## RESEARCH

**Open Access**

# Using online search activity for earlier detection of gynaecological malignancy

Jennifer F. Barcroft[1], Elad Yom-Tov[2], Vasileios Lampos[3], Laura Burney Ellis[1], David Guzman[3], Víctor Ponce-López[3], Tom Bourne[1], Ingemar J. Cox[3,4] and Srdjan Saso[1*]

### Abstract

**Background**  Ovarian cancer is the most lethal and endometrial cancer the most common gynaecological cancer in the UK, yet neither have a screening program in place to facilitate early disease detection. The aim is to evaluate whether online search data can be used to differentiate between individuals with malignant and benign gynaecological diagnoses.

**Methods**  This is a prospective cohort study evaluating online search data in symptomatic individuals (Google user) referred from primary care (GP) with a suspected cancer to a London Hospital (UK) between December 2020 and June 2022. Informed written consent was obtained and online search data was extracted via Google takeout and anonymised. A health filter was applied to extract health-related terms for 24 months prior to GP referral. A predictive model (outcome: malignancy) was developed using (1) search queries (terms model) and (2) categorised search queries (categories model). Area under the ROC curve (AUC) was used to evaluate model performance. 844 women were approached, 652 were eligible to participate and 392 were recruited. Of those recruited, 108 did not complete enrollment, 12 withdrew and 37 were excluded as they did not track Google searches or had an empty search history, leaving a cohort of 235.

**Results**  The cohort had a median age of 53 years old (range 20–81) and a malignancy rate of 26.0%. There was a difference in online search data between those with a benign and malignant diagnosis, noted as early as 360 days in advance of GP referral, when search queries were used directly, but only 60 days in advance, when queries were divided into health categories. A model using online search data from patients ($n = 153$) who performed health-related search and corrected for sample size, achieved its highest sample-corrected AUC of 0.82, 60 days prior to GP referral.

**Conclusions**  Online search data appears to be different between individuals with malignant and benign gynaecological conditions, with a signal observed in advance of GP referral date. Online search data needs to be evaluated in a larger dataset to determine its value as an early disease detection tool and whether its use leads to improved clinical outcomes.

**Keywords**  Ovarian neoplasms, Endometrial neoplasms, Early detection of cancer, Cancer screening test, Internet, Health

---

*Correspondence:
Srdjan Saso
srdjan.saso01@imperial.ac.uk
Full list of author information is available at the end of the article

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 2 of 13

## Introduction

Ovarian Cancer remains the most lethal gynaecological malignancy, largely because most women (75%) present with advanced stage disease [1]. Detection of ovarian cancer at an early stage is associated with a seven-fold increase in five-year survival (93 vs 13%), compared to advanced disease [1]. Endometrial cancer is the commonest gynaecological malignancy in the UK with a rapidly increasing incidence driven by the obesity epidemic [2, 3]. Despite advances in diagnostics only 38% of ovarian and 79% of endometrial cancers are detected at an early stage (stage I & II) [4]. Earlier cancer detection relies upon prompt referral for investigation of symptomatic individuals and periodic screening of the asymptomatic population.

The non-specific, vague symptoms associated with ovarian cancer, coupled with a poor awareness of its presentation among the general population, contributes to delayed diagnoses [5]. Primary care physicians (GPs) are responsible for identifying high risk patients and arranging urgent referral into specialist services for further investigation via an "urgent cancer pathway" [6, 7]. Referral barriers within primary care are known to be a key contributor to delayed cancer diagnoses, as women typically present to their primary care provider three times before being referred [6–8]. Reluctance among primary care clinicians to refer patients for further investigation is closely correlated to poor cancer survival rates [6].

An effective screening strategy can address primary care referral barriers, but requires engagement in the value of screening, which can vary significantly [9]. Whilst screening has proved effective in facilitating the earlier detection of cervical cancer and preventing 70% deaths in England, a cost-effective screening program for ovarian and endometrial cancer does not exist [10]. Two large randomised controlled trials (UKTOCs and PLCO), assessed the value of combined imaging (pelvic ultrasound) and a tumour marker (CA-125) to improve the earlier detection of ovarian cancer [11, 12]. Neither trial resulted in a mortality benefit, therefore were not deemed to be a cost-effective screening strategy by the National Institute for Clinical Excellence (NICE, 2011) and were not integrated into clinical practice. More recently, the Refining Ovarian Cancer Test Accuracy Scores (ROCkeTS) trial, which incorporates symptom questionnaires, serology, with ultrasound-based diagnostic models, is being evaluated in a prospective study, to determine their role in the diagnosis of ovarian cancer [13]. The value of routine pelvic ultrasound in endometrial cancer screening was evaluated in a systematic review of 11,000 asymptomatic women. It concluded that assessment of endometrial thickness has no value in endometrial cancer detection because of the

inherent diagnostic inaccuracy in its measurement and was therefore not recommended [14, 15].

Widespread access to the Internet globally (97.8% in the UK) has facilitated vast numbers of online services [16, 17]. Users of these online services generate digital footprints either directly, through posting on social media platforms, or indirectly, through online search histories stored by service providers, such as Google. Google search has 86.3% of the UK online search engine market [18]. Worldwide, Google processes approximately 9 billion searches per day, of which 630 million are health-related (7%) [19, 20]. Digital footprints have been shown to be useful in disease surveillance [21, 22], and for generating individualised health risk profiles [23]. The non-episodic, temporally dense nature of digital footprints complements the conventional method of disease detection using sparse, episodic healthcare records. Online search data has identified individuals at risk of developing common health conditions, including myocardial infarction, allergies and Human Immunodeficiency virus, and in highlighting novel disease risk factors, useful in the prevention of disease [23]. The use of online search data to facilitate the earlier detection of cancer was demonstrated by *White* et al. [24], which showed that 58% of individuals, thought to have lung cancer (based on their online searches), could be identified using a predictive model (AUC of 0.89) up to 39 weeks prior to being diagnosed [24]. A similar finding was reported in individuals thought to have pancreatic cancer (based on their online searches) *by Paparrizos* et al. [25] Whilst the presence of a signal in online search patterns to enable the earlier identification of disease is exciting, previous studies used a 'proxy' diagnosis of cancer, based on an individual's online searches, not a clinically confirmed diagnosis. This assumption may be invalid, as no study has clinically validated experiential searches in individuals with confirmed diagnoses. Furthermore, the interpretation of the disease timeline is limited by a lack of robust disease diagnosis time points.

Existing research in this field has focused on comparing a malignant cohort to 'healthy population' controls, rather than using a cohort with an underlying benign diagnosis. This has not only limited the comparison between benign and malignant conditions but is also likely to have resulted in overly optimistic findings.

We aimed to appraise online search patterns in symptomatic individuals with known gynaecological diagnoses, to determine (1) if there is a difference in online search patterns between individuals with a malignant and benign diagnosis and (2) if this can enable the identification of individuals with a gynaecological malignancy at an earlier stage.

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 3 of 13

## Methods

### Recruitment and inclusion criteria

This pilot study was conducted at a tertiary London University Hospital between December 2020 and July 2022. Women (aged 18 or older) who were referred to the hospital by their primary care physician (GP) with gynaecological symptoms, had a Google account and were English speakers were eligible to participate. Patients with a personal history of ovarian and endometrial cancer (previously treated) were also eligible for inclusion.

