BMJ Open

Google Search Histories of Patients Presenting to an Emergency Department

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024791
Article Type:	Research
Date Submitted by the Author:	13-Jun-2018
Complete List of Authors:	Asch, Jeremy; Hospital of the University of Pennsylvania, Emergency Medicine Asch, David; University of Pennsylvania, Center for Health Care Innovation; Philadelphia Veterans Affairs Medical Center, The Center for Health Equity Research and Promotion Klinger, Elissa; University of Pennsylvania, Center for Digital Health Marks, Justine; University of Pennsylvania, Center for Digital Health Sadek, Norah; University of Pennsylvania, Emergency Medicine Merchant, Raina; University of Pennsylvania, Center for Digital Health
Keywords:	World Wide Web technology < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

Google Search Histories of Patients Presenting to an Emergency Department

Jeremy M. Asch, BA[1,2], David A. Asch, MD, MBA[2,3], Elissa V. Klinger, SM[2], Justine Marks, BS[1,2], Norah Sadek, BS[1], Raina M. Merchant, MD, MSHP[1,2]

[1] Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA 19104

[2] Center for Digital Health, Penn Medicine, Philadelphia, PA 19104

[3] The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA 19104

Corresponding author:

Jeremy M. Asch

3400 Civic Center Blvd, fl. 14, Philadelphia, PA 19104

Phone: (610) 348-1279

Jeremy.asch@uphs.upenn.edu

Word Count: 2099

Keywords: Google, search history, electronic medical record, health care utilization

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 2016-2017.

Setting

Academic medical center 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. 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 seven days of the ED visit, and associations between patients' clinical and demographic characteristics and their internet search volume and search content.

Results

The 119 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 utilization in advance of ED visits.

Strengths and Limitations of this study

- By combining search histories with the patients' health record, we were able to create a much more robust correlation between online queries and participant health outcomes.
- The search database we collected is incomplete. We do not know what patients searched for if they used other media (e.g. Bing or a friend's device) or used private browsing.
- This study was done in at an emergency department in an urban environment. That along with the young population which consented may make the results not applicable to other hospitals.

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 behaviors, symptoms, or diagnoses.[1,2] But while these social media posts reflect what 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.[3] More than 72% of US adults report using the internet for health-related questions (How is Zika transmitted?), symptoms (itchy rash on elbow), or behavioral intentions (quitting smoking).[1,4] 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.[5] Unfortunately, these reports often overestimated regional influenza burden.[6,7] Other work has demonstrated association between recipes people search for and what they actually consume.[8] 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 analyzed in

association with their electronic health records, [2] is there variation in health related searches leading up to an emergency department (ED) visit, [3] what do patients search for?

Methods

Participants

From March 2016 - 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 (e.g,. gunshot wounds). Eligible patients were 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 $40 gift card.

Consenting participants self-reported demograhic information. Consenting participants' EMR data were extracted, including the chief complaint upon 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. Interrater agreement was high (kappa = 0.83). Disagreements were adjudicated by a third reviewer (JA). Searches occuring seven days before their recruitment visit to the ED were further coded into more granular themes. Two physicians independently coded the health-related searches along side the patients' chief complaint to determine if searches were related to the clinical presentation at the ED (kappa = 0.75) or to logistic aspects of the visit, such as the address of the ED, adjudicating differences with a third reviewer (JA).

Analysis

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

This study was approved by the University of Pennsylvania Institutional Review Board.

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 119 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 gynecological 20 (17%), and neurological 17 (14%).

