Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results

John Paparrizos, MSc, Ryen W. White, PhD, and Eric Horvitz, MD, PhD

QUESTION ASKED: Can signals mined from large-scale anonymized Web search logs about symptom queries over time be harnessed to build a valuable screening methodology for pancreatic adenocarcinoma?

SUMMARY ANSWER: Search logs can provide valuable signals to predict the later appearance of first-person queries on disease management that are strongly suggestive of a professional diagnosis of pancreatic carcinoma. Performance of the risk stratification holds many weeks in advance and improves when conditioned on the presence of specific symptoms or risk factors found in people’s search histories.

WHAT WE DID: We performed a statistical analysis of the web queries of millions of anonymized searchers. We identified experiential searchers who issued a first-person diagnostic query for pancreatic cancer (e.g., “I was just diagnosed with pancreatic cancer”; Fig.) and we constructed statistical models that can be applied to predict in advance the appearance of such experiential queries from signals derived from the search activity of individuals.

WHAT WE FOUND: Early detection from log data can recall 5% to 15% of the positive cases at extremely low false-positive rates (0.00001 to 0.0001). We identified specific query terms and inferred demographic factors that provide significant boosts in predicting the rise of experiential queries.

BIAS, CONFOUNDING FACTOR(S), DRAWBACKS: Results are based on retrospective analysis of search logs, where experiential queries are used as a proxy for pancreatic cancer diagnoses in the absence of direct reporting from patients. We do not directly consider false negatives associated with missed diagnoses.

REAL-LIFE IMPLICATIONS: The results highlight the promise of using Web search logs as a new direction for screening for pancreatic carcinoma. The methods suggest that low-cost, high-coverage surveillance systems can be deployed to passively observe search behavior and to provide early warning for pancreatic carcinoma, and with extension of the methodology, for other challenging cancers. Surveillance systems could also provide for automated capture and summarization of data and landmarks over time so as to provide patients with talking points in their discussion with medical professionals. Real-world deployment of the methods would need to carefully convey the uncertainties associated with detection outcomes based on consideration of the evidential findings and prevalence rates, while also balancing such issues as searcher anxiety and cost of potentially unnecessary consultation and screening.

See the figure on the following page.
FIG. Venn diagram depicting the sets of searchers used in the search log analysis: pancreatic adenocarcinoma searchers (A), pancreatic adenocarcinoma searchers with experiential diagnostic queries (B), and those who searched for pancreatic adenocarcinoma symptoms (C). |A ∪ C| (ie, the total number of searchers in our original, prefiltered data set) was 9.2 million. Positives are sourced from B ∩ C and negatives are sourced from C \ A. Relative set sizes in the diagram are not to scale.

Searchers in A who query for symptoms but have no experiential queries. Excluded from analysis because a positive or negative label cannot be reliably determined.
Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results

John Paparrizos, MSc, Ryen W. White, PhD, and Eric Horvitz, MD, PhD

Abstract

Introduction
People’s online activities can yield clues about their emerging health conditions. We performed an intensive study to explore the feasibility of using anonymized Web query logs to screen for the emergence of pancreatic adenocarcinoma. The methods used statistical analyses of large-scale anonymized search logs considering the symptom queries from millions of people, with the potential application of warning individual searchers about the value of seeking attention from health care professionals.

Methods
We identified searchers in logs of online search activity who issued special queries that are suggestive of a recent diagnosis of pancreatic adenocarcinoma. We then went back many months before these landmark queries were made, to examine patterns of symptoms, which were expressed as searches about concerning symptoms. We built statistical classifiers that predicted the future appearance of the landmark queries based on patterns of signals seen in search logs.

Results
We found that signals about patterns of queries in search logs can predict the future appearance of queries that are highly suggestive of a diagnosis of pancreatic adenocarcinoma. We showed specifically that we can identify 5% to 15% of cases, while preserving extremely low false-positive rates (0.00001 to 0.0001).

Conclusion
Signals in search logs show the possibilities of predicting a forthcoming diagnosis of pancreatic adenocarcinoma from combinations of subtle temporal signals revealed in the queries of searchers.

