BMJ Open

Google search histories of patients presenting to an emergency department: an observational study

Jeremy M Asch,[1,2] David A Asch,[2,3] Elissa V Klinger,[2] Justine Marks,[1,2] Norah Sadek,[1] Raina M Merchant[1,2]

¹Department of Emergency Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
²Center for Digital Health, Penn Medicine, Philadelphia, Pennsylvania, USA
³The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania, USA

Correspondence to
Jeremy M Asch;
jeremy.asch@uphs.upenn.edu

ABSTRACT

Objective To test patients' willingness to share and link their prior Google search histories with data from their electronic medical record (EMR), and to explore associations between search histories and clinical conditions.

Design Cross-sectional study of emergency department (ED) patients from 2016 to 2017.

Setting Academic medical centre ED.

Participants A total of 703 patients were approached; 334 of a volunteer sample of 411 (81%) reported having a Google account; 165 of those (49%) consented to share their Google search histories and EMR data; 119 (72%) were able to do so. 16 (13%) of those 119 patients had no data and were not included in the final count. Patients under the age of 18 or with a triage level of 1 were considered ineligible and were not approached.

Main outcome measures Health relatedness of searches in the remote past and within 7 days of the ED visit, and associations between patients' clinical and demographic characteristics and their internet search volume and search content.

Results The 103 participants yielded 591 421 unique search queries; 37 469 (6%) were health related. In the 7 days prior to an ED visit, the percentage of health-related searches was 15%. During that time, 56% of patients searched for symptoms, 53% for information about a hospital and 23% about the treatment or management of a disease. 53% of participants who used Google in the week leading up to their ED visit searched for content directly related to their chief complaint.

Conclusions Patients were willing to allow researchers simultaneous access to their Google search histories and their EMR data. The change in volume and content of search activity prior to an ED visit suggests opportunities to anticipate and improve health care utilisation in advance of ED visits.

Digital media capture and document an increasing segment of our personal lives in the tracks left from online or in-store purchases, wearable devices or engagement with social media. Many of these digital traces reflect health. Facebook, Twitter and Instagram posts can reveal health-related behaviours, symptoms or diagnoses.[1–3] But while these social media posts reflect what

Strengths and limitations of this study

► By combining search histories with the patients' health record, we were able to create an individual-level correlation between online queries and participant health conditions.
► Using retrospective data from patients, we eliminated various forms of respondent bias, providing a clearer picture of what people search for prior to visiting an emergency department.
► This is a novel data set which offers promise to predict and anticipate emergency department visits and perhaps hospitalisations; however, its clinical value remains uncertain.
► The search database we collected is incomplete. We do not know what patients may have searched for using other media (eg, Bing or a friend's device) or using private browsing.
► This study was performed in an emergency department in an urban environment. That along with the young population who consented may make the results less applicable to other settings.

people want others to see, internet search histories reflect something potentially more revealing, which is what people want to know. Google processes approximately 100 billion searches per month.[4] More than 72% of US adults report using the internet for health-related questions (How is Zika transmitted?), symptoms (itchy rash on elbow) or behavioural intentions (quitting smoking).[1 5] Because posts are largely public and directed outward and searches are largely private and directed inward, posts and searches might reveal different insights about health.

The Google Flu Trends project, launched in 2008, aimed to link influenza outbreaks to searches of flu-related symptoms.[6] Unfortunately, these reports often overestimated regional influenza burden.[7 8] Other work has demonstrated association between recipes people search for and what they actually consume.[9] These investigations associate

aggregated search data with community-level outcomes and so they remain ecological.

There is enormous promise in moving to individual-level associations between searches and health, both for public health surveillance and perhaps for the good of individual patients. Deriving individual-level associations is logistically harder, because individual search histories and health conditions are private and require individual-level consent for observation.

To test this promise, we aimed to answer three questions among patients presenting to a large academic health system: (1) would patients share their search data so that it could be analysed in association with their electronic health records, (2) is there variation in health-related searches leading up to an emergency department (ED) visit and (3) what do patients search for?