The requirement of a Google email account was necessary for this study, as the participant's online search history was obtained through Google Takeout. Online search histories could be obtained using alternative search engines however Google was used for this study as it is the most common search engine in the UK. Informed written consent was obtained by J.B. or L.B.E to complete enrolment.

Within the National Health Service (NHS) women who attend their General Practitioner (primary care provider) and are felt to be high risk of a suspected cancer (based on symptom presentation) are referred to hospital (secondary care) for an urgent appointment within two weeks. Patients referred via this 'urgent cancer pathway' were eligible for recruitment into this study. Patients that consented to study participation completed a clinical questionnaire and extracted their Google takeout file.

Clinical outcomes (benign or malignant) for the study participants were extracted from the medical records, either clinical (ultrasound-based) or histological diagnosis, for those who underwent surgery. We analysed 24 months of Google takeout data prior to the GP referral date (see also below) and correlated it with the clinical outcomes (benign or malignant).

### Patient and public involvement

We involved patients in the study design, which informed changes to the questionnaire and terminology used in patient information sheets. It was a useful process to gain insight into patient opinion on using patient data to facilitate the earlier detection of disease.

### Online search data acquisition

Online search queries were extracted as a Google takeout file, which was shared with the research team via secure email. The Google takeout file was pseudo-anonymised. An automated health filter was applied to extract specific health-related queries for a 24-month period prior to the date of GP referral. The automated filter used a previously developed list of medical terms, comprising symptoms, disease, and drugs [23]. Only search queries which contained one or more of the keywords defined in the medical terms list remained in the filtered Google takeout file.

Note that since this filter considered each word in a medical phrase, such as "club foot", independently, the filtered output contained many, possibly irrelevant, queries, e.g., queries for "club". The manual filtering process excluded queries pertaining to pets (e.g., "cat bleeding pain") and to irrelevant medical conditions and body organs (e.g., "finger bleeding pain"). Queries that were repeated verbatim in consecutive searches were also eliminated.

Recognising the impact of the GP consultation on online search patterns, we evaluated online search data up to the day before GP referral i.e., excluded the date of GP referral from the analysis. Patients who were not referred by their GP i.e., presented to the emergency department or another specialty were included, but the date of consultation was used as a substitute for the GP referral date.

The key outputs included: (1) the time of first search query (2) the number of queries before filtering, and (3) the number of queries after the health filter was applied. Patients with Google takeout files (post filtering) that were empty were excluded from the analysis (Fig. 1).

### Clinical questionnaire data

A clinical questionnaire (Supplementary Fig. 1, Supplementary Table 1) was completed by patients to extract relevant clinical data, including demographics (age, BMI, ethnicity), medical and family history, and the nature of their symptoms at presentation to the clinic. Patients were asked if they had experienced up to 20 common gynaecological symptoms, their frequency, severity, and duration in the 12 months prior to presentation to the Gynaecology clinic. Relevant information was extracted from the medical records, including clinical (ultrasound-based) or histological diagnosis, for those who underwent surgery. Further details about staging and grade were available for malignant cases.

### Preliminary analysis of keyword categories

A list of health-related keywords was defined using known medical and colloquial terms, in English, French and Spanish, to allow inclusion of Google takeout file's which were in French, Spanish, or English. The keywords were categorised manually by J.F.B and S.S into 14 categories (not mutually exclusive) (bleeding, bloating, diagnostics, fatigue, gastrointestinal, gynaecological, menopause, nutrition, other conditions, pain, pelvic organs, symptoms, urinary, vagina or pelvic organs) and non-relevant keywords were removed (Supplementary Table 2). Preliminary analysis of the online search data (search terms and clinical data) was performed by grouping search queries into categories (categories model) or individual words/word pairs (that have been used at least five times) and analysed their appearance over time. The
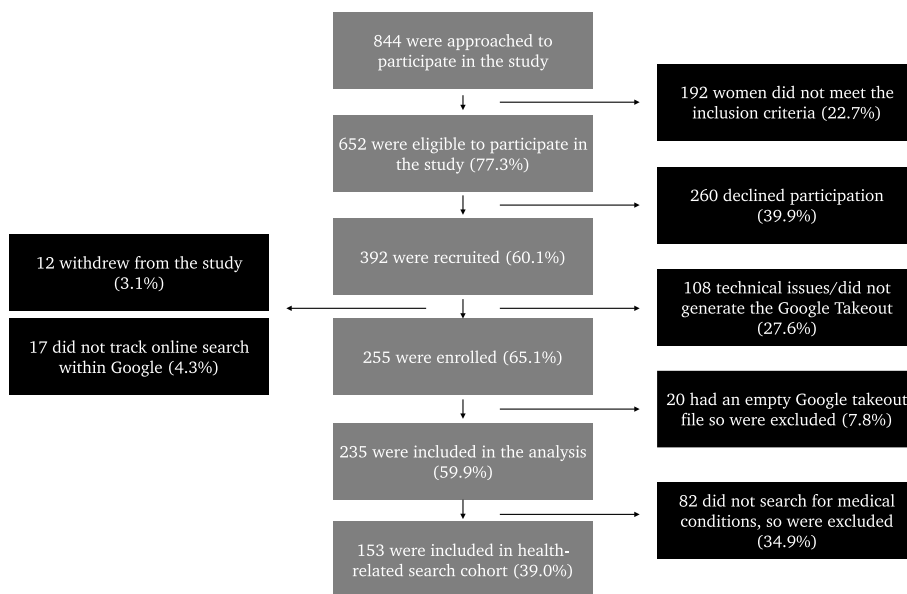
Barcroft *et al. BMC Public Health*  (2024) 24:608

Page 4 of 13



**Fig. 1** Summary of patient enrolment flowchart. The flowchart outlines the enrolment process for the study cohort ($n = 235$) and health-related search cohort ($n = 153$), from individuals referred to a London University Teaching Hospital with a suspected cancer between December 2020-June 2022. It outlines the reasons for incomplete enrolment and exclusion from the study

resulting lists of query categories, keywords, and most common queries for each condition are presented in Supplementary Table 2.

**Prediction of outcomes**
Patient outcomes (malignant or benign) were predicted using query data between $T_1$ and $T_2$ days prior to the GP referral date ($T_1 < T_2$), for patients that made online search queries during that time window. We considered all pairs of $T_1$ and $T_2$ between -700 days and -1 days to create different time windows. The input representation to the model was either based on categories or search terms:

1. Categories model: The number of search terms in each of the keyword categories, as described above.
2. Terms model: A vector-space model [26] of all words and word pairs (consecutive words) which were used by at least five patients. This representation is commonly used for text, especially when the terms with high information pertaining to differentiating between classes is unknown.

The categories model is considered to be based on expert knowledge, whereas the terms model is a data-driven approach including any terms used by more than five individuals within our cohort.

Predictor performance was superior in the vector-space model [26] and we therefore focus on this representation.

Patient representations were used as input to a gradient boosting model, considered one of the state-of-the-art models for structured data such as the vector-space representation [27, 28], with 50 weak learners. The models were evaluated using leave-one-out cross-validation [29] to reduce the likelihood of overfitting. Leave-one-out is appropriate due to the size of the dataset [30]. In line with previous work [25, 31–34], the performance measure reported is the Area Under the receiver operating Curve (AUC). The output of the model is a real value number in the range of 0–1, where zero indicates that the patient is unlikely to have cancer and 1 indicates that the patient is likely to have cancer.

