From Web Search to Healthcare Utilization: Privacy-Sensitive Studies from Mobile Data

Ryen W. White, Ph.D. and Eric Horvitz, M.D., Ph.D.
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
{ryenw, horvitz}@microsoft.com

Keywords: Healthcare utilization, privacy-sensitive analytics, search log analysis, mobile health, predictive modeling

Word count: 5177

ABSTRACT

Objective

We explore relationships between health information seeking activities and engagement with healthcare professionals via a privacy-preserving analysis of geo-tagged data from mobile devices.

Materials and Methods

We analyze logs of mobile interaction data stripped of individually identifiable information and location data. The data analyzed consists of time-stamped search queries and distances to medical care centers. We examine search activity that precedes the observation of salient evidence of healthcare utilization (EHU) (i.e., data suggesting that the searcher is using healthcare resources), in our case taken as queries occurring at or near medical facilities.

Results

We show that the time between symptom searches and observation of salient evidence of seeking healthcare utilization depends on the acuity of symptoms. We construct statistical models that make predictions of forthcoming EHU based on observations about the current search session, prior medical search activities, and prior EHU. The predictive accuracy of the models varies (65-90%) depending on the features used and the timeframe of the analysis, which we explore via a sensitivity analysis.

Discussion

We provide a privacy-preserving methodology that can be used to generate insights about the pursuit of health information and healthcare. The findings demonstrate how large-scale studies of mobile devices can provide insights on how concerns about symptomatology lead to the pursuit of professional care.

Conclusion

We present new methods for the analysis of mobile logs and describe a study that provides evidence about how people transition from mobile searches on symptoms and diseases to the pursuit of healthcare in the world.

BODY

Background and Significance

The Web is a first stop for many when concerning symptoms come to the fore [1]. Indeed, a recent poll found that 80% of Web users have looked online for medical information [2]. A recent large survey on healthcare search behavior found that information obtained from medical searches can influence peoples' concerns, their decisions about when to engage a physician for assistance with diagnosis or therapy, and their overall approach to maintaining their health or the health of a family member [3]. Approximately one quarter of survey respondents reported that they were "put over the threshold to engage with a medical professional based on Web content." Respondents reported that, in most cases (72%), their encounter with a physician eased their concerns. Such findings provide evidence that interaction with information on the Web can affect the anxiety of consumers and stimulate engagements with healthcare professionals. Accurate prediction of when users will seek in-world medical resources based on the review of health information on the Web could allow for accurate pre-visit intervention to suggest alternative courses of action, recommend appropriate medical content to prepare the patient to have a constructive dialog with medical professionals, or even provide route guidance for the patient or early alerting for the medical facility if the destination can be inferred [4].

Prior studies have shown correlations between Web usage and decisions to seek medical care [5,6,7,8], and how the Web can help people monitor symptoms and adapt their lines of medical inquiry [9]. Researchers have investigated the search behavior of medical domain experts [10,11,12,13] to better understand the search behavior of those with specialist domain knowledge and how their behavior differs from domain novices. Log data gathered by search engines offers an opportunity to study Web search and browsing behaviors at scale in a naturalistic setting. These data can be used, among other things, to better understand how people search the Web [14], predict their future interests [15], and improve the performance of search engines [16]. However, there has been little work to date on using logs to better understand how people seek health-related information and inform the development of tools to help people find and understand medical information online. In previous work, we performed analysis of signs of medical anxiety revealed in search logs [1], and studied how exploring content online can lead to an escalation of the medical severity of Web searches, including the generation of queries showing *healthcare utilization intent* (HUI) [17]—the pursuit of in-world professional care (e.g., queries such as [dermatologist 98033] and [neurologist decatur, il]).

Analyses of escalations can provide clues on likely anxieties and transitions to resource usage, but they do not let us link Web searching directly to healthcare utilization. Location information useful for personalizing search results has been extracted from terms in the query text [18] and obtained via consideration of Web pages typically visited by people at a location [19]. Such methods typically result in coarse representations of location, usually at the city level. Statistics about healthcare utilization are gathered by individual hospitals as well as government agencies such as the Centers for Disease Control and Prevention, and include patient counts and revisitation rates. Reliable statistics can be gathered once patients reach the medical facility, but do not include behaviors before or after the visit. Important clues regarding medical intent, including symptoms, conditions, and diagnosis, may be visible in search log data [20]. We investigate geocoded mobile log data about searches performed by consenting users on mobile devices. Since location in mobile logs is based on Global Positioning System (GPS) coordinates, these logs allow for a finer-grained analysis of medical search, including the provision of evidence of healthcare utilization.

Objective

The objective of this study is to characterize and predict aspects of in-world healthcare utilization from evidence derived from logs of users of mobile devices who have consented to sharing their data. We seek to infer user information goals before, during, and after utilization. We are motivated by the opportunity to better understand the interaction of healthcare-related Web searches with in-world utilization, and for predicting utilization given features extracted from concerned searchers' query histories. We believe that applications might one day be developed for mobile devices that help users make more informed triaging decisions regarding when and how to utilize healthcare resources or provide them with relevant information that might be helpful in dialog with a medical professional.