METHODS
Participants
From March 2016 to March 2017, research assistants (RAs) approached patients seeking acute care in the ED of a large urban academic health system. We aimed to recruit at least 100 patients, a number we estimated would be sufficient to judge the feasibility of the approach and orient future directions. Patients were excluded if younger than 18 years, in police custody, triage class 1 (eg, gunshot wounds). Eligible patients were approached while they were a patient at the ED and asked if they were interested in participating in research and if they had a Google account. They were informed that only past searches would be retrieved, that they could review their Google search data before sharing it with the research team, and that the team could not access searches made following recruitment. Participants were eligible for a lottery drawing with a 2% chance of receiving a US$40 gift card. All patients provided written consent to be considered eligible.

Consenting participants self-reported demographic information. Consenting participants' electronic medical record (EMR) data were extracted, including the chief complaints on arriving to the ED and discharge diagnoses. A sample of patients visiting the same ED (n=61 841) during the same period of recruitment and meeting the same exclusion criteria was used to compare the characteristics of participants sharing data with the overall ED population.

Extracting search data
Using a laptop, RAs navigated to takeout.google.com where participants were asked to download zip files of their Google usage data. The downloaded search files contain time-stamped reports of what users typed into their Google search bar, independent of the browser used. Of note, many users stay logged into Google because they use Gmail or other Google applications that are facilitated through log in. Searches made when not logged are not included in the takeout file. Takeout files do not show

what appeared on the page of search results, nor what the user may have clicked on within those pages.

Consenting participants able to download their takeout files received these files in their email account from Google. Participants were then able to view their own data and determine if they wanted to share the file with the research team.

Coding search data
Searches conducted after the time of ED triage were excluded. Two RAs (NS, JM) independently coded each search term as health related or not. Disagreements were adjudicated by a third reviewer (JMA). Searches occurring within 7 days before their recruitment visit to the ED were further coded into more granular themes. Two physicians independently coded the health-related searches alongside the patients' chief complaint to determine if searches were related to the clinical presentation at the ED or to logistic aspects of the visit, such as the address of the ED, adjudicating differences with a third reviewer (JMA).

Analysis
Summary statistics were used to characterise sociodemographic data, EMR data (eg, chief complaint, discharge diagnoses) and search queries (time, date). STATA V.14 was used for summary statistical analysis.

Patient and public involvement
No patients were involved in the study design nor in the recruitment plans for this research. There are no plans for dissemination of the outcomes of this research to the patients.

RESULTS
Participants
Of 703 patients approached in the ED, 411 (58%) agreed to participate in research of whom 334 (81%) reported having a Google account; 165 of those patients (49%) consented to share access to all EMR data in our health system and all prior Google searches. The final cohort included 103 participants; none censored any of their search history data (figure 1).

Table 1 reports characteristics of the participants and study-eligible patients seeking care in the ED during the study period. Study participants were younger and more likely to be women. The most common ED chief complaints of these patients were gastrointestinal 44 (37%), obstetrics and gynaecological 20 (17%), and neurological 17 (14%).

