Screening for Cancer Using a Learning Internet Advertising System

ELAD YOM-TOV, Microsoft Research

Studies have shown that search engine queries are indicative of future diagnosis of several types of cancer. These studies were based on self-identification of illness and were limited in that diagnostic information could not be shared with screened individuals. Here I report on two studies that overcome these limitations.

Advertisements were displayed on the Bing and Google ads systems to people who sought to self-diagnose one of three types of cancer. People who clicked on these ads were provided with clinically verified questionnaires and the outcomes of these questionnaires.

A classifier trained to predict suspected cancer, inferred from questionnaire responses, from past Bing queries reached an area under the curve of 0.64. People who received information that their symptoms were consistent with suspected cancer increased searches for healthcare utilization.

In a second study, questionnaire responses provided to the conversion optimization mechanism of the Google advertisement system enabled it to learn to identify people who were likely to have suspected cancer. Following a training period of approximately 10 days, 11% of people selected for showing of targeted campaign ads were found to have suspected cancer.

These results demonstrate the utility of using modern advertising systems to identify people who are likely suffering from serious medical conditions.

1 INTRODUCTION

Data generated while people browse the Internet, especially when using Internet search engines, has been shown to reflect the experiences of people in the physical world [30]. Indeed, most Internet users refer to search engines when they have a medical concern [6]. For this reason, search queries have been used to address a range of medical questions, including tracking infectious diseases such as influenza [12, 17] and answering questions on the relationship between diet and chronic pain [7], and to identify precursors to disease [32].

Recently, search engine queries were shown to be a useful signal for evaluating whether people are likely to be suffering from different cancer types, including ovarian [24], cervical [24], pancreatic [15], and lung [29]. The underlying assumption in all of these studies is that these cancers manifest themselves in externally recognizable

symptoms that are either unfamiliar or relatively benign, meaning that people do not immediately turn to professional medical consultation, even though early detection is known to be an important predictor of survival [26].

A basic shortcoming in studying the connection between web searches and disease onset is that web searches are almost always anonymous when accessible to researchers. This means that linking actual medical information, such as exact disease diagnosis and its date, is limited to indirect inference. An exception to this rule is studies where patients consent to contribute their medical records and their online activity for study [14]. However, such studies are necessarily limited by their scale and the diversity of the cohort they recruit. Thus, as noted earlier, most studies collect data from anonymous sources and need to infer the medical status of users. This can be done, for example, using search queries of self-identified users (SIUs).

SIUs are people who, in their queries, identify themselves as having a condition of interest through experiential queries, such as "I have breast cancer" (see [15, 29]). Unfortunately, people who self-identify are few and are drawn from an unrepresentative population [24, 31], consisting of more females and younger people than in the general Internet user and patient population. Others [24, 32] used both SIUs and information on geographic variability in disease incidence to infer which users are suffering from specific medical conditions, based on the queries they made. These methods provided a larger (and more diverse) cohort than is possible with SIUs but are still lacking in that no clinical information about the users is known. Thus, it is impossible to have a definite clinical indicator of disease for those users.

The most widely adopted approach to obtaining clinical indicators of disease is through clinical questionnaires. For example, de Choudhury et al. [2] used questionnaires to assess the level of depression of crowdsourced workers as a basis for using their social media posts to distinguish depressed from non-depressed individuals. Here, this approach is adopted to focus on cancer and correlate the score of a clinically validated questionnaire to medical symptom searches. In contrast to depression, the incidence of cancer is significantly lower than that of depression (creating a recruitment challenge). Moreover, people are more likely to ask for medical symptoms, especially those of a personal nature, on search engines rather than on social media [16]. Therefore, here I used a targeted advertising campaign and search engine logs as the primary data source.

Operating in conjunction with all major search engines is a highly developed advertising system. Advertising systems show ads to people when they use a search engine to query for terms defined by the advertisers. These ads display a short text (and sometimes an image) and provide a link to the advertiser's website. Advertisers commonly pay whenever a user clicks on the ad, and therefore the ads platform is optimized to show a certain ad only when it is likely to be clicked. More recently, advertising systems have begun allowing advertisers to signal the system when a user purchases the advertised product. This indication, known as a conversion, allows the search engine to use past searches and other parameters to identify people who are likely to purchase the product, not just to click on an ad. Over time, a system can learn to identify such people from the feedback provided by the advertiser. Ipeirotis and Gabrilovich [9] used this mechanism to find people who can correctly answer questions on topics of interest by providing a conversion signal when people answered several test questions correctly.

Internet advertising systems have long been used to elicit information on people's health and to modify behavior for better health. Thus, Eysenbach [4] ran an advertising campaign to measure the incidence of influenza in Canada. More recently, advertising on search engines was used to try and create behavior change in people with anorexia [33], who were smoking [35], and who are obese [36]. In those studies, search behavior before and after ads were shown was used to measure if behavior change had occurred.

The current work demonstrates for the first time that the ads serving platform can be used to target population at risk for early diagnosis. This is done by utilizing the conversion tracking and optimization mechanism of the ads systems to automatically learn to screen for three types of cancer: lung, breast, and colon. This could provide a major public health benefit, especially in countries with under-developed access to healthcare.

Thus, here the conversion optimization mechanism of ads systems is utilized as follows. Users asking if they have specific cancers will be referred through an ad to a clinically validated questionnaire that calculates the

likelihood of the user having a specific type of cancer. This score will be provided to users, and if the score is such that the person is likely to have cancer, it will also be provided as a conversion signal to the advertisement system. Therefore, the questionnaire score is utilized as a conversion signal. The hypothesis is that over time the system will learn to identify more people who will score high on the questionnaires, indicating that more people at risk of cancer are identified.