Materials and Methods

We used datasets collected with consent from Bing for Mobile applications for Android and iPhone. Users agreed to share their search activity and location information. The published Privacy Statement for Bing for Mobile applications addresses the storage, use, sharing, and retention of data collected in the course of the operation of the service. In particular, the statement indicates that Microsoft may employ this data for analyzing trends and for operating and improving its products and services, as we aim to do with this work. The original Privacy Statements do not state that the data will be shared publically for other purposes. Thus, these datasets are not available publically for further research. This research and its findings are not part of Bing for Mobile applications, and the analyses that we describe are a research effort only. Our goal was not to identify or study specific individual activities, but rather to understand the patterns of the aggregated activities and to explore their implications. The datasets were stripped of individually identifiable information. In particular, location and user identification were removed early in the analytical pipeline. Queries were analyzed using automated key word and phrase spotting. The study was reviewed and approved by our organization's privacy policy officer.[1]

Evidence of Healthcare Utilization

On methods, we started with a log of several hundred thousand search entries stored on a secure server. The logs were gathered over a one-year period from 34,540 consenting users. Each entry contains a unique numerical identifier and GPS location where a query was issued. Location information was recorded only if and when a search query was issued. As we describe below in more detail, we removed absolute locations of users and of care centers in an automated manner, and only considered information about distances to care centers.

Location information was not used to track users' locations over time. Instead we only considered whether searchers queried Bing from within 200 meters of a medical facility (e.g., hospital, medical

[1] Our organization does not have an institutional review board of the type employed at universities.

center) at some point within the period of time covered by the log. We considered a non-periodic approach to a medical facility as being evidence of pursuit of healthcare utilization (EHU). This provided an automated, scalable mechanism to identify medical visits retrospectively from log data. Given the nature of the logs, we were unable to distinguish between users visiting the medical facility for treatment, to visit an admitted patient, or simply traveling near the facility but not using its services. Immediately after obtaining the logs, all GPS coordinates were stripped from them. The only information retained was a bit representing whether a query was issued proximal to an anonymized medical facility, and relative distances between the user and a medical facility for each Web query leading up to the queries made near a medical facility.

We obtained the GPS coordinates of medical facilities across the United States through a crawl of the Bing Web search engine's business listings, which includes data from a variety of sources, including the Yellow Pages (yp.com), and experimented with variants of searcher proximity to the coordinates of the facilities. The constructed database contained 34,750 sites. Site types included emergency medical and surgical services (54.7%), hospitals (34.1%), medical centers (10.5%), and outpatient services (0.6%). As part of formulating a definition for evidence of seeking healthcare engagement, we experimented with a number of different radii from the location of the hospital (from 20m to 300m) to the locations of devices hosting queries. Of options tested, we found that a radius of 200 meters from sites was sufficient to encompass most medical locations without extending too far into the surrounding area (verified for a sample of hospitals by visual inspection using mapping software), and provided a good tradeoff between precision and recall.

All users had search activity on at least 90 different days. 5239 of those users (14.8%) queried from within 200 meters of the medical facility at least once in those 90 days.

The default assumption is that each EHU maps to a visit to a healthcare provider and that they reflect situations where people search about their own symptoms (or those of a family member) using their own device and then later search from that same device from the medical facility. However, we note that we are looking at the world through a keyhole; the log data is a noisy and sparse sensor and we cannot resolve uncertainty about whether the EHUs as defined actually represent a visit. Potential errors in the analyses include:

- **Type I errors (false positives):** Examples include: (i) people who drive past or work near or live near one of the medical facilities, and that happen to search via their mobile device; and (ii) people who happen to be passing by one of the facilities over the course of the logs, and search via their mobile device while passing.

- **Type II errors (false negatives):** Examples include: (i) people who go to medical facilities for a real problem or concern but whom do not search in advance with their mobile devices-for one or more reasons (e.g., using desktop); and (ii) people who search in advance and who also go to medical facilities for a real problem but who do not search near medical facilities.

Note that there are also people who do not search and who do not go to a hospital (i.e., true negatives). These users will not appear in our log data and therefore we cannot observe or validate them.

We performed explicit filtering to reduce the likelihood of Type I and Type II errors in our data. We required that users had previously searched for at least one of a set of tracked symptoms or synonyms from the Merck medical dictionary[2] in the 90 days before the EHU. In total, 4006 of the 5239 users (76.5%) met this requirement. To improve the likelihood that observed EHU events were related to treatment and consultation, we removed the 885 users (22.1%) who had queried from within 200 meters of the facility in the time period before the first symptom query. These users may live nearby the facility, may work at or near the facility or be receiving long-term care at the facility, and would hence not be relevant to our focus on transitions from Web search to healthcare utilization. To improve the reliability of this analysis of proximity patterns, we required that users had at least seven days of logs before the first symptom query. This resulted in a set of 3121 users, who generated 5642 EHUs, who we study in the remainder of this paper.

Analysis

The first phase of analysis involved characterizing the transition between Web search and EHU. This was performed using the 5642 EHUs identified as described in the previous section.