Search histories
The 103 participants had 591 421 unique search queries (mean 5742; range 2–51 751). Inter-rater agreement between the coders was high (kappa=0.83); 6% were health related. Eighty-six participants made searches in the 7 days prior to the ED visit (3978 searches; mean 33; range 1–220); 15% were health related. Among these
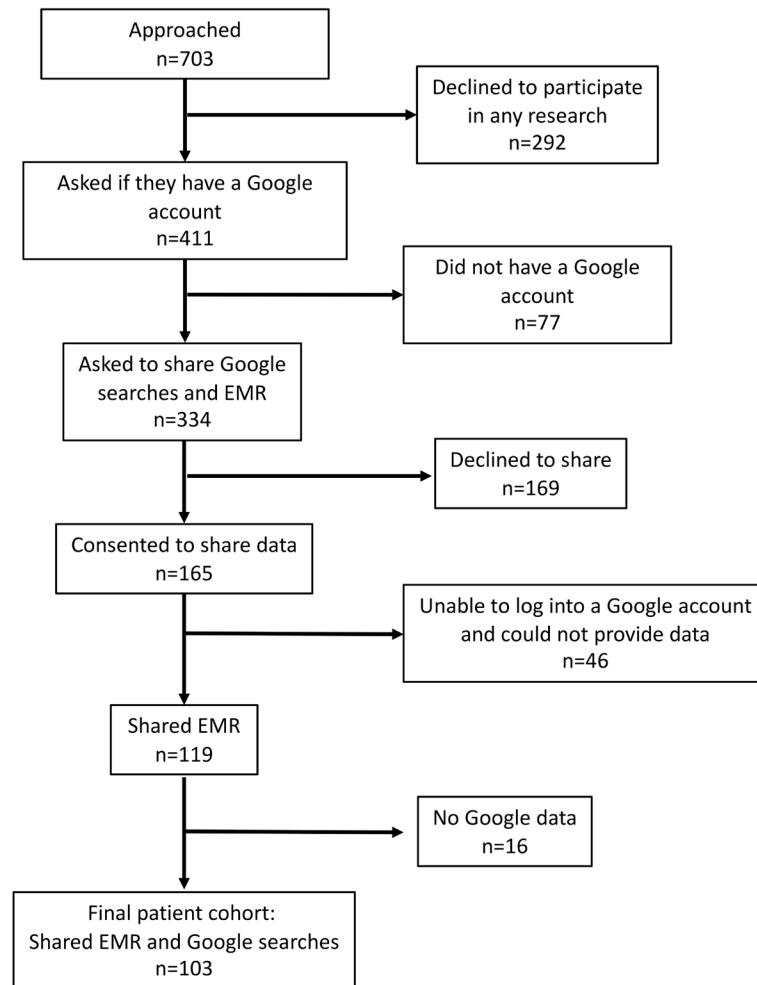
Figure 1 Study enrolment: this figure details inclusion and exclusion criteria applied to identify the final patient cohort. EMR, electronic medical record.

86 patients, 54 (63%) searched proportionately more for health-related topics in the 7 days prior than in the earlier period; 46 (53%) searched for information clinically related to their chief complaint (eg, 'How to relieve sinus pressure' with a chief complaint of 'Headache') (kappa=0.75); and 13 (15%) searched for directions or other logistic information about the ED or other healthcare facilities. Most searches prior to an ED visit reflected non-health-related topics (table 2); for example, 67% of participants searched for something related to entertainment.

DISCUSSION

This study has three main findings. First, 23% of patients approached in a large ED, and 49% of willing research participants with Google accounts were willing to share access to their online search data and allow those data to be linked, for research purposes, with the clinical information in their EMR. While several investigations link internet search data with health and healthcare at the population level, we believe this is the first effort to

link internet search histories and EMR at the level of individual patients. The high rates of participation may be surprising against a backdrop of expressed concern for privacy in internet settings and in healthcare settings after the Equifax breaches and issues regarding Facebook and Google's handling of user data.[10] In prior work, we demonstrated that 27% of patients approached and 71% of patients eligible were willing to share access to their social media (eg, Facebook) and allow their posts to be linked with their EMR data.[2] Search history data seem potentially more private, because while social media posts reflect what people want to project about themselves, search data reflect what presumably they want to know. For that reason, search data may be particularly revealing.

Second, many of the participants' searches were health related, suggesting an opportunity to better understand patients' knowledge, attitudes and behaviours about health—an understanding that might be hard to gain other ways. A report from the Pew Research Center found that only 53% of 'online diagnosers' talked with a clinician about the health information they were searching

