

Could data broker information threaten physician prescribing and professional behavior?

Marco D. Huesch, MBBS PhD, *University of Southern California*

Michael K. Ong, MD, PhD, *University of California, Los Angeles; VA Health Care System*

Barak D. Richman, JD, PhD, *Duke University*

Abstract

Privacy is threatened by the extent of data collected and sold by consumer data brokers.

Physicians, as individual consumers, leave a ‘data trail’ in the offline (e.g. through traditional shopping) and online worlds (e.g. through online purchases and use of social media). Such data could easily and legally be used without a physician’s knowledge or consent to influence prescribing practices or other physician professional behavior.

We sought to determine the extent to which such consumer data was available on a sample of more than 3,000 physicians, healthcare faculty and healthcare system staff at one university’s health units. Using just work email addresses for these employees we cheaply and quickly obtained external data on nearly two thirds of employees on demographic characteristics (e.g. income, top 10% national wealth, children at home, married), purchases (e.g. baby products, cooking, sports), behavior (e.g. charitable donor, discount shopper) and interests (e.g. automotive, health and wellness).

Consumer data brokers have valuable, cost-effective and detailed information on many healthcare professionals, including data that could be used to segment, target, detail and generally market to physicians in ways that seem under-appreciated. We call for greater attention to this potential aspect of physician-industry relationships.

Introduction

In the Internet age, privacy is threatened on numerous fronts. National security interests have led to government capabilities to obtain and integrate the communications data and online activities of many Americans.¹ Analogous commercial interests lead many firms to monetize the data that consumers create offline, online or on smartphones. A large gray market of data brokers has grown to meet the demands for detailed information on purchasing behavior and interests.² We claim that such data could pose under-appreciated threats to physician's prescribing and professional behavior.

To see this, consider first the sources, the prevalence, and then the potential uses of such data. Consumer healthcare websites create the capabilities to track users and leak sensitive search information to third parties.³ More generally, it has been estimated that currently an average of 42 data collection events occur on *each page* of each of the top 50 ad-supported websites.⁴ Smartphone wellness applications share users' data with advertising and analytics companies.⁵

Historically using offline data, but increasingly online data as well, data aggregators and data brokers have constructed massive proprietary databases. These contain years' worth of inferred interests, demographics, household data, and purchasing behavior. One leader in this field, Acxiom, processes more than 50 trillion data points yearly, and has about 1,500 data points per person on the majority of adults in the country.²

Such databases offer cheap, real-time information on individuals. To illustrate this, we obtained a convenience sample of the work email addresses of USC healthcare faculty and staff in a USC Institutional Review Board approved study.

Data and Methods

Using random searches of USC's internal online directory, we obtained 3,009 work email addresses sampled from physicians, researchers and support staff within the Children's Hospital of Los Angeles, USC Health, and from all departments and divisions at Keck School of Medicine.

We contracted with Rapleaf, one of 9 leading data brokers,⁶ whose specialty is enhancing email lists with external data and claims to have at least one data point on 80% of all US consumer email addresses. Data of this sort is collected and collated from offline and online store shopping behavior captured by commercial entities, credit cards or store loyalty cards, from public records, and from surveys and questionnaires.

We paid Rapleaf less than \$300 to provide up to 34 data points linked to each of these work email addresses. We uploaded a text file containing the email addresses to Rapleaf's website and they supplied a spreadsheet data file containing results in approximately 10 minutes.

Rapleaf guarantees to keep all uploaded emails private, and warrants that all data was collected in line with prevailing privacy laws and good business practices, and consumers explicitly or implicitly consented to such collection. In our setting of a university and associated health system sites, use of work email addresses as log-in credentials to commercial websites (e.g. Amazon) may have contributed to the richness of data we were able to purchase.

Findings

For around two thirds of the emails a rich set of information was available identifying personal information spanning economic, family, interests and purchases data (Table 1). Where data was

available, a typical health unit employee was a married early middle-aged male, in a household earning \$112,000 income, with approximately 1 in 6 being in the top 10% nationally by net worth, very likely to own a home, typically worth \$632,000. A variety of interests were captured by Rapleaf and associated with these work email addresses. For example, 14% had documented cooking interests, nearly 7% had documented charitable donation interests, 14% had purchasing interests in health and wellness, and while nearly 8% were identified as high-end brand shoppers, 1.3% were observed by Rapleaf and its data partners as having purchased through discount shopping channels.

While we are prohibited by our Institutional Review Board from revealing individual USC employee data, it was possible in our results to characterize individuals in quite fine categories based on the distinct “true” fields for interests and purchases, and lifecycle stage among the 34 fields reported by the data broker. Had we matched on individual name and residential address and contracted with a market leader like Acxiom, we could potentially have obtained many times the number and detailed values of fields (Table 2).

Discussion

Other industries routinely and exhaustively utilize such data. Financial services firms increasingly integrate external social media data, blog posts, emails, transcripts of interviews, and other unstructured data sources with their large internal quantities of transactional data.⁷ Banks do this as they seek to break down their consumer market into finely segmented groups of customers and to predict the behavior (e.g. response to offers, willingness to pay for services) of customers in these groups.