INTRODUCTION
Pancreatic adenocarcinoma poses a difficult and resistant challenge in oncology. It is the fourth leading cause of cancer death in the United States and is the sixth leading cause of cancer death in Europe.1 The illness is frequently diagnosed too late to be treated effectively2,3 and can progress from stage I to stage IV in just over 1 year.4 Approximately 75% of patients with pancreatic adenocarcinoma who are not candidates for surgery will die within 1 year of diagnosis, and only 4% will survive for 5 years postdiagnosis.5

Early signs and symptoms of pancreatic adenocarcinoma are subtle and often
present as nonspecific symptoms that appear and evolve over time. The symptoms often do not become salient until the disease has metastasized. We studied a nontraditional, yet promising direction for the early detection of pancreatic adenocarcinoma. The approach centers on the analysis of signals from Web search logs. Specifically, we examined the feasibility of detecting “fingerprints” of the early rise of pancreatic adenocarcinoma via population-scale statistical analyses of the activity logs of millions of people performing searches on sets of relevant symptoms.

People frequently turn to Web searches to locate health-related information. For example, searchers concerned about one or more related symptoms of interest, but show no interest in searching and

Surveillance and screening programs for pancreatic adenocarcinoma face the challenges of engagement and coverage, especially for detecting and addressing subtle, yet important symptoms. We believe that search logs can serve as a new kind of large-scale, widely distributed sensor for capturing concerning temporal patterns of the onset and persistence of queries about symptoms. The sequences of terms that searchers input to search engines over time can capture symptoms as the illness progresses from its early stages to increasingly salient and frank symptoms.

Patterns of onset and persistence of symptoms for pancreatic adenocarcinoma include back pain, abdominal discomfort, unexplained loss of weight and appetite, light-colored stools, generalized pruritus, darkening urine, and yellowing sclera and skin. From the perspective of traditional screening, there are few salient symptoms in early stages of the disease, and the symptoms are not sufficiently specific to raise a suspicion of pancreatic adenocarcinoma. Symptoms may not even concern patients enough to schedule an appointment with their physician.

We present a feasibility study of the early identification of pancreatic adenocarcinoma based on symptom-centric search queries over time, and the temporal relationships and patterns among queries from multiple sessions over several months. Our experiments center on the early prediction of the future appearance in search logs of special queries that we term experiential diagnostic queries. Experiential diagnostic queries are terms inputted into search engines that provide evidence of searchers having recently been professionally diagnosed. These are distinct from exploratory queries, including searches on symptoms or diseases, which appear to be less intensive, more casual searches for information. Experiential queries for pancreatic adenocarcinoma are identified via consideration of the query structure and patterns of information gathering over many searchers in search logs. We specifically sought evidence of credible, first-person assertions such as the query, “I was just diagnosed with pancreatic adenocarcinoma,” which, when associated with prior queries about symptoms, identifies searchers who we label as positive for pancreatic adenocarcinoma. Searchers who inquire about one or more related symptoms of interest, but show no evidence over time of searches for pancreatic adenocarcinoma, constitute the negatives.

METHODS

Search services track characteristics of people’s searching and clicking activities to capture intentions, improve their responses, and personalize content. Searching activities provide streams of data to construct a statistical model that can be used to risk-stratify searchers for screening. Every interaction corresponds to a log entry containing the query, the results
selected, and a timestamp. A unique, anonymized identifier linked to the Web browser is also included, enabling the extraction of search log histories for up to 18 months. The anonymous identifier is tied to a single machine. On shared machines, it may represent the search activity of multiple searchers. The identifier does not enable the consolidation of activity from a single searcher across multiple machines. We used proprietary logs from Bing.com for searchers in the English-speaking United States locale, from October 2013 to May 2015 (inclusive).

