavior shared voluntarily by a large number of users of Web search engines. We focus particularly on the association between the input of search terms that describe common symptoms and shifts of focus of attention to serious illnesses—illnesses that are rarely the causes of such common complaints. We contrast medical search sessions that show a trajectory from basic symptoms to a review of content that may induce or increase anxiety with sessions that do not lead to such potentially troubling information. We supplement the log analysis where appropriate with findings from a survey of 515 individuals' health-related search experiences. Our study's log-based methodology lets us examine at scale how people interact with medical information and represents an initial step toward understanding cyberchondria. Its findings, and the implications drawn from them, highlight a nascent set of opportunities for researchers in academia and industry to help people wrestling with the access, comprehension, and interpretation of healthcare information.

Two research objectives guided our exploration: (i) *Characterizing cyberchondria*: We characterize the nature and frequency of the escalation of concern about what are likely to be common, innocuous symptoms to concerns about more serious illnesses, and (ii) *Studying the effects of cyberchondria over time*: We investigate whether medical concerns linked to common symptoms persist over multiple sessions, following a shift of focus of attention to serious illnesses, and characterize the extent to which they interfere with subsequent user activities. Identifying the recurrence of concerns about a rare disorder—especially when the recurrence occurs during another search task—may indicate that earlier escalations extend over time, and that anxieties or heightened awareness continues to interrupt users' online activities over prolonged time periods. Such findings may be proxies for the rise and persistence of deep concerns that may

disrupt other aspects of daily life. Findings of these explorations have implications for the design of supportive user interface features and specialized indexing and ranking algorithms, including the use of explicit probabilistic inference about the likelihoods of different disorders given the sets of symptoms input by users. Findings about long-term concerns and behaviors associated with medical anxiety induced or heightened by interactions with the Web have implications for the design of personalized systems that can offer tailored support for individual searchers over time.

We analyzed interaction logs of searching and browsing activities of consenting users with automated tools. We temper our results by stressing that our utmost attention to user privacy makes it impossible and unreasonable to know details about the rationale and influence of searches. We did not have access to information about peoples' non-Web search behaviors (e.g., interactions with physicians, or patients with similar symptoms or diagnoses), and cannot be certain that observed search engine users were actually becoming more anxious during interactions with medical content on the Web. We also do not have evidence about online users' predispositions to anxiety, and to their medical anxiety more particularly. People with heightened awareness or a priori interest in serious illnesses given basic concerns may also be more likely to experience unnecessary anxiety. Such a predisposition may be associated with unfounded medical concerns regardless of online interactions, thus further confounding the induction of causal arguments about the influence of searching and browsing on medical anxiety.

Given the nature of our study, and our paramount respect of user privacy, it is difficult to identify and assess frank anxiety. However, we can analyze with confidence the focus of attention of people performing online searches. Thus, we broaden the scope of cyberchondria to include the heightened awareness, attention, and interest surrounding serious medical conditions. We believe that our work serves as an important step toward gaining better understanding of how people search for medical information online, how the severity of their concerns may change over the course of a search session, and, more generally, the challenges that cyberchondria presents for search engine designers, and how these challenges might be addressed.

We structure the remainder of this article as follows. We discuss related research in Section 2. In Section 3, we motivate this research through an empirical study of the potential for escalation from examining Web search results. In Section 4, we describe key aspects of the data and analyses employed in our study. Section 5 describes the findings of our investigation into within-session escalations, and Section 6 covers longer-term persistence of anxieties and interruptions. In Section 7, we discuss our findings and describe techniques that may help alleviate inappropriate health anxiety or unwarranted interest in serious medical conditions given symptoms. We summarize and conclude in Section 8.

2. RELATED RESEARCH

The wealth of medical information discovered by Web search engines creates a potential for users to conduct their own diagnosis and healthcare assessment based on limited knowledge of diseases and interpretation of their symptoms.

Hypochondriasis is often characterized by fears that minor bodily symptoms may indicate a serious illness, constant self-examination and self-diagnosis, and a preoccupation with one's body. The small fraction (1-5%) of the general population afflicted with the disorder *hypochondria* are particularly predisposed to the emergence of unfounded concerns, especially since they are often undiscerning about the source of their medical information [Barsky and Klerman 1983]. Studies have shown that hypochondriacs express doubt and disbelief in their physicians' diagnosis, report that doctors' reassurance about an absence of a serious medical condition is unconvincing, and may pay particular attention to diseases with common or ambiguous symptoms [Barsky and Klerman 1983]. The Web is fertile ground for those with hypochondria to conduct detailed investigations into their perceived conditions.

The diagnosis and treatment of hypochondria has received attention in the medical community [Barsky and Klerman 1983; Barsky and Ahern 2004]. These studies have generally targeted the development and diagnosis of hypochondria, the self-perceptions of hypochondriacs, and the use of techniques such as cognitive behavioral therapy to treat hypochondriasis. We use the term *hypochondria* in the traditional manner, as a disorder associated with a tendency to have unfounded medical fears. *Cyberchondria* as we define it is an unfounded medical fear, or a heightened attention to serious disorders, based on the review of Web content. The term *escalation* defines specific instances of cyberchondria, within a single search session.

Beyond frank hypochondria as characterized by definitions in the *Diagnostic and Statistical Manual of Mental Disorders* [American Psychiatric Association 1994] or diagnoses by psychologists or psychiatrists, peoples' tendencies to become anxious about unlikely medical disorders may sit on a spectrum of concern. Medical experts have argued for action to lessen the likelihood of unnecessary health anxiety for all consumers of health information, regardless of whether they are diagnosed as suffering from hypochondria (e.g., [Asmundson et al. 2001]). Asmundson and colleagues [2001] describe research on the clinical features and current theoretical understanding of health anxiety, with a particular focus on hypochondriasis. There have also been studies on problems with the review of health-related Web content (e.g., [Cline and Haynes 2001; Eysenbach and Köhler 2002; Baker et al. 2003; Sillence et al. 2004; Eastin and Guinsler 2006; Lewis 2006]). Cline and Haynes [2001] present a review of work in this area that suggests that public health professionals should be concerned about online health seeking, consider potential benefits, synthesize quality concerns, and identify criteria for evaluating online health information. Eysenbach and Köhler [2002] used focus groups and naturalistic observation to study users attempting assigned search tasks on the Web. The investigators found that the credibility of Websites (in terms of source, design, scientific or official appearance, language used, and ease of use) was important in the focus group setting but appeared less important in practice, with many participants largely ignoring the source of their medical information. Baker and colleagues [2003] measured the extent of Web use for healthcare among a representative sample of the United States population, to examine the prevalence of e-mail use for health care, and to examine the effects that Web and e-mail use has on users' knowledge about health care matters and their use of the health care system. They base their findings on self-reported rates of Web and e-mail use gathered through telephone interviews. They found that users rarely use email to communicate with physicians and that the influence of the Web on the utilization of external healthcare is uncertain. Sillence and colleagues [2004] studied the influence of design and information content on the trust and mistrust of online health

sites. They conducted an observational study of a small number of subjects engaged in structured and unstructured search sessions over a four-week period. They found that aspects of design appeal engendered mistrust, whereas the credibility of information and personalization of content engendered user trust. Eastin and Guinsler [2006] investigated the relationship between online health information seeking and healthcare utilization such as visiting a general practitioner. Their findings suggest that an individual's level of health anxiety moderates the relationship between online health information seeking and health care utilization decisions. Lewis [2006] discusses the growing trend towards the general population accessing information about health-related matters online. She performed a qualitative study into young peoples' use of the Web for health material that showed that in fact they are often skeptical consumers of the material they encounter. The findings of these studies demonstrate some of the conflicting opinions around the effect of healthcare information on human behavior. This may be attributable to differences in the goals of the studies, the samples used, and the experimental methodologies.

Studies on unfounded medical concerns associated with the review of Web content, including many of those cited above, typically rely solely on responses to questionnaires, in-person interviews, telephone surveys, or monitor interaction behavior for assigned tasks. These data-gathering methods are not amenable to the following of behavior in the world as assessments are often captured after the fact and depend on participant self-reporting, which may be biased. The log-based methodology employed in our study provides a window into Web searchers' natural information-seeking behaviors over a sustained period of time, allowing for a more accurate description of how people search for health-related information.

Web interaction logs have been used previously to study medical Web search behavior (e.g., [Bhavnani et al. 2003; Spink et al. 2004]). Bhavnani and colleagues [2003] explored the timing and numbers of pages visited by experts and non-experts, and demonstrated that term co-occurrence counts for medical symptoms and disorders on Web pages can be a reasonable predictor of the degree of influence on user search behavior. Spink and colleagues [2004] characterized healthcare-related queries issued to Web search engines, and showed that users were gradually shifting from general-purpose search engines to specialized Web sites for medical- and health-related queries. Ayers and Kronenfeld [2007] employed a similar methodology and utilize log data on Web use, and perform a multiple regression analysis to explore the relationship between chronic medical conditions and frequency of Web use, as well as changes in health behavior due to frequency of Web use. Their findings suggest that it was not the presence of one particular chronic illness, but rather the total number of chronic conditions that determines the nature of Web use. They also found that the more frequently a person uses the Web as a source of health information, the more likely they are to change their health behavior. However, unlike our investigation, the authors did not study Web search behavior or examine the escalation of seemingly innocuous concerns to more serious illnesses during Web search sessions. Our focus on Web search is an important differentiator between our work and previous research. Web search is especially important for many users given their reliance on search engines to locate Web content.

Information retrieval (IR) and information science researchers have investigated the search behavior of medical domain experts [Hersh et al. 1998, 2002; Bhavnani 2002; Wildemuth 2004], with a view to better understanding the search behavior of those with specialist domain knowledge. Hersh and colleagues [1998] review research in the

medical informatics and information science literature on how physicians use IR tools to support clinical question-answering and decision-making. They found that retrieval technology was inadequate for this purpose and generally retrieved less than half of the relevant articles on a given topic. They follow-up this review with a study of how medical and nurse practitioner students use MEDLINE to gather evidence for clinical question answering [Hersh et al. 2002]. Their findings show that these users were only moderately successful at answering clinical questions with the assistance of literature searching. Bhavnani [2002] observed healthcare and online shopping experts while they performed search tasks inside and outside their domains of expertise. The findings of the study identified domain-specific search strategies in each domain, and that such search knowledge is not automatically acquired from using general-purpose search engines. Wildemuth [2004] performed a longitudinal study examining the tactics of medical students searching a factual database in microbiology. Findings showed that over the course of the study changes in students' search tactics were observed as their domain knowledge increased.

Despite the broad range of previous work in this area, none of the prior studies have addressed the important issue of the links between online activity and medical anxiety, and the potential escalation of medical concerns during Web search and browsing. In this article, we take a first step towards tackling this important challenge through an exploratory study of medical escalation in the Web search domain.