Table 1	Patient characteristics	
	ED demographics during study duration (n=61 841)	**Participants sharing Google and EMR data (n=103)**
	n (%)	**n (%)**
Age group		
18–30	20 504 (33)	74 (72)
31–40	10 204 (17)	24 (23)
41–50	8537 (14)	4 (4)
51–60	9682 (16)	1 (1)
>60	12 914 (21)	0 (0)
Gender		
Female	37 036 (60)	76 (73)
Race		
White	18 634 (30)	33 (32)
Black	36 959 (60)	56 (54)
Asian	2242 (4)	2 (3)
Other	2275 (4)	3 (3)
Ethnicity		
Hispanic/Latino	2471 (4)	10 (8)
Chief complaint*		
Gastrointestinal		41 (37)
Obstetrics and gynaecology		17 (17)
Neurological		16 (14)
Respiratory		11 (11)
Cardiovascular		11 (10)
Genitourinary		11 (9)
Musculoskeletal		10 (8)
ENT		8 (8)
Infection		7 (6)
Injury		3 (3)
Other		6 (5)
Search query quantity		
≥5000		31 (26)
1000–4999		39 (33)
1–999		33 (28)

*Percentages sum to more than 100% because some patients have multiple diagnoses.
ED, emergency department; EMR, electronic medical record; ENT, ear, nose and throat.

for.[1] These searches also reveal gaps in traditional health communication. One participant searched 'how big is a walnut', followed by 'what is a fibrous tumour?' A review of the EMR revealed the patient had been told of a walnut-sized fibrous tumour at a previous hospital encounter, but the later search histories suggest the message may not

have been understood on its own. Search histories reveal otherwise invisible gaps in patients' understanding.

Third, health-related searches doubled prior to an ED visit, suggesting an opportunity to better understand patients' concerns before seeking care in an acute care setting. More specifically, in the week before visiting the ED, 15% of participants searched for location or logistical information and 53% searched for clinical information relevant to their visit. Participants often searched for health-related topics multiple times prior to making the decision to visit the hospital. These findings suggest an ability to anticipate demand even for patients visiting the ED. In retail contexts, search terms currently direct targeted advertisements, taking advantage of anticipated demand. One can imagine predicting the demand of hospital services in the same way advertisers predict sales.[11] Search terms predictive enough of serious illness to suggest emergency care or predictive enough of benign illness to suggest avoiding such care. By knowing what patients search for prior to a hospitalisation, we can gain a better understanding of how to respond to what matters most to patients. Although the stakes and costs of false positive and false negative errors would be considerably higher than with misdirected advertising, Google already provides information about suicide prevention services when search terms suggesting that suicidal intent is entered.[12]

This study has several limitations. First, our findings reflect a convenience sample of patients seeking care in one urban academic ED and may not be representative of other patient populations. Our sample was young and predominantly female, and not representative of all patients presenting to that ED who may have different health conditions and search patterns. Second, while Google search queries provide a detailed view of questions patients ask about their health, they capture only those searches performed when patients are logged into their Google account. Due to the autofill function of Google, some searches may have been influenced by the auto population of terms that a person may not have intended to search for. They may also reflect the searches of other individuals sharing the same device. Finally, search histories do not reveal what the participant was shown after the query nor what the participant may have selected from the choices shown. Originally when the study began, the Google Takeout platform did not discriminate between users with no data from those who had searches available, which lead to 16 patients being recruited who were eventually deemed ineligible. Google has changed their Takeout tool to no longer allow the download of null data, which would reduce the likelihood of recruiting patients with no available data in future studies.

We believe this is the first study to associate internet search data with EMR data at the level of the individual. Several studies have used searches as a proxy to indicate patient health, or have had patients self-report their internet searches.[13] The kind of dose–response relationships we observe between the content and frequency of