Symptoms and Risk Factors
We reviewed the signs, symptoms, and risk factors associated with pancreatic adenocarcinoma. We developed a symptom set covering the following concerns: yellowing sclera or skin, blood clot, light stool, loose stool, enlarged gall bladder, dark urine, floating stool, greasy stool, dark or tarry stool, high blood sugar, sudden weight loss, taste changes, smelly stool, itchy skin, nausea or vomiting, indigestion, abdominal swelling or pressure, abdominal pain, constipation, and loss of appetite. Synonyms for each symptom were identified (eg, symptom: yellowing sclera or skin, synonym: jaundice; symptom: abdominal pain, synonyms: belly pain, stomach ache). We also identified risk factors (eg, pancreatitis, alcoholism) and their associated synonyms (see Lowenfels and Maisonneuve34), describing attributes or characteristics that may increase the likelihood of developing pancreatic adenocarcinoma. The symptoms and the risk factors were mapped to terms in search queries.

Extracting Pancreatic Adenocarcinoma Searchers and Symptom Searchers
To identify positive and negative cases in generating a learned model, we built a data set of searchers from two groups (Fig 1A). Pancreatic adenocarcinoma searchers (A) includes all searchers who inputted one or more queries matching the expression [(pancreas OR pancreatic) AND cancer]. We considered searchers with a diagnosis of pancreatic adenocarcinoma (B) as the subset of searchers (A) who issued one or more experiential diagnostic queries. Symptom searchers (C) includes all searchers with one or more queries related to pancreatic adenocarcinoma symptoms or synonyms (see Symptoms and Risk Factors).

The full search histories of 9.2 million searchers comprise the union of (A) and (C) in Figure 1A. We used a statistical topic classifier developed for use by the Bing search service to identify all health-related queries. We also applied statistical classifiers developed by Bing to make inferences about searchers’ ages and gender. Using these statistical models as filters, we identified searchers for whom >20% of their queries were health related. We excluded those searchers, given the high likelihood that they were health care professionals.35 A total of 7.4 million searchers remained, among whom 479,787 were pancreatic adenocarcinoma searchers. As additional features for statistical analysis, we used a classifier that provides distributions of topics for queries and clicked results.36 We also considered the dominant geolocation for each searcher using a table that links their Internet provider address to locations.

Positive and Negative Cases
We created query timelines for those labeled as experiential diagnostic searchers and exploratory symptom searchers, and drew sets of observations from these timelines to construct a risk-stratification model. Figure 1B summarizes the strategies for identifying positives and negatives. Query timelines are aligned across searchers based on the point when people issued the first experiential diagnostic query. To ensure sufficient data about each searcher, we removed from the study those with fewer than five search sessions (comprising a sequence of search actions with no more than 30 minutes between actions)15,17 spanning five different days. This reduced the population to 6.4 million searchers, with a mean total duration (time between first and last queries) of 210.32 days (standard deviation of 182.93 days and interquartile range of 120 days).

Positive cases
To identify experiential pancreatic adenocarcinoma searchers, we defined first-person diagnostic queries for pancreatic adenocarcinoma (Expd) based on an exploration of logs. Queries admitted as experiential diagnostic queries included such phrases as “Just diagnosed with pancreatic cancer,” “How did I get cancer in pancreas,” and “I was told I have pancreatic cancer, what to expect.” From the set of pancreatic adenocarcinoma searchers, 3,203 matched the diagnostic query patterns. Experiential searchers must have searched for at least one symptom before Expd. This generated 1,072 query timelines of experiential searchers containing periods of symptom lookup followed by the diagnostic query (33.5% of all experiential diagnostic searches). The symptom lookup period starts when the first symptom is detected in our symptom set (mean duration mean = 109.34 days, standard deviation = 49.66 days). For positives, the symptom lookup period terminates at least 1 week before diagnosis (Σ = 1 week) to reduce the likelihood of overlap between them (which could
add noise to model training and testing), while allowing us to understand predictive performance with minimal lead times.

**Negative cases**

To generate a set of searchers we considered negative for pancreatic adenocarcinoma, we sampled from those who searched for pancreatic adenocarcinoma symptoms but who did not search for pancreatic adenocarcinoma directly anywhere in their timeline. We reduced the number of negatives via a sampling procedure to include only those with symptom lookup durations within three standard deviations of the mean of the positives (n = 3,025,046). The resultant positive and