3. POTENTIAL FOR ESCALATION

At the outset of our studies of cyberchondria, we explored general statistical clues that could provide insights into how Web content might typically link searches focused on common symptoms to content describing relatively rare, serious illnesses versus more common, benign explanations. Searchers may often seek information (implicitly or explicitly) on the probability of different disorders given perceived symptoms. Thus, we have been particularly interested in how the distribution of medical content and links between content and symptoms may diverge from a distribution that is representative of the prior and posterior probabilities of medical disorders. We sought to compare these statistical results from three different corpora: (i) a large random sample of the Web, (ii) results from a general-purpose Web search engine, and (iii) results from a specialized medical search engine.

We retrieved a 40-million page random sample of Web content based on a breadth-first crawl of all categories in the Open Directory Project (ODP) (<http://dmoz.org>), a human-edited directory of the Web. Following the crawl, for each of three common symptoms (headache, muscle twitches, and chest pain), we compared the co-occurrence statistics for the symptom and the corresponding most likely benign explanations with the co-occurrences of the symptom and serious, but less likely disorders. We excluded co-occurrence instances if a negation appeared within five words of the symptom in the page (e.g., "...headache *not* malignant...").

We also computed similar sets of term co-occurrence statistics from the following two sources:

- Web search engine: Microsoft's *Live Search* engine provided Web search results.

- Domain search engine: *MSN Health and Fitness* provided medical search results. MSN Health and Fitness (<http://health.msn.com>) is a Web-based provider of health-related information that offers access to a large number of articles from authoritative sources (e.g., <http://www.mayoclinic.com>). Such specialized engines have access to a range of authoritative medical resources that are typically not available through a single Web site or Web search engine.

We issued a query comprising solely of the symptom name to each of these sources and computed term co-occurrence statistics in content contained on the pages of the top-100 search results. We used synonyms of the conditions where appropriate, e.g., for amyotrophic lateral sclerosis we also included its acronym, ALS, Lou Gehrig’s disease, and motor neuron disease.

In Table I, we list symptoms, some common non-serious explanations, and more serious concerns, along with associated probabilities, from each of the random crawl, Web search, and specialized domain search.

Table I. Probability of Mention of Cause Given Symptom.

Symptom	Cause	Web crawl	Web search	Domain search
headache	caffeine withdrawal	.29	.26	.25
	tension	.68	.48	.75
	brain tumor	.03	.26	.00
muscle twitches	benign fasciculation	.53	.12	.34
	muscle strain	.40	.38	.66
	ALS	.07	.50	.00
chest pain	indigestion	.28	.35	.38
	heartburn	.57	.28	.52
	heart attack	.15	.37	.10

As can be seen in Table I, the estimates for Web search differ dramatically from those of Web crawl or for domain search, with more weight being given to serious conditions. For example, the co-occurrence statistics for the Web crawl may be interpreted naïvely by a searcher as indicating that there is a probability of 0.03 that “headache” is associated with “brain tumor,” 0.29 for “caffeine withdrawal,” and 0.68 for “tension.” In reality, the probability of a brain tumor, given the chief complaint of headache, is much smaller than 0.03. Headaches are exceedingly common and the background chance per year of a brain tumor, based on the U.S. annual incidence rate, is 0.000116 (around 1:10,000). A naïve probability estimate of “brain tumor” given “headache” based on co-occurrence statistics in the top-10 Web search results was 0.26, more than eight times the Web estimate, and significantly higher than the general incidence rate. In comparison, co-occurrence statistics from domain search were roughly in line with the Web estimate.

Other examples follow a similar pattern. Muscle twitches may herald the onset of ALS. However, the twitching of muscles does not definitively mean someone has this serious condition. U.S. annual incidence rates for ALS are approximately 1:55,000, or a

background likelihood of ALS of 0.0000186. Although the latter incidence rate is for the overall population, not for people who report the rise of twitching (or of the awareness of twitching), the incidence rate provides a clue as to the low probability of ALS given muscle twitches. In fact, benign twitches are quite common in the population, being associated with such benign causes as muscle fatigue, stress, and caffeine. Beyond the intermittent twitching of muscles (e.g., common eyelid twitches) that come and go, are more salient but still benign presentations of twitching based in poorly understood phenomena that are grouped by physicians into the phrase *benign fasciculation syndrome*. Experts in neuromuscular disorders report that they can often discriminate between the potential subtle differences between benign muscle twitches and more concerning twitching, especially in the context of other clues. However, the subtleties in interpretation and implication that come with expertise are lost in web content that simply refers to the link between “twitches” or “fasciculations” and the onset of ALS.

As another example, let us consider the frequency of observing the topic “heart attack” in Web search results relative to other explanations for queries about “chest pain.” We shall focus a bit more deeply on the complaint of chest pain, given that heart disease is the leading cause of death in the United States. Results of our co-occurrence analyses for the complaint of chest pain are displayed in Table I. On the broad crawled Web content, “heart attack” co-occurs with chest pain 15 percent of the time. “Heart attack” co-occurs with chest pain in 37 percent of the content drawn from the top-ranked search results for a broad Web search and 10 percent of content drawn from medical domain search.

The onset of chest pain is a worrying sign as it can indicate the rise of a coronary event in a previously healthy person. Early intervention that brings rapid access to a medical team and hospital-based care can be important in the survival of a patient with an acute coronary syndrome. However, multiple non-cardiac factors can be at the root of chest pain. Chest pain can often be an indication of less serious esophageal, gastrointestinal, and musculoskeletal problems, some that will disappear over time without any special treatment.

From an expert’s perspective, the *a priori* likelihood of the onset of a first acute cardiac event in a previously health person depends on several factors. Considerations include the age and gender of the person, and details about the nature of the pain—nuances that are not necessarily captured or reported in Web queries and web content that simply refer to “chest pain.” Non-cardiac chest pain is common in patients presenting to hospital emergency departments. One study estimated that as many as 25% of people complaining of chest pain who are concerned enough to seek care at a hospital emergency department have non-cardiac chest pain that is associated or amplified by a panic disorder [Fleet, et al., 1996; Huffman and Pollack, 2003]. For people who have not yet been diagnosed with cardiac disease, a meta-analysis identified several key factors as indications that the patient is primarily grappling with anxiety [Huffman and Pollack, 2003]. These factors include atypical quality of chest pain, a high degree of self-reported anxiety, and younger age.

The probability of the rise of an acute coronary event in a previously healthy person is sensitive to age and gender and these factors can be made salient to worried searchers. Heart attacks are rare in people under 35. The average annual rates of the first major cardiovascular event have been reported to be 0.003 in men at ages 35 to 44 rising to 0.074 at ages 85 to 94. Comparable rates in women are seen about ten years later, with

the gap between the rates in women and men getting smaller with advances in age [Hurst, 2002]. Another study found that the incidence rate of hospitalization for myocardial infarction, for people in the group 35 to 74 years of age is 0.004 for males and 0.002 for females [Rosamond et al., 1998]. A study of the annual incidence rate of heart disease in women found an incidence of disease for women 49 years of age or younger to be 0.00013, 0.00053 for women 50 to 54 years of age, 0.00149 for women 55 to 59 years of age, 0.00214 for women 60 to 64 years of age, and 0.00244 for women 65 years of age or older [Hu et al., 2000].

We note that the cited incidence rates for the onset of heart disease are not conditioned on the existence of chest pain. They also do not consider such known risk factors as having diabetes mellitus or having a parent who experienced a cardiac problem early in life. However, concerns about the onset of an acute heart problem in a healthy, young person can be tempered with an appreciation for the background incidence rates and knowledge that various types of chest pain can be caused by non-cardiac and frequently benign processes.

In summary, expert clinicians often probe subtleties of symptomatology and fuse together multiple findings, including demographic considerations such as the gender and age of a patient, in assessing the rough likelihoods of different explanations for a patient's concerns and symptoms. The subtleties of presentation and insightful fusion of demographics, and multiple signs and symptoms are not easily accessible by people seeking diagnostic support with Web search. The tendency of Web searchers to start with symptoms that are coarsely reported and also coarsely referred to in Web content can stimulate potentially unwarranted anxiety.

Our findings suggest that there is inappropriate escalatory risk associated with using general Web search to support differential diagnosis, and that more valuable information may come via search within expert medical sites, as results align better with statistical estimates. However, unwarranted anxieties may come even with review of the specialized sites. In the next section, we will describe a study aimed at characterizing the escalation of health concerns (as observed through queries) both within single search sessions and across multiple search sessions.

4. STUDY

In the second phase of our analysis, we performed a log-based study of health-related Web searching behavior. The aim was to characterize the nature of within-session escalations in querying and browsing behavior, and the longer-lasting effects of these escalations. To study the escalation of health concerns, we formulated a list of relatively common symptoms and associated benign and more serious illnesses to represent the source and destination of escalations. Table II displays the list of symptoms and serious illnesses that we considered. These lists were based on the *International Classification of Diseases 10th Edition* (ICD-10) published by the World Health Organization, and pruned based on common concerns expressed in commercial Web search engine query logs. In our log-centric analysis, we also employed synonyms of symptoms and conditions to increase coverage (e.g., including "tiredness" in addition to "fatigue"). In addition, we reviewed content on the U.S. National Library of Medicine's *PubMed* service and other Web-based medical resources to create a set of common explanations for each of the

medical symptoms. For example, likely explanations for “insomnia” include “stress,” “caffeine,” and “jet lag.” These were verified and expanded by one of the authors (EH), who received formal medical training within an MD/PhD program. Table II shows the set of all medical symptoms, common explanations, and serious illnesses used in this study. Note that for reference the explanations and serious illnesses for all of the 12 medical symptoms are pooled and sorted alphabetically in Table II.

4.1 Medical Escalation

For the purposes of this investigation, we define escalations to be observed increases in the severity of concerns represented by the search terms within a single *search session*. We define a search session as a chronologically ordered set of Web pages initiated with a query to a commercial Web search engine and terminating with a session inactivity timeout of 30 minutes. A similar timeout has been used to demarcate search sessions in previous work [Downey et al. 2007; White and Drucker 2007]. Query escalations are revealed by queries issued by the user to a commercial search engine such as Google, Yahoo!, or Live Search where query terminology is related to the serious illnesses defined in Table II and/or associated with modifiers used to express grave concern (e.g., “chronic,” “fatal”).

It is also possible to study *navigational* escalations (i.e., escalations revealed by access to potentially escalatory Web content rather than queries containing escalatory terms). We experimented with term occurrence measures as a way to determine escalations automatically by examining Web pages visited. For example, pages containing serious illness names could be regarded as escalatory evidence, even if no escalation was evident in the query stream. However, we encountered numerous challenges in extracting such evidence from Web pages (e.g., pages containing lists of all possible explanations for a given symptom may or may not be escalatory). Since queries are explicit indications of user search intent, they are a more reliable source of escalatory evidence than implicit evidence garnered from the content of visited Web pages. For this reason, we focus on query escalations in our analysis.