The contributions of this paper are as follows. First, clinically verified questionnaires are used to calculate the likelihood of a user having a suspected cancer diagnosis in lieu of queries that indicate a cancer diagnosis (SIUs or users inferred through other, indirect means). Second, individual questionnaire scores of people are shown to correlate with their past search engine queries to show that suspected cancer could have been predicted based on these queries. Finally, I show that the learning capabilities of advertising systems can be used to find more people who are likely to have suspected cancer. These results demonstrate that the proposed methods can assist in finding people with suspected cancer in an accurate and economic manner, and can help people obtain clinically valid assessments even when medical systems are lacking.

2 METHODS

2.1 Overview

This study focused on three types of cancer: lung, breast, and colon. These three were chosen for their relatively high incidence [23] and because the symptoms of these cancers (as described in the relevant questionnaires) were assessed to be understood by laypersons.

Users were recruited through ads shown when they searched for information on diagnosis of these three specific cancers. People who clicked on these ads were referred to a specially designed website where clinically validated questionnaires on whether they should see a specialist oncologist were administered to them. The scores of the completed questionnaires were provided to users.

Users were requested to provide their data for the experiment. In the first study, the scores were correlated with searches of users who consented. In the second study, a conversion indicator was given to the advertising system for users with high scores.

The first study was conducted using the Bing ads system and required privileged access to the search system data to obtain past user queries. The second study was conducted using the Google ads system, with no such privileged access to past queries. The latter was done to demonstrate that public health organizations with no privileged access could also utilize these systems. In both studies, the campaign budget was set to $15 per day, which meant that not all people who issued the relevant queries could be shown the ads. This was done so as to allow the ads system in the second study to select relevant participants from all users issuing relevant queries.

This study was approved by the Microsoft Institutional Review Board (IRB9672).

2.2 Recruitment

Recruitment was similar in both studies: users were recruited through ads displayed using the respective ads system. Recruitment ads were shown when people searched for "symptoms of <cancer type>," "signs of <cancer type>," "<cancer type> diagnosis," "<cancer type> quiz," or "<cancer type> questionnaire."

The ads contained one of the following three titles: "<cancer type> - Do you have it?," "<cancer type> - Think you have it?," or "<cancer type> - Worried you have it?" The text of the ads was "Click here to check if you should see a doctor" (or physician). All ads were shown with equal probability.

2.3 Questionnaires

People who clicked on these ads were referred to a specially designed website. At this website, they were shown a questionnaire developed based on the UK National Institute for Health and Care Excellence (NICE) [5]. The Suspected Cancer Recognition and Referral questionnaires were designed to assist general practitioners to decide

---

ALGORITHM 1: A simplified flow chart of the breast cancer questionnaire, based on NICE guidelines

---

Result: Suspected cancer score (SCS)

if *you are 30 years or over* **then**

    if *you have an unexplained breast lump with or without pain* **then**

        SCS = High;

    else

        SCS = Low;

    end

end

if *you are 50 years or over* **then**

    if *you have discharge, retraction, or other changes of concern in one nipple only* **then**

        SCS = High;

    else

        SCS = Low;

    end

end

---

if a patient should see a specialist oncologist, balancing true-positive and false-negative rates. After answering the questions on the questionnaire, physicians were advised whether or not to refer people for an appointment with the specialist within 2 weeks. The output of the questionnaire is referred to as a suspected cancer score (SCS). People with a high SCS are advised to consult an oncologist within 2 weeks. A simplified version of the breast cancer questionnaire (omitting a few of the questionnaire options) is shown in Algorithm 1. The complete questionnaire for breast cancer can be accessed at https://aka.ms/BreastCancerQuestionnaire.

In this work, users who responded to the questionnaire on the website and received a high SCS were advised to consult with a doctor immediately (in the clinical use of the questionnaire, patients with a high score are referred to an oncologist within 2 weeks). If the SCS was low, users were advised that their symptoms were not commonly associated with cancer but that they should see a medical doctor if the symptoms were persistent or worrying.

Users were asked for their consent to participate in the study both at the beginning of the questionnaire and after the results of the questionnaire were provided to them. Only people who consented in both times were included in the study.

2.4 Study 1: SCS and Their Correlation with Search Engine Queries

People were recruited through ads and asked to complete questionnaires. The ads were shown between December 7, 2017, and April 13, 2018, to people in the United States. I extracted all queries made on Bing by users who completed a questionnaire, consented to contribute their data to the study, and were logged into Bing at the time of ad display. The queries were extracted from 3 months before the questionnaire completion and until 20 days after that date. Queries of each person were represented by the following:

(1) The number of times a medical symptom was mentioned in the queries. Symptoms comprised a list of 195 symptoms and their layperson description as developed in Yom-Tov and Gabrilovich [34].

(2) The words and word pairs (excluding stopwords) in the queries, if these phrases appeared in use by at least 5% of people in the sample.