---

**FIG 1.** (A) Venn diagram depicting the sets of searchers used in the search log analysis: pancreatic adenocarcinoma searchers (A), pancreatic adenocarcinoma searchers with experiential diagnostic queries (B), and those who searched for pancreatic adenocarcinoma symptoms (C). |A ∪ C| (i.e., the total number of searchers in our original, prefiltered data set) was 9.2 million. Positives are sourced from B ∩ C and negatives are sourced from C \ A. Relative set sizes in the diagram are not to scale. (B) Schematic illustrating the query timelines used in the selection of positive and negative cases. S₀ refers to the first symptom query and Exp₀ is the first experiential diagnostic query. α is the duration of the symptom lookup period, which is meant to be approximately equal in the aggregate for the positives and negatives. β is the duration of the period of diagnosis, set to 1 week in the current study.
negative distributions are statistically indistinguishable using two-sample Kolmogorov-Smirnov tests for temporal duration (\(D = 0.005, P = .7017\)) and number of queries (\(D = 0.003, P = .7681\)), even though the latter was not a filtering criterion.

**Early Detection**

We framed early detection as a binary classification challenge using a statistical classifier. We trained the classifier on features from query timelines of experiential pancreatic adenocarcinoma searchers and symptom-only searchers. Given concerns about false positives and the rarity of pancreatic adenocarcinoma, we focused on maintaining low false-positive rates (FPRs; ie, one wrong prediction in 100,000 correctly identified cases), while retaining a high imbalance ratio of positives and negatives (ie, 1,000 positives \(\times\) millions of negatives).

The set of observations or features extracted from the symptom lookup period are grouped into five categories as follows: (1) searcher demographic information, including age/sex predictions and dominant location (Demographics); (2) session characteristics, query classes, and URL classes, including activity characteristics and the topics of queries issued and resources accessed (Search Characteristics); (3) characteristics about symptoms searched, including generic symptom searching (eg, number of distinct symptoms; Symptom General) and features for each symptom (Symptom Specific); (4) features that capture the temporal dynamics of the features (eg, increasing/decreasing over time, rate of change; Temporal), and (5) risk factors, including their presence in queries (Risk Factors).

The learned statistical model is based on the gradient boosted trees\(^3^7\) method. Regularization methods were used to minimize the risk of overfitting. See Paparrizos et al\(^3^8\) for details on the construction of the classifier. We used the statistical classifier to study our ability to perform early identification of searchers who would later make experiential diagnostic queries for pancreatic adenocarcinoma. To characterize the predictive power, we used the area under the receiver operator curve (AUROC) and the recall (true-positive rate [TPR]) at low FPRs as evaluation metrics. Model generalizability was assessed using 10-fold cross validation, stratified by searcher.

**RESULTS**

Performance of the statistical classifier using data up to the period of diagnosis (ie, \(Exp_0\) — 1 week) was strong (AUROC = 0.9003). Because low error rates are important when applying our model, the TPR (ie, fraction of positives recalled) at low FPRs (ie, 0.0001 or 0.00001) is also of interest. Focusing on FPRs in the range 0.00001 to 0.01, the model recalls 5% to 30% of the positives, depending on the FPR.

**Performance by Week**

Prediction performance can change as we increase the lead time between prediction and diagnostic query. We selected

<table>
<thead>
<tr>
<th>No. of Weeks Before (Exp_0^*)</th>
<th>TPR at FPRs Ranging From 0.00001 to 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00001</td>
</tr>
<tr>
<td>1</td>
<td>7.122</td>
</tr>
<tr>
<td>5</td>
<td>7.122</td>
</tr>
<tr>
<td>9</td>
<td>7.122</td>
</tr>
<tr>
<td>13</td>
<td>7.122</td>
</tr>
<tr>
<td>17</td>
<td>6.825</td>
</tr>
<tr>
<td>21</td>
<td>6.528†</td>
</tr>
</tbody>
</table>

NOTE: Values are averaged across the 10 folds of the cross-validation. Weeks denotes the week before first experiential diagnostic query when the prediction is made (eg, “5 weeks” means to train the model using data up to 5 weeks before the first experiential diagnostic query [\(Exp_0\)].

Abbreviations: AUROC, area under the receiver operating characteristic curve; \(Exp_0\), first-person diagnostic queries for pancreatic adenocarcinoma; FPR, false-positive rate; TPR, true-positive rate.