4.2 Research Objectives

We specifically sought to explore the extent to which pursuing information on common, innocuous symptoms can escalate into the review of content on serious, often rare conditions that may be associated with the common symptoms. Our study aimed to characterize the nature of query-based escalation from common symptoms to more serious illnesses within a session, and the emergence of longer-term medical anxieties. More broadly, we investigate increases in the focus of attention on serious medical conditions, following the identification of an escalation in our logs.

As we mentioned, while anonymized interaction logs allow for studying actual behaviors at a large scale, we cannot confirm with certainty a causal association between exposure to Web search results and unfounded escalation of anxiety (e.g., users may simply be curious about a condition). The findings presented in Section 3 demonstrate that Web search has the potential to bias medical information toward more serious illnesses, and as we will show in this log-based study and survey findings reported in this article, users often gravitate toward serious illnesses for seemingly innocuous symptoms. Even if this gravitation is a result of curiosity not anxiety, it is worthy of attention since interest may

Table II. Symptoms, Explanations, and Serious Illnesses.

Medical symptoms	Serious illnesses
breathlessness chest pain dizziness fatigue fever headache insomnia lump nausea rash stomach pain twitching	acute coronary syndrome AIDS Alzheimer's disease anemia angina appendicitis arthritis asthma balance disorder bipolar disorder brain hemorrhage bronchitis cancer cerebral vascular accident chronic fatigue syndrome clot coronary artery disease Crohn's disease diabetes embolism emphysema encephalitis epilepsy glaucoma heart attack heart block heart disease heart failure hepatitis Huntington's chorea hypertension irritable bowel syndrome kidney disease labyrinthitis leukemia liver disease Lou Gehrig's disease lupus lymphoma malaria Meniere's disease meningitis motor neuron disease multiple sclerosis muscular dystrophy myopathy narcolepsy obstructive pulmonary disease osteoarthritis osteoporosis Parkinson's disease pneumonia polymyositis rheumatoid arthritis sexually transmitted disease sleep apnea spinal muscular atrophy stroke tuberculosis tumor ulcer
Common explanations	
acne allergy angina anxiety benign fasciculation benign paroxysmal positional vertigo boil bruise caffeine withdrawal callus common cold constipation corn cyst dehydration dermatitis dysphasia ear infection eczema esophagitis exercise eyestrain fatigue food allergy food poisoning gastroenteritis heartburn hunger indigestion influenza insect bite irritation jet lag lactose intolerance laryngitis lipoma migraine mole motion sickness obesity panic attack pregnancy sleep disorder stress sunburn tension throat infection tiredness tonsillitis underactive thyroid urinary tract infection wart	

evolve into concern and frank anxiety. We now describe data collected to meet our research objectives.

4.3 Data Collection

We automatically mined the anonymized interaction logs of hundreds of thousands of consenting Windows Live Toolbar users during an 11-month period. The Windows Live Toolbar is a plug-in to the Internet Explorer browser that provides additional browser functionality in return for users providing consent for their page-level interactions to be logged. During installation of the toolbar users were invited to consent to their interaction with Web pages being recorded (with a unique identifier assigned to each client) and used to improve the performance of future systems. The information contained in our logs included a client identifier, a timestamp for each page view, a unique browser window identifier (to resolve ambiguities in determining which browser a page was viewed), and the URL of the page visited. We stress again that user privacy and confidentiality was paramount: no personal information was elicited, no attempt was made to identify or study an individual, and *findings were aggregated over multiple users*. Logs contained interaction with all major Web search engines such as Google, Yahoo!, or Live Search and the pages that followed a result click. This provided us with a significant amount of data on querying and browsing behavior. These data differ from that described in Section 3 in that we now study user interaction logs rather than search results and Web crawls.

Medical queries were identified in the logs based on string matching with a list of terminology comprising the union of a consumer health vocabulary (described in detail in [Zeng et al. 2007]), a list of drug names from the United States Food and Drug Administration, and the lists of medical symptoms, common explanations, and serious illnesses shown in Table II. Queries were labeled as *medical* if any of their constituent terms matched a term in these collections. To improve coverage, we also included spelling variants, inflections, and synonyms where appropriate (e.g., “malignant” and “malignancy” for “cancer”). We sought to minimize false-positives in identifying medical queries. To this end, we manually analyzed a sample of ten-thousand queries tagged as medical and created a list of stop words, stop phrases, and parsing rules designed to exclude non-medical queries from the logs. For example, we sought to avoid labeling as human medical queries pet ailments or non-medical queries containing medical symptoms, e.g., “saturday night fever.”

We found that approximately 2% of all queries were health-related, and approximately 250 thousand users (around one quarter of our original user sample) engaged in at least one medical search in the duration of the study. As our term list was limited, we believe that this represents a conservative estimate of the likely larger number of medical queries and concerned users in our logs. We focus on a subset of these users that submitted a query with at least one of the medical symptoms shown in Table II. Since these searchers, associated with the machines that served as sources of volunteered data, expressed medical concerns and are involved in our study, we refer to these users as *concerned subjects* in the remainder of this article. We now describe some relevant attributes of the search interactions of these subjects.

4.4 Concerned Subjects

Of particular interest given our research objectives, were subjects that issued queries containing any of the 12 medical symptoms within the period of time captured by the

duration of our logs. In total, 8,732 subjects issued queries containing at least one of those symptoms and issued more than one query of any sort in the duration of the study, providing an opportunity for observing sessions with an escalation.

In Table III, we present the mean average (M) and the standard deviation (SD) for relevant aspects of the interaction behavior of these concerned subjects. Computed attributes include: the number of queries issued, the number of search sessions per searcher, the percentage of queries that contain a medical symptom, the number of search sessions with a query containing a medical symptom, the number of unique concerns in the queries they issue, the proportion of pages visited whose URL appears in the “Health” category of the ODP,¹ and the proportion of queries that are health-related.

Table III. Summary Statistics (per concerned subject).

Feature	<u>M</u>	<u>SD</u>
Number of queries	978.3	1065.2
Number of sessions	170.6	167.6
Number of unique symptoms	1.3	0.5
Number of queries with ≥ 1 symptom	10.6	13.6
Number of sessions with ≥ 1 symptom	2.3	2.4
Percentage of pages that are health-related	15.4	28.0
Percentage of queries that are health-related	3.6	6.0

The statistics show that, within the culled set of subjects, a small number of symptoms are investigated, that approximately one in seven of the pages they visit is health-related, and about one in thirty queries is health-related. Our analysis also indicates that 78.3% of all queries related to a medical symptom occur within two weeks of the initial query for that symptom. This suggests that searches for symptoms may occur in a bursty manner, with periods of calm punctuated with periods of intense medical search activity.

Statistics such as these may be useful in determining whether some subjects may be potentially predisposed to escalate (e.g., those that query for broad medical symptoms regularly or those that visit a large number of consumer health sites). Later in this article, we study whether there is any relationship between these features and the likelihood of escalation (or non-escalation). Understanding such relationships could provide insight on personalizing medical search in a way that could reduce the likelihood of inappropriate escalation for a particular user or group of users.

4.5 Survey

In addition to the log-based approach outlined in this section, we also composed a survey to elicit peoples’ perceptions of online health-related information, their experiences in searching for health-related information online, and the influence of the Web on their healthcare concerns and interests. We review relevant findings from the large survey. We

¹ Matching URLs to the “Health” category was conducted using incremental backoff up to the top-level domain. The approach we use is similar to that proposed by Shen and colleagues [2005].

distributed the survey within Microsoft Corporation to 5,000 randomly selected employees. Although Microsoft employees are not necessarily representative of the online population, we have no evidence that the employees' experiences with medical Web search differ significantly from those of the general user population.

In the invitation to take the survey, we requested participation of people who had performed at least one search for health-related information. Of the 5,000 people invited to take the survey, 515 volunteers (350 males and 165 females) completed the survey for a participation rate of 10.3%. The average age of respondents was 36.3 years (median = 35 years, $SD = 8.2$ years). The survey contained open and closed questions and covered a broad range of issues in the health domain, including medical history and engagement with healthcare professionals. Five-point scales were used to measure frequency, with the following response options: *always*, *often*, *occasionally*, *rarely*, and *never*.

In Table IV (overleaf), we summarize responses to background questions regarding respondent health-related search habits and their levels of health-related anxiety. The findings show that participants believe that they perform approximately two health-related searches per week and one search for a professionally undiagnosed medical condition every two weeks. They primarily search for themselves or family members and target information on symptoms and serious medical conditions. Around four in ten respondents reported being concerned about having a serious medical condition based on their own observations, when no condition was present. Nearly nine out of ten respondents reported at least one instance where a Web search for the symptoms of basic medical conditions led to their review of content on more serious illnesses; one in five responded that this had happened to them frequently (i.e., responses were *often* or *always*). We find these to be remarkable findings, especially given that respondents were not overly anxious about medical concerns (i.e., only 3-4% of respondents reported that they consider themselves to be "a hypochondriac," and the average health anxiety rating was around three out of ten). The reported prevalence by people surveyed of the review of serious disorders following searches on basic medical symptoms underscores the importance of characterizing and learning more about the escalation of medical concerns in online environments.

Table IV. Summary Statistics on Health-Related Search/Anxiety (per survey respondent).

Health-related search habits (N=515)		
On average, how many health-related Web searches do you perform per month?	<u>M</u> =10.22, <u>SD</u> =45.58, Median=2	
On average, how many health-related Web searches for <i>professionally undiagnosed medical conditions</i> do you perform per month?	<u>M</u> =2.12, <u>SD</u> =5.84, Median=1	
Who are your health-related Web searches primarily for?	Yourself	58.1%
	Relative	36.9%
	Friend or work colleague	3.5%
	Other	1.6%
When you seek health-related information online you generally search for? (multiple responses permitted)	Information on symptoms (e.g., headache, chest pain)	85.8%
	Information on serious medical conditions (e.g., cancer, myocardial infarction)	49.1%
	Medical diagnoses	41.7%
	Forums or pages describing others' experiences with similar conditions to your own	38.1%
	Other	6.2%
Health-related anxiety (N=515)		
On a scale of 1 to 10, how would you rate your overall anxiety about potential medical conditions that are not present or currently undiagnosed (1 = don't worry about health issues, 10 = severe anxiety)	<u>M</u> =2.78, <u>SD</u> =1.71, Median=2	
Do you think that you are a hypochondriac?	Yes	3.5%
	No	96.5%
Have you ever been called a "hypochondriac" by friends, family, or a health professional (e.g., a physician)?	Yes	4.7%
	No	95.3%
Have you ever been concerned about having a serious medical condition based on your own observation of symptoms <i>when no condition was present</i> ?	Yes	39.4%
	No	60.6%
How often do your Web searches for symptoms / basic medical conditions lead to your review of content on serious illnesses?	Always	1.9%
	Often	19.0%
	Occasionally	42.3%
	Rarely	28.5%
	Never	8.2%

5. STUDYING WITHIN-SESSION MEDICAL ESCALATION

We now investigate the *escalation* of medical concerns where an initial focus on common symptoms appears to shift to a focusing of attention on serious illnesses within a single search session. As described earlier, we consider an escalation as occurring when a user initially queries for or visits pages that contain innocuous medical symptoms, and then searches for or browses to pages that contain more serious illnesses. Escalations may arise from exposure to search results, pages that users visit from search results, or external sources such as physician consultations, medical textbooks, or interactions with others that share their symptoms. To minimize the influence of external factors, we focus on search sessions containing a medical symptom in the query—queries that suggest that users have an immediate focus on medical information.