Search histories

The 119 participants had 591,421 unique search queries (mean 4,970 ;range 0-51,751); 6% were health-related. Eighty six participants made searches in the seven days prior to the ED visit (3,978 searches; mean 33; range 1-220); 15% were health-related. Among these 86 patients, 54 (63%) searched proportionately more for health-related topics in the seven days prior than in the earlier period; 46 (53%) searched for information clinically related to their chief complaint (e.g., “How to relieve sinus pressure” with a chief complaint of “Headache”); and 13 (15%) searched for directions or other logistic information about the ED or other health care 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 percent of patients approached in a large emergency department, and 49 percent 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 health care at the population level, we believe this is the first effort to link internet search histories and medical records at the level of individual patients. The high rates of participation are surprising against a backdrop of expressed concern for privacy in internet settings and in health care settings. In prior work we demonstrated that 27 percent of patients approached and 71 percent of patients eligible were willing to share access to their social media (e.g., 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 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 behaviors 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 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 tumor?" A review of the medical record revealed the patient had been told of a walnut sized fibrous tumor 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. These findings suggest an ability to anticipate demand. In retail contexts, search terms currently direct targeted advertisements, taking advantage of anticipated demand. One can imagine constellations of search terms predictive enough of serious illness to suggest emergency care or predictive enough of benign illness to suggest avoiding such care. 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 suicidal intent are entered.[9]

This study has several limitations. First, our findings reflect a convenience sample of patients seeking care in one urban academic emergency department 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.

This study also has strengths. We believe this is the first study to associate internet search data with medical record data at the level of the individual. The kind of dose-response relationships we observe

between the content and frequency of 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 health care use and understand health-related knowledge, attitudes, and behaviors 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' use of social media 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, and perhaps more immediately accessible and usable.[10]

Funding: This project was funded by a Robert Wood Johnson Foundation Pioneer Award (72695). No sponsor of funding source played a role in: "study design and the collection, analysis, and interpretation of data and the writing of the article and the decision to submit it for publication."

All researchers are independent from funders

Disclosures: None

Acknowledgements: We thank Katherine Choi, MD and Eugene Gitleman, MD for assistance with clinical coding.

References

1. Health Online 2013. @pewinternet, 2013. (Accessed November, 2017, at 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 Quality and; Safety 2016;25:414-23.

3. Google Search Statistics - Internet Live Stats. 2017. (Accessed November, 2017, at http://www.internetlivestats.com/google-search-statistics/.)

4. What ails America? Dr. Google can tell you. @pewresearch, 2013. (Accessed November, 2017, at http://www.pewresearch.org/fact-tank/2013/12/17/what-ails-america-dr-google-can-tell-you/.)

5. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales. PLOS Computational Biology 2013;9:e1003256.

6. Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. Science 2014;343:1203-5.

7. Butler D. When Google got flu wrong. Nature News 2013;494:155.

8. 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.

9. Gillberti M. Learning more about clinical depression with the PHQ-9 questionnaire. 2017.

10. Asch DA, Rader DJ, Merchant R. Mining the Social Mediome. Trends Mol Med 2015;21:528-9.

Figure Legend

Figure 1: Study enrollment: This figure details inclusion and exclusion criteria applied to identify the final patient cohort