\(^*P < .01, †P < .001, \text{and} \, §P < .0001.\)

\(b\) in Figure 1B.

\(\dagger P < .01, ‡P < .001, \text{and} \, §P < .0001.\)
Table 2. Top 10 Features, Ranked in Descending Order by Evidential Weight

<table>
<thead>
<tr>
<th>Observation Type</th>
<th>Weight</th>
<th>Direction</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of distinct symptoms searched</td>
<td>1.0000</td>
<td>Positive</td>
<td>Symptom general</td>
</tr>
<tr>
<td>Fraction of search queries that are health related</td>
<td>0.8253</td>
<td>Positive</td>
<td>Query topic</td>
</tr>
<tr>
<td>No. of distinct symptom synonyms searched</td>
<td>0.6899</td>
<td>Positive</td>
<td>Symptom general</td>
</tr>
<tr>
<td>Probability that searcher’s age is 50-85 years</td>
<td>0.6889</td>
<td>Positive</td>
<td>Demographic</td>
</tr>
<tr>
<td>Searcher has searched for back pain</td>
<td>0.6622</td>
<td>Negative</td>
<td>Symptom specific</td>
</tr>
<tr>
<td>Searcher has searched for indigestion</td>
<td>0.6432</td>
<td>Negative</td>
<td>Symptom specific</td>
</tr>
<tr>
<td>Searcher has searched for indigestion, then abdominal pain</td>
<td>0.6349</td>
<td>Positive</td>
<td>Temporal</td>
</tr>
<tr>
<td>Gradient of best-fit line for no. of distinct symptoms searched</td>
<td>0.6154</td>
<td>Positive</td>
<td>Temporal</td>
</tr>
<tr>
<td>Searcher has searched for back pain, then yellowing sclera or skin</td>
<td>0.6004</td>
<td>Positive</td>
<td>Temporal</td>
</tr>
<tr>
<td>Probability that searcher’s age is &lt; 18 years</td>
<td>0.5869</td>
<td>Negative</td>
<td>Demographic</td>
</tr>
</tbody>
</table>

NOTE. Weights are relative to the top-weighted feature, “No. of distinct symptoms searched,” which was assigned a weight of 1.0000. Direction of positive or negative means that the feature correlates positively or negatively with ground truth.

Table 2 shows the observation types (features) with the highest evidential weight. Direction is based on correlations between the feature and training data labels. The number of distinct pancreatic adenocarcinoma symptoms searched is most important, representing a high level of concern. Also important are temporal features, including sequence ordering of symptom pairs, inferred age, and searches for back pain and indigestion (which are common ailments and have many explanations).

Observations also varied in predictive power at FPR = 0.00001, for example, temporal dynamics (AUROC = 0.8391, TPR = 0.2985%), specific symptoms (AUROC = 0.8176, TPR = 2.800%), and demographic information (AUROC = 0.6565, TPR = 0.2800%), differing significantly from the full model (at P < .01 using paired t tests).

Contributions by Observation Type

Symptoms and Risk Factors

The presence of specific symptoms and risk factors in searchers’ query timelines could affect early detection performance. Risk factors include pancreatitis, smoking, and obesity, as well as cancer syndromes such as hereditary intestinal polyposis syndrome or familial atypical multiple mole melanoma syndrome, which can all increase the likelihood of developing pancreatic adenocarcinoma.26,39-43

We applied cross-validation. For training, we learned a model on searchers in the nine folds allocated to training. For testing, we iterated through symptoms and risk factors and isolated searchers in the test fold who searched for those symptoms or risk factors at Exp0 – 1 week or earlier. In each case, the number of positives and negatives is less than the full set. Appendix Table A1 (online only) presents statistics on the performance for each model with ≥ 10 positives (to help ensure that AUROC calculations were meaningful). TPRs at different FPRs are shown, as are the percentage of positives or negatives with symptom or risk factor searches. The last three columns present the estimated number of true positives (TPs) or false positives (FPs) that would be observed at FPR = 0.00001, and capture cost estimates in terms of numbers of searchers correctly and falsely alerted. Ideal targets for rates of capture versus cost in a deployed service can be derived via a decision analysis that considers the net expected value of the early detection and the expected costs of unnecessary anxiety and rule-out. Such an optimization would leverage a careful characterization of the value of early intervention and details of designs of methods for engaging people.