Given a symptom occurring within a session, we noted one of three possible outcomes as follows:

Escalation: Session escalates to an uncommon, serious explanation for the medical condition, e.g., queries for “headache” escalate to queries for “brain tumor.” We were interested in escalations to serious concerns given an initial innocuous complaint. For example, consider the following session:

Query [headache]
Visit <http://pennhealth.com/ency/article/007222.htm>
Query [headache tumor]
Query [brain tumor treatment]

A brain tumor is a concerning possibility when a searcher experiences headache. However, the probability of a brain tumor given a general complaint of headache is typically quite low.

Non-escalation: Session progresses to a non-serious and high-likelihood explanation for the medical condition, e.g., queries for “headache” become queries for “caffeine withdrawal.” Non-escalations are seemingly appropriate given the initial complaint. For example:

Query [headache]
Visit <http://www.headaches.org/consumer/educationalmodules/caffeine/fast.html>
Query [headache coffee]
Query [caffeine withdrawal symptoms]

No change: Session does not escalate or does not continue; either same query is issued repeatedly, another unrelated or non-medical query is issued, or session is abandoned.

Certainly, the review of information about unlikely, yet serious medical possibilities is reasonable, when couched in the appropriate language, with appropriate caveats. From a decision-analytic perspective, consideration of the possible presence of an unlikely disorder can be a rational exercise, given the expected cost of delayed diagnosis and therapy. However, the absence of clear likelihood information or the implicit relay of inappropriate likelihoods can shift rational review to irrational anxiety. Escalations in terms of increased focus of attention and concern may also be reasonable given sets of symptoms combined with details about a searcher’s medical background and family history. Unfortunately, rich sets of symptoms and detailed background information are rarely provided to search engines given the short queries input during a session. Even if

such information was available, search engines do not have the ability to interpret and respond with accurate assessments. Web search engines base ranking decisions on sparse information on symptoms and on various measures of informational relevance. They are not designed to not perform coherent diagnostic reasoning, which would require probabilistic reasoning methods. Thus, for many single or small sets of symptoms input to search engines, several factors may come together—including the informational linkage among common symptoms and rare disorders, the quantity of Web content on rare disorders, the prevalence of the symptoms in healthy people, and the low probability of rare diseases conditioned on those symptoms—to foster unfounded medical anxiety.

Multiple symptoms can be input within a single search session. As we wanted to capture as many concern + escalation/non-escalation pairs as possible, we employed a simple method for associating escalations and non-escalations with symptoms. For each of the symptoms defined in Table II, we took the common explanations, identified by the medical information described earlier, and an equal number of top-ranked serious illnesses ranked in descending order based on their per term co-occurrence statistics. We generated via this procedure a list of common explanations and a list of the top serious illnesses for each of the common symptoms listed in Table II.

For each session, we stored each symptom as it appeared in the logs. Each follow-on query in the session was assessed automatically to determine whether it included a common, benign explanation or a top-ranked serious illness for a symptom. To do this we used the set of serious illnesses and common explanations for each of the 12 symptoms described in Table II. Recall that these possible outcomes were associated with each symptom based on the review of content from the U.S. National Library of Medicine's *PubMed* service and other Web-based medical resources. Serious illnesses and common explanations were verified and expanded by one of the authors (EH).

If the session contained a symptom and an associated top-ranked serious illness or common explanation, the concern + escalation/non-escalation pair (as well as associated information such as time and number of Web interaction events in-between) were stored and the symptom was temporarily retired until the next instance within the current session or a future session. This allows us to contrast escalation from general symptoms with sessions where the concern progresses to the more common, non-escalatory explanation. It is worth noting that search sessions where users escalated and then de-escalated were not common in our logs. Once a concern escalates to a more serious condition this generally persists for the duration of the session.

We now describe some characteristics of query escalations. In particular, we target query escalation and the effect on escalation of subject predisposition. To determine the statistical significance of differences in features we use parametric statistical testing ($p < .05$) and logarithmic transforms as appropriate.

5.1 Query Escalations

Across the logs of all 8,732 concerned subjects, we selected search sessions where the user had submitted a query containing a symptom listed in Table II that then proceeded to escalate either to include a serious illness or a grave concern that was indicative of an increase in the level of severity or subject worry. From the 11,158 sessions that contained a concern, 593 (5.3%) led to a query escalation, 831 (7.4%) resulted in a non-escalation,

and 9,734 (87.3%) led to no change. We note that the estimated escalation and non-escalation frequencies based on our limited, focused vocabulary are a lower bound; higher values are likely with a broader vocabulary that contains more entities and variants for each condition. We investigated why “no change” was so prevalent, and performed detailed hand labeling of a set of 250 randomly selected no-change sessions. Figure I displays the distribution of labels assigned to those sessions. Multiple labels were assigned a session where appropriate. In addition, we divided labels based on whether an escalation or non-escalation occurred. For example, 17% of no-change sessions contained an escalation missed by our automated analysis because the escalatory condition was unspecified for that symptom.

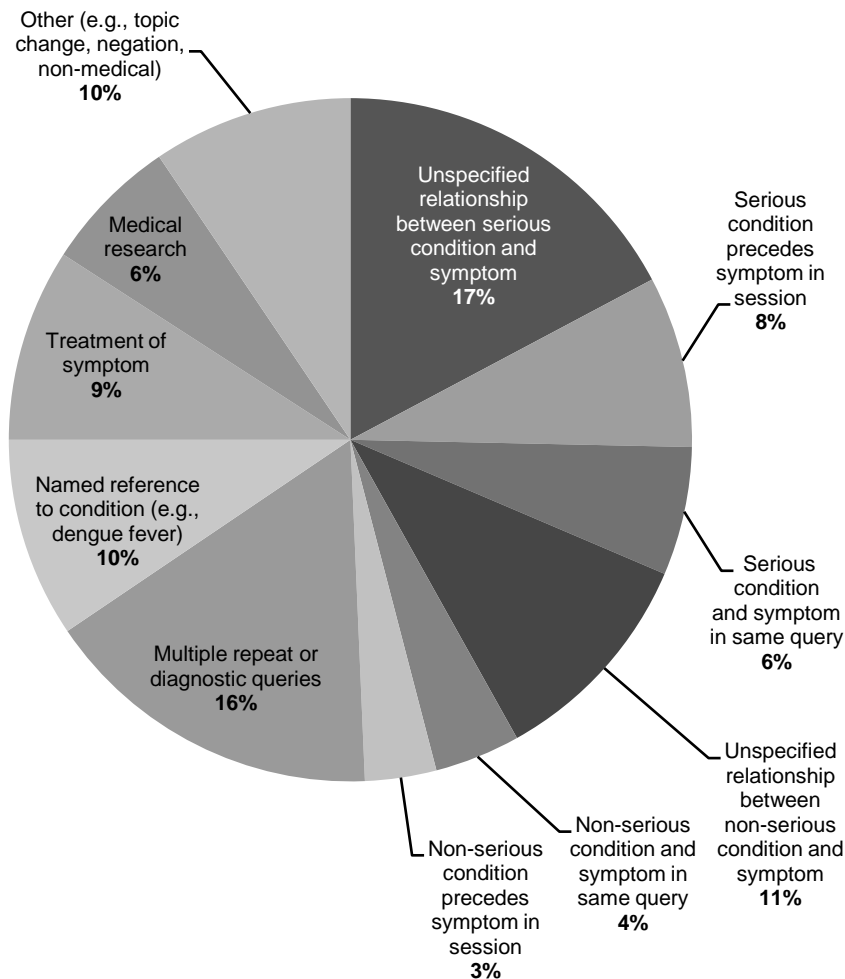


Figure I. Distribution of labels assigned to set of hand-labeled no-change sessions.

In Figure I, we show that many no-change sessions are explained by: (i) unspecified relationships between serious/non-serious condition and the symptom (28%);² (ii) the symptom appearing after the serious/non-serious condition in the session (11%), or; (iii) the symptom appearing in the same query as the serious/non-serious condition (10%).³ These three types of escalation or non-escalation were not recognized by our automated analysis. The remaining no-change sessions (51%) had no escalation or non-escalation, and comprised: (i) multiple repeat or diagnostic queries (16%); (ii) an initial named reference to a particular condition (e.g., dengue fever) and then searches for more information about that condition (10%); (iii) searches for treatment options for a symptom (9%), and; (iv) medical research for journals and specific studies (6%). The other no-change sessions (10%) included topic shifts following symptom input, searches for drug names and symptoms associated with them, negations (e.g., “not fever”), and non-medical sessions that had not been filtered out by our automated tools.

Of the sessions in our logs that led to a query escalation, 91.6% were caused by the inclusion of the name of a serious illness in the query and 8.4% by the inclusion of an accelerating or grave concern in the query (e.g., the query “chest pain” escalating to “severe chest pain”). Out of the 700 subjects for whom we observed an escalation or non-escalation, 230 subjects (32.9%) escalated and 491 (70.1%) did not escalate. There was an overlap of only 21 subjects between these two groups, suggesting that concerned subjects may be somewhat predisposed to escalate or not escalate, something we study in more detail later in this article.

5.1.1 Session. We were interested in whether there were differences in interactions by searchers during sessions where escalation occurred, versus where users tended towards a common explanation, or when there was no significant change in the semantics of their medical queries in the session containing the medical symptom. In Table V, we present summary statistics on the sessions where at least one of three types of event occurs. In the last row of the table we also include the proportion of medical pages from trusted source (i.e., .edu, .gov, and .org domains), used as a proxy for the reliability/complexity of Web content viewed.

² For example, a rash may be indicative of meningitis, but meningitis was not one of the possible escalations or common explanations for rash considered in our automated log analysis.

³ Note that 31.4% of no-change sessions showed escalation and 17.9% of sessions had non-escalation. If we assume that these percentages provide approximate likelihoods for all no-change sessions and include the automated log analysis, the percentage of sessions with escalations/non-escalations is 32.7%/25.3% respectively.

Table V. Summary Statistics (per search session).