We also pursued the construction of a predictive model to forecast forthcoming visits to medical facilities given features extracted from a user's search history and/or prior observations of evidence of healthcare utilization. We sought to predict the following: Given a randomly-selected non-terminal, non-initial query in a search session (defined as a sequence of queries in [14]), predict if one of the queries in the session issued after the selected query will be issued proximal to a medical facility. We constructed a logistic regression-based classifier that was trained using a 50/50 split of positive and negative examples, with only one example selected per user.

Positive examples for the prediction task were drawn from the data described in the previous section, and comprised the set of 3121 EHU events (randomly selected, one per user), which had a query near a medical facility. We also randomly selected 3121 sessions as negative examples for training purposes (again, one per user). For these "negative" users, we required that there was at least 90 days of log data that a symptom search occurred within the 90 days of the termination of the log data, that they had no EHU events before the first symptom query, and that they had an overall distribution of initial Merck symptoms that matched the "positive" users. Negative sessions were randomly selected from the time period after the first symptom search and before the end of the log. After the initial symptom search, users for whom we selected negative examples did not visit the medical facility in the randomly-chosen session or at any time after the initial symptom search (i.e., their log data terminated without an EHU event).

To improve query coverage, synonyms of the Merck symptoms were also used. Synonyms for each symptom were identified through a two-step walk on the search engine result-click graph (constructed

[2] http://www.merckbooks.com/symptoms

from one year of Bing search engine query and click logs) using an approach similar to prior research in the data mining community [21]. The number of queries we observed for a particular user was likely to influence the likelihood that we would observe a query proximal to a medical care facility. To normalize this effect, we balanced each user, for whom we observed EHU, with a corresponding user with the same number of queries who had searched for the same symptom but did not later show EHU.

The features used in the prediction are described in Table 1. Three classes of features were used: (i) whether the user issued a search demonstrating healthcare utilization intent, (ii) the extent to which users queried for medical symptoms or benign or serious medical conditions, and (iii) whether we identified EHU earlier in the log. Since longitudinal EHU data are potentially most sensitive, involving multiple user visits over time, access to that data is not assured and we break out our later analysis by whether or not we have access to such data. Also we note that the healthcare utilization features are not computed for the current session and only computed for search behavior beyond 12 hours before the current session for the other *timeframes* (Day, Week, etc.) from which features were generated. We do this to help prevent feature contamination from current medical facility visits, which may still be evident in features of the current session.

Table 1: Features used in predicting healthcare utilization from Web search behavior.

Feature Classes and Names	**Description**
Healthcare utilization intent (HUI)	
Query with HUI	*Whether at least one healthcare utilization intent in query history*
Number of HUIs	*Number of queries in query history showing healthcare utilization intent*
Time Between HUIs	*The average time between HUIs*
Has HUI Refinement	*Whether post HUI refinement (e.g., [doctor] to [doctor seattle wa])*
Medical search	
Num Symptom Searches	*Number of searches for one of the Merck symptoms*
Num Serious Condition Searches	*Number of searches for serious conditions*
Num Benign Condition Searches	*Number of searches for benign conditions*
Num Unique Symptoms	*Number of unique symptom searches*
Num Unique Serious Conditions	*Number of unique serious conditions searched*
Num Unique Benign Conditions	*Number of unique benign conditions searched*
Prior evidence of healthcare utilization (EHU)	
Has Previous Hospital Visit	*Whether user has been near a hospital previously (beyond 12 hours)*
Num Hospital Visits	*Number of approaches to hospital (if any) (beyond 12 hours)*
Time Since Last Visit	*Time elapsed since last approach to hospital (if any) (beyond 12 hours)*

The features were derived based on word lists developed in previous work [1,17]. Benign and serious conditions were identified using a list of commonly-occurring conditions from the *International Classification of Diseases 10th Edition* (ICD-10) published by the World Health Organization and used in previous log analysis [1]. HUIs were defined using the methodology described in [17], which required the queries to contain evidence of intent to seek professional medical care (e.g., [physicians 98072]).

Results

For the remainder of the article, we focus on our users' search behavior for at least 90 days before the EHU, extending our previous work on HUIs [18] to target EHU directly. Parametric statistical testing is used with significance (α) set to 0.05 unless otherwise stated.

Characterizing Transitions from Web Search to Evidence of Healthcare Utilization

We first examined the rate at which search queries with utilization intent transitioned to EHUs. Given our logs, we measured the time between the queries and any observed EHU that followed. For this analysis, we removed from consideration queries occurring more than 90 days before the EHU since we believed that they may be unrelated to the visit (although they could be connected to scheduling of appointments or long-term treatment planning). We found that 12.8% of the HUIs observed in our logs transitioned to EHUs within 30 days, a five-fold lift over the base rate of 2.3% for transitions between a sampling of any query to EHUs within the same timeframe. In addition, 3.6% of EHUs were preceded at some point by a query showing health utilization intent. The histogram of the times between the EHU and the most proximal query with utilization intent, grouped by timeframe, are summarized in Figure 1.