337 positives and 945,394 negatives who were still observed in the logs many weeks before Exp0, and reported results for β = 1 to 21 weeks. Because feature generation requires 4 weeks of data, for inclusion at Exp0 − 21 weeks, a searcher needs to be observed at Exp0 − 25 weeks.

We trained a model for the filtered set of searchers as for all searchers. Table 1 reports the TPR at different FPRs for this same set of searchers at different 4-week increments, as well as the AUROC. Performance dropped consistently with increased lead time, but even at 21 weeks before Exp0, the predictive performance was still strong (AUROC = 0.8315, TPR [at FPR = 0.00001] = 6.528%).
Appendix Table A1 shows that considering only searchers seeking information related to risk factors such as smoking, hepatitis, and obesity leads to better overall performance. Fewer than 10 searchers searched for each cancer syndrome (eg, hereditary nonpolyposis colorectal cancer), and these cases were excluded from Appendix Table A1. We found terms for symptoms and risk factors that are more likely to occur in positives (eg, pancreatitis is six times more likely, smoking is four times more likely). If we fixed FPR = 0.00001, we would correctly detect 52 searchers (TPs) but would mistakenly alert 30 searchers (FPs; capture cost ratio = 1.72). Appendix Table A1 also shows that conditionalizing on specific symptoms/risk factors markedly improves the capture cost ratio. For example, for alcoholism or obesity, we found 20 to 30 times more TPs than FPs.

**DISCUSSION**

Web search logs may offer a useful source of signals for pancreatic adenocarcinoma screening, with significant lead time (eg, 5 months before the diagnostic query, TPR is 6% to 32% at extremely low FPRs). Because pancreatic adenocarcinoma may progress from stage I to stage IV in just over 30% at extremely low FPRs). Because pancreatic adenocarcinoma screening, with significant lead time (eg, hereditary nonpolyposis colorectal cancer), and these cases were excluded from Appendix Table A1. We found terms for symptoms and risk factors that are more likely to occur in positives (eg, pancreatitis is six times more likely, smoking is four times more likely). If we fixed FPR = 0.00001, we would correctly detect 52 searchers (TPs) but would mistakenly alert 30 searchers (FPs; capture cost ratio = 1.72). Appendix Table A1 also shows that conditionalizing on specific symptoms/risk factors markedly improves the capture cost ratio. For example, for alcoholism or obesity, we found 20 to 30 times more TPs than FPs.

To understand how particular symptoms or risk factors influence model performance, we excluded searchers who lacked supporting evidence for each symptom or risk factor in their search histories. An alternative is to train a separate model for each symptom or risk factor. However, there were insufficient positive examples in each data set with which to train a robust model. In addition, training a generic model and conditioning its application on the presence of symptoms and risk factors is more similar to how the model would be applied in practice.

Our approach leverages low-cost passive observation rather than active screening. This could be generalized to other diseases where noticeable symptoms appear and evolve over periods of time before diagnoses are made. Active screening is not cost effective unless there is a reasonable probability of detecting invasive or preinvasive disease (eg, at least 16% 45). Search log–based (retrospective) methodologies support the characterization of individuals’ longitudinal behaviors at a scale that is infeasible in other studies, which are typically much smaller, for example, Huxley et al46 and Renehan et al.47 Comparisons against baselines, where suspicions about the presence of pancreatic adenocarcinoma are raised via direct screening, are needed to determine changes in screening costs associated with our method. Clinical trials are necessary to understand whether our learned model has practical utility, including in combination with other screening methods.

Alerting people about the potential value of seeking medical care can be challenging. Surveillance systems need to convey the uncertainties associated with detection outcomes while balancing other issues such as alarm and anxiety for searchers and liability for search providers. Systems could summarize historic symptom search activity as talking points for discussion with medical professionals or alert physicians separately from patients.