Measure	Escalation		Non-escalation		No change	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Duration (seconds)	3801	2806	3412	2633	2806	2391
Number of query iterations	24.8	18.5	16.6	14.5	10.3	9.6
Number of pages	29.2	16.3	16.1	13.4	13.6	12.2
Number of unique domains	9.8	7.2	6.4	6.3	4.6	4.8
Percentage of medical pages	39.1	23.8	39.2	25.2	18.4	16.5
Percentage of medical pages from trusted sources	25.1	20.7	19.1	18.3	10.1	8.7

We performed a one-way independent measures analysis of variance (ANOVA) to determine whether the observed differences between sessions were significant. To reduce the number of Type I errors, i.e., rejecting null hypotheses that were true, we set the alpha level (α) to .008 i.e., .05 divided by 6, the number of tests performed. Our findings suggest that sessions that escalate last longer (in terms of time and pages visited), contain more queries, and include visits to more unique domains and trusted sources (all $F(2, 11155) \geq 7.27$, all $p \leq .007$; Tukey's *post-hoc* tests: all $p \leq .005$, $\alpha = .008$). It appears that the exposure to additional Web content, different perspectives from multiple domains, and perhaps detailed information from trusted sources may contribute to the likelihood that escalation will occur. In addition, it is worth noting that some escalating users engaged in extremely long sessions lasting over three hours. Visual inspection of aggregated representations of these concerned subjects' search sessions ruled out session demarcation errors in our log parsing in all but two cases; those cases were removed from the data prior to analysis. It is also worth noting that sessions with any large change in health-related semantics (i.e., an escalation or non-escalation) were not only longer than those with no change but included around twice as many medical pages, and of those pages, twice as many came from government or academic sources. The volume and type of medical information viewed may also contribute to escalation or non-escalation likelihood.

5.1.2 Distance Between Symptom and Escalation / Non-escalation. We explored the distance between the submission of a query containing the initial symptoms and the escalation or non-escalation occurring within a single session. A better understanding of the onset of escalations may allow us to predict when they are going to occur and to build tools that can adapt interfaces or ranking algorithms to minimize the likelihood of escalation given common symptoms. We measured distance in three ways:

- Time in seconds: The number of seconds between the query for the symptom and the escalation or non-escalation.
- Number of queries: The number of queries between the symptom and the escalation or non-escalation.

- Number of page views: The number of non-search pages viewed between the submission of the query containing the initial symptom and the escalation or non-escalation.

To study escalation, we examined sessions containing at least one escalatory query and measured distance from the first symptom-related query to the first escalation. To characterize non-escalation, we examined sessions containing only symptoms or non-escalations, and measured the distance from the first symptom-related query to first non-escalation. Table VI shows the average and the standard deviation for the distances of each of these three measures between the symptom and the escalation or non-escalation.

Table VI. Escalation/Non-escalation Distances.

Distance Measure	Escalation		Non-escalation	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Time in seconds	132.7	140.2	92.3	73.7
Number of queries	2.3	2.2	1.2	1.1
Number of page views	2.2	1.9	1.1	1.0

As can be seen from Table VI, distances between symptom and serious illness or grave concern (escalation) are larger than between symptom and non-serious common explanation (non-escalation), as verified with independent measures t -tests (all $t(1422) \geq 2.58$, all $p \leq .01$). In the additional time between query and escalation, users appear to be submitting more queries and viewing more pages than between query and non-escalation. The high variance of each of the distance measures suggested that they may not be evenly distributed over time. In Figures II, III, and IV we illustrate graphically the frequencies of actions indicative of escalations and non-escalations as functions of the variables shown in Table VI. Times between query and escalation/non-escalation are considered at 30 second intervals with a maximum timeout of 600 seconds. Since non-escalations outnumber escalations, the lines depict a percentage of the total number of escalations or non-escalations, rather than the actual frequency values.

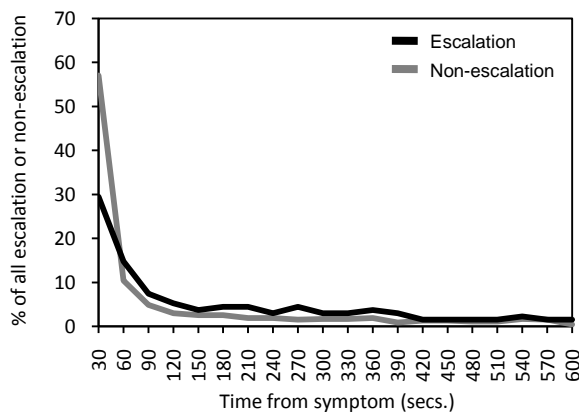


Figure II. Temporal distance from initial input of symptom (within session).

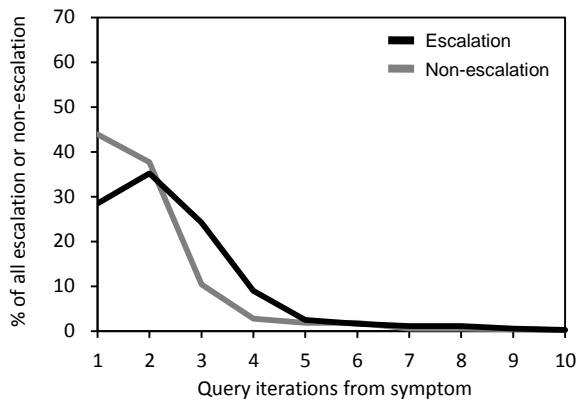


Figure III. Query distance from initial input of symptom (within session).

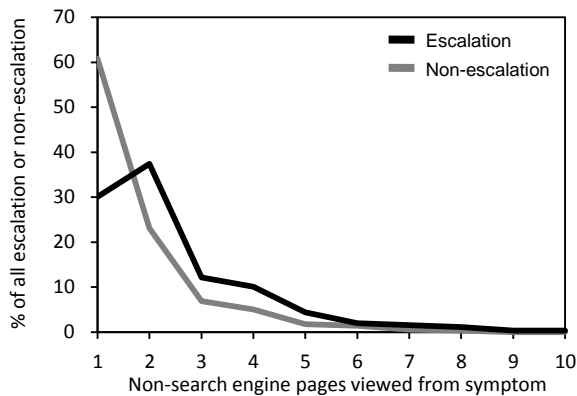


Figure IV. Navigational distance from initial input of symptom (within session).

The graphs show that: (i) escalations occur more gradually throughout the search sessions than non-escalations, (ii) escalations occur less frequently immediately after the first follow-on query, and (iii) escalations occur more frequently for a few non-search pages after the query and then tail off. These observations might be explained by a domain sampling model (Nunnally, 1967) where a sufficient pool of available evidence of Web content about a symptom of interest is collected by users in return for some assumed reasonable allocation of search and browsing effort. The pool of data is considered to be a sufficiently representative sample of all relevant data on the Web for deliberating about the explanation for the symptoms. In the context of such a sampling model, we might expect the observations displayed in Figures II, III, and IV, if each page visited in pool of evidence has a probability of causing an escalation via containing information about a serious explanation; the probability of an escalation occurring would increase with multiple views within a bound of the evidence set.

5.2 Query Escalations and Subject Predisposition

In addition to viewing pages containing the names of serious illnesses, some users may simply be predisposed to experience escalations in their medical searches. For each of the

700 subjects that experienced an escalation or a non-escalation we sought to determine whether there were differences in the medical searching behavior or source selection of these users. In particular we studied behaviors relating to the average number of medical queries per day, the proportion of overall pages viewed that were medical, the average number of medical page views per day, the number of unique symptoms, and the proportion of medical pages viewed that came from “trusted” medical sources (e.g., .edu, .gov, and .org). In Table VII, we present the average values for each of these features for concerned subjects with query escalations, subjects with query non-escalations, and those subjects that searched for a medical symptom but did not experience any increase or noticeable change in the nature of information sought (i.e., no change). Again, these values are a lower bound based on the ability to detect medical query instances given the partial list of medical terminology used.

Table VII. Interaction Features from Subject Groups.

Measure	Escalators		Non-escalators		No change	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Num. of medical queries per day	0.6	0.7	0.4	0.7	0.2	0.4
Num. of unique symptoms	1.8	1.5	1.4	0.7	1.1	1.0
Num. of medical page views/day	0.6	0.7	0.4	0.5	0.2	0.3
% of all pages medical	5.5	5.7	5.1	5.2	2.3	2.1
% medical pages from trusted	47.9	27.9	40.7	28.0	36.9	33.6

There were differences in the medical search behavior of all three user groups, across all features, suggesting that subjects could be in some way predisposed to escalation (one-way independent measures ANOVA: $F(2,11155) \geq 7.55$, all $p \leq .006$, $\alpha = .01$). In addition, for the escalators and non-escalators, we performed a multiple regression analysis with the five features listed in Table VII as independent variables and the proportion of sessions containing an escalation or non-escalation as the dependent variable. The multiple correlation coefficients for the escalators and the non-escalators are .32 and .24 respectively, both of which significantly differ from zero (*Escalators*: $F(5,224) = 5.22$, $p < .001$; *Non-escalators*: $F(5,485) = 5.93$, $p < .001$). Although the correlation coefficients are in the low-moderate range, they do suggest that it may be possible to infer escalation likelihood given only information about searchers’ medical Web search interaction history, especially for subjects who escalate. Details of the frequency or content of related user activities beyond Web search behavior (e.g., interactions with physicians, perusal of medical textbooks or medical articles in the popular press, discussions on symptoms and conditions with other patients with similar or related ailments), may help us estimate escalation likelihood even more reliably. This also suggests that factors beyond the exposure to medical information in Web search results—in this instance a user’s predisposition to escalate—can influence the likelihood that an escalation can occur. These findings highlight the complexity of this challenge that systems trying to alleviate inappropriate medical anxiety face. However, the findings do suggest that in light of limited information about a user’s interaction history, we may

be able to compute an escalation likelihood that could be factored into tailored ranking algorithms for users. In the next section, we study the persistence of health concerns.

6. PERSISTENCE OF HEALTH CONCERNS

One way in which psychological disorders, such as different forms of anxiety and depression, are diagnosed is through characterization as impairing functioning (i.e., how psychological symptoms interfere with peoples' normal daily activities). The *Diagnostic and Statistical Manual of Mental Disorders* introduced earlier provides guidelines to psychologists and psychiatrists for classifying such mental disorders. The manual states that a person must have a set of characterizing symptoms that are significant enough to cause impairment for them to have a disorder. The persistence of unfounded medical concerns can be debilitating and lead to a reduced quality of life for those afflicted. Persistence of medical anxiety has been studied previously [Asmundson et al. 2001] but not in the context of Web search and not in particular following inappropriate escalations such as those described in the previous section.