The SCS was predicted from query data of participants for whom at least 14 days of query data prior to questionnaire completion were available. The independent attributes for prediction included the query terms, as
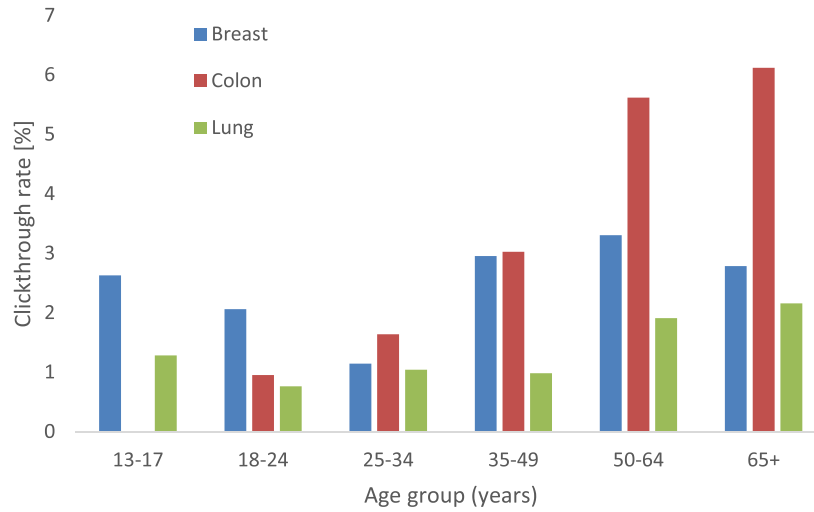
Fig. 1. Clickthrough rates on ads by cancer type and age group.

described earlier, as well as age and gender of the user, as provided during registration with Microsoft. A random forest model [1] was trained and its performance evaluated using leave-one-out estimation [3].[1]

2.5 Study 2: Advertising Systems Can Learn to Identify People with Suspected Cancer

Ads were shown on the Google ads system between May 16 and June 12, 2018. Once each questionnaire was completed, and if users consented to participate in the study, a conversion signal was fed to the Google ads system for those users whose SCS was high. I report the conversion rate over time—that is, the percentage of people who saw the ads, clicked on them, and were found to have a high SCS. If the system can learn to identify people with high SCS, it is expected that this rate will rise over time.

3 RESULTS

3.1 Study 1: SCS and Their Correlation with Past Search Engine Queries

Campaign performance. Recruitment ads were shown 159,170 times and clicked 2,899 times during this experiment. Clickthrough rates for different conditions were similar, ranging from 1.2% (breast cancer) to 4.8% (colon cancer). Females and males were similarly likely to click on the ads for colon and lung cancer, but females were 2.0 times more likely to click on ads for breast cancer.

Clickthrough rates on ads, by cancer type and age group, are shown in Figure 1. As the figure shows, although the range of clickthrough rates are similar, older people tended to click more on ads, with the exception of breast cancer, which was also clicked by younger people.

Throughout the experiment, 1,285 questionnaires were started and 681 were completed (53% completion rate). It took an average of 126 seconds to complete the questionnaires. After excluding people who did not consent to the participate in the study and people who did not have a query history of at least 14 days, the data from 288 remaining people (185 lung, 81 colon, and 22 breast) were analyzed.

Prediction of questionnaire outcome. Since the number of people who were screened for the different cancers were not equal, and some cancers had relatively few examples, SCS was first modeled for all cancers together.

---

[1]A regression tree and a support vector machine were also tested but exhibited the worst performance. Hence, only the performance of the random forest model is reported.
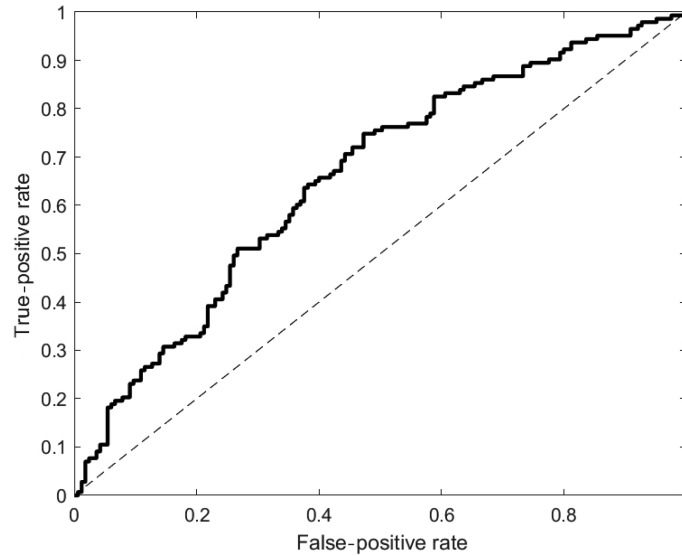
Fig. 2. ROC for detecting people with high likelihood of suspected cancer. The AUC is 0.66.

Figure 2 shows the receiver operating curve (ROC) [3] for the detection of all three cancers from the queries. As the figure shows, the area under the ROC curve (AUC) is 0.66, indicating that it is possible to identify those people who are very likely to have the suspected cancer.

Trained separately, AUCs for the different cancers are 0.74 (colon), 0.56 (lung), and 0.50 (breast). Thus, colon cancer is the easiest one to identify, followed by lung cancer. The data do not allow the prediction of breast cancer outcome.

Attribute importance was calculated according to the average increase in prediction error if the values of that variable are excluded, divided by the standard deviation over the entire forest ensemble. The attributes most indicative of likely cancer were medically related words ("remedies," "colon," "pain," and "diet") and words with no clear link to cancer ("college," "joe," "north," "boots," "watch," and "movie").

Figure 3 shows the likelihood to ask about each of the first four terms at each day prior to the date of the questionnaire (denoted by "relative time"), divided by the likelihood of making any query on that day [20]. As the figure shows, all four terms tend to increase over time. The Spearman correlation between the time and the values shown in the figures are remedies: 0.54 ($p = .004$); colon: 0.48 ($p = .006$); pain: 0.28 ($p = .128$); and diet: 0.50 ($p = .004$).