Volume 12 / Issue 8 / August 2016   jop.ascopubs.org 743

---

**Footnotes:**

1. Some searchers sought information related to symptoms and risk factors (eg, Whipple procedure, pancreaticoduodenectomy, neoadjuvant therapy) and > 20% searched for related medications (eg, gemcitabine, fluorouracil). In contrast, only 0.5% and 0.02% of our negative cases searched for treatments and medications, respectively, at any point in their query timeline. The impact of additional risk factors such as race, family history, and medical history needs to be understood. Oncologists and patients also need to be directly involved in future studies.

2. To understand how particular symptoms or risk factors influence model performance, we excluded searchers who lacked supporting evidence for each symptom or risk factor in their search histories. An alternative is to train a separate model for each symptom or risk factor. However, there were insufficient positive examples in each data set with which to train a robust model. In addition, training a generic model and conditioning its application on the presence of symptoms and risk factors is more similar to how the model would be applied in practice.

3. Our approach leverages low-cost passive observation rather than active screening. This could be generalized to other diseases where noticeable symptoms appear and evolve over periods of time before diagnoses are made. Active screening is not cost effective unless there is a reasonable probability of detecting invasive or preinvasive disease (eg, at least 16% 45). Search log–based (retrospective) methodologies support the characterization of individuals’ longitudinal behaviors at a scale that is infeasible in other studies, which are typically much smaller, for example, Huxley et al46 and Renehan et al.47 Comparisons against baselines, where suspicions about the presence of pancreatic adenocarcinoma are raised via direct screening, are needed to determine changes in screening costs associated with our method. Clinical trials are necessary to understand whether our learned model has practical utility, including in combination with other screening methods.

4. Alerting people about the potential value of seeking medical care can be challenging. Surveillance systems need to convey the uncertainties associated with detection outcomes while balancing other issues such as alarm and anxiety for searchers and liability for search providers. Systems could summarize historic symptom search activity as talking points for discussion with medical professionals or alert physicians separately from patients.
References

7. Paparrizos et al.
AUTHORS’ DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO’s conflict of interest policy, please refer to www.asco.org/rwc or jop.ascopubs.org/site/misc/ifc.xhtml.

John Paparrizos  
No relationship to disclose

Ryen W. White  
No relationship to disclose

Eric Horvitz  
No relationship to disclose
### Table A1. Performance of the Models Conditioned on a Variety of Symptom and Risk Factors

