BMJ 2015;351:h5974 doi: 10.1136/bmj.h5974 (Published 16 November 2015)

ANALYSIS

What web browsing reveals about your health

Timothy Libert *PhD candidate*[1], David Grande *assistant professor of medicine*[2], David A Asch *professor of medicine*[2 3 4]

[1]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA; [2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA ; [3] Wharton School, University of Pennsylvania; [4]Philadelphia VA Medical Center, Philadelphia, PA 19104, USA

Abstract

Advertisers can learn a lot from our digital footprints, including about our health. **Timothy Libert and colleagues** argue that we need to tighten restrictions on such data for marketing purposes while looking at their potential to improve health

In January 2014, the Privacy Commission of Canada declared that Google had violated Canadian privacy law when it used information about individuals' online activities to target them with health related advertisements.[1] Someone who had searched online for medical devices to treat sleep apnoea found himself followed by advertisements for such devices as he visited unrelated websites. His search had resulted in a browser cookie that triggered ads for sleep apnoea devices when he visited websites that used Google's advertising services. Google acknowledged that some of the advertisers using its service do not comply with its policy against advertising relating to sensitive issues such as race, religion, sexual orientation, or health.[1]

How "remarketing" strategies work

When we seek health information online, the visited websites often include code that initiates network connections (known as HTTP requests) to third parties such as online advertisers. It is these HTTP requests that effectively share browsing histories (figure⇓). Records of visits to pages for sleep apnoea, depression, or addiction treatment can be resold to organisations that want to know who is interested in these topics. Such information may be as sensitive as that contained in electronic health records, and yet little legal oversight regulates how this health information is collected, how long it is kept, and how it is used.

Scale of information leakage

A recent study using 1986 health related search terms showed that 91% of more than 80 000 commonly visited web pages initiated such requests, most of which transmit user information to third parties, largely to promote marketing. Commercial websites require advertising revenue to operate; it is not surprising, therefore, that they initiate third party requests. By contrast, non-profit, government, and education websites might have less need to rely on advertising revenue. Nevertheless, 92% of .org, 86% of .gov, and 76% of .edu pages initiated third party requests.[2] Although most of the research is in the United States, the problem appears global. In the United Kingdom, the NHS homepage generates third party requests to seven domains owned by American companies, including NetIQ, Amazon, and Google. We also examined 25 health related French web pages and discovered 72% generated third party cookies and that third party requests sent user data to Google, Facebook, Twitter, AOL, and other companies. While the way this information is used is often a trade secret, these results show substantial leakage.

Most often websites include third party request codes not because there is a business interest in sharing user information but because web developers routinely use free services provided by online advertisers.[3] Indeed, 45% of pages we examined included a hidden "tracking pixel" that is part of the Google Analytics service. This service provides web masters with information on traffic to their websites while simultaneously providing Google with information useful in targeting advertisements.[2] These requests often disclose the address of the page the user is visiting. The address itself may be uninformative (such as, www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001911/) or it may contain suggestive health information (such as, www.cdc.gov/cancer/breast/). An analysis of page addresses shows that 70% included information related to specific symptoms, conditions, or treatments.

Most web browsing reveals little about any given person and is largely benign. However, once aggregated, patterns of behaviour can be attributed to specific people. For example, Facebook requires real names and also receives HTTP requests from 31% of the sites analysed. This potentially allows Facebook (and similar companies) to pair real names with real health conditions. In addition, data brokers who sell identified user data, such as Experian and Acxiom, track views on health related pages.[4] Companies do not clearly disclose how they use the data, but there are few limits on what they can do. By contrast, there are strict controls to protect personal health information

Correspondence to: D A Asch asch@wharton.upenn.edu

Glossary of terms

Browser cookie—A small text file saved to users' computers that can be used to identify and correlate their visits among many websites. Companies using cookies can track users until the user deletes the cookie

Browser fingerprint—A new method that uses the characteristics of computers (eg, operating system, screen size, installed fonts) to identify and correlate a person's visits across many websites. Unlike cookies, browser fingerprints cannot easily be deleted

Behavioural advertising—The process by which online advertisers and data brokers use information about online behaviour (eg, the history of which websites have been visited) to target specific advertisements and deals at people with a certain behavioural profile

HTTP request—Hyper-Text Transfer Protocol is the type of network request by which users download websites. These requests may be made to a "first party," which is the address listed in a browser's address bar. In addition, "third party requests" may be made in the background to advertisers and data brokers outside the user's view; often this is called the hidden web

generated during clinical encounters. It is as if the front door is barricaded and the back door is wide open.

Potential for harm

Web browsing may reveal as much or more about patients' health as clinical records. Web browsing by someone with diabetes, for example, could show not only that they have diabetes but actions they are taking or considering to control their diabetes. Marketers mining health information from web browsing patterns can create harm when, for example, a browser on someone's computer reveals embarrassing health information through advertisements that appear when that computer is shared or used in a public setting such as a workplace. People care a great deal about privacy,[5] and many find it objectionable that their information is used without permission.