If the beginning of increase in query likelihood is defined as the first time that the value of likelihood is greater than the 80th percentile of the values across terms and days, then "remedies" and "diet" increase earlier (55 days and 70 days before the questionnaire date, respectively), whereas "pain" and "colon" rise closer to the day of the questionnaire (25 and 5 days, respectively), possibly triggering the searches for the possibility of cancer.

It is hypothesized that the second set of words typifies specific demographics or behaviors, and that they are relevant because of these factors and not because of a direct connection to the disease. To support this hypothesis, I estimated the trend shown in Figure 3 by calculating the Spearman correlation of the likelihood over (relative) time. The average correlation for the preceding medically related terms was 0.45, whereas for the six terms in the second group, it was −0.12.

Effect of the campaign on user behavior. The goal of the proposed methodology is to enable people who are at high risk of having cancer to reach a medical facility faster while helping people with non-critical symptoms
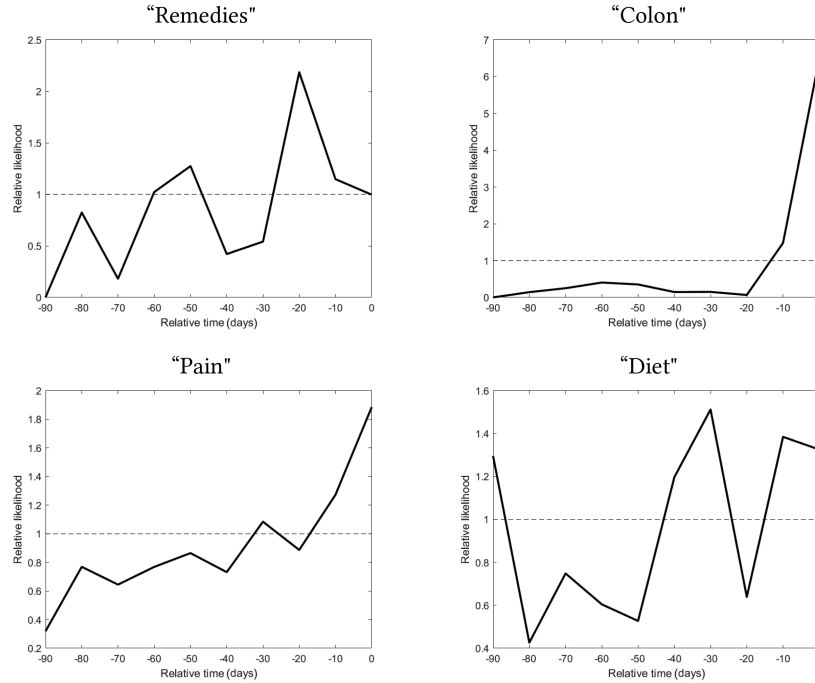
Fig. 3. Likelihood of asking about four words among those with the highest attribute importance. Likelihood is computed as the probability to use the word in a query on a specific day, divided by the probability to use any term in a query that day. The dotted line shows an equal probability to ask either about the term or about any other term. Time is relative to the day of the questionnaire.

allay their worries. Thus, one measure of success is whether people with a high SCS sought to utilize a medical facility, as evident in queries with healthcare resource utilization intent [28], or to find more information about their possible cancer after receiving their questionnaire result. Conversely, people who receive a low SCS should reduce their search for such information.

To measure this behavioral change, relevant queries, defined as those containing the word "cancer" or those referring to a medical facility (containing one or more of the words "clinic," "hospital," "doctor," "physician," or "referral") were identified. The distribution of these queries was examined from 20 days before the questionnaire was completed and until 20 days after it, stratified by the SCS.

Figure 4 shows the distribution of relevant queries as a function of time, relative to the probability of issuing any query [20]. As the figure demonstrates, people with a low SCS return to the baseline rate of relevant queries after approximately 2 days. This is in contrast with people who were given a high SCS score, who sustained a high level of relevant queries for 9 days after completing the questionnaire. The latter also queried more for relevant terms: prior to completing the questionnaire, 2.0% of queries made by people with a low SCS contained relevant terms compared to 3.1% after it (a 53% increase in the short time such queries are made). This is in contrast with people who received a high SCS score, for whom 1.4% of queries contained relevant terms prior to receiving their score compared to 3.5% after it (a 153% increase) (statistically significant, chi^2 test, p-value $< 10^{-10}$). Thus, the questionnaire scores seemingly had the desirable influence on people who completed them.
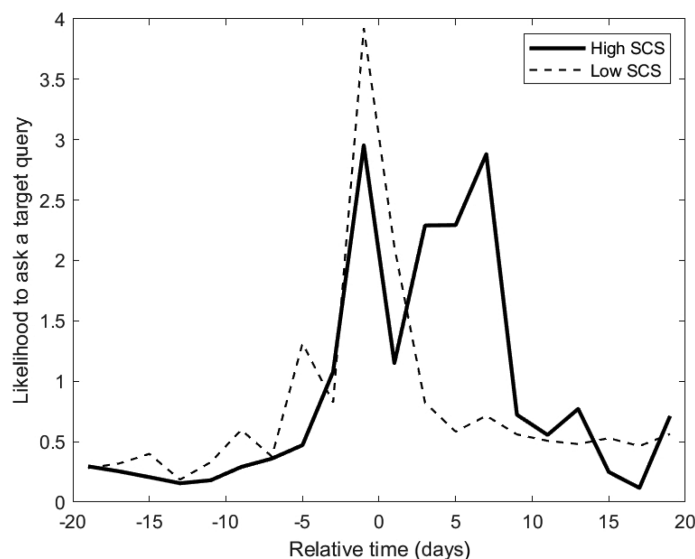
Fig. 4. Probability to ask a query containing the word "cancer" or a health utilization term compared to all queries over time, stratified by SCS. A likelihood of 1 represents equal likelihood to use a relevant term compared to all terms. Time (in days) is relative to the date on which the questionnaire was completed (denoted as day zero). People with a high SCS search for a much longer period following the completion of a questionnaire.