**Effect of sample size on prediction accuracy**
The limited sample size and large number of features meant removing patients from the dataset could negatively impact prediction accuracy. To evaluate the effect of sample size on accuracy, we randomly selected subsets of patients (without replacement) and trained a model on each subset. This process was repeated five times for each subset size. A linear model was used to analyse the relationship between the area under the curve (AUC) and sample size.

**Removal of patients who do not query for health-related topics**
Not all internet users use Google for health information. It is hypothesized that prediction of health outcomes for

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 5 of 13

these users will be more challenging. Therefore, we investigated the effect on prediction accuracy of removing users who did not query health-related online searches.

For a given time period commencing at $T_1$ and ending at $T_2$, patients were removed if they did not mention any of 5521 medical conditions [23] in the first half of the period, i.e. from $T_1$ to $(T_1+(T_2-T_1)/2)$. We only considered the first half, as we hypothesized that filtering on the second half might remove more benign patients and bias our results. A model was trained and evaluated as discussed previously. AUC is dependent on sample size (for small datasets) [35, 36]. To compensate for the decreased training set size, we normalise the resulting AUC according to the regression parameters described in supplementary Fig. 2 (following the approach described in Floares et al. [37]) We refer to this value as the sample-corrected AUC.

### Symptoms questionnaire as outcome predictors

The overlap between a patient querying about a symptom and mentioning it in the symptom questionnaire was evaluated by first mapping a list of query terms to each symptom. This is listed in the Supplementary information. Then, for each symptom, a $2\times2$ contingency table was calculated enumerating the number of patients that indicated (did not indicate) the symptom in the questionnaire and the number of patients that searched (did not search) for the symptom. The association was evaluated through the chi-squared test.

An indication of whether each patient mentioned a symptom was used as input to a prediction model of patient outcome, with and without the vector-space features. Thus, we compared the performance of a predictive model, using search queries only, then questionnaire symptom data only, and finally search queries and symptom questionnaire data combined. By comparing the performance of these three approaches, we can determine whether search activity trends are a superior risk indicator compared to questionnaires, as well as assess the potential added value of modelling both information sources together. The prediction model (gradient boosting) was used for all models.

Note that the questionnaire-based approach is less sensitive to sample size (owing to its low dimensionality): A regression model of AUC as a function of sample size was not statistically significant ($P>0.05$). Therefore, the AUC of this model was not corrected for sample size.

### Comparison to a general population

To compare our results to a general population of UK-based internet users, we examined the queries of 1.8 million UK-based online users of the Microsoft Bing search engine. The Bing cohort consisted of users who searched for at least one keyword from the medical key word list (Supplementary Table 2) and made at least one search query each month between October 2021 and September 2022. Since this control group was anonymous, it included both males and females. Bing users who asked about gynaecological cancers ten or more times during the data period (October 2021-June 2022) or the three-month period immediately following (July–September 2022), were excluded from the analysis to safeguard against including people with a pre-existing or new gynaecological diagnosis [38]. The remaining Bing population was assumed not to have an active gynaecological cancer diagnosis. There are no known statistically significant differences in the demographics between Google and Bing users [39].

The control (Bing) group consisted of users who searched for health terms that could be relevant to gynaecological cancers but were unlikely to have a gynaecological cancer. A model with $T_1=-270$ and $T_2=-1$ (trained on data from all participants- benign and malignant cases) was applied to search queries made in the data period (9-month period, Oct 2021-June 2022) by users in the Bing control group. The end point of the data period (June 2022) in the Bing group represents the day before GP referral in the Gynaecology group.

### Ethical considerations

Institutional review board approval was granted in May 2020, by the North of Scotland ethics committee (REC approval 20/NS/0063). All patients signed informed consent to participate in the study. The filtered Google takeout files were pseudo-anonymised, and the original Google takeout file (non-filtered) was deleted. Data was processed in line with GDPR regulations. Permission was granted to utilise Bing data by the Microsoft Ethics Review Board (approval number 10532).

### Results

#### Google use among gynaecology patients and acceptability of online search data use

77.3% (652/844) of individuals approached to participate in this study had a Google account. Of those who met the study inclusion criteria, 60.1% consented to study participation (392/652). Complete enrollment (Google Takeout file and completion of questionnaire) was achieved in 65.1% of individuals (255/392). Of those who completed enrollment ($n=255$), 7.8% (20/255) were excluded due to insufficient online search data (i.e., no searches which passed a filter of health-related queries), resulting in a final dataset of 235 women (Fig. 1).

The 235 women in the cohort made 519,048 health-related queries (an average of 2208 searches per patient). The rate of malignancy was 26.0% (61/235),

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 6 of 13

with predominantly ovarian ($n=42$, 68.9%), followed by endometrial ($n=15$, 24.6%) cancer. The cohort consisted of pre-dominantly post-menopausal women ($n=136$, 57.9%) with a median age of 53 years old (20–81) (Table 1).

The reasons for incomplete enrolment ($n=137$, 34.9%) included: technical issues exporting the Google takeout file (27.6%, $n=108$), study withdrawal ($n=12$, 3.1%) and not tracking Google searches, so a takeout file ($n=17$, 4.3%) could not be generated (Fig. 1). The excluded cohort ($n=137$) had a median age of 51 years old (range: 20–89), which is comparable to the enrolled cohort. The excluded cohort had a rate of malignancy of 18.2% ($n=25$), with 52% ovarian ($n=13$), 40% endometrial ($n=10$) and 8% non-gynaecology cancers ($n=2$) respectively.

## Clinical symptoms present at different points in gynaecological Cancer

To evaluate the pattern of symptom presentation, the frequency of queries per week for each of the 14 categories in Supplementary Table 2 were evaluated and stratified by clinical outcome (malignant vs benign). Figure 2 shows the number of search queries within each category according to clinical outcomes. Gastrointestinal and pain-related symptoms presented up to 365 days before referral by the GP, whereas urinary and bleeding related symptoms presented later, at 140 days prior to GP referral. Around 70 days prior to GP referral, symptoms relating to bloating, gynaecological organs (vagina, pelvis) and menopause become more prevalent. The same pattern in symptom presentation was not seen within the benign group, thus suggesting a pattern that may be specific to malignancy.

Queries (represented as either search terms or categories) were evaluated for various time windows. The start time ($T_1$) was progressively increased from 30 to 700 days prior to the time of GP presentation. For each start time, the end time ($T_2$) was progressively decreased i.e., length of the window was increased (Fig. 3). The model's ability to differentiate between benign and malignant cases was evaluated based on the AUC. Across all time frames ($T_1$, $T_2$), the AUC for models using search terms representation was higher than models using categories representation by an average of 0.06 ($P < 10^{-5}$, sign test) (0.11 when only including models with an AUC greater than 0.55, $P < 10^{-5}$, sign test). Although both representations (search terms and categories) had comparable qualitative performance metrics (Fig. 3), we focused on the search terms representation given its superior quantitative performance.

Focusing on the 'search terms' model with a window starting 630 days before presentation to the GP, we observed the AUC surpass random decision (AUC 0.50) 360 days before the GP referral date ($T_2=360$) (AUC 0.64). This suggests there may be a difference between health-related queries in benign and malignant cases as early as a year prior to GP presentation. The closer to the GP referral date, the better the model performed, as demonstrated by an AUC of 0.74 for a time window up to 60 days in advance of GP referral ($T_1=630$, $T_2=60$) (Fig. 3).