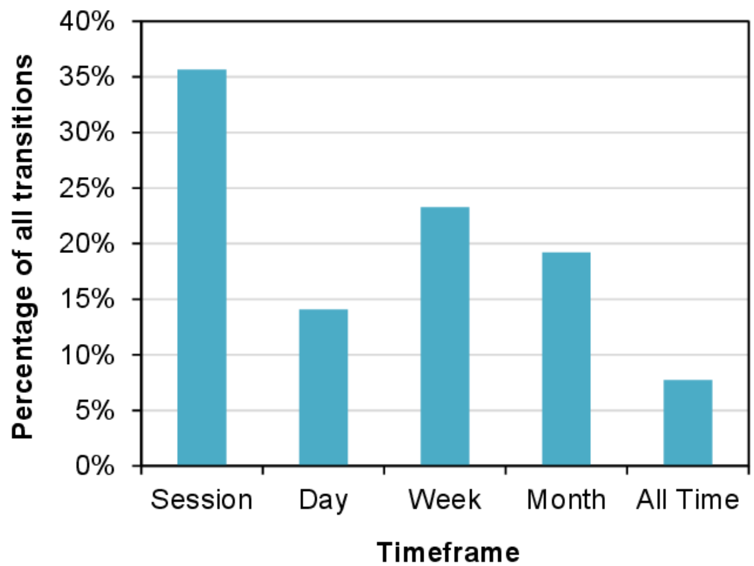

Figure 1: Time from most recent query demonstrating intention of healthcare utilization to observed evidence of care utilization

Figure 1 shows that most of the transitions between queries with utilization intention to EHU occur within the same search session, perhaps associated with people searching for the medical location prior to traveling there or when they were en route to the medical facility. Note that it was less common to observe these transitions within the same day, suggesting that searchers either transition more or less immediately, or wait multiple days (perhaps associated with appointment scheduling). There was little difference in the distributions for the four types of medical facility used in our analysis.

We explored the relationship between the types of medical symptoms searched and the time taken to visit a medical facility. In Figure 2, we display the time between an EHU and the most recent symptom searched by users for transitions that occur within 36 hours of the symptom search.

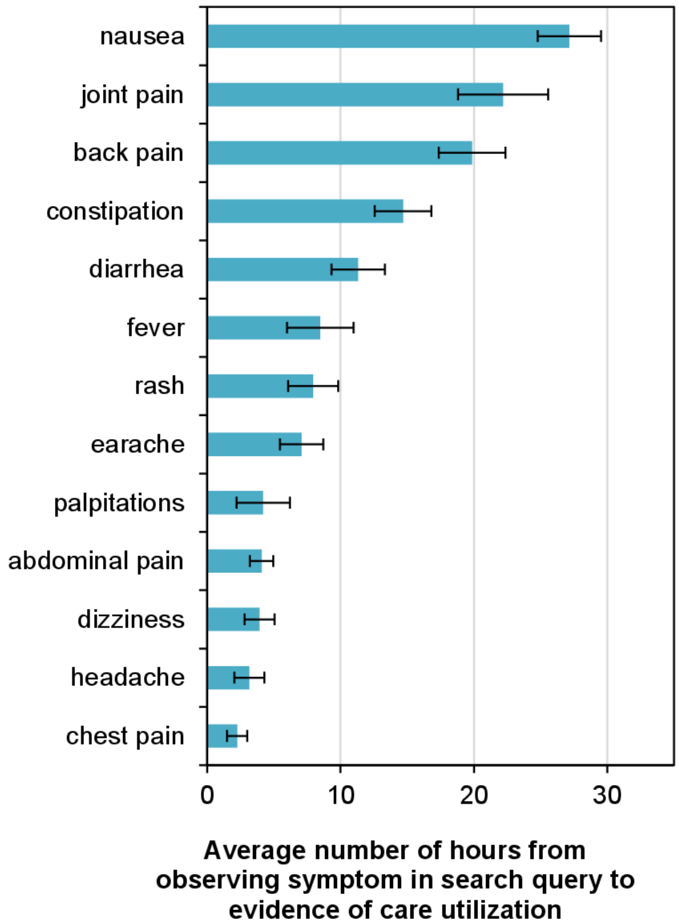

Figure 2: Time between most recent symptom query and observation of evidence of healthcare utilization within 36 hour horizon. Error bars denote standard error of the mean (± SEM).

Figure 2 shows a marked difference in the transition time depending on the symptom used. The conversion time is significantly lower for symptoms which may be more worrying to users, such as chest pain (concern about myocardial infarction) and headache (concern about a hemorrhage or tumor). In contrast, for symptoms that may not be viewed as severe by users, such as joint pain and nausea, the time between the occurrence of those symptoms and the EHU is significantly larger.

Predicting Healthcare Utilization

We explored the use of predictive to models to forecast whether a user would approach a medical facility in the current search session given features extracted from their search history. We considered three groups of features: on healthcare utilization intent shown in queries, medical search, and prior EHU. We introduced prior EHUs as health statistics show that revisitation to medical facilities is common; a recent study of Medicare claims data placed readmission rates with 30 days at 19.6% [22]. In our logs we observed that 21.5% of our users with an EHU showed another EHU within 30 days, strikingly similar to the published findings. We also measured prediction accuracy based on interactions within the same session, day, week, month, and over all time (up to 90 days before the current session). We performed five-fold cross validation over 100 randomized experimental runs.