3.2 Study 2: Advertising Systems Can Learn to Identify People with Suspected Cancer

Campaign performance. The ads were shown to 70,586 people between May 16 and June 12, 2018. During that time, 6,484 people clicked on these ads. Of those, 2,917 people began the questionnaire and 1,049 completed it (36% completion rate). The conversion rate over time is shown in Figure 5. As can be seen in the figure, the conversion rate rises from a negligible level to an average of 11% (s.d. 3%) in the last 10 days of the study. This means that the advertising system learned to identify people who would score high on the questionnaire such that about 1 person in 9 who click on the ads were suspected to have cancer. The average conversion rates at the last 10 days of the study for individual cancer types was 11% for breast cancer, 9% for colon cancer, and 9% for lung cancer. The number of people shown the ad each day was, on average, 2,681 (s.d. 467), indicating that a relatively constant number of people saw the ads each day, but those people who saw the ads were increasingly more likely to have suspected cancer. Similarly, the rate of clicks on ads remained approximately constant at 10% after the first 10 days of the study.

Advertising systems are largely a "black box," as their operators do not provide details on how they focused the campaign on relevant people (i.e., people who will "convert"). Thus, no information is directly provided on attribute values that the system identified as leading to more conversions. However, analysis of the aggregate reports provided by the advertising system should hint at how the system performed its optimization. Thus, if the system is improving over time, it should be expected that when the advertising system identifies an attribute value (or combination of values among several attributes) that leads to a higher conversion rate, the number of times that ads shown to people with that attribute value (e.g., the impression rate) should rise over time. Additionally, the conversion rate should be higher than average. Conversely, an attribute value that has a poor conversion rate should receive fewer impressions over time.

Given this insight, the keywords for which the campaign ads were presented, and (separately) the demographic (age and gender) of users were analyzed. Note that a combined analysis was impossible because users of the ads
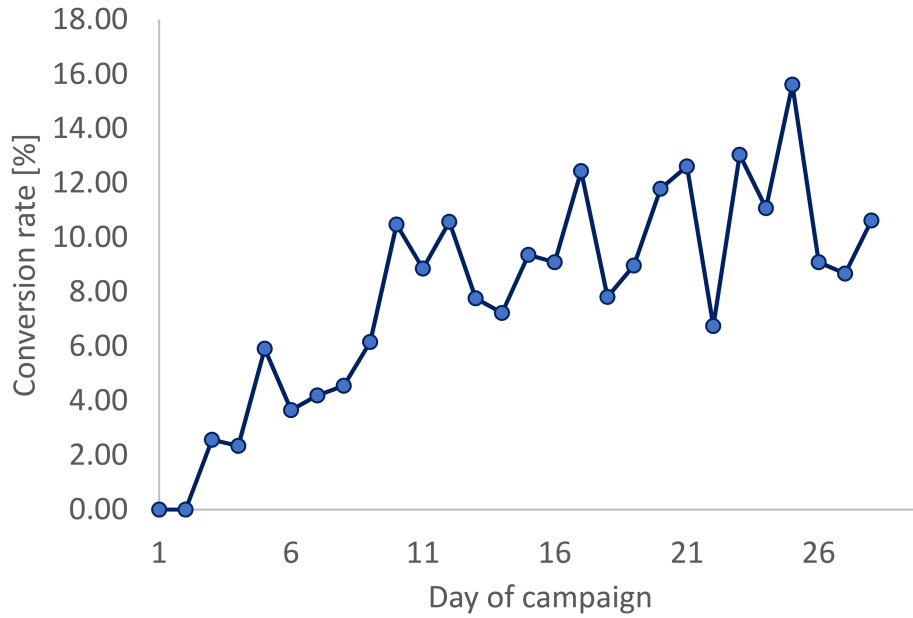
Fig. 5. Conversion rate over time.

system have limited access data from these systems. A rank regression model was fit to the impression rate per day over the duration of the campaign. The model fit (R^2), slope of the impression rate over time, its p-value, and the average impression rate are reported in Table 1. As the table shows, two of the five keywords used to trigger colon cancer ads saw decreased impression rates over time. As expected, their average conversion rates (3.0% and 4.4%) were well below the overall conversion rate of the campaign. Thus, the system identified these keywords as leading many people who did not have suspected cancer to click on the ads. Conversely, the impression rate increased for two keywords related to breast cancer, and the impression rate for these was higher than the first two keywords. Interestingly, the impression rate for one demographic group (females 18–24 years old) increased during the campaign and had a high conversion rate. It is noted that NICE for breast cancer suggests referring women in this age group for medical attention only if they have an unexplained breast lump with or without pain.

Understanding the reason people access the questionnaires. People might search for diagnostic information for different reasons, including as a way to decide if they should visit a medical doctor, to validate the findings given to them by a medical professional, or in lieu of a medical professional. The latter can happen when access to the medical system is restricted either by its availability or by cost (e.g., if the patient does not have medical insurance). Thus, to understand who made use of the campaign ads, I modeled the data from 40 countries where ads were shown at least 150 times.

The clickthrough rate per country was modeled as a function of three independent demographic attributes: gross domestic product (GDP) per person (log transformed), Internet penetration, and life expectancy. GDP is a measure of economic performance of each country. Internet penetration was used as a proxy for the ease of access to information from the Internet. Life expectancy is a known surrogate measure for the quality of health systems [25].