Marketers in fast moving consumer goods industries similarly use such external data both to inform advertising campaigns in the offline world and to qualify prospects in the online world.⁸ Especially among older consumers, validated clustering of consumers into groups defined by common lifecycle, purchasing behavior and demographics is routine.⁹

On the one hand, consumers benefit by being presented with better suited offers closely related to their own interests. On the other hand, marketers can use this data to price discriminate, to quality and service discriminate, and even to make decisions on whether to offer the online opportunity to particular types of consumers without falling foul of non-discrimination laws. Similarly, researchers can use publicly disclosed social media to identify consumers who hold unorthodox opinions on common topics such as vaccinations.¹⁰

This rich data also has potentially problematic uses in the healthcare arena especially in physician-industry interactions. In an era where sales representatives have tougher, briefer, and more controlled access to physicians, industry has understandable interests in best targeting scarce resources at the most likely prescribers or users of devices.

Marketing teams in industries which seek to affect prescribing and treatment practices clearly have the ability to duplicate our approach. Purchasing data on physician interests could successfully inform direct-to-physician marketing and in-person detailing. Obtaining information on lifestyle, recent financial or personal stresses could be used to define susceptible segments of physicians.

For example, Audience Partners uses its own large-scale consumer data to allow marketers to target healthcare professionals based on using “geographic, demographic, attitudinal, and behavioral attributes.”¹¹ Experian and other credit bureau will sell “life event data” such as purchases of a new home and names of expectant parents to marketers.¹² Epsilon will sell data on who is likely to read the Bible at home and who prefers what music,¹³ and notes that individual’s social media tweets, likes, posts, comments, shares and recommendation are collected, analyzed and sold.¹⁴

When a customer of Sears or Best Buy shares his email address with the retailer, a world of profitable data mining possibilities opens up. Similarly, no barriers — especially, no regulatory barriers — prevent full exploitation of the personal information owned by data brokers by marketers in healthcare’s supplier industries. None of these practices need to be disclosed to physicians and hence physicians would not know whether their data had been ‘harvested’ or ‘mined’ and whether as a result they had been ‘targeted’.

Indeed, none of this information is covered by existing law. Data brokers might eventually be subject to regulatory oversight under privacy laws enforced by federal agencies such as the Federal Trade Commission,⁶ but currently consumer awareness of the risks posed by data aggregators remains low. Physician’s rights to have their prescription practices remain confidential and not be sold to marketers such as IMS Health have been litigated. We are not aware of any recent or upcoming cases in which the use of data broker data has been litigated.

Of course, ethical marketers and pharmaceutical and medical device industry professional culture may limit such controversial uses. However, the technology and the stock of data exists and the commercial interests to make use of it are clear.

Accordingly, our recommendation is for greater clarity in what uses are being made of physician's private transaction data, inferred purchase interests, and other potentially sensitive information. Asking a sales representative to explain exactly what he or she knows about you may be illuminating and could herald a need for the profession to be more vigilant about the brave new world of data.

===

Acknowledgements:

Contributors: We acknowledge clarification and advice from Rupleaf on their sources of data.

Funders: We acknowledge financial support by an unrestricted research allowance provided to the first author by the University of Southern California.

Conflicts of interest: Dr Huesch reports employment by USC and Duke, grant support by AHRQ, Lockheed Martin and Baxter, and consulting to the IOM Committee on Geographical Variation, Parkland Center for Clinical Innovation and Precision Health Economics. Dr Ong reports serving as Deputy Editor for the Journal, and also reports separate grant support by AHRQ and NIH, and consulting to the IOM Committee on Geographical Variation and the Public Health Institute. Dr

Richman made no disclosures. This study was not related to any funded projects, was not reviewed by funders' representatives, and no endorsement is intended or should be inferred.

Corresponding Author: Marco D. Huesch, MBBS, PhD; USC Price School of Public Policy, 3335 S. Figueroa St, USC Gateway Unit A, Los Angeles, CA 90089-7273. (e-mail: huesch@usc.edu).

References

-
- ¹ **Risen J, Poitras L.** NSA gathers data on social connections of US citizens. New York Times. Available: <http://www.nytimes.com/2013/09/29/us/nsa-examines-social-networks-of-us-citizens.html> Accessed November 4, 2013
 - ² **Singer N.** Mapping, and sharing, the consumer genome. New York Times. 2012 Available <http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html> Accessed November 4, 2013
 - ³ **Huesch MD.** Patient privacy risks when seeking online health information: analysis of 20 popular websites. JAMA Internal Medicine. 2013;173(19):1838
 - ⁴ **Krux.** The state of data collection on the web. 2013 Krux cross industry study. Available: http://www.krux.com/pro/broadcasts/krux_research/CIS2013/ Accessed November 4, 2013
 - ⁵ **Kahl A.** A healthy data set. The Evidon Blog. Available: <http://www.evidon.com/blog/healthy-data-set> Accessed November 4, 2013