Figure 1. Study Enrollment

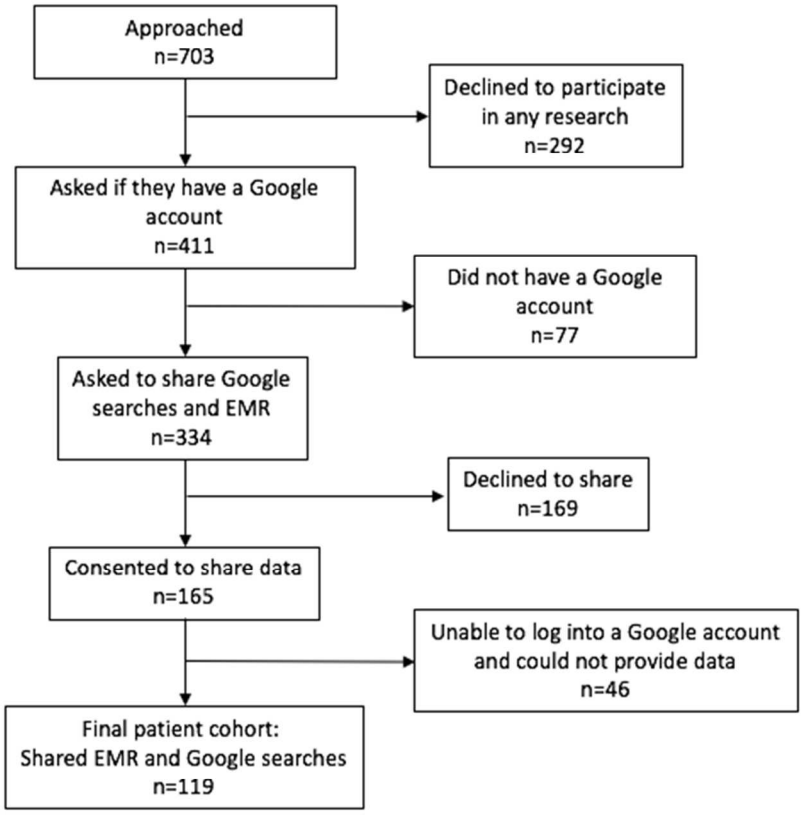

Table 1. Patient characteristics

	ED demographics during study duration (n=61,841) n (%)	Participants sharing Google and EMR data (n=119) n (%)
Age Group		
18-30	20,504 (33)	85 (71)
31-40	10,204 (17)	28 (24)
41-50	8,537 (14)	5 (4)
51-60	9,682 (16)	1 (1)
>60	12,914 (21)	0 (0)
Gender		
Female	37,036 (60)	88 (74)
Race		
White	18,634 (30)	38 (32)
Black	36,959 (60)	64 (54)
Asian	2,242 (4)	4 (3)
Other	2275 (4)	3 (3)
Ethnicity		
Hispanic/Latino	2471 (4)	10 (8)
ED Discharge Diagnosis*		
Diseases of the digestive system	6,535 (11)	35 (29)
Symptoms; signs; and ill-defined conditions and factors influencing health status	9,140 (15)	33 (28)
Endocrine; nutritional; and metabolic diseases and immunity disorders	3,794 (6)	26 (22)
Complications of pregnancy; childbirth; and the puerperium	2,294 (4)	23 (19)
Diseases of the respiratory system	8,127 (13)	22 (18)
Diseases of the musculoskeletal system and connective tissue	6,778 (10)	22 (18)
Diseases of the genitourinary system	6,403 (10)	21 (18)
Diseases of the circulatory system	8,625 (11)	20 (17)
Injury and poisoning	8,070 (13)	20 (17)
Diseases of the nervous system and sense organs	6,120 (10)	19 (16)
Diseases of the skin and subcutaneous tissue	2,628 (4)	8 (7)

Mental Illness	2,532 (4)	4 (3)
Residual codes; unclassified; all E codes [259. and 260.]; Other	10,229 (17)	18 (15)
Search Query Quantity		
>= 5000		31 (26)
1000 - 4999		39 (33)
1 - 999		33 (28)
0		16 (13)

*Percentges sum to more than 100% because some patients have multiple diagnoses.

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/Exersize	"Is it possible to lose 80lbs in a year" "crossfit gym"	11 (13)	41 (1)
Medical Accessories	"Syringe cvs" "hemorrhoid pillows"	11 (13)	19 (0)
Other	"Why wont hot water heater stay lit"	80 (93)	1043 (31)