The cohort contained some individuals ($n=82$, 34.9%) who did not query any of 5521 pre-defined health

**Table 1** Summary of demographics, clinical diagnoses and symptom presentation of individual's referred with suspected gyncaecological cancer

| Total *N* = 235 | Median, (range) | | |
|---|---|---|---|
| **Age** | 53 (20–81) | | |
| **BMI** | 25.9 (16.5–50.0) | | |
| | | **Number** | **%** |
| **Ethnicity** | White European | 141 | 60.0 |
| | Asian | 25 | 10.6 |
| | Black/African | 28 | 11.9 |
| | Hispanic | 6 | 2.6 |
| | Mixed | 9 | 3.8 |
| | Indian | 12 | 5.1 |
| | Arab | 8 | 3.4 |
| | Other | 6 | 2.6 |
| **Parity** | 0 | 88 | 37.5 |
| | 1 | 48 | 20.4 |
| | 2 | 58 | 24.7 |
| | 3 | 41 | 17.4 |
| **Menopausal status** | Pre | 99 | 42.1 |
| | Post | 136 | 57.9 |
| **Diagnosis** | Benign | 82 | 34.9 |
| | Benign (histological) | 92 | 39.1 |
| | Malignant | 61 | 26.0 |
| **Malignancy type** | Ovarian | 42 | 68.9 |
| | Endometrial | 15 | 24.6 |
| | Cervical | 1 | 1.6 |
| | Gastrointestinal | 3 | 4.9 |
| **Symptom presentation** | Pelvic pain | 150 | 63.8 |
| | Bloating | 119 | 50.6 |
| | Post-menopausal bleeding | 99 | 70.7 |
| | Dyspareunia | 62 | 26.4 |
| | Weight loss | 43 | 18.3 |
| | Constipation | 65 | 27.7 |
| | Appetite loss | 65 | 27.7 |
| | Diarrhoea | 44 | 18.7 |
| | Urgency | 63 | 26.8 |
| | Frequency | 91 | 38.7 |

Barcroft *et al. BMC Public Health* (2024) 24:608
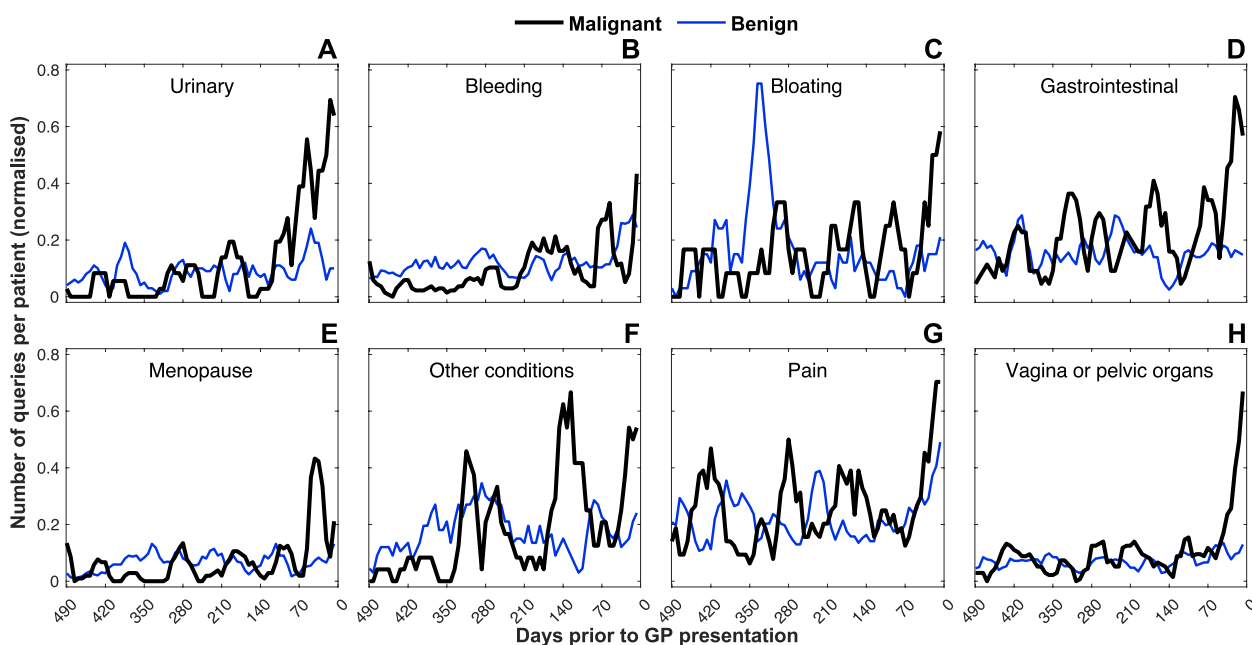
Page 7 of 13



**Fig. 2** The time series chart outlining the number of online search queries per discrete category within the study cohort. The time series chart outlines the number of online search queries made per patient, within each distinct symptom category: menopause, urinary, bleeding, bloating, gastrointestinal, vagina, pain etc. stratified by outcome (benign/malignant) up to 490 days in advance of GP referral. The time series are smoothed using a 4-week moving average. Online search activity can identify symptomatic individuals with gynaecological cancer at an earlier stage

conditions, i.e., they did not use online search for health purposes. These users most likely reduce the model's performance. After removing 82 users who did not query any health condition, the AUC for the remaining 153 users (malignant $n = 41$, 26.8% and benign $n = 112$, 73.2%) for the time window up to 60 days in advance of GP presentation ($T_1 = 630$ to $T_2 = 60$) reached a sample-size adjusted AUC of 0.82. The sample size adjustment accounts for the fact that the AUC declines with sample size (for small samples) and is described in detail in the supplementary information.

The most common symptom in the patient survey was pain ($n = 150$, 63.8%), followed by bloating ($n = 119$, 50.6%) Table 1, which is in accordance with the NICE guidance for the typical clinical presentation of ovarian cancer [40]. In post-menopausal individuals ($n = 136$), the most common symptom was post-menopausal bleeding ($n = 99$, 70.7%).

Interestingly, no correlation was noted between online search and clinical symptom patterns when a chi squared test ($p < 0.05$ with Bonferroni correction) was applied. The lack of correlation between online search and clinical symptoms suggests that online search data is not a mere online record of their clinical symptoms. Instead, it appears to represent other important health data, that is not reflected in the healthcare questionnaire. The difference may also be attributable to the fact that

questionnaire data, unlike online search data, relies upon an individual's recall ability.

To assess the value of symptom questionnaire data in the detection of gynaecological malignancy, a model utilising: (1) questionnaire symptom data only and (2) questionnaire and online search queries (combined) was developed. The questionnaire-only model performance reached an AUC (sample corrected) of 0.62, when combined with online search terms the AUC improved to 0.77 (sample corrected, $T_1 = 630$, $T_2 = 0$). The addition of questionnaire data to the model appears to slightly decrease the performance of the online search query-based model from an AUC (sample corrected) of 0.82 to 0.77 respectively.