The model reached an R^2 of 0.21 and found that Internet penetration was statistically significantly correlated with the clickthrough rate (slope: 0.001; p-value: .04), as was life expectancy (slope: −0.002; p-value: .02). Thus,

Table 1. Modeling the Increase in Impression Rate
for Different Attribute Combinations

Keywords	R^2	Slope	p-Value	Avg. Conversion Rate
Colon cancer diagnosis	0.33	−0.60	$8 \cdot 10^{-4}$	3.0%
Signs of colon cancer	0.27	−0.54	$3 \cdot 10^{-3}$	4.4%
Signs of breast cancer	0.59	0.78	10^{-6}	8.5%
Symptoms of breast cancer	0.57	0.76	$2 \cdot 10^{-6}$	7.9%
Female, 18–24 years	0.26	0.53	$3 \cdot 10^{-3}$	8.6%

Note: Only statistically significant ($p < .05$ with Bonferroni correction) attribute combinations are shown. The first four rows are campaign keywords and the bottom row a demographic combination.

people living in countries with better Internet access and worse health outcomes were more likely to use the ads to seek diagnosis.

4 DISCUSSION

The potential for using search engine queries to screen for different cancer types has been demonstrated in past work [15, 24, 29], albeit through the use of biased populations who indicated their condition in queries or for people whose condition was inferred. Here, a clinically verified questionnaire was used to identify people with suspected cancer and to correlate their queries with the outcomes of the questionnaires. Moreover, the learning capabilities of advertising systems were utilized to focus on the people most likely to have suspected cancer, thus assisting these people to obtain clinically valid information for themselves.

The outcomes of the first study indicate that queries are predictive of SCS for two of the cancer types (colon and lung), especially for those people for which the strength of the prediction is highest. It is noted that it is difficult to identify whether the dearth of data or a lack of relevant information in the data was the cause for the low performance in the case of lung and breast cancer.

Importantly, the terms used to predict SCS are associated with medical symptoms, suggesting that the questionnaires were completed by people who suspected they have cancer, not those who sought a "second opinion" on an existing diagnosis. If the latter would have been the case, it is expected that the terms most indicative of cancer would have been treatments or treatment facilities.

People who completed the questionnaires seem to have utilized the information they received from them, subsequently searching less and for a shorter time if they received a low SCS compared to people with a high SCS, who searched more and for a much longer period. These findings suggest that providing people with screening information can lead them to the desirable behavioral change.

Having demonstrated that queries are indicative of a high SCS, the second study demonstrated that SCS can serve as a conversion signal to the ads system, enabling the system to learn to identify people suspected of having these conditions based on their past queries and other information available to it. This learning allows the system to focus on the people most likely to have cancer and inform them of possible risk through the combination of ads and questionnaires.

As shown in Section 3, older people tended to click more on ads, with the exception of breast cancer, which was also clicked by younger people. Although the former is understood through the higher incidence of cancer in older ages, the latter is attributed to puberty, whose symptoms might be misinterpreted as a serious disease by younger adults [27], especially when they lack the necessary social support.

Results suggest that the proposed method, which is economical for medical authorities to operate, was used most by people with reduced access to the health system. It can thus be utilized for screening of serious medical

conditions in disadvantaged populations, leveling access to medical knowledge. Moreover, this method can likely reduce stress to people without suspected cancer but are worried about their symptoms. However, further studies are required to verify that such stress reduction is efficacious and that the rate of people with a high SCS who have cancer makes these questionnaires cost effective. Additionally, since people are seeking information to assist them in self-diagnosis even in countries with developed health systems, the proposed method could help alleviate some of the effort required in this endeavor.

Our study has several limitations. First, although the questionnaires are designed to identify people with suspected cancer, our data does not contain diagnostic information. Obviously, not all people who received a high SCS will be diagnosed with cancer. NICE questionnaires are designed to balance errors (false positives and false negatives) in a way that is cost effective (e.g., see Table 3 of Section 4.4 at https://www.nice.org.uk/guidance/ng12/evidence/appendices-ae-pdf-74333342). However, the fact that a high SCS is not proof of cancer should not detract from the usefulness of the SCS in screening. Indeed, accepted screening tests have relatively modest true-positive rates: it is estimated that around 6.3% of women in their 50s with an abnormal mammography have invasive breast cancer [11], and studies report that around 4.6% of people with positive fecal occult blood test are verified as having colon cancer [21]. Nevertheless, these screening tests are considered useful and are recommended by health authorities. Future work will focus on a follow-up with people who completed their questionnaire until after diagnosis to measure their detection rates.

Another limitation of our study is the availability of questionnaires. Currently, NICE does not offer guidelines to all types of cancer, and some guidelines require either medical tests or a trained physician to assess symptoms such as skin lesions. Therefore, the proposed approach is limited by the number of cancers to which it could be applied. Future work is intended to develop new disease questionnaires, for example, by analyzing past searches and social media postings by people who were later diagnosed with a disease of interest.

A conversion in this study requires the satisfaction of three constraints: completion of the questionnaire, receiving a high SCS, and consenting to participate in the study. Thus, the actual rate of people with a high SCS could be different from the reported conversion rate, and the ads system is optimizing for a goal that is not simply people with a high SCS. However, since the goal of this work is to identify people who are willing to complete their questionnaire and consent to receive their SCS, this conversion indicator is a useful one.