BMJ Open

Google Search Histories of Patients Presenting to an Emergency Department: an Observational Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024791.R1
Article Type:	Research
Date Submitted by the Author:	14-Nov-2018
Complete List of Authors:	Asch, Jeremy; Hospital of the University of Pennsylvania, Emergency Medicine Asch, David; University of Pennsylvania, Center for Health Care Innovation; Philadelphia Veterans Affairs Medical Center, The Center for Health Equity Research and Promotion Klinger, Elissa; University of Pennsylvania, Center for Digital Health Marks, Justine; University of Pennsylvania, Center for Digital Health Sadek, Norah; University of Pennsylvania, Emergency Medicine Merchant, Raina; University of Pennsylvania, Center for Digital Health
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Emergency medicine, Health informatics
Keywords:	World Wide Web technology < BIOTECHNOLOGY & BIOINFORMATICS, Information technology < BIOTECHNOLOGY & BIOINFORMATICS, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

Google Search Histories of Patients Presenting to an Emergency Department: an Observational Study

Jeremy M. Asch, BA[1,2], David A. Asch, MD, MBA[2,3], Elissa V. Klinger, SM[2], Justine Marks, BS[1,2], Norah Sadek, BS[1], Raina M. Merchant, MD, MSHP[1,2]

[1] Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA 19104

[2] Center for Digital Health, Penn Medicine, Philadelphia, PA 19104

[3] The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA 19104

Corresponding author:

Jeremy M. Asch

3400 Civic Center Blvd, fl. 14, Philadelphia, PA 19104

Phone: (610) 348-1279

Jeremy.asch@uphs.upenn.edu

Word Count: 3482

Keywords: Google, search history, electronic medical record, health care utilization

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 2016-2017.

Setting

Academic medical center 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 seven 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 utilization in advance of ED visits.

Strengths and Limitations of this study

- By combining search histories with the patients' health record, we were able to create an individual-level correlations 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 hospitalizations; 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 (e.g. Bing or a friend's device) or using private browsing.
- This study was performed in at an emergency department in an urban environment. That along with the young population who consented may make the results less applicable to other settings.

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 behaviors, symptoms, or diagnoses.(1-3) But while these social media posts reflect what 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 behavioral 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 analyzed in

association with their electronic health records, [2] is there variation in health related searches leading up to an emergency department (ED) visit, [3] what do patients search for?

Methods

Participants

From March 2016 - 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 (e.g,. gunshot wounds). Eligible patients were approached while they were a patient at the Emergency Department 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 $40 gift card. All patients provided written consent to be considered eligible.

Consenting participants self-reported demograhic information. Consenting participants' EMR data were extracted, including the chief complaints upon 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 (JA). Searches occurring within seven days before their recruitment visit to the ED were further coded into more granular themes. Two physicians independently coded the health-related searches along side 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 (JA).

Analysis

Summary statistics were used to characterize sociodemographic data, EMR data (e.g. chief complaint, discharge diagnoses), and search queries (time, date). STATA 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.

This study was approved by the University of Pennsylvania Institutional Review Board.

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 gynecological 20 (17%), and neurological 17 (14%).

Search histories

The 103 participants had 591,421 unique search queries (mean 5,742 ;range 2-51,751). Interrater agreement between the coders was high (kappa = 0.83); 6% were health-related. Eighty six participants made searches in the seven days prior to the ED visit (3,978 searches; mean 33; range 1-220); 15% were health-related. Among these 86 patients, 54 (63%) searched proportionately more for health-related topics in the seven days prior than in the earlier period; 46 (53%) searched for information clinically

related to their chief complaint (e.g., "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 health care 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 percent of patients approached in a large emergency department, and 49 percent 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 health care 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 health care settings after the Equifax breaches and issues regarding Facebook and Google's handling of user data.(10) In prior work we demonstrated that 27 percent of patients approached and 71 percent of patients eligible were willing to share access to their social media (e.g., 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 behaviors 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 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 tumor?" A review of the EMR revealed the patient had been told of a walnut sized fibrous tumor 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 hospitalization, 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 suicidal intent are entered.(12)

This study has several limitations. First, our findings reflect a convenience sample of patients seeking care in one urban academic emergency department 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 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 health care use and understand health-related knowledge, attitudes, and behaviors 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)

Funding: This project was funded by a Robert Wood Johnson Foundation Pioneer Award (72695). No sponsor of funding source played a role in: "study design and the collection, analysis, and interpretation of data and the writing of the article and the decision to submit it for publication."