Predicting Using Search Behavior Only

We first focus only on the use of information about medical keywords and healthcare utilization intent (HUI) seen in search. These features do not require logging and remembering user locations at multiple points in time (as with prior EHUs), which may not be generally acceptable in a real-world version of a predictive system. Mean average accuracy numbers for the prediction task, and their associated standard deviations, are shown in "HUIs + Medical Search" column of Table 2 (in dark gray) for each of the five timeframes: *Session*, *Day*, *Week*, *Month*, and *All Time*.

Table 2: Mean accuracy and standard deviations (parenthesized), for variations in prediction timeframe and feature classes used. Bold=significantly different from "All Features" (α=.0125). N=100.

	Feature Sets						
Timeframe ▼	**All Features**	**HUIs + Medical Search**	**Medical Search + Prior EHU**	**HUIs + Prior EHU**	**Prior EHU**	**Medical Search**	**HUIs**
Session	n/a	0.70 (0.21)	n/a	n/a	n/a	0.64 (0.29)	0.71 (0.29)
Day	0.89 (0.31)	0.89 (0.31)	0.89 (0.32)	**0.63** (0.26)	**0.65** (0.29)	0.86 (0.30)	**0.63** (0.30)
Week	0.76 (0.29)	0.75 (0.29)	0.81 (0.30)	**0.66** (0.31)	**0.62** (0.31)	0.77 (0.30)	**0.66** (0.30)
Month	0.78 (0.29)	0.72 (0.31)	0.74 (0.33)	**0.71** (0.33)	**0.68** (0.32)	**0.69** (0.31)	**0.69** (0.32)
All Time	0.89 (0.29)	**0.69** (0.31)	0.86 (0.32)	**0.75** (0.33)	**0.74** (0.33)	**0.71** (0.31)	**0.69** (0.31)

The dark gray column in the table shows that using features from the 24-hour period preceding the current query (*Day*) leads to the most accurate prediction, whereas using features gathered across the full 90 days of search history (*All Time*) is least accurate. A one-way analysis of variance (ANOVA) revealed that *Day* performed significantly better than the other timeframes and *All Time* performed worse than others apart from *Session* ($F(4,495)=4.33$, $p=.002$; Tukey tests: *Day* vs. Other, $p<.001$; *All Time* vs. Other not *Session*, $p<.01$). Search behavior may be noisy and less directly related to EHU over long time periods. Predictions from within the same session are also less reliable than most other timeframes (other than *All Time*), in part because there may be very few previous actions in the session from which to generate features.

We visualize the performance of our classifiers graphically with a receiver-operator characteristic (ROC) curve for each timeframe across all features, and averaged across all experimental runs. This is shown in Figure 3. The ROC curves demonstrate strong performance of the predictors based on *Day* features across the range of the discrimination threshold.

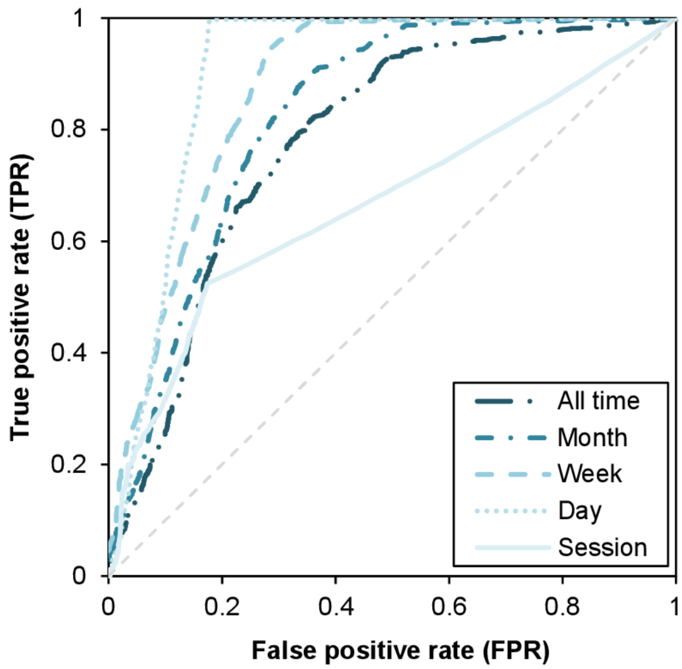

Figure 3: ROC curve for classifier not using prior EHU information.

These results are based on all users with available search behavior. However, the observed differences between timeframes may therefore be caused by differences in the user distribution. To ameliorate the possible effects of this, we restricted our analysis to a subset of users for whom we had examples of queries at all five timeframes. A two-way ANOVA revealed significant differences in accuracy between timeframes ($\underline{F}(4,990)=3.63$, $\underline{p}=.006$) but not between user groups ($\underline{F}(1,990)=.09$, $\underline{p}=.77$). Since there was no effect from group, we used all users with available data at each timeframe for the remaining analysis.

Influence of Prior Healthcare Utilization Estimates

In addition to using features of search behavior, we experimented with using information on prior EHU. These EHU refer to noting geographic proximity to medical facilities in sessions within 90 days prior to the current session. Note that although we are filtering users who are proximal to the medical facility before the first symptom query, multiple EHUs (and hence *prior* EHUs) are still possible after the first symptom query. The ROC curve for those experiments is in Figure 4 and the accuracy numbers are reported in the "All Features" column in Table 2. Note that we do not include *Session* in this analysis, since EHUs within the same session may be attributable to the same visit, making the predictions meaningless.