Another technical limitation of the study is in the ability to understand what signals are used by the advertising systems to identify people with a high SCS. The reported analysis, although novel, cannot be considered exhaustive, since the advertising system provides only limited information on people, ads, and campaign performance. This also makes it difficult to perform an analysis of attribute combinations. Moreover, some of the information the ads system uses (e.g., user location) might not be reported by it at all. However, the analysis shows that the advertising system optimized the campaign over time according to (at least) user demographics and keyword matches.

Finally, it could be that the advertising system, rather than advertising to representative populations, is focusing on a specific sub-population. This is both because advertisements were shown only to people whose queries indicated that they were worried that they are suffering from cancer or because specific populations respond better to campaign advertisements. People may also query for close family members, in which case their past queries might not be indicative of cancer. All of these may detract from the ability of the advertising system to focus on people with suspected cancer.

The focus of the campaign was on people who inquired about cancer diagnosis and not, for example, on people who queried for relevant symptoms or those who could be identified through previous methods [24, 28] as having cancer. There are several reasons for this. First, advertising to people who did not specifically query for cancer diagnosis could unnecessarily worry. Second, advertising to symptoms could be too unspecific, and ads system do not, at present, support advertising to people who queried for a confluence of symptoms, as would be required to advertise to people who were identified using the previously proposed methods.

Beyond these technical reasons, advertising to people who did not specifically query for diagnosis raises a larger issue of ethics known as 'unsolicited medical diagnosis [18]. Ratzan [19] suggested that physicians have an obligation to provide such diagnosis to a person who is a stranger to them when three conditions are met: (1) a high probability of potentially serious disease, (2) the potential disease is not recognized by the person, and (3) when the physician has the medical knowledge appropriate to interpret this information. However, offering such unsolicited diagnosis is controversial mostly because of the possibility of false positives and due to privacy concerns. Showing the proposed advertising campaign to people who are not seeking diagnosis (e.g., only searching for a possibly relevant symptom) is similar in that people may not be seeking diagnosis. Therefore, although not ruling out the possibility of widening the campaign to show ads to people who are deemed likely to have cancer even when they do not directly seek diagnostic information, it is perhaps necessary to achieve wider social acceptance of the proposed method before running such a campaign.

If such a campaign is executed, additional study is required to verify if advertisements can be shown without causing undue stress and false indications. Such a study should address the ethical aspects of working at a lower precision at the cost of higher recall, the technical challenges, and the societal expectations of such advertisements.

There is significant evidence of the advantages in survival [13] and reduced cost of treatment [10] for cancers such as the ones analyzed in this study. This study suggests new ways in which diagnosis can be made earlier and provided to people, possibly improving medical outcomes and survival of devastating diseases. This is especially true for the people who made use of the proposed system and live in countries where access to healthcare is difficult and sometimes prohibitively expensive.

Studies using Internet data have mostly concentrated on improving population health [22, 33, 35] and showing the feasibility to detect illness [8, 24, 28] but without providing this information in actionable form to people. The current study opens the door for health authorities to (indirectly) use Internet data for screening and to provide medical advice and nudges in a way that empowers users to choose if they would like to utilize it.

REFERENCES

[1] Leo Breiman. 2001. Random forests. *Machine Learning* 45, 1 (2001), 5–32.

[2] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM'13)*.1–10.

[3] Richard O. Duda, Peter E. Hart, and David G. Stork. 2012. *Pattern Classification*. John Wiley & Sons.

[4] Gunther Eysenbach. 2006. Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. In *AMIA Annual Symposium Proceedings*, Vol. 2006. American Medical Informatics Association, Bethesda, MD, 244.

[5] National Institute for Health and Clinical Excellence. 2017. Suspected Cancer: Recognition and Referral. Retrieved February 15, 2020 from https://www.nice.org.uk/guidance/ng12.

[6] Susannah Fox. 2013. Health Online 2013. *Pew Research Center*. Retrieved February 15, 2020 from http://www.pewinternet.org/2013/01/15/health-online-2013/.

[7] Eitan Giat and Elad Yom-Tov. 2018. Evidence from web-based dietary search patterns to the role of B12 deficiency in non-specific chronic pain: A large-scale observational study. *Journal of Medical Internet Research* 20, 1 (2018), e4.

[8] Irit Hochberg, Deeb Daoud, Naim Shehadeh, and Elad Yom-Tov. 2019. Can Internet search engine queries be used to diagnose diabetes? Analysis of archival search data. *Acta Diabetologica* 56, 10 (2019), 1149–1154.

[9] Panos Ipeirotis and Evgeniy Gabrilovich. 2014. Quizz: Targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd International World Wide Web Conference*.

[10] Zura Kakushadze, Rakesh Raghubanshi, and Willie Yu. 2017. Estimating cost savings from early cancer diagnosis. *Data* 2, 3 (2017), 30.

[11] Karla Kerlikowske and John Barclay. 1997. Outcomes of modern screening mammography. *JNCI Monographs* 1997, 22 (1997), 105–111.

[12] Vasileios Lampos, Elad Yom-Tov, Richard Pebody, and Ingemar J. Cox. 2015. Assessing the impact of a health intervention via user-generated Internet content. *Data Mining and Knowledge Discovery* 29, 5 (2015), 1434–1457.

[13] Sean McPhail, Sam Johnson, David Greenberg, Mick Peake, and Brian Rous. 2015. Stage at diagnosis and early mortality from cancer in England. *British Journal of Cancer* 112, s1 (2015), S108.