All researchers are independent from funders

Data Sharing Statement: Because of the potential identifiability of participants through either their search histories or their EMR information, data will not be shared.

Contributors: JA and RM designed the study. JA, NS, and JM recruited patients and assisted in data analysis. DA, RM, and EK provided mentorship. All authors contributed to the editing and writing of the paper.

Disclosures: None

Acknowledgements: We thank Katherine Choi, MD and Eugene Gitleman, MD for assistance with clinical coding.

References

1. Fox S, Duggan M. Health Online 2013: @pewinternet; 2013 [updated 2013-01-15. Available from: http://www.pewinternet.org/2013/01/15/health-online-2013/. Last accessed 11-8-18
2. Padrez KA, Ungar L, Schwartz HA, Smith RJ, Hill S, Antanavicius T, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. BMJ Quality & Safety. 2016;25(6):414-23.
3. Smith RJ, Crutchley P, Schwartz HA, Ungar L, Shofer F, Padrez KA, et al. Variations in Facebook Posting Patterns Across Validated Patient Health Conditions: A Prospective Cohort Study. Journal of Medical Internet Research. 2017;19(1):e7.
4. Google Search Statistics - Internet Live Stats 2017 [Available from: http://www.internetlivestats.com/google-search-statistics/. Last accessed 11-8-18
5. Fox S. What ails America? Dr. Google can tell you Fact Tank: @pewresearch; 2013 [updated 2013-12-17. Available from: http://www.pewresearch.org/fact-tank/2013/12/17/what-ails-america-dr-google-can-tell-you/.
6. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales. PLOS Computational Biology. 2013;9(10):e1003256.
7. Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. Science. 2014;343(6176):1203-5.
8. Butler D. When Google got flu wrong. Nature News. 2013;494(7436):155.
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. 2488510: ACM; 2013. p. 1399-410.
10. Mahmoodi J, Čurdová J, Henking C, Kunz M, Matić K, Mohr P, et al. Internet Users' Valuation of Enhanced Data Protection on Social Media: Which Aspects of Privacy Are Worth the Most? Frontiers in psychology. 2018;9:1516-.
11. Libert T, Grande D, Asch DA. What web browsing reveals about your health. BMJ. 2015;351.
12. Gillberti M. 2017. Available from: https://www.blog.google/products/search/learning-more-about-clinical-depression-phq-9-questionnaire/. Last accessed 11-8-18
13. McCarthy DM, Scott GN, Courtney DM, Czerniak A, Aldeen AZ, Gravenor S, et al. What Did You Google? Describing Online Health Information Search Patterns of ED patients and Their Relationship with Final Diagnoses. The western journal of emergency medicine. 2017;18(5):928-36.
14. Asch DA, Rader DJ, Merchant R. Mining the Social Mediome. Trends Mol Med. 2015;21(9):528-9.
15. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences. 2018;115(44):11203-8.
16. Patt DA, Bernstam EV, Mandel JC, Kreda DA, Warner JL. More Medicine, Fewer Clicks: How Informatics Can Actually Help Your Practice. American Society of Clinical Oncology Educational Book. 2017(37):450-9.

17. Merchant RM, Elmer S, Lurie N. Integrating Social Media into Emergency-Preparedness Efforts. New England Journal of Medicine. 2011;365(4):289-91.

Figure Legend

Figure 1: Study enrollment: This figure details inclusion and exclusion criteria applied to identify the final patient cohort

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	8,537 (14)	4 (4)
51-60	9,682 (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	2,242 (4)	2 (3)
Other	2275 (4)	3 (3)
Ethnicity		
Hispanic/Latino	2471 (4)	10 (8)
Chief Complaint		
Gastrointestinal		41 (37)
Ob/Gyn		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.