**Comparison to control (Bing) population**

The distribution of model scores for users in the control (1.8 million UK-based online users of the Microsoft Bing search engine), benign, and malignant groups are depicted in Fig. 4. The scores for the benign and malignant populations are computed using leave-one-out cross-validation, whereas a model trained on all benign and malignant patients was used for computing the model scores of the control population. The control group closely mirrors the benign study population (Fig. 4). This supports the potential generalisability of
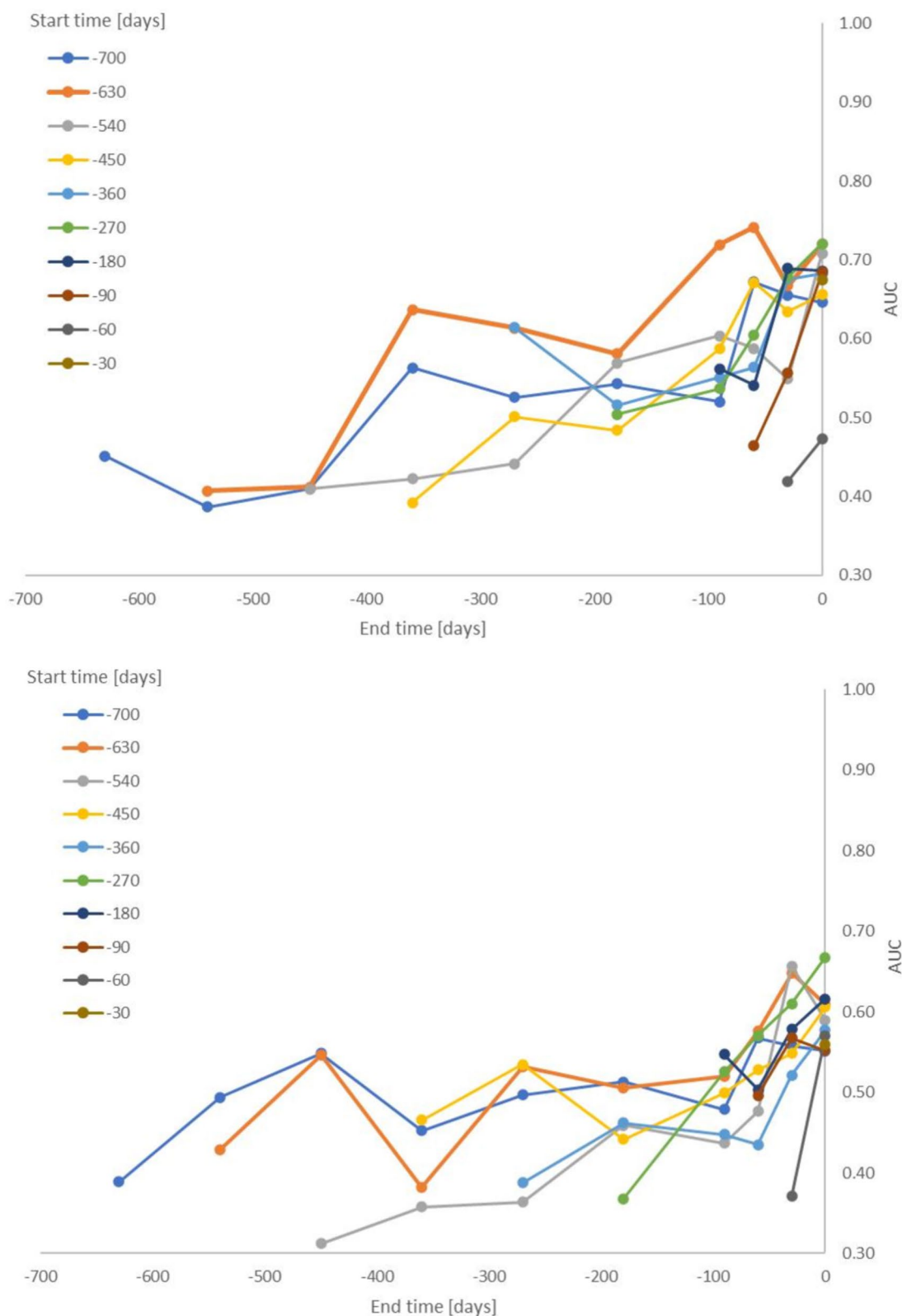
Barcroft *et al. BMC Public Health*  (2024) 24:608

Page 8 of 13



**Fig. 3** Model performance as a function of start and end times. The top figure shows the AUC for the terms model and the bottom figure for the categories model. The start and end times correspond to the duration of time in advance of GP referral date. Different lines correspond to different start ($T_1$) times and the dots on each line correspond to different end ($T_2$) times. Each dot represents the average of 10 runs. Standard deviation is equal, on average, to 0.01 (1.8% of the average AUC)
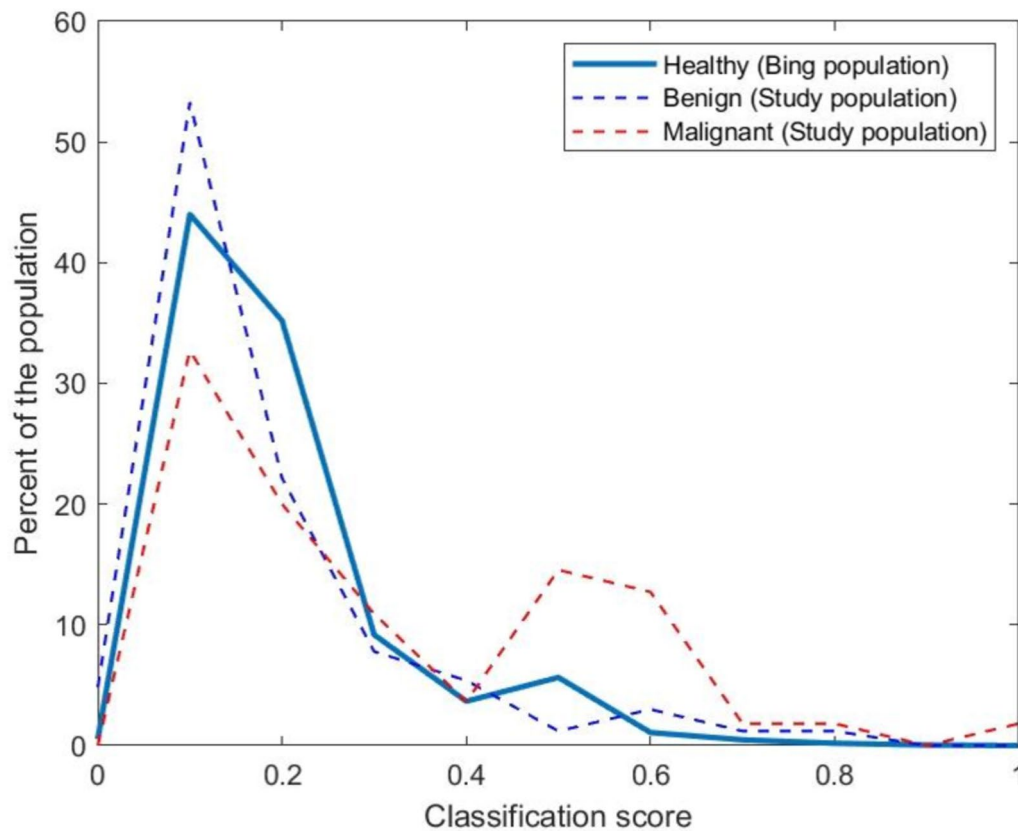
Barcroft *et al. BMC Public Health*  (2024) 24:608

Page 9 of 13



**Fig. 4** A histogram (10 bins) of model classification scores, when applied to our sample population (*n* = 235) and Bing users (*n* = 1.8 million). The histogram demonstrates the classification score for individual users. A high classification score indicates an increased likelihood of a malignant diagnosis. The Bing user population is distributed towards lower classification scores, in line with benign sample population and a lower likelihood of malignancy

the model and its ability to discriminate individuals with benign and malignant diagnoses.