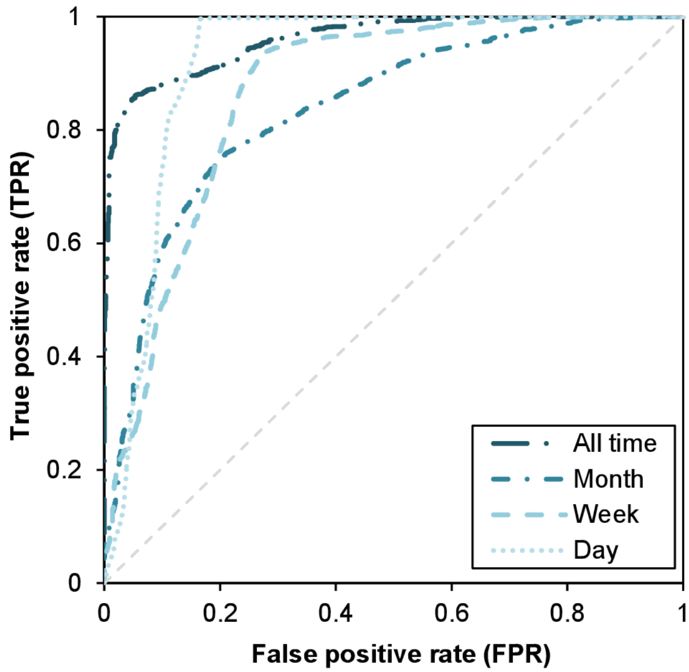

Figure 4: ROC curve with use of information on prior EHU.

In comparing Figures 3 and 4, we see that in most cases, the curves are pushed up and to the left, suggesting that prior EHU helps, especially for *All Time*. Although there are only slight gain in maximum predictive accuracy (still 90%), there are improvements in the true positive rate (+5%) from using prior EHU features. This may be expected given that people often need to return to medical facilities over time for follow-up treatment, so a prior visit may suggest that the user is seeking treatment and increase their likelihood of re-visiting.

Analyzing Feature Class Contributions

Table 2 illustrates the effect on accuracy of performing feature ablation, where we examine all combinations of our feature classes.

Within each of the timeframes, we performed one-way ANOVAs with post-hoc testing to compare the feature values with each of the feature subsets with "All Features." A Bonferroni correction was applied to control for false positives and set α to .0125. The findings revealed significant differences within each feature timeframes (all $F(6,694) \geq 3.83$, all $p \leq .001$) and Tukey post-hoc testing identified cases that were significantly different from "All Features" (all $p < .01$). These cases are bolded in Table 2.

We found the medical search features to be more predictive than the health utilization intent features, suggesting that there may be underlying medical search behaviors that imply impending healthcare visits that are not associated with intermediate queries on accessing professional care. As only 12.8% of queries with evidence of healthcare utilization intention transition to EHU, further exploration of the search behavior in the other sessions is needed to determine whether there are indicators of EHU in search behavior.

As shown, the EHU features bring benefit, but only in the longer term (*Month* and beyond – see "All Features" vs. "HUI + Medical Search"), perhaps since they are related to repeat visits to medical facilities, which are unlikely to happen less than monthly unless there are serious ailments, which may already be captured by search activity (HUI and Medical Search feature classes). The findings suggest that even without prior EHU, we are still able to predict future EHU with good accuracy. The best performance *without prior EHU* comes from using features from within the last 24 hours (*Day*). There may be aspects of the search behavior that are evident inside a day (e.g., concerns about serious conditions) that might be directly predictive of healthcare utilization that perhaps would not be visible inside a session (where the focus might be on identifying professional healthcare resources) or in the week or beyond (where the relationship between search and visitation might be less direct).

Analyzing Individual Feature Contributions

To understand the role that individual features played in the predictions, we computed their evidential weights. In Table 3, we list the features that have the highest evidential weight with and without access to historic healthcare utilization information for the best-performing predictive time frame (*Day* for non-EHU, *All Time* for EHU), as well as the feature class to which each feature belongs.

Table 3: Top five features from models *with and without prior EHU* ordered by absolute evidential weight relative to the most predictive feature.

Use Prior EHU features?	Feature name	Feature Class	Weight
Yes	*Has Previous Hospital Visit*	Prior EHU	1.00
	Has HUI Refinement	HUI	0.60
	Num Unique Symptoms	Medical Search	0.13
	Number of HUIs	HUI	0.10
	Num Serious Condition Searches	Medical Search	0.09
No	*Num Unique Serious Conditions*	Medical Search	1.00
	Number of HUIs	HUI	0.46
	Num Symptom Searches	Medical Search	0.41
	Has HUI	HUI	0.37
	Has HUI Refinement	HUI	0.33

Table 3 shows that when prior EHU features are used, they dominate the predictive models. However, when EHU features are unavailable, searches related to the number of *distinct* serious medical conditions searched for *Num Unique Serious Conditions*[3] and the frequency of searches with healthcare utilization intent are most predictive of a forthcoming approach to a location where professional care is available.