[14] Kevin A. Padrez, Lyle Ungar, Hansen Andrew Schwartz, Robert J. Smith, Shawndra Hill, Tadas Antanavicius, Dana M. Brown, Patrick Crutchley, David A. Asch, and Raina M. Merchant. 2015. Linking social media and medical record data: A study of adults presenting to an academic, urban emergency department. *BMJ Quality and Safety* 25, 6 (2015), 414–423.

[15] John Paparrizos, Ryen W. White, and Eric Horvitz. 2016. Screening for pancreatic adenocarcinoma using signals from web search logs: Feasibility study and results. *Journal of Oncology Practice* 12, 8 (2016), 737–744.

[16] Dan Pelleg, Elad Yom-Tov, and Yoelle Maarek. 2012. Can you believe an anonymous contributor? On truthfulness in Yahoo! Answers. In *Proceedings of the 2012 International Conference on Privacy, Security, Risk, and Trust (PASSAT'17) and the 2012 International Conference on Social Computing (SocialCom'17)*. IEEE, Los Alamitos, CA, 411–420.

[17] Philip M. Polgreen, Yiling Chen, David M. Pennock, Forrest D. Nelson, and Robert A. Weinstein. 2008. Using Internet searches for influenza surveillance. *Clinical Infectious Diseases* 47, 11 (2008), 1443–1448.

[18] Gustav Preller and Sabine Salloch. 2018. Melanoma in the shopping mall: A utilitarian argument for offering unsolicited medical opinions in informal settings. *Bioethics* 32, 3 (2018), 193–198.

[19] Richard M. Ratzan. 1985. Unsolicited medical opinion. *Journal of Medicine and Philosophy* 10, 2 (1985), 147–162.

[20] Matthew Richardson. 2008. Learning about the world through long-term query logs. *ACM Transactions on the Web* 2, 4 (2008), 21.

[21] T. Rubeca, S. Rapi, M. Confortini, M. Brogioni, G. Grazzini, M. Zappa, D. Puliti, G. Castiglione, and S. Ciatto. 2006. Evaluation of diagnostic accuracy of screening by fecal occult blood testing (FOBT). Comparison of FOB Gold and OC sensor assays in a consecutive prospective screening series. *International Journal of Biological Markers* 21, 3 (2006), 157–161.

[22] Adam Sadilek, Stephanie Caty, Lauren DiPrete, Raed Mansour, Tom Schenk Jr., Mark Bergtholdt, Ashish Jh, Prem Ramaswami, and Evgeniy Gabrilovich. 2019. Machine-learned epidemiology: Real-time detection of foodborne illness at scale. *NPJ Digital Medicine* 1 (2019), Article 36.

[23] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. 2017. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* 67, 1 (2017), 7–30.

[24] Luca Soldaini and Elad Yom-Tov. 2017. Inferring individual attributes from search engine queries and auxiliary information. In *Proceedings of the 26th International Conference on World Wide Web*. 293–301.

[25] Matthew C. Stiefel, Rocco J. Perla, and Bonnie L. Zell. 2010. A healthy bottom line: Healthy life expectancy as an outcome measure for health improvement efforts. *Milbank Quarterly* 88, 1 (2010), 30–53.

[26] C. S. Thomson and David Forman. 2009. Cancer survival in England and the influence of early diagnosis: What can we learn from recent EUROCARE results? *British Journal of Cancer* 101, S2 (2009), S102.

[27] Susan P. Weinstein, Emily F. Conant, Susan G. Orel, Julia A. Zuckerman, and Richard Bellah. 2000. Spectrum of US findings in pediatric and adolescent patients with palpable breast masses. *Radiographics* 20, 6 (2000), 1613–1621.

[28] Ryen W. White and Eric Horvitz. 2010. Web to world: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. In *AMIA Annual Symposium Proceedings*, Vol. 2010. American Medical Informatics Association, Bethesda, MD, 882.

[29] Ryen W. White and Eric Horvitz. 2017. Evaluation of the feasibility of screening patients for early signs of lung carcinoma in web search logs. *JAMA Oncology* 3, 3 (2017), 398–401.

[30] Elad Yom-Tov. 2016. *Crowdsourced Health: How What You Do on the Internet Will Improve Medicine*. MIT Press, Cambridge, MA.

[31] Elad Yom-Tov. 2019. Demographic differences in search engine use with implications for cohort selection. *Information Retrieval Journal* 22, 6 (2019), 570–580.

[32] Elad Yom-Tov, Diana Borsa, Andrew C. Hayward, Rachel A. McKendry, and Ingemar J. Cox. 2015. Automatic identification of web-based risk markers for health events. *Journal of Medical Internet Research* 17, 1 (2015), e29.

[33] Elad Yom-Tov, Anat Brunstein-Klomek, Or Mandel, Arie Hadas, and Silvana Fennig. 2018. Inducing behavioral change in seekers of pro-anorexia content using Internet advertisements: Randomized controlled trial. *JMIR Mental Health* 5, 1 (2018). e6.

[34] Elad Yom-Tov and Evgeniy Gabrilovich. 2013. Postmarket drug surveillance without trial costs: Discovery of adverse drug reactions through large-scale analysis of web search queries. *Journal of Medical Internet Research* 15, 6 (2013), e124.

[35] Elad Yom-Tov, Peter Muennig, and Abdulrahman M. El-Sayed. 2016. Web-based antismoking advertising to promote smoking cessation: A randomized controlled trial. *Journal of Medical Internet Research* 18, 11 (2016), e306.

[36] Elad Yom-Tov, Jinia Shembekar, Sarah Barclay, and Peter Muennig. 2018. The effectiveness of public health advertisements to promote health: A randomized-controlled trial on 794,000 participants. *NPJ Digital Medicine* 1, 1 (2018), 24.