In addition to characterizing medical escalations as they occur, we also wished to characterize the extent to which anxieties persist across multiple search sessions, potentially spanning multiple days, weeks, or months, and the extent to which they interrupt users' search activities, based on logs of healthcare-related querying and post-query browsing history. Prior to doing so, we determined the prevalence of persistence and interruption related to medical escalations among our survey respondents. We asked those who had had experienced medical escalation (per the question in the last row of Table IV) to respond to three attitude statements about the persistence and impact of searches for serious illnesses following an initial escalation. A summary of the findings is presented in Table VIII.

Table VIII. Responses to Survey Questions Regarding Persistence and Interruption.

Attitude statement	Responses (N=472)	
Following an initial escalation from querying for symptom / basic medical condition to querying for a serious illness, your queries for that serious illness persist over weeks, months, or years	Always	0.4%
	Often	6.7%
	Occasionally	25.8%
	Rarely	39.8%
	Never	27.3%
Following an initial escalation from querying for a symptom / basic medical condition to querying for a serious illness interrupted your <i>online</i> activities	Always	0.2%
	Often	3.6%
	Occasionally	19.3%
	Rarely	35.0%
	Never	41.9%
Following an initial escalation from querying for a medical symptom / basic medical condition to querying for a serious illness interrupted your <i>other</i> activities	Always	0.2%
	Often	3.0%
	Occasionally	20.3%
	Rarely	36.9%
	Never	39.6%

The responses summarized in Table VIII suggest that seven out of ten respondents searched for serious illnesses post escalation at all (6-7% of respondents did so frequently). The online and other activities of around six out of ten survey respondents were affected at least once by interruptions related to prior medical escalations (3-4% of respondents were affected frequently). Post-escalation persistence and interruption affected a significant number of our respondents. We investigated these issues further in our log-based study.

In this section, we extend our log-based analysis beyond a single search session to focus on the *reoccurrence* of medical conditions over extended periods of time such as weeks and months, and *interruptions* in other searches and activities that are caused by an urge to perform medical searches about a worrying disorder following a detected escalation. Re-occurrence and associated interruption implies significant anxiety and cost that might be overcome with enhanced awareness and technological innovation.

6.1 Re-occurrence

We seek to understand how escalations can lead to persistent concerns over longer periods of time. Beginning with the first occurrence of an escalation, via noting terms representing a serious illness, we determined with an automated procedure how often the serious illnesses associated with that concern reappeared until the end of the interaction logs for each subject. The concern may continue to reoccur beyond the end of our log sample, but we have sufficient information to characterize its onset and its reoccurrence to a reasonable level. We did the same for non-escalations and symptoms. Again, privacy considerations were central. *We tracked no other aspects of subjects' interaction behavior, only whether the condition reoccurred again in queries issued.* We envision that search services could be personalized to provide information relevant to a recurring condition—or anxiety about a condition, based on search history, given the appropriate addressing of privacy concerns.

In total, there were 2,542 re-occurrence events in our logs, affecting 1,177 subjects (13.5%). Re-occurrence seems to form an important part of escalatory behavior. Of this total, 1,290 (50.8%) were from symptom reoccurrences (i.e., searching for the same symptom across multiple sessions), 580 (22.8%) from the reoccurrence of querying on serious illnesses (note that 65% of these re-occurrences were for “cancer”), and 672 (26.4%) from the reoccurrence of common explanations. In Table IX, we show the number of search sessions and the number days between the re-occurrence events for each of these three types.

Table IX. Distance Between Medical Re-occurrences.

Distance measure	Symptoms		Serious illnesses		Common explanations	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Session	22.8	29.2	20.5	20.6	12.6	17.0
Day	18.9	23.3	19.0	25.6	11.4	10.9

Given the durations shown in Table IX, it appears that concern about medical conditions can persist over multiple sessions and multiple days. The significance of the difference in reoccurrence frequencies between serious illnesses and common explanations may be because non-serious ailments such as eyestrain or migraine can be related to multiple more serious conditions, so are likely to occur as queries more frequently (one-way independent measures ANOVA: *Session*: $F(2,2540) = 3.92$, $p = .02$, Tukey's *post-hoc* test: $p = .02$; *Day*: $F(2, 2540) = 4.61$, $p = .01$; Tukey's *post-hoc* test: $p = .01$, $\alpha = .025$). We noted a high degree of variance in each of these metrics when broken out by groups of subjects or condition. As suggested earlier in the article, it seems that re-occurrence is staccato in nature, with periods of relative calm followed by intense medical searching; these may align with periods of medical anxiety, although more research with human subjects is required to test this.

6.2 Persistent Anxieties as Interruptions

Medical conditions can profoundly affect the daily activities of those concerned. To be diagnosed with a disorder such as hypochondria individuals need to not only demonstrate the symptoms of medical anxiety but also that their concerns impair their normal daily activities. Interruption has been studied in detail in the human computer interaction and psychology literature [Ovsiankina 1928; Czerwinski et al. 2004; Ibaq and Horvitz 2007]. However, these studies have focused on experimental interruptions or on in situ investigations of the costs of alerts from electronic communications and telephones, and on self-interruptions to switch among work tasks. We shall define an interruption instance in our study as a situation where:

- (i) We have already observed a user escalating from a common condition to a more serious illness at some point in their search history;
- (ii) The same user engages in another session at some future time (later that hour, later that day, the next day, the next week, etc.) that starts with at least one non-healthcare query;
- (iii) That same session evolves to then contain healthcare-related queries, and;
- (iv) Those same healthcare-related queries describe the same serious illness as the escalation in (i).

In total, there were 885 instances of interruption in our logs, affecting 480 concerned subjects (5.5%). The validity of these interruptions was verified by visual inspection of a sampling of the sessions by one of the authors. Interruption mainly arose from searching for symptoms repeatedly across multiple medical sessions (62.7%), rather than serious illnesses or common explanations. Queries related to "cancer" and "pregnancy" interrupted users most for escalations and non-escalations respectively. For some users interruption represented a potentially significant hindrance on their search activities, with some medical concerns interrupting over 15% of their search sessions. Although there were only a small fraction of concerned subjects (less than 20) for which the situation was as serious, their presence at all highlights the opportunity to modify search engines and content so as to help people to manage their medical concerns more effectively.

7. DISCUSSION

We have investigated medical search behavior and focused on the potential for Web search and navigation to lead to escalations of medical concerns. Via a survey and large-scale log analyses, we found evidence that such escalations can occur and may lead to both short- and longer-term anxieties and unnecessary costs in time, distraction, and engagements with medical professionals. We believe that our studies are a call for additional research on the use of online search and retrieval for self diagnosis. Our initial explorations underscore the potential value of focusing attention on designs and mechanisms to address the challenges identified. In this section, we discuss pertinent results and offer recommendations for the design of information-retrieval systems to support more effective medical searching and a reduction of cyberchondria.

7.1 Judgment Biases

Beyond potential problems with the quality of medical content described earlier, we believe that cyberchondria is based more centrally on intrinsic problems with the *implicit use of Web search as a diagnostic engine*. In such a usage, disorders described in a ranked list of results, following a query containing symptoms, may be coarsely interpreted by users as diagnostic entities sorted by likelihood. To test the validity of this claim we asked our survey respondents about their interpretation of health-related Web search results. A summary of responses to relevant questions is included in the first two rows of Table X. The last two rows contain responses to questions about respondent engagement with health professionals.

The survey responses summarized in the first two rows of Table X show that three in four respondents have at least once interpreted the *ranking* of Web search results as indicating the *likelihood* of the illnesses, with links to pages describing more likely diseases appearing higher up on the result page. Just under one quarter of all respondents interpreted search results in this way frequently, and approximately the same proportion had used Web search engines as though Web search functioned as a medical expert system. The last two rows of Table X show that one in five survey respondents were convinced to seek medical attention based on the review of online medical content. However, only one in four of the respondents that sought medical attention had a medical condition that warranted them doing so.

Table X. Responses to Survey Questions Regarding Searches for Diagnoses.

Questions	Responses	
If your queries contain medical symptoms, how often do you consider the ranking of Web search results as indicating the likelihood of the illnesses, with more likely diseases appearing higher up on the result page(s)? (N=515)	Always	2.7%
	Often	20.8%
	Occasionally	27.4%
	Rarely	26.8%
	Never	22.3%
Have you ever used Web search as a medical expert system where you input symptoms and expect to review possible diseases ranked by likelihood? (N=515)	Yes	24.5%
	No	75.5%
Do you believe you have been in the situation where Web content “put you over the threshold” for scheduling an appointment with a health professional, when you would likely have not sought professional medical attention if you had not reviewed Web content? (N=515)	Yes	23.7%
	No	76.3%
Did the appointment reassure you that your worries were not justified? (N=122)	Yes	73.0%
	No	27.0%

These results demonstrate the effect of Web content on non-Web behaviors and show that a significant portion of the user population are using search results as a proxy for what physicians refer to as the *differential diagnosis*—the list of diseases under consideration ranked by their corresponding likelihoods, given a patient’s history and symptoms. Such usage of Web search as diagnostic inference is natural for people, yet is not typically considered in the design and optimization of general-purpose ranking algorithms. For example, ranking methods employed by search services may take user clicks and dwells on Web pages as an indication that the page is relevant to the adjacent query [Agichtein et al. 2006]. If the “worried well” are more drawn to content about potentially serious concerns than about more likely but less worrisome explanations, the ranking of Web pages on rare but serious disorders could be skewed towards the top of ranked lists. Such a bias could be an important source of erroneous but self-reinforcing feedback; studies have demonstrated that users tend to click on the top-ranked results of Web pages [Joachims et al. 2005]. Thus, anxious click-throughs on items appearing on search result pages, in response to queries about common symptoms, may lead to ranking refinements that push rare but concerning health problems increasingly higher in the list over time.

Beyond self-sustaining anxiety-driven click-throughs, other core biases may play an active role with the use of the Web search as medical diagnosis. Cognitive psychologists who study human judgment and decision making have presented evidence that people often employ heuristics in assessing the likelihoods of events that can lead to biases in judgment, as compared with normative probabilistic updating [Tversky and Kahneman 1974]. We believe that previously studied heuristics and biases of human judgment likely play a significant role in cyberchondria. Beyond their influence on people pursuing medical information on the Web, the biases likely also directly influence the indexing and ranking of medical content, as the search methodologies are not designed to perform coherent probabilistic updating. We focus on two well-studied biases: (i) *base-rate neglect*—the failure to adequately consider background or *prior* probabilities of events and (ii) the *availability bias*—the influence of recent exposure to events on a subject’s assessments of probabilities of the events.