## Discussion

This is the first clinical pilot study to demonstrate the feasibility of using online search histories in individuals with known gynaecological diagnoses, as a potential disease detection tool. Our study suggests that screening based on online search data may provide a signal of disease, up to 360 days prior to primary care (GP) referral with a suspected malignancy (AUC 0.64) and gradually improves closer to the GP referral date. The best performing model had a sample size-adjusted AUC of 0.82 in users (*n* = 153) who engaged in health-related searches, up to 60 days prior to GP referral. Furthermore, online search data provided insight into the presentation of gynaecological cancer, with an increased frequency and severity of urinary and gastrointestinal symptoms noted around 140 days and menopausal symptoms and pain at around 70 days in advance of GP referral. The presence of symptoms up to a year in advance of a diagnosis of ovarian cancer

is consistent with previous studies and challenges the concept that it is a 'silent killer' where most women are asymptomatic [41, 42]. This study is novel as it (1) evaluated online search data in individuals with known malignant and benign gynaecological conditions to ensure it is clinically robust and reproducible (2) used a symptomatic, benign cohort as a control group (3) and has a digital timeline of GP referral to diagnosis, to extract relevant time-specific data. Previous studies [25, 32–34] used proxy indicators such as experiential (or self-identifying) queries to identify individuals with the disease of interest, selected non-matched population controls and had no record of the diagnosis timeline.

We considered classification based on known health categories (categories model) and found that they were predictive. This is a validation of our data against medical knowledge, which is important. However, the model which used all search terms (terms model) outperformed the health categories model, which suggests that an individual's 'search terms' are likely to encode additional variables, that may be important in disease detection

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 10 of 13

and beyond our existing knowledge of disease. Furthermore, the lack of correlation between the questionnaire and the search term model reinforces the hypothesis that online search data is not a mere online representation of an individual's symptoms, but incorporates other variables (i.e., nutrition, health behaviour) that are relevant in the risk of disease. The difference may in part be attributable to an individual's ability to recall symptoms for up to 12 months in advance of disease presentation. Clinically, online search data may be a useful diagnostic adjunct in primary care to identify those at risk of disease and appropriately triage patients. Future work should focus on evaluating the performance of the search terms model in an independent population to understand its generalisability and potential clinical value as a diagnostic tool. Further validation will add to existing knowledge of clinical disease presentation and how it differs between benign and malignant conditions.

An individual's online search data is an example of a digital footprint i.e., information people knowingly or unknowingly generate when using electronic services, including mobile telephone data, social media posts, credit and loyalty card use. The recent CLOCS study [43] evaluated loyalty card purchases within two retailers in 153 women with ovarian cancer, compared to healthy controls, to determine if shopping habits could be used as proxy symptoms indicators, to facilitate the earlier detection of disease. Whilst an association between ovarian cancer and purchases of over-the-counter indigestion (AUC 0.65) and pain (AUC 0.63) medications were identified up to 13 months before diagnosis, the results must be interpreted with an element of caution, given the control group consisted of healthy population rather than symptomatic women with a benign condition. Furthermore, online search data, likely hold more promise than loyalty card purchase data due to its relative accessibility, breadth of topic coverage and its frequency of use [43, 44].

Systematic delays in the referral pathway need to be addressed in order to facilitate earlier detection of gynaecological cancer [7]. This study provides a unique insight into the disease trajectory of gynaecological cancer and its typical presentation, which is invaluable for improving disease detection at a patient and primary care level, particularly given GP's typically see an ovarian cancer case every five years [8, 45, 46]. Understanding the 'triggers' for accessing primary care is useful from a health promotion perspective, given we know existing cancer campaigns do not generate sustained behavioural changes [47, 48]. Screening programs identify 'at risk individuals' and feed them directly into specialist services, thus circumnavigating referral delays [10]. The best performing online search-based model reached a sample-corrected

AUC of 0.82, which is comparable to other established cancer screening programs in place to detect cervical (HPV, AUC: 0.87) and breast (Mammography AUC: 0.88) cancer [49, 50]. The test sensitivity is dependent on the operating point along the ROC. For example, the model with $T_1 = 0$, $T_2 = 270$ detects 36% of the positive cases at the cost of 8% false positives, while at another operating point, it can detect 62% of the positive cases at a cost of 38% false positives. The selection of the specific operating point is always a trade-off between the true positives and false positives and it depends on the specific clinical scenario where the model is used. For this reason, we focused on the AUC, which is an overall measure of performance, which takes all possible operating points into account.

This is the first study to evaluate online search data in individuals with known gynaecological conditions and linked symptom data. Substantial efforts were made to develop a generalisable model and reduce the risk of overfitting, through methodical leave one-out-cross validation and evaluation of the model's performance in an independent control population. The next step should be to test the model in an independent test set of symptomatic women with linked gynaecological diagnoses to evaluate its clinical value as a diagnostic tool.

The date of GP referral into secondary care was used as the last date of online search data, given it was available and did not rely upon patient recall. However, the GP referral date may not reflect previous presentations to the GP (for the same condition), so could have introduced a degree of bias to the search patterns. Future work should collaborate with primary care to use the first date of presentation to the GP, to control for this bias. Furthermore, the use of multiple Google accounts or private browsing may have affected the quality of an individual's online search data. Finally, the prevalence of malignancy within this cohort is not reflective of the UK population, given ovarian is less common (2%) than endometrial (2.78%) cancer [1, 3].

An online search-based model has potential as an accessible real-time screening tool, providing individualised risk profiles, which addresses barriers to screening uptake. We must acknowledge the physical and psychological morbidity and costs associated with a screening program triggering further investigations and treatment including surgery, particularly for those without the disease (false positive cases), when evaluating the value of an online search-based screening model within the health service [11, 49, 50]. There are several issues associated with using a model based on online search data. First, the digital divide may exacerbate health inequalities within disease screening programs [9]. Second, data anonymity and confidentiality are vital given the sensitive nature of

Barcroft *et al. BMC Public Health* (2024) 24:608

Page 11 of 13

online search data but could be addressed using cryptographic methods. Third, the psychological implications associated with receiving information suggesting a 'high risk of cancer'. Further research into behavioural psychology is required to better understand how to manage these issues before clinical integration can be considered.

Finally, we have shown that online search data may be able to identify individuals with gynaecology cancer at an earlier point than standard care, which is comparable to the findings from previous ovarian cancer screening trials (UKTOCs and PLCO) [12, 51]. Whether earlier detection of disease translates to improved clinical outcomes, i.e., mortality benefit, needs to be evaluated in a sufficiently powered clinical study, with adequate malignant cases to understand its clinical value as a diagnostic support tool. Furthermore, validation studies will contribute to existing knowledge about clinical disease presentation, thus supporting the discrimination between malignant and benign conditions.