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/Exersize	"Is it possible to lose 80lbs in a year" "crossfit gym"	11 (13)	41 (1)
Medical Accessories	"Syringe cvs" "hemorrhoid pillows"	11 (13)	19 (0)
Other	"Why wont hot water heater stay lit"	80 (93)	1043 (31)

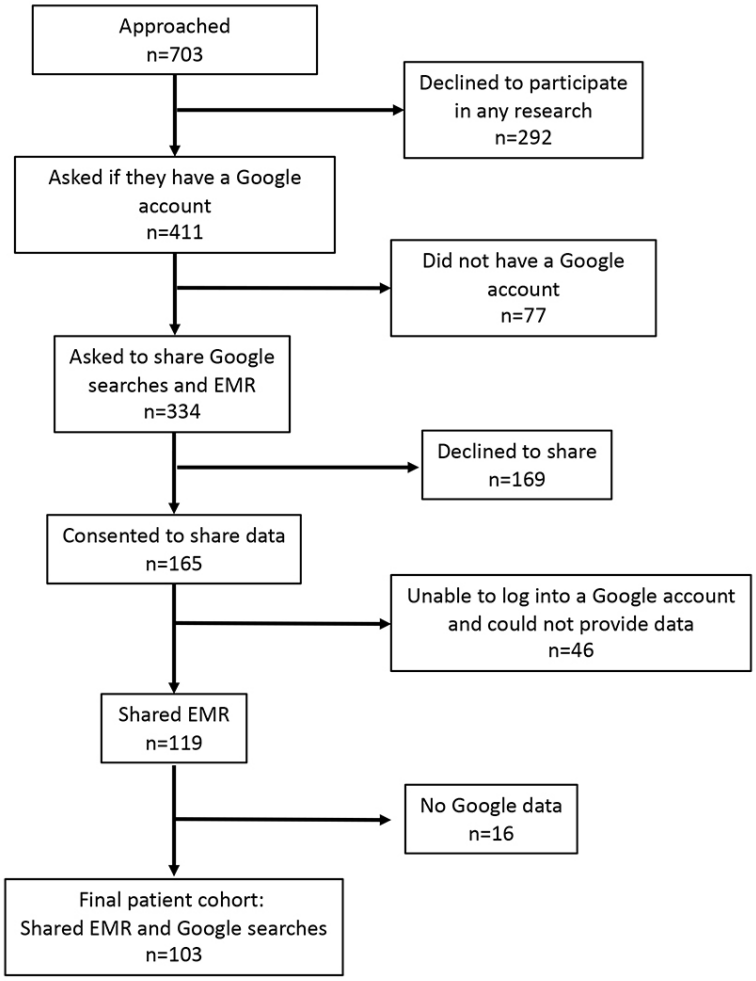

Figure Legend

Figure 1: Study enrollment: This figure details inclusion and exclusion criteria applied to identify the final patient cohort

90x90mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(pg.1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(pg.2)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(pg.5)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(pg. 5)**
Methods		
Study design	4	Present key elements of study design early in the paper **(pg.6)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(pg. 7)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(pg. 6)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(pg. 7)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(pg. 7)**
Bias	9	Describe any efforts to address potential sources of bias **(pg. 4)**
Study size	10	Explain how the study size was arrived at **(pg. 6)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(pg. 7)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(pg. 7)**
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed **(pg. 11)**
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(pg. 8)**
		(b) Give reasons for non-participation at each stage **(pg. 8)**
		(c) Consider use of a flow diagram **(fig 1.)**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (**pg. 8 and table 1)**
		(b) Indicate number of participants with missing data for each variable of interest **(pg. 11)**
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time **(pgs 8,9)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were

		adjusted for and why they were included **(pg. 8)**
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses (**pgs. 8, 10)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(pgs. 9-11)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(pgs. 4, 10,11)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(pg. 11)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(pgs. 10,11)**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(pg. 13)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.