<table>
<thead>
<tr>
<th>Symptom or Risk Factor</th>
<th>TPR at FPRs Ranging From 0.00001 to 0.1</th>
<th>AUROC</th>
<th>No. Positive</th>
<th>No. Negative</th>
<th>% All Positive</th>
<th>% All Negative</th>
<th>FPR = 0.00001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark or tarry stool</td>
<td>Condition</td>
<td>0.0001</td>
<td>0.001</td>
<td>0.01</td>
<td>0.1</td>
<td>0.713*</td>
<td>13</td>
</tr>
<tr>
<td>Abdominal swelling/pressure</td>
<td>Condition</td>
<td>0.0001</td>
<td>0.001</td>
<td>0.01</td>
<td>0.1</td>
<td>0.773*</td>
<td>24</td>
</tr>
<tr>
<td>Ulcers</td>
<td>Risk factor</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>7.895</td>
<td>50.000</td>
<td>0.789*</td>
</tr>
<tr>
<td>Dark urine</td>
<td>Symptom</td>
<td>0.0000</td>
<td>5.556</td>
<td>16.667</td>
<td>27.778</td>
<td>50.000</td>
<td>0.8129*</td>
</tr>
<tr>
<td>Pancreatitis</td>
<td>Risk factor</td>
<td>6.061</td>
<td>9.091</td>
<td>12.121</td>
<td>24.242</td>
<td>54.566</td>
<td>0.8220</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>Symptom</td>
<td>5.385</td>
<td>10.000</td>
<td>16.923</td>
<td>32.308</td>
<td>60.000</td>
<td>0.8343*</td>
</tr>
<tr>
<td>Enlarged gall bladder</td>
<td>Symptom</td>
<td>0.885</td>
<td>2.655</td>
<td>9.735</td>
<td>25.664</td>
<td>53.982</td>
<td>0.8359*</td>
</tr>
<tr>
<td>Constipation</td>
<td>Symptom</td>
<td>3.529</td>
<td>7.059</td>
<td>9.412</td>
<td>22.353</td>
<td>57.647</td>
<td>0.8469*</td>
</tr>
<tr>
<td>Smoking</td>
<td>Risk factor</td>
<td>3.846</td>
<td>3.846</td>
<td>7.692</td>
<td>15.385</td>
<td>53.846</td>
<td>0.8585</td>
</tr>
<tr>
<td>Blood clot</td>
<td>Symptom</td>
<td>4.494</td>
<td>10.112</td>
<td>14.607</td>
<td>31.461</td>
<td>61.798</td>
<td>0.8599</td>
</tr>
<tr>
<td>High blood sugar</td>
<td>Symptom</td>
<td>6.135</td>
<td>8.896</td>
<td>16.564</td>
<td>31.595</td>
<td>60.429</td>
<td>0.8611</td>
</tr>
<tr>
<td>Nausea or vomiting</td>
<td>Symptom</td>
<td>3.200</td>
<td>8.800</td>
<td>17.600</td>
<td>30.400</td>
<td>63.200</td>
<td>0.8703</td>
</tr>
<tr>
<td>Chills or fever</td>
<td>Risk factor</td>
<td>3.636</td>
<td>7.273</td>
<td>20.909</td>
<td>30.909</td>
<td>65.455</td>
<td>0.8727</td>
</tr>
<tr>
<td>Loose stool</td>
<td>Symptom</td>
<td>4.615</td>
<td>7.692</td>
<td>18.462</td>
<td>35.385</td>
<td>72.308</td>
<td>0.8756</td>
</tr>
<tr>
<td>Indigestion</td>
<td>Symptom</td>
<td>7.547</td>
<td>12.264</td>
<td>20.755</td>
<td>38.679</td>
<td>68.868</td>
<td>0.8932</td>
</tr>
<tr>
<td>Itchy skin</td>
<td>Symptom</td>
<td>18.750</td>
<td>25.000</td>
<td>25.000</td>
<td>25.000</td>
<td>75.000</td>
<td>0.8982</td>
</tr>
<tr>
<td>Back pain</td>
<td>Symptom</td>
<td>7.801</td>
<td>14.184</td>
<td>19.858</td>
<td>34.752</td>
<td>69.504</td>
<td>0.9047</td>
</tr>
<tr>
<td>Yellowing sclera or skin</td>
<td>Symptom</td>
<td>2.174</td>
<td>5.439</td>
<td>19.565</td>
<td>38.044</td>
<td>73.913</td>
<td>0.9217</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>Risk factor</td>
<td>7.692</td>
<td>10.256</td>
<td>20.513</td>
<td>38.462</td>
<td>71.795</td>
<td>0.9275</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>Risk factor</td>
<td>12.500</td>
<td>16.667</td>
<td>27.083</td>
<td>41.667</td>
<td>89.583</td>
<td>0.9494*</td>
</tr>
<tr>
<td>Obesity</td>
<td>Risk factor</td>
<td>20.690</td>
<td>20.690</td>
<td>37.931</td>
<td>62.069</td>
<td>82.7590</td>
<td>0.9572*</td>
</tr>
<tr>
<td>Overall</td>
<td>None</td>
<td>4.851</td>
<td>8.302</td>
<td>17.258</td>
<td>36.474</td>
<td>72.015</td>
<td>0.9003</td>
</tr>
</tbody>
</table>

**NOTE:** Values below the dashed line have a higher AUROC than Overall. Capture represents the number of true-positive cases in the cohort of positives \[\cup\] negatives at FPR = 0.00001. Cost represents the number of false-positive cases in that same set at FPR = 0.00001. A capture cost ratio > 1.0 means that more people could benefit from an alert than could be mistakenly alerted. Statistically significant differences with Overall model using DeLong’s test [DeLong ER, et al: Biometrics 44:837-845, 1988] are marked using *P < .0001 and †P < .001 (where the significance threshold following a Bonferroni correction is .002). Abbreviations: AUROC, area under the receiver operating curve; FPR, false-positive rate; TPR, true-positive rate.