## Conclusion

This is the first study to demonstrate the potential role of online search data in facilitating the earlier detection of clinically confirmed disease, specifically, though not limited to gynaecological cancer. Predictive performance varied depending on whether categorical or uncategorized 'search terms' were used. The best search-terms based model had a comparable performance to established disease screening programs [49, 50]. However, further research is required to evaluate the performance of the online search model within a larger cohort. Our results demonstrate the feasibility and acceptability of utilising online search data for health screening, which highlights its potential application in other diseases. To further evaluate the diagnostic capability of online search data in the earlier detection of disease, we aim to do a multi-centre study, to improve the overall performance and generalisability of the model to the general population, thus supporting its translation into clinical practice.

## Abbreviations

| | |
|---|---|
| OC | Ovarian cancer |
| EC | Endometrial cancer |
| OSD | Online search data |
| GP | General practitioner /primary care physician |
| AUC | Area under the ROC curve |
| UK | United Kingdom |
| UKCTOCS | UK collaborative trial of ovarian cancer screening |
| CA-125 | Cancer antigen-125 |
| NICE | National institute for clinical excellence |
| ROCkeTS | Refining ovarian cancer test accuracy scores |
| GT | Google takeout |
| NHS | National health service |
| BMI | Body mass index |
| REC | Regional ethics committee |
| GDPR | General data protection regulation |
| T1 | Start time |

| | |
|---|---|
| T2 | End time |
| CLOCS study | Cancer loyalty card study |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-17673-0.

---

**Additional file 1: Supplementary Figure 1.** Clinical questionnaire (Page 1-3) to extract clinical data, including symptom presentation, medical, family history and social history. **Supplementary Table 1**. Outlines the specific questionnaire symptoms that link to defined keywords identified in the online search query data. **Supplementary Table 2**. Outlines the list of online search query keyword categories, the number of queries containing the specific keywords and the three most common queries in each category in English, Spanish, and French. **Supplementary Figure 2.** Outlines the dependence of model AUC on sample size. The dotted line is a linear regression curve whose parameters are shown in the figure. This regression curve was used to assess the sample-size-adjusted AUC.

---

### Authors' contributions

S.S, I.J.C, V.L, E.Y-T, J.B and T.B were responsible for the original study design, drafting and protocol revision. J.B. and L.E were responsible for patient recruitment and clinical data extraction. D.G was responsible for the development and revision of software to extract specific online search data. E.Y-T was responsible for data science and development of machine learning models. E.Y-T, V.P.L, I.J.C and V.L were responsible for data extraction and statistical analysis aspect of the manuscript. J.B, I.J.C. S.S, E.Y-T, V.P.L, V.L, were involved in writing and preparation of the manuscript, including tables and figures. I.J.C, E.Y-T, V.L, V.P.L (focus: statistical analysis) provided technical input into the development of machine learning models and addressed statistical-related issues within the manuscript. J.B, L.E, T.B and S.S (focus: Gynaecology) provided valuable input into study design and addressed gynaecology-related issues within the manuscript. I.J.C, E.Y-T, V.L, V.P.L, D.G, J.B, L.E, T.B and S.S were all involved in revision of the manuscript and approved the final version for submission. S.S is the guarantor for this paper and accepts full responsibility for the work and/or the conduct of the study. SS affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

### Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available as patient permission for open access was not obtained but are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

Institutional review board approval was granted in May 2020, by the North of Scotland ethics committee (REC approval 20/NS/0063). All patients signed informed consent to participate in the study. The filtered Google takeout files were pseudo-anonymised, and the original Google takeout file (non-filtered)

Barcroft *et al. BMC Public Health*  (2024) 24:608

Page 12 of 13

was deleted. Data was processed in line with GDPR regulations. Permission was granted to utilise Bing data by the Microsoft Ethics Review Board (approval number 10532).

### Consent for publication
Not applicable.

### Competing interests

### Author details
[1]Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK. [2]Microsoft Research, Hoshaya, Israel. [3]Department of Computer Science, University College London, London, UK. [4]Computer Science, University of Copenhagen, Copenhagen, Denmark.

### References

1. Cancer Research UK [Internet]. 2015 [cited 2023 Jan 27]. Ovarian cancer statistics. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer
2. Sundar S, Balega J, Crosbie E, Drake A, Edmondson R, Fotopoulou C, et al. BGCS uterine cancer guidelines: Recommendations for practice. Eur J Obstet Gynecol Reprod Biol. 2017;213:71–97.
3. Cancer Research UK [Internet]. 2015 [cited 2023 Mar 1]. Uterine cancer statistics. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer
4. Henson KE. National Disease Registration Service: Case-mix adjusted percentage of cancers diagnosed at stages 1 and 2 in England, by Clinical Commissioning Group. 2020 [cited 2023 Feb 21]. National Disease Registration Service: Case-mix adjusted percentage of cancers diagnosed at stages 1 and 2 in England, by Clinical Commissioning Group. Available from: https://www.gov.uk/government/statistics/case-mix-adjusted-percentage-cancers-diagnosed-at-stages-1-and-2-by-ccg-in-england/national-disease-registration-service-case-mix-adjusted-percentage-of-cancers-diagnosed-at-stages-1-and-2-in-england-by-clinical-commissioning-group
5. Rennison R. Pathfinder England: Transforming futures for women with ovarian cancer. [cited 2023 Feb 21]. Pathfinder England: Transforming futures for women with ovarian cancer. Available from: https://targetovariancancer.org.uk/sites/default/files/2020-07/Pathfinder%202016%20-%20England%20report.pdf
6. Rose PW, Rubin G, Perera-Salazar R, Almberg SS, Barisic A, Dawes M, et al. Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. BMJ Open. 2015;5(5):e007212–e007212.
7. Rampes S, Choy SP. Early diagnosis of symptomatic ovarian cancer in primary care in the UK: opportunities and challenges. Prim Health Care Res Dev. 2022;23: e52.
8. Mendonca S, Abel G, Lyratzopoulos G. Pre-referral GP consultations in patients subsequently diagnosed with rarer cancers: a study of patient-reported data. Br J Gen Pract. 2016;66(644):e171–81.
9. Tanton C, Soldan K, Beddows S, Mercer CH, Waller J, Field N, et al. High-Risk Human Papillomavirus (HPV) Infection and Cervical Cancer Prevention in Britain: Evidence of Differential Uptake of Interventions from a Probability Survey. Cancer Epidemiol Biomark Prev. 2015;24(5):842–53.
10. Landy R, Pesola F, Castañón A, Sasieni P. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case–control study. Br J Cancer. 2016;115(9):1140–6.
11. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, et al. Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. The Lancet. 2021;397(10290):2182–93.
12. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. JAMA. 2011;305(22):2295–303.
13. Sundar S, Rick C, Dowling F, Au P, Snell K, Rai N, et al. Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. BMJ Open. 2016;6(8): e010333.
14. Breijer MC, Peeters J a. H, Opmeer BC, Clark TJ, Verheijen RHM, Mol BWJ, et al. Capacity of endometrial thickness measurement to diagnose endometrial carcinoma in asymptomatic postmenopausal women: a systematic review and meta-analysis. Ultrasound Obstetrics Gynecol. 2012;40(6):621–9.
15. Sundar S, Balega J, Crosbie E, Drake A, Edmondson R, Fotopoulou C, et al. BGCS uterine cancer guidelines: Recommendations for practice. Eur J Obstetrics Gynecol Reprod Biol. 2017;1(213):71–97.
16. One in two EU citizens look for health information online [Internet]. 2021 [cited 2023 Jan 27]. Available from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20210406-1
17. Countries with the highest internet penetration rate 2023 | Statista [Internet]. [cited 2023 Feb 20]. Available from: https://www.statista.com/statistics/227082/countries-with-the-highest-internet-penetration-rate/
18. Leading search engines ranked by market share UK 2021 | Statista [Internet]. [cited 2023 Mar 5]. Available from: https://www.statista.com/statistics/280269/market-share-held-by-search-engines-in-the-united-kingdom/
19. Statista [Internet]. 2022 [cited 2023 Jan 27]. Online search usage. Available from: https://www.statista.com/topics/1710/search-engine-usage/
20. Dr Google will see you now: Search giant wants to cash in on your medical queries [Internet]. 2019 [cited 2023 Jan 27]. Available from: https://www.telegraph.co.uk/technology/2019/03/10/google-sifting-one-billion-health-questions-day/
21. Yom-Tov E, Lampos V, Inns T, Cox IJ, Edelstein M. Providing early indication of regional anomalies in COVID-19 case counts in England using search engine queries. Sci Rep. 2022;12(1):2373.
22. Lampos V, Majumder MS, Yom-Tov E, Edelstein M, Moura S, Hamada Y, et al. Tracking COVID-19 using online search. npj Digit Med. 2021;4(1):17.
23. Yom-Tov E, Borsa D, Hayward AC, McKendry RA, Cox IJ. Automatic Identification of Web-Based Risk Markers for Health Events. J Med Internet Res. 2015;17(1): e29.
24. White RW, Horvitz E. Evaluation of the Feasibility of Screening Patients for Early Signs of Lung Carcinoma in Web Search Logs. JAMA Oncol. 2017;3(3):398.
25. Paparrizos J, White RW, Horvitz E. Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results. JOP. 2016;12(8):737–44.
26. Vijayan VK, Bindu KR, Parameswaran L. A comprehensive study of text classification algorithms. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) [Internet]. Udupi: IEEE; 2017 [cited 2023 Mar 1]. p. 1109–13. Available from: http://ieeexplore.ieee.org/document/8125990/
27. Wainberg M, Alipanahi B, Frey BJ. Are Random Forests Truly the Best Classifiers? J Mach Learn Res. 2016;17(110):1–5.
28. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? J Mach Learn Res. 2014;15(90):3133–81.
29. Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley; 2001. p. 654.
30. Bishop CM. Pattern Recognition and Machine Learning [Internet]. 2006 [cited 2023 Jun 21]. Available from: https://link.springer.com/book/9780387310732
31. Shaklai S, Gilad-Bachrach R, Yom-Tov E, Stern N. Detecting Impending Stroke From Cognitive Traits Evident in Internet Searches: Analysis of Archival Data. J Med Internet Res. 2021;23(5): e27084.
32. Youngmann B, Allerhand L, Paltiel O, Yom-Tov E, Arkadir D. A machine learning algorithm successfully screens for Parkinson's in web users. Ann Clin Transl Neurol. 2019;6(12):2503–9.
33. Hochberg I, Daoud D, Shehadeh N, Yom-Tov E. Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. Acta Diabetol [Internet]. 2019 May 15 [cited 2019 May