Base-rate neglect has been detailed in the literature on the psychology of judgment [Kahneman et al. 1982], and, more specifically, in the literature on medical decision making [Elstein et al. 1978]. Base-rate neglect has been invoked to explain the failure of people to accurately take the low prior probabilities of rare events into consideration in reasoning about outcomes. It is critical, in effective medical diagnosis from symptoms, to take into account both the prior probability of illnesses and the probabilistic updates provided by sets of observed symptoms. For rare diseases, even multiple evocative symptoms may not raise the likelihood of an illness enough to be a significant concern. Even medical experts are not immune to overestimating the likelihood of rare disorders because of base-rate neglect, or more generally, because of an inadequate folding in of the small prior probabilities of rare disorders. Base-rate neglect likely plays a central role in self-diagnosis by laypeople engaged in search and navigation on the Web.

Beyond the failure by people and search engines to integrate a consideration of prior probabilities, cyberchondria may be additionally stimulated by the influence of the quantity of content about rare disorders in results and browsing on the cognitive availability of the disorders. Psychologists of judgment and decision making have provided evidence that the density and recency of events makes them more “available” to people when they reflect about likelihoods and that this increased availability leads people to expect that the events will occur with higher probabilities. The reliance of people on the cognitive availability of events in the process of generating estimates of probability has been referred to as the *availability heuristic* within the psychology of judgment and decision making [Tversky and Kahneman 1974]. Studies have demonstrated how subjects’ probability assessments can be manipulated by changing the recency and density of events that they are exposed to. On the Web, larger amounts of indexed content about serious disorders can make these disorders more available to both search engines and to people who search and browse content. Similar or larger quantities of content may be devoted to rare, yet serious illnesses compared to content on more common explanations for symptoms. For example, headaches are far more often caused by caffeine withdrawal than by cerebral hemorrhage or brain tumors, but there is a great deal written about the link between headaches and the more serious, albeit rare ailments. Although it may be reasonable for more attention, and thus, literature, to be devoted to discussion of serious but rare disorders than to common, benign causes of symptoms, the abundance of content on rare diseases can lead search engines and people astray.

In summary, base-rate neglect and availability bias are well-known biases in judgment associated with the failure to integrate the relevance of low prior probabilities and the erroneous linking of the availability of information to likelihood of events likely play a role in cyberchondria. These phenomena influence people directly, but also can act on search engines themselves, leading to the generation of search result lists that contain low probability but highly concerning items near the top of results pages. In addition, click-through and dwell on serious disorders may lead to self-sustaining boosts in the ranking of the rare but troubling disorders.

Although query escalations have been our primary focus, it is also worth considering post-query navigation to Websites containing serious explanations and escalatory terminology as sources of escalatory evidence. To establish the extent to which interaction with Websites could reveal medical escalations we asked the 198 survey respondents (38.4%) who had experienced an increase in anxiety from searching health information online, to provide more information about the source of their anxiety. For

those who suggested that the source was content-related, we asked for more information about the nature of the content. The findings are summarized below in Table XI.

Table XI. Sources of Health-Related Anxiety and Contribution of Content Features.

Questions	Responses (N=198)	
What was your anxiety related to? (multiple responses permitted)	The content of pages visited from a result click	70.7%
	The content of the Web search result pages (e.g., page titles, captions, URLs)	31.8%
	The content of pages visited on the browse trail following a result click	27.8%
	The rank order of the returned pages	11.1%
	Other	4.5%
What was it about the content of those pages that contributed to your anxiety? (multiple responses permitted)	Mention of serious explanations	64.1%
	Presence of escalatory terminology (e.g., grave, fatal, life-threatening, serious)	41.4%
	Mention of serious explanations and no (or very few) non-serious explanations	36.4%
	Reliability of the source	28.3%
	Presence of complex medical terminology	18.7%
	Other	10.1%

The responses show that search engine result pages, the contents of the pages visited directly from the result pages, and pages visited thereafter, may all contribute to health-related anxiety to different extents. On those pages, it was the mention of serious explanations and escalatory terminology that contributed most to respondents' distress. Interaction with pages containing serious explanations or escalatory terminology could therefore serve as a proxy for medical escalation if no further query evidence was available, or to add support to query-garnered evidence if it was available. So-called *navigational* escalations could involve users migrating from queries about common symptoms to: (i) pages with text on related serious explanations, (ii) a more conservative estimate of (i) where pages must have text on related serious explanations and no mention of related non-serious illnesses, and (iii) pages whose URL contains a serious illness name (e.g., www.cancer.org). However, more research is needed on the nature of pages that reveal medical escalation versus, say, those that merely list possible causes for a medical condition. Once we have a better understanding of such implicit evidence we can incorporate navigational escalations into our characterization of cyberchondria and predictive models, such as those described in the next section.

7.2 Design Recommendations and Future Opportunities

A more complete understanding of potential biases and of characteristics associated with how people search for common symptoms can lead to the design of search systems that can reduce user distress and support more informed medical education and decision making. In one area of innovation, medical searches may be recognized and specially handled. Specialized ranking algorithms have been studied for medical domains (e.g., [Luo et al. 2007]) and for classifying queries as health related. Algorithms tailored to the

medical domain may be able to handle longer search queries (including natural language descriptions of symptoms with little medical terminology), with the aim of returning comprehensive lists of relevant search results. Comprehensiveness is important since patients or physicians do not want to miss important documents that may contain useful diagnosis or treatment information.

We are particularly interested in techniques that promise to reduce the likelihood that users will become inappropriately concerned. Methods for reducing cyberchondria include developing techniques for recognizing health-related queries, and for considering such evidence as the nature and timing of the review of medical content, as in Tables VI and VII. Opportunities for addressing cyberchondria include the following:

Detection of diagnostic intent: There is an opportunity to detect if a searcher is employing Web search to perform diagnosis, much as they might use a medical expert system [Heckerman et al. 1992] if such a system were available—where symptoms are input and a list of a reasonable explanations ranked by their likelihood is reviewed. We are pursuing the creation of classifiers that indicate when a user is likely using Web search as a diagnostic system. As shown in the survey findings reported in Table X and our log analysis, this is a common user activity. Estimates of user's predisposition to escalating might be obtained through log analyses of prior medical escalations. Such information might be used to predict or detect escalations, and then action might be taken to reduce the likelihood of unnecessary anxiety. For example, given detection of "diagnostic intent," search services could provide a list of diseases sorted by likelihood, along with assistance and caveats in interpreting the results. There is opportunity for online services to forward users to diagnostic systems or to seamlessly shift their operation to an explicit probabilistic diagnostic modality, including the use of methods that engage users in an active dialog that is driven by computations of the expected value of acquiring different types of information. Online services that assist people with interpreting symptoms would not necessarily need to serve as frank online diagnostic systems. There is great room for enhancing online search and content to recognize and respond appropriately when people seek diagnostic support for common symptoms and sets of symptoms. As an example, a search service might display above search results the overall incidence rates of concerning entities, as well as incidence rates of related benign explanations linked to detected symptoms. Rates conditioned on different age groups and on common symptoms linked with the disorders, could also be included. Given our finding that "trusted" sources are viewed in sessions associated with escalations more than sessions without escalations, information gleaned largely from these sources could be presented in a way that is more understandable to non-expert users. Such an approach may reduce the potential for escalation by providing analyses similar to the kind of higher quality information seen with domain search versus Web search, as captured in Table I.

Providing expertise: Table XI showed results of survey questions on sources of health-related anxiety. The results highlight how unreliability of Web sources and the content of Web search results contributes to the heightened anxiety of around three in ten survey respondents. Search engine engineers as well as authors of sites with information on rare, serious concerns should remain aware that searches on common symptoms and concerns may be entry points to content on serious disorders. Authors can provide discussion that about the likelihoods of more common, less concerning illnesses and links to discussions of the more common, benign explanations. To improve the reliability of the information present in search results, expert sources of medical information might be consulted by

search providers in automated and handcrafted analyses. This could ensure that frequent searches about medical symptomatology are linked with reasonable lists of results that are unlikely to induce unfounded concerns about more serious illnesses. The labor costs required to create these lists for a small set of the most popular queries would be small compared to the possible benefit to users in feeling assured that the results were reliable. More generally, insightful flowcharts or decision trees displayed early on in the pursuit of an online diagnosis may be of great help to people who might otherwise become needlessly anxious. More details can be provided to searchers about the potentially low incidence rates when factors such as age, gender, and other evidence that is easy to observe. Symptoms and signs can be described in more detail and in terms that searchers can understand, especially when subtleties of a presentation are important in distinguishing un concerning versus concerning variants of symptoms. We understand that the latter can be very difficult and that subtleties are sometimes not even appreciated by physicians outside of specialties. For example, surgeons with a great deal of experience with appendicitis may be more skillful than an emergency department physician at interpreting abdominal pain; a generalist's interpretation of "rebound tenderness" may need to be confirmed by a consulting surgeon.

Debiasing search results and searchers: The findings reported in Section 3 demonstrated the potential the Web offers for escalation. In addition, the survey findings reported in Table X and Table XI shows that the rank order of search results made one in ten respondents more anxious. Biases in medical information on the Web might be studied directly and methods, employing reliable human and digitally encoded medical expertise, could be used to de-bias results. For example, the salience of a serious disorder may lead to more content being generated and available about the serious concern, and, thus, to higher-ranked and more available results when common symptoms are explored. If such availability is interpreted as probabilities, in line with studies by psychologists of how people may use and misuse the availability heuristic, searchers may be misled about likelihoods. Such bias might be handled with insightful filtering and de-biasing analysis. We note that analogous biases, based in the mismatch between the quantities of content available on the web and the likelihoods of different explanations, and also the related misinterpretation of page relevance as likelihood, may pose similar problems for search and retrieval in non-medical areas.

Evaluating search results: Frequent and stereotypical escalations and related behaviors might be detected as heralds for potential problems with search results. Features such as those used in the analysis presented in this article (e.g., Table V and Table IX) could form the basis of detection algorithms developed for this purpose. Queries flagged as candidates for escalation could be assigned to a domain expert for the creation of a handcrafted list as described previously. Web pages frequently present in escalatory events could be down-weighted in the ranking algorithm or marked for subsequent expert review.

Click-through tuning: We mentioned that standard application of rank optimization methods, that take as input click-through and dwell data as indications of appropriate and inappropriate result lists, might lead to special problems if the worried well were clicking on results that described less likely but more serious concerns associated with symptoms. Such methods might need to be adjusted to handle medical queries in a special manner, such that the escalatory potential of a page is also considered alongside interaction features such as the click-through frequency and dwell time when ranking search results.

In all these cases, tailoring search support offered by a system to a particular user, or group of users, based on their estimated escalation likelihood (e.g., some representation of their level of predisposition) may help reduce instances of cyberchondria. There is also opportunity to develop methods for detecting anxiety based on escalations and frank hypochondriases based on short-term interactions such as user click-through or, where privacy concerns have been addressed, over longer-term interactions. We are actively working toward the goal of automatically detecting cyberchondria using Bayesian inference networks and machine learning algorithms, with the aim of reducing the number of users affected by the phenomenon using alerting mechanisms in Web browser plug-ins or on search engine result pages. We are also investigating the use of query chains (similar to [Radlinski and Joachims 2005]) to study series of queries and escalations/non-escalations rather than individual instances as described in this article.