[3] Note that this is a different feature from *Num Serious Condition Searches* (ranked fifth most important when *Prior EHU* is used), which is the raw count of the number of queries containing serious conditions.

Sensitivity Analysis

In the analysis we have described, identify EHU we set the user's required distance, *d*, from the GPS coordinate of the hospital to 200 meters. We set this parameters with intuitions accrued during a manual exploration of the data. However, to pursue an understanding of the sensitivity of the results to this parameter setting, we performed a sensitivity analysis. In the analysis, we re-ran the experiments in an end-to-end manner for All Features and *All Time* (as they were the best performing) with a parameter sweep across *d* (range: [20m,300m], increment: 10m). Figure 5 presents the findings. Error bars denote standard error of the mean.

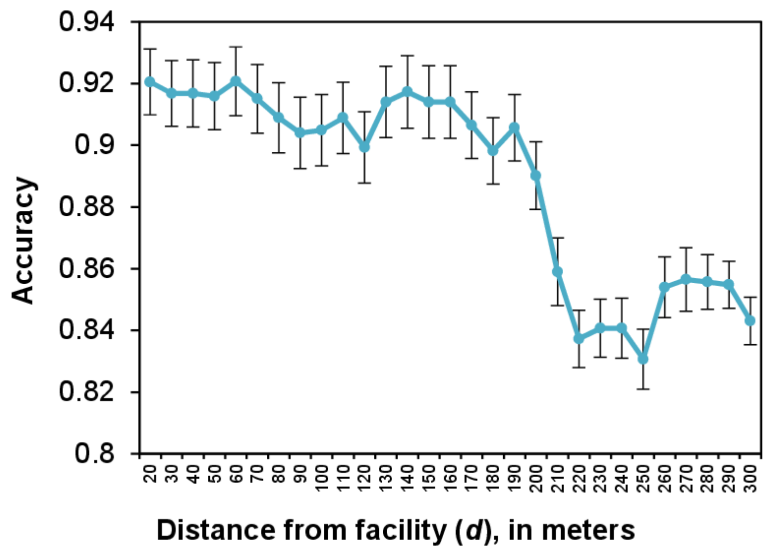

Figure 5: Sensitivity analysis for variations in *d*, the distance from the facility required to be noted as EHU.

Figure 5 suggests that accuracy is higher when we define EHU as closer to the medical facility. There may be less noise from focusing on smaller distances. However, requiring that EHUs be too close to the facility limits the coverage of our method and the number of examples available for analysis (e.g., there were only 211 EHU events at 20m vs. over five thousand for the 200m value we selected in this study).

Discussion

Our findings highlight the opportunities for using sparse logs of behavioral data to learn about interactions between the pursuit of online healthcare information and the seeking of professional care. We computed mean times between the onset of searches on symptoms and logged evidence of being near a healthcare provider. We also demonstrated the ability to construct classifiers to predict that a user doing searches on symptoms would later enter queries at or near the location of a healthcare provider, considering evidence about search activity and prior proximity to care centers. The strong performance of our predictive models may be related in part to the source of the data: those searching for medical information on mobile devices might be more willing and able to visit a medical facility than those operating a desktop computer. Indeed, 39.5% of EHUs were preceded by a search for a serious medical condition on a mobile device in the last 24 hours.

Limitations

We leveraged log data stripped of individually identifiable information and location to study aspects of transitions from web search to in-world activities. We cannot confirm the intention behind queries or whether searches at locations within 200 meters of the coordinates of medical facilities mean that users are visiting the facility in pursuit of care. That said, we are reassured by the predictive performance of medically oriented search features, which provides evidence that the visits were labeled correctly. As location data is only provided when searches are executed, identifying EHU depends on users performing a search in the vicinity of the medical location. The likelihood of search queries being associated with a healthcare utilization event may depend on multiple factors, including the frequency with which someone relies on search in daily life, the background health status of the user, the perceived severity of a condition, and type of treatment that is required. More work is needed to understand the impact of these factors, and on how these considerations influence the evidence and meaning of logged events. The absence of location data when there are no queries could be addressed in future studies by gaining access to streaming location data with consent from users and study review boards.

Implications

This research has implications for the design of search technology to predict when searchers transition toward engagement with healthcare professionals, some of which may be potentially costly and unnecessary. Systems lack sufficient information to advise such users against seeking medical advice, but they can perform a range of functions on people's behalf, such as proactively retrieving Web pages about their symptoms or illness that might help them in discussions with medical professionals.[4] Information could also be provided on the prior or conditional probabilities of causes of symptoms along with key information helpful for understanding ideal next steps. We could also study search behaviors once users are at care centers to look for signals in how they react to prognosis (e.g., via queries such as ["what is dialysis"]) and help them comprehend medical information in situ as they have discussions with medical professionals.