Barcroft *et al. BMC Public Health*  (2024) 24:608

Page 13 of 13

23]; Available from: http://link.springer.com/https://doi.org/10.1007/s00592-019-01350-5

34. Soldaini L, Yom-Tov E. Inferring Individual Attributes from Search Engine Queries and Auxiliary Information. In: Proceedings of the 26th International Conference on World Wide Web - WWW '17 [Internet]. Perth, Australia: ACM Press; 2017 [cited 2019 May 23]. p. 293–301. Available from: http://dl.acm.org/citation.cfm?doid=3038912.3052629

35. Al-Mekhlafi A, Becker T, Klawonn F. Sample size and performance estimation for biomarker combinations based on pilot studies with small sample sizes. Communications in Statistics - Theory and Methods. 2022;51(16):5534–48.

36. Tanaka K, Nakada T aki, Takahashi N, Dozono T, Yoshimura Y, Yokota H, et al. Superiority of Supervised Machine Learning on Reading Chest X-Rays in Intensive Care Units. Frontiers in Medicine [Internet]. 2021 [cited 2023 Jun 21];8. Available from: https://www.frontiersin.org/articles/https://doi.org/10.3389/fmed.2021.676277

37. Floares AG, Ferisgan M, Onita D, Ciuparu, Andrei, Calin, George A., Manolache, Florin B. The Smallest Sample Size for the Desired Diagnosis Accuracy. Int J Oncol Cancer Therapy. 2017;2:13–9.

38. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of Information-Seeking for Cancer on the Internet: An Analysis of Real World Data. Holme P, editor. PLoS ONE. 2012;7(9):e45921.

39. Rosenblum S, Yom-Tov E. Seeking web-based information about attention deficit hyperactivity disorder: where, what, and when. J Med Internet Res. 2017;19(4): e126.

40. Ovarian cancer: Recognition and initial management [Internet]. NICE; 2011 [cited 2023 Mar 31]. Available from: https://www.nice.org.uk/guidance/cg122

41. Goff BA. Frequency of Symptoms of Ovarian Cancer in Women Presenting to Primary Care Clinics. JAMA. 2004;291(22):2705.

42. Olson S. Symptoms of ovarian cancer. Obstet Gynecol. 2001;98(2):212–7.

43. Brewer HR, Hirst Y, Chadeau-Hyam M, Johnson E, Sundar S, Flanagan JM. Association Between Purchase of Over-the-Counter Medications and Ovarian Cancer Diagnosis in the Cancer Loyalty Card Study (CLOCS): Observational Case-Control Study. JMIR Public Health Surveill. 2023;26(9): e41762.

44. Brewer HR, Hirst Y, Sundar S, Chadeau-Hyam M, Flanagan JM. Cancer Loyalty Card Study (CLOCS): protocol for an observational case–control study focusing on the patient interval in ovarian cancer diagnosis. BMJ Open. 2020;10(9): e037459.

45. Rampes S, Choy SP. Early diagnosis of symptomatic ovarian cancer in primary care in the UK: opportunities and challenges. Prim Health Care Res Dev. 2022;23: e52.

46. Rose PW, Rubin G, Perera-Salazar R, Almberg SS, Barisic A, Dawes M, et al. Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. BMJ Open. 2015;5(5): e007212.

47. The Angelina Jolie effect – Impact on breast and ovarian cancer prevention A systematic review of effects after the public announcement in May 2013 - Gianmarco Troiano, Nicola Nante, Mauro Cozzolino, 2017 [Internet]. [cited 2023 Sep 2]. Available from: https://journals.sagepub.com/doi/abs/https://doi.org/10.1177/0017896917712300

48. Cohen SA, Cohen LE, Tijerina JD. The impact of monthly campaigns and other high-profile media coverage on public interest in 13 malignancies: a Google Trends analysis. Ecancermedicalscience. 2020;14:1154.

49. Wu Z, Li T, Han Y, Jiang M, Yu Y, Xu H, et al. Development of models for cervical cancer screening: construction in a cross-sectional population and validation in two screening cohorts in China. BMC Med. 2021;19(1):197.

50. Yang L, Wang S, Zhang L, Sheng C, Song F, Wang P, et al. Performance of ultrasonography screening for breast cancer: a systematic review and meta-analysis. BMC Cancer. 2020;20(1):499.

51. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, et al. Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet. 2021;397(10290):2182–93.

## Publisher's Note