Constructive use

But they care even more about the purpose of use. In a recent study of a representative sample of Americans, people were more tolerant of research uses of personal health information without consent than they were of commercial marketing uses of the same information with consent.[6] The concern that individuals don't want to share their health related information is further challenged by the uptake of wireless health devices aimed at promoting fitness, weight loss, medication adherence, or management of chronic disease.[7] Patients increasingly use this kind of information sharing to provide self motivation or to engage others, including friends and physicians, in their quest for better health.

Information from browsing histories has the potential for benefit by helping to identify patients with treatable conditions. Current healthcare use of browsing information operates on a population level—examining regional symptoms to forecast influenza trends, for example. But individualised browsing histories might also be useful for individual healthcare. For people with undiagnosed depression, for example, web browsing patterns may reveal the diagnosis even when they, or their doctors, are unaware. By electing to share appropriately protected web browsing data people could obtain valuable insights about strategies to achieve health goals. The same data used in aggregate might inform societal research goals.

Moving forward

What is needed now is a tightening of access to information used for marketing while simultaneously re-examining how patients want their digital information used for other purposes. Organisations like the US Centers for Disease Control and Prevention (CDC) should examine the coding within their websites to ensure that sharing of browsing information is consistent with the goals and expectations of the organisation and its users. Regulators should think past their focus on information source in designing protections of health related information, since that information is no longer produced only during clinical encounters. In the US, the Health Insurance Portability and Accountability Act regulates health information derived from clinical settings, but no regulations constrain the use of information from web browsing, social media, and wearable devices. In the European Union, the EU Data Protection Directive applies to health information created and accessed outside of healthcare settings. However, implementation and enforcement of the directive has been inconsistent.

Perhaps most importantly, healthcare and public health organisations should engage the public in discussions about how they might use some of the same clever methods of marketers to understand the health needs of patients and the public and respond to them. For instance, by observing and analysing web browsing patterns, a healthcare organisation might be able to identify patients who are ready and motivated to quit smoking and target interventions at them. What could make these new strategies work for individual clinical care is offering patients clear and easy opportunities to control the information they elect to share. Societal research purposes could be supported without such opt-in provisions, so long as the information was securely protected and its use monitored. These processes deserve discussion as much for the opportunities they might provide as for the protections they will require. But even as we decry the methods used by marketers, health and healthcare organisations might learn from them, adapting the same tools for social purpose.

We thank Maxime Marie Jean Cordier for a list of health related French web pages and Carolina Garzon Mrad for the figure.

Contributors and sources: DAA has expertise in health policy and DG has conducted research into the influence of marketing in medicine and how individuals view privacy of their electronic health information. TL has experience in computer science and expertise in the technology, policy, and business aspects of online advertising. DAA is the guarantor.

Competing interests: We have read and understood BMJ policy on declaration of interests and declare the following interests: DAA is a principal at VAL Health and DG is a volunteer board member of Healthy Philadelphia (non-profit).

1 Office of the Privacy Commissioner of Canada. Google ads sparked by web surfing on health sites violate privacy rights, investigation finds. Press release, 15 Jan 2014. www.priv.gc.ca/media/nr-c/2014/nr-c_140115_e.asp.
2 Libert T. Privacy implications of health information seeking on the web. *Comm ACM* 2015;58:68-77.
3 Roesner F, Kohno T, Wetherall D. Detecting and defending against third-party tracking on the web. Proceedings of the 9th USENIX conference on networked systems design and implementation 2012:12.
4 Turow J. The daily you. Yale University Press, 2013.
5 Hoofnagle C, Urban J, Li S. Privacy and modern advertising: most US internet users want "do not track" to stop collection of data about their online activities. Amsterdam Privacy Conference, 2012.
6 Grande D, Mitra N, Shah A, Wan F, Asch DA. The importance of purpose: moving beyond consent in the societal use of personal health information. *Ann Intern Med* 2014;161:855-62.
7 Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 2015;313:459-60.

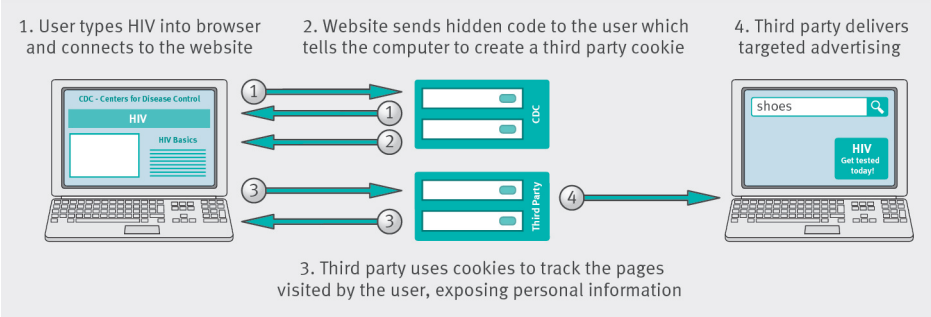
Key messages

- Digital footprints from activities like internet browsing inadvertently reveal a great deal about our health
- Despite strict regulation of clinical information, health information obtained through digital activities is now widely available to marketers with little protection
- Such information also has potential to advance individual and population health
- We need to tighten access to information used for marketing while discussing with patients how these methods could be borrowed by healthcare and public health organisations

Accepted: 23 October 2015

Cite this as: *BMJ* 2015;351:h5974

Figure



How online searching leads to targeted web advertising