Table 2 Google Searches in the 7 days prior to an emergency department visit

Query subject	Example search	# of participants where search occurred at least once. n=86 n (%)	Frequency of query n=3978 n (%)
Entertainment	'Bon iver concert' 'xbox one'	58 (67)	726 (18)
Directions to non-clinical locations	Current location ->non-clinical location	55 (64)	449 (11)
Retail (online shopping, physical location of stores)	'Amazon' 'Walmart'	47 (55)	279 (7)
Information about hospital/care	'Penn Medicine ER' 'medical tower philly'	46 (53)	240 (6)
Symptoms and disease info	'What is diabetes' 'cramping 12 weeks agony'	48 (56)	223 (6)
Unrelated/unintelligible	'asj' '344'	42 (49)	198 (5)
Social media	'Facebook'	39 (45)	167 (4)
Food/drink	'How long to cook pizza'	44 (51)	153 (4)
Financial questions	'How to reach a person at the irs'	35 (41)	116 (3)
Treatment and disease management	'cure food poisoning' 'nosebleed treatment'	20 (23)	94 (2)
Pornographic materials	'Pornhub' 'xvideos'	16 (19)	78 (2)
Legal questions	'What are the traffic laws if there is no stop sign'	18 (21)	77 (2)
Drug use	'How many mgs of Advil to take' 'Vicoden vs percocet '	20 (23)	75 (2)
Diet/exercise	'Is it possible to lose 80lbs in a year' 'crossfit gym'	11 (13)	41 (1)
Medical accessories	'Syringe cvs' 'haemorrhoid pillows'	11 (13)	19 (0)
Other	'Why wont hot water heater stay lit'	80 (93)	1043 (31)

health-related searches and participants' clinical events supports their validity.

CONCLUSION

This study reveals patients' willingness to share access to their Google searches and EMR with researchers, offering opportunities to use this information to better predict healthcare use and understand health-related knowledge, attitudes and behaviours of more general populations. At a time when diagnosis increasingly occurs at the molecular level and when precision medicine aims to tailor treatments based on largely genetic characteristics, this study adds to our understanding of the health relevance of individuals' search histories and other digital resources. The Social Mediome may offer predictions about health and opportunities to target interventions as informative as what comes from the genome.[14–17]

REFERENCES

1. Fox S, Duggan M. Health Online 2013. 2013 http://www.pewinternet.org/2013/01/15/health-online-2013/
2. Padrez KA, Ungar L, Schwartz HA, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Qual Saf* 2016;25:414–23.
3. Smith RJ, Crutchley P, Schwartz HA, et al. Variations in Facebook Posting Patterns Across Validated Patient Health Conditions: A Prospective Cohort Study. *J Med Internet Res* 2017;19:e7.
4. Google Search Statistics. Internet Live Stats. 2017 http://www.internetlivestats.com/google-search-statistics/ (Accessed 11 Aug 2018).
5. Fox S. What ails America? Dr. Google can tell you. 2013 http://www.pewresearch.org/fact-tank/2013/12/17/what-ails-america-dr-google-can-tell-you/
6. Olson DR, Konty KJ, Paladini M, et al. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013;9:e1003256.
7. Lazer D, Kennedy R, King G, et al. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014;343:1203–5.
8. Butler D. When Google got flu wrong. *Nature* 2013;494:155–6.
9. West R, White RW, Horvitz E. *From cookies to cooks: insights on dietary patterns via analysis of web usage logs. Proceedings of the*

22nd international conference on World Wide Web. Rio de Janeiro, Brazil: ACM, 2013:1399–410.

10. Mahmoodi J, Čurdová J, Henking C, *et al*. Internet Users' Valuation of Enhanced Data Protection on Social Media: Which Aspects of Privacy Are Worth the Most? *Front Psychol* 2018;9:1516.

11. Libert T, Grande D, Asch DA. What web browsing reveals about your health. *BMJ* 2015;351:h5974.

12. Gillberti M. Learning more about clinical depression with the PHQ-9 questionnaire. 2017 https://www.blog.google/products/search/learning-more-about-clinical-depression-phq-9-questionnaire/

13. McCarthy DM, Scott GN, Courtney DM, *et al*. What Did You Google? Describing Online Health Information Search Patterns of ED patients and Their Relationship with Final Diagnoses. *West J Emerg Med* 2017;18:928–36.

14. Asch DA, Rader DJ, Merchant RM. Mining the social mediome. *Trends Mol Med* 2015;21:528–9.

15. Eichstaedt JC, Smith RJ, Merchant RM, *et al*. Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A* 2018;115:11203–8.

16. Patt DA, Bernstam EV, Mandel JC, *et al*. More medicine, fewer clicks: how informatics can actually help your practice. *Am Soc Clin Oncol Educ Book* 2017;37:450–9.

17. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. *N Engl J Med* 2011;365:289–91.