Conclusion

We presented a population study linking Web search activity to actions in the world, using logs of mobile search activity stripped of user identity and location information, and abstracted to tokens extracted from queries about symptoms and illnesses and the pursuit of professional assistance. We found that, for approximately one out of eight searches, users entering queries that demonstrate health utilization intent will later approach a medical facility. We found that the time between a symptom search and visit to a care facility was generally influenced by the acuteness of symptoms contained in

[4] Previous work has shown that 65% of those who elected to bring Web content to consultations found that it was helpful in actively participating in the conversation with the health professional, and that many more physicians appeared pleased that their patient had performed Web research than were discontent or irritated [17].

queries. In addition to characterizing aspects of the transition from search activity to evidence of in-world healthcare utilization, we examined the prediction of forthcoming utilization based on information in user's search history for different feature classes and at varying levels of temporal granularity.

We showed that we can predict seeing an approach to a medical care facility using features of search behavior and prior history of such evidence of healthcare utilization. Our findings have implications for better understanding the processes by which people transition from Web search to engaging with medical resources. Deepening understanding of this transition promises to help with the design of applications and services that enhance access to health information. As examples, a service might point people to the nearest medical facility or display the best route, provide information on the likelihoods of conditions and related triage information, provide direct contact with healthcare providers given inferences of raised likelihood of the need for urgent attention, or provide questions they might pose to their physician. Healthcare providers may also use aggregate anonymized mobile data (perhaps collected through a smartphone application) to track trends in healthcare utilization (e.g., changes in readmission rates over time) that may take time to gather using other methods such as analysis of admission records and claims data. Finally, we hope to stimulate interest and discussions about the possibilities for doing new kinds of privacy-sensitive studies from mobile data and for drawing inferences about people from incomplete and sparse data. We highlighted the feasibility of studying an aspect of human behavior on a population level from mobile data, with a methodology that removes potentially sensitive data about identity, location, and query content. In particular, we showed how we can work to gain insights by abstracting location information to relative distances and narrowing the focus of attention to cases of interest.

REFERENCES

1. White, R.W. and Horvitz, E. (2009). Cyberchondria: Studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems*, 27(4): 23.
2. Fox, S. (2011). Health topics: 80% of internet users look for health information online. *Pew Internet and American Life Project*.
3. White, R.W. and Horvitz, E. (2010). Web to World: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. *Proc. AMIA*, 882–886.
4. Krumm, J. and Horvitz, E. (2006). Predestination: Inferring destinations from partial trajectories. *Proc. UbiComp*, 243–260.
5. 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.
6. Kassirer, J.P. (2000). Patients, physicians, and the internet. *Health Affairs*, 19(6): 115–123.
7. Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., Lee, K., Zapert, K. and Turner, R. (2003). The impact of health information on the internet on health care and the physician-patient relationship. *Journal of Medical Internet Research*, 5(3).
8. Nicholson, W., Gardner, B., Grason, H.A., and Powe, N.R. (2005). The association between women's health information use and health care visits. *Women's Health Issues*, 15: 240–248.

9. Ayers, S. and Kronenfeld. J. (2007). Chronic illness and health-seeking information on the internet. *Health*, 11(3): 327–347.
10. Bhavnani, S.K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *Proc. ACM SIGCHI*, 610–611.
11. Hersh, W.R. and Hickam, D.H. (1998). How well do physicians Use electronic information retrieval systems? *Journal of the American Medical Association*, 280: 1347.
12. Hersh, W.R., Crabtree, M.K., Hickam, D.H., Sacherek, L., Friedman, C.P., Tidmarsh, P., and 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: 283–293.
13. Wildemuth, B.M. (2004). The effects of domain knowledge on search tactic information. *Journal of the American Society for Information Science and Technology*, 55(3): 246–258.
14. White, R.W. and Drucker, S.M. (2007). Investigating behavioral variability in web search. *Proc. WWW*, 21–30.
15. Dupret, G. and Piwowarski, B. (2008). A user browsing model to predict search engine click data from past observations. *Proc. ACM SIGIR*, 331–338.
16. Joachims, T. (2002). Optimizing search engines using clickthrough data. *Proc. ACM KDD*, 133–142.
17. White, R.W. and Horvitz, E. (2009). Experiences with web search on medical concerns and self-diagnosis. *Proc. AMIA*, 696–700.
18. Jones, R., Hassan, A., and Diaz, F. (2008). Geographic features in Web search retrieval. *ACM GIR Workshop on Geographic Information Retrieval*, 57–58.
19. Bennett, P.N., Radlinski, F., White, R.W., and Yilmaz, E. (2011). Inferring and using location metadata to personalize web search. *Proc. ACM SIGIR*, 135–144.
20. Cartright, M., White, R.W., and Horvitz, E. (2011). Intentions and attention in exploratory health search. *Proc. ACM SIGIR*, 65–74.
21. Beeferman, D. and Berger, A. (2000). Agglomerative clustering of a search engine query log. *Proc. ACM KDD*, 407–416.
22. Jencks, S.F., Williams, M.V., and Coleman, E.A. (2009). Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360: 1418–1428.

Competing interests:

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: all authors had financial support (salary) from Microsoft Corporation for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Funding:

No direct funding for this study was provided. The authors' salary was paid by Microsoft Corporation.

Contributorship statement:

RW planned the study, mined and analyzed the data, developed and evaluated the predictive models, and drafted and revised the paper. EH planned the study, advised on analysis and modeling strategies, and drafted and revised the paper.