-
- ⁶ **Federal Trade Commission.** FTC to study data broker industrys collection and use of consumer data. 2012. Available <http://www.ftc.gov/opa/2012/12/databrokers.shtm>
Accessed November 4, 2013
- ⁷ **Hickins M.** Banks using Big Data to discover ‘new silk roads’. Wall Street Journal. 2013.
<http://blogs.wsj.com/cio/2013/02/06/banks-using-big-data-to-discover-new-silk-roads/>
Accessed June 14, 2013
- ⁸ **Valentino-Devries J, Singer-Wine J.** They know what you’re shopping for. Wall Street Journal. 2012.
<http://online.wsj.com/article/SB10001424127887324784404578143144132736214.html>
Accessed June 14, 2013
- ⁹ **Axiom Corporation.** Buy data – enhance your list.
<http://mymailingleads.com/products/enhance-your-list.html> Accessed June 14, 2013
- ¹⁰ **Huesch MD, Ver Steeg G, Galstyan A.** Analyzing vaccine skepticism in social media. Proceedings of the Association for the Advancement of Artificial Intelligence-13 Workshop on Expanding the Boundaries of Health Informatics using AI. Summer 2013.
<https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/.../7094/6502> Accessed November 4, 2013
- ¹¹ **Audience Partners.** Healthcare audiences.
<http://www.audiencepartners.com/audiences/#healthcare> Accessed November 15, 2013
- ¹² **Experian.** Life event triggers. <http://www.experian.com/marketing-services/life-event-marketing.html> Accessed November 15, 2013

¹³ **Epsilon.** Customer data and data cards.

<http://lists.epsilon.com/market;jsessionid=6A817DF16F9D9DABB743DA921DBF46A1?page=research/datacard&id=313339> Accessed November 15, 2013

¹⁴ **Propublica.** Yes, companies are harvesting – and selling – your Facebook profile.

<http://www.propublica.org/article/yes-companies-are-harvesting-and-selling-your-social-media-profiles> Accessed November 15, 2013

Table 1. Data Broker Information Linked to USC Health Faculty and Staff Email Addresses

| Demographics and household data | | | | | | | | | | |
|--|------------------|----------------|---------------|-------------------|-----------------------|--------------|---------------------|-----------------|------------------|----------|
| | Age | Male | Zipcode | Top 10% net worth | Home value | Home owner | Household income | Time in address | Married | Children |
| Available, % * | 22.3 | 62.1 | 19.6 | 17.9 | 22.1 | 21.5 | 19.8 | 22.2 | 20.7 | 21.5 |
| Mean, value † | 50.7% | 42.4% | nm | 17.9% | \$632k | 85.5% | \$112k | 10.7 y | 54.7% | 21.9% |
| Interests and purchase history data | | | | | | | | | | |
| | Arts & crafts | Books | Business | Health & wellness | News & current events | Baby product | Beauty | Auto-motive | Charitable donor | Cooking |
| True, % ‡ | 5.6 | 18.2 | 1.5 | 14.1 | 10.4 | 0.8 | 6.8 | 2.6 | 6.5 | 14.1 |
| | Discount shopper | High end brand | Home & garden | Home improvement | Luxury goods | Magazines | Outdoor & adventure | Sports | Tech | Travel |
| True, % ‡ | 1.3 | 7.6 | 12.9 | 18.6 | 1.0 | 13.2 | 4.3 | 10.3 | 4.9 | 7.7 |

nm = not meaningful; k = ,000 dollars; y = years of residence

* Calculated for all 3,009 faculty and staff work email addresses sampled randomly from USC Health, Children's Hospital of Los Angeles, and Keck School of Medicine. For 1,093 (36.3%) email addresses, Rappleaf held no information for any of the 34 data fields requested.

† Calculated only for email addresses with the respective data field available.

‡ Calculated for email addresses with at least one available data field.

Table 2. Available Broker Information Linked to Consumer Name+Residence*

| Income and expenditures | Stressors |
|--|--|
| Net worth, income and home value Credit card new issue Credit: range of new credit Retail activity : date of last Credit card: frequency of purchase Retail purchases: ____ category Retail purchases: most frequent category Online purchasing indicator for ____ category Mail order buyer, donor Vehicle: make, year 1st Vehicle | Size of household Tenure in residence, e.g. new mover Net worth and income Home loan amount, home loan-to-value Smoking and tobacco use Guns and ammunition purchases Gaming inerests and casino visits Recent divorce New parent Child near high school graduation |
| Social exclusion | Health behaviors and interests |
| Recent divorce Empty nester Parenting Community involvement in 'causes' Text messaging use Cellplan use Computer and internet use Documented interests in children, grandchildren | Smoking, tobacco, cigars Cooking: ____ category Foods: ____ category Exercise: ____ category Dieting, weight loss Allergy Homeopathic Organic |

(*) Selected categories of over 1,500 data fields potentially available at Acxiom, linked to name and residence