The problems identified, lessons learned, and solutions for enhancing medical search described so far in this section will likely be relevant to other specialty searches where concerns are likely to escalate. For example, in auto repair an engine noise may relate to a faulty oil pump or a more serious, but also more unlikely, cylinder head problem. Beyond the application of caveats and ideas learned in healthcare search to domains with analogous notions of an escalation, the challenges and opportunities for enhancements via special indexing, analysis, design, and user interfaces may more generally point to the need for special handling of specialty searches. This recommendation differs from the trend of indexing and ranking of results with methodologies that are applied universally across domains.

An advantage of the methodology we employed in our study is the scale that is available via interaction logs. The inclusion of the survey helped to bolster some of our claims and provide ideas for future research. However, we feel that studies with groups of live subjects doing health searches would also be valuable. An aspect of this research that will form part of future work is to perform user studies with actual patients to deepen our understanding of medical escalation and the costs involved in such escalation (e.g., resources expended and unnecessary interactions with healthcare providers).

8. CONCLUSIONS

We have presented a log-based study of medical Web-search behavior. The study carves out a nascent set of research challenges for the IR community centered around cyberchondria, focused on the unfounded escalation of medical concerns. We analyzed the escalation of concerns about common symptoms into queries on serious, rare illnesses within a session. We conducted a large-scale survey to support our claims and highlight opportunities for future work. We found that escalation of medical concerns is potentially related to the amount and distribution of medical content viewed by users, the presence of escalatory terminology in pages visited, and a user's predisposition to escalate or seek more reasonable explanations for ailments. We also demonstrated that the persistence of concerns following an escalation and the effect that such ongoing concerns could have on interrupting users' activities over a prolonged period of time. We discussed several potential factors contributing to the rise of inappropriate concern, including biases of judgment studied in cognitive psychology. Beyond affecting people directly, the biases of availability and base-rate neglect may be directly influencing the ranking of results by search engines. Finally, we discussed several methods and designs that hold opportunity

for improving the search and navigation experience for health seekers. There are algorithmic challenges in incorporating likelihood estimates and de-biasing search results, evaluation challenges in determining the probability that a set of search results will lead to unfounded escalation, and interface challenges in when and how we should alert users that an escalation is imminent or has already occurred. Search engine architects have a responsibility to ensure that searchers do not experience unnecessary concern generated by the definitions of relevance and the ranking algorithms their engines use. They must be cognizant of the potential challenges of cyberchondria, and focus on serving medical search results that are reliable, complete, and timely, as well as topically relevant. Directly tackling cyberchondria is an opportunity to leverage readily available expertise in the information-retrieval and medical informatics communities in areas such as document ranking, user modeling, machine learning, and user interface design for the direct benefit of the many people turning to the Web to interpret common medical symptoms.

REFERENCES

- AGICHTEIN, E., BRILL, E. AND DUMAIS, S. (2006). Improving Web search ranking by incorporating user behavior information. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 19-26.
- AMERICAN PSYCHIATRIC ASSOCIATION. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- ASMUNDSON, J.G., TAYLOR, S. AND COX, B.J. (2001). *Health anxiety: Clinical and research perspectives on hypochondriasis and related conditions*. Wiley.
- AYERS, S. AND KRONENFELD, J. (2007). Chronic illness and health-seeking information on the internet. *Health*, 11(3): 327-347.
- BAKER, L., WAGNER, T.H., SINGER, S. AND BUNDORF, M.K. (2003). Use of the internet and e-mail for health care information. *Journal of the American Medical Association*, 289(18): 2401-2406.
- BARSKY, A.J. AND AHERN, D.K. (2004). Cognitive behavioral therapy for hypochondriasis. *Journal of the American Medical Association*, 291(12): 1464-1470.
- BARSKY, A.J. AND KLIERMAN, G.L. (1983). Overview: hypochondriasis, bodily complaints, and somatic styles. *American Journal of Psychiatry*, 140: 273-283.
- BENIGERI, M. AND PLUYE, P. (2003). Shortcomings of health-related information on the internet. *Health Promotion International*, 18(4): 381-387.
- BERLAND, G.K., ELLIOTT, M.N., MORALES, L.S., ALGAZY, J.I., KRAVITZ, R.L., BRODER, M.S., KANOUSE, D.E., MUÑOZ, J.A., PUYOL, J.-A., MARIELENA, L., WATKINS, K.E., YANG, H. AND MCGLYNN, E.A. (2001). Health information on the internet: Accessibility, quality, and readability in spanish and english. *Journal of the American Medical Association*, 285(20): 2612-2621.
- BHAVNANI, S.K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 610-611.
- BHAVNANI, S.K., JACOB, R.T., NARDINE, J. AND PECK, F.A. (2003). Exploring the distribution of online healthcare information. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 816-817.
- BIERMANN, J.S., GOLLADAY, G.J., GREENFIELD, M.L. AND BAKER, L.H. (1999). Evaluation of cancer information on the Internet. *Cancer*, 86(3): 381-390.
- CLINE, R.J. AND HAYNES, K.M. (2001). Consumer health information seeking on the Internet: the state of the art. *Health Education Research*, 16(6): 671-692.
- CZERWINSKI, M., HORVITZ, E. AND WILHITE, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 175-182.
- DOWNEY, D., DUMAIS, S. AND HORVITZ, E. (2007). Models of searching and browsing: Languages, studies and application. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2740-2747.
- EASTIN, M.S. AND GUINSLER, N.M. (2006). Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. *Cyberpsychology and Behavior*, 9(4): 494-498.

- ELSTEIN, A.S., SHULMAN, L.S. AND SPRAFKA, S.A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press.
- EYSENBACH, G. (1998). Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. *British Medical Journal*, 317: 1496-1502.
- EYSENBACH, G. AND KOHLER, C. (2002). How do consumers search for and appraise health information on the world wide Web? Qualitative studies using focus groups, usability test, and in-depth interviews. *British Medical Journal*, 324: 573-577.
- EYSENBACH, G., POWELL, J., KUSS, O. AND SA, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide Web, a systematic review. *Journal of the American Medical Association*, 287(20): 2691-2700.
- FLEET, R.P., DUPUIS, G., MARCHAND, A. BURELLE, D, ARSENAULT, A., BEITMAN, BD (1996). Panic disorder in emergency department chest pain patients: prevalence, comorbidity, suicidal ideation, and physician recognition. *Am J Med.*, 101(4): 371-380.
- HECKERMAN, D.E., HORVITZ, E.J. AND NATHWANI, B.N. (1992). Toward normative expert systems: part I – the pathfinder project. *Methods of Information in Medicine*, 31: 90-105.
- HERSH, W.R. AND HICKAM, D.H. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *Journal of the American Medical Association*, 280, 1347.
- HERSH, W.R., CRABTREE, M.K., HICKMAN, D.H., SACHEREK, L., FRIEDMAN, C.P., TIDMARSH, P., MOSBAEK, C. AND KRAEMER, D. (2002). Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association*, 9, 283-93.
- HU F.B., STAMPFER, M.J., MANSON, J.E., GORDSTEIN, F., COLDITZ, G.A., SPEIZER, F.E., WILLETT, W.C. (2000). Trends in the incidence of coronary heart disease and changes in diet and lifestyle in women. *New England Journal of Medicine*. 343(8): 530-7.
- HUFMANN, J.C. AND POLLACK, M.H. (2003). Predicting panic disorder among patients with chest pain: an analysis of the literature. *Journal of the American Medical Informatics Association*, 9, 283-93.
- HURST, W. *The Heart, Arteries and Veins* (2002). 10th ed. New York, NY: McGraw-Hill
- IBQAL, S.T. AND HORVITZ, E. (2007). Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 677-686.
- KAHNEMAN, D., SLOVIC, P. AND TVERSKY, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- KELLNER, R. (2001). Diagnosis and treatments of hypochondriacal syndromes. *Pyschosomatics*, 33:278-289
- JADAD A.R. AND GAGLIARDI, A. (1998). Rating health information on the Internet: navigating to knowledge or to Babel? *Journal of the American Medical Association*, 279(8): 611-614.
- JOACHIMS, T., GRANKA, L., PAN, B., HEMBROOKE, H. AND GAY, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 154-161.
- LEWIS, T. (2006). Seeking health information on the internet: lifestyle choice or bad attack of cyberchondria? *Media, Culture & Society*, 28(4): 521-539.
- LUO, G., TANG, C., YANG, H. AND WEI, X. (2007). MedSearch: A specialized search engine for medical information. In *Proceedings of 16th Annual World Wide Web Conference*, pp. 1175-1176.
- NUNNALLY, J.C. (1967). *Psychometric Theory*. New York, NY: McGraw-Hill Book Company.
- OVSJANKINA, M. (1928). Die wiederaufnahme unterbrochener handlungen. *Psychologische Forschung*, 11: 302-379.
- PEW INTERNET AND AMERICAN LIFE PROJECT. *Online Health Search 2006*. Accessed September 1, 2007. Available at: http://www.pewinternet.org/PPF/r/190/report_display.asp.
- RADLINSKI, F. AND JOACHIMS, T. (2005). Query chains: Learning to rank from implicit feedback. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 239-248.
- ROSAMOND, W.D., CHAMBLESS, L.E., FOLSOM, A.F., COOPER, L.S., CONWILL, D.E., CLEGG, L., WANG, C., AND HEISS, G. (1998). Trends in the Incidence of Myocardial Infarction and in Mortality Due to Coronary Heart Disease, 1987 to 1994, *New England Journal of Medicine*, 339(13): 861-867.
- SHEN, X., DUMAIS, S. AND HORVITZ, E. (2005). Analysis of topic dynamics in Web search. In *Proceedings of the World Wide Web*, 1102-1103.
- SILLENCE, E. BRIGGS, P., FISHWICK, L. AND HARRIS, P. (2004). Trust and mistrust of online health sites. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computer Systems*, 663-670.
- SPINK, A., YANG, Y., JANSEN, J., NYKANEN, P., LORENCE, D.P., OZMUTLU, S. AND OZMUTLU, H.C. (2004). A study of medical and health queries to Web search engines. *Health Information and Libraries Journal*, 21, 44-51.
- TVERSKY, A. AND KAHNEMAN, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157): 1124-1131.

- WHITE, R.W. AND DRUCKER, S.M. (2007). Investigating behavioral variability in Web search. In *Proceedings of the World Wide Web Conference*, 21-30.
- WILDEMUTH, B.M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, 55(3): 246-258.
- ZENG, Q.T., TSE, T., DIVITA, G., KESELMAN, A., CROWELL, J., BROWNE, A.C., GORYACHEV, S. AND NGO, L. (2007). Term identification methods for consumer health vocabulary development. *Journal of Medical Internet Research*, 9(1): e4.