Published in final edited form as:

Trends Mol Med. 2015 September ; 21(9): 528–529. doi:10.1016/j.molmed.2015.06.004.

Mining the Social Mediome

David A. Asch, MD MBA,

Penn Center for Healthcare Innovation, 423 Guardian Drive, Philadelphia, PA 19104; Tel: 215-746-0019, asch@wharton.upenn.edu

Daniel J. Rader, MD, and

Department of Genomics and Computational Biology, Department of Cell and Molecular Biology, 3400 Civic Center Blvd, Philadelphia, PA, 19104; Tel: 215-615-4949, rader@mail.med.upenn.edu

Raina Merchant, MD MSHP

Penn Social Media and Health Innovation Lab, 423 Guardian Drive, Philadelphia, PA 19104; Tel: 215-746-8681, Twitter: @RainaMerchant, raina.merchant@uphs.upenn.edu

Abstract

The experiences and behaviors revealed in our everyday lives provide as much insight into health and disease as any analysis of our genome could ever produce. These characteristics are not found in the genome, but may be revealed in our online activities which make up our social mediome.

Keywords

communication; health information technology; health services research; social media

“Wouldn’t it be great,” the thinking goes, “if we collected everyone’s DNA and linked it to the phenotypes of health and disease that people really care about?” We would learn and do a lot with the information this approach would reveal. The process needs care and attention to ethical considerations, because it occasionally generates unwanted information and can threaten privacy for individuals and groups, for known situations now, and unknown situations in the future. While the promise and perils of DNA biobanking and mining the genome are more complicated than that, this tension between discovery and privacy is central.

One fifth of the world’s population uses Facebook or Twitter and many of these users share information daily. Posts, photos, videos, “likes,” check-ins, and other digital footprints of social media reveal a rich diary of everyday life.[1] Health information is revealed explicitly when someone tweets about a GI bug[2] or an inpatient posts about a hospital experience.[3] But health is also revealed indirectly, by the cigarettes seen in the foreground of a shared

Correspondence: Raina Merchant, MD MSHP, Penn Social Media and Health Innovation Lab, 423 Guardian Drive, Philadelphia, PA 19104; Tel: 215-746-8681, Twitter: @RainaMerchant, raina.merchant@uphs.upenn.edu.

Prior presentations: none

Disclosures: Asch: US Government employee; Rader: none; Merchant: research grants: NIH, NHLBI R01 (HL122457-01A1), NIH, NHLBI K23 (109083-01)

family photograph, or by changes in the tone, depth, or frequency of postings, or a drop off in followers or friends. DNA biobanking offers a window into our patients' genome, and it is easy to see the health care value of that. Systematic surveillance of our patients' social media data has the potential to provide clinicians and hospitals something they may have never realized they were missing, which is a window into their patients' lives.

However, much of this social media content sits beyond the view of clinicians, who could never devote the time to surveilling it and are unlikely to know exactly what to look for anyway.[4] While most Twitter posts are publicly available, most other forms of social media require invitations that limit their use for individual clinical care, research, or public health. Ironically, perhaps, the same people apparently willing to broadcast strikingly personal information on social media sites also care about their privacy—a contradiction explained by the different audiences and uses of the information different contexts bring.[5] One result is that most health-related analyses of social media speak only at the level of ecological groups. At the University of Pennsylvania Health System, we are asking some patients to let us search their past and future Facebook and Twitter posts and link that information to what we have on hand about them in our electronic health records. The project hints at the opportunities these linkages might provide, and also the privacy and social acceptability challenges they must address. Early findings are not only that a large proportion of these patients use Facebook and Twitter regularly, but also that many of them are willing to open these accounts up to us and let us link their social media data to their electronic medical records for research purposes. Other groups have used surveys to assess symptomatology compared with self-reported social media behaviors or observed postings and noted associations (neutral, positive, negative) between social media use and depressive symptoms.[6-9] In this context, using highly calibrated algorithms, social media platforms could be used to automatically detect posts about depression or other conditions, and anonymously provide individuals with related health resources.[8, 10] Considering the potential for false alarms and unintended labeling or targeting of patients, these behavioral health focused interventions need to be implemented carefully and thoughtfully with close attention to issues of patient privacy. These risks can potentially be mitigated through data transparency, consent, and patient engagement in implementation and data use.

And so while many scientists, pharmaceutical companies, and health care providers aim to make medicine more personalized or precise by looking into our genome, we were taught in medical school that “high powered microscopy is low powered pathology.” In contrast, social mediomics might allow us to zoom out and see people's lives and behaviors in addition to their DNA variation.

Mining social media data is a new frontier for personalized or precision[11] medicine. It was unreachable ten years ago. Now people are revealing enormous amounts about their situational and behavioral phenotypes in electronic formats that are aggregatable, searchable, and analyzable in scales that the study of genomes hasn't yet achieved. The rough equation goes something like this:

$$\text{Social Media} \times \text{Big Data Analytics} \approx \text{Biobanking} \times \text{Computational Genetics}.$$

This approach isn't new to the retailers. Buy something on Amazon and you will be told that those who bought that product also bought some other suggested products. Those who buy Big Macs might also enjoy Atorvastatin. Those who tweet or search for the word "heart disease" might also be at risk of it.[12] If it takes widespread biobanking and computational genetics to open up a window into our genome and benefit human health, what will it take to open up the potential of social media data so that it is also useful for health and health care?

First, we need to enhance our ability to collect and interpret information from social media sources.[4] Whereas genetic information is collected from individuals, social media information is already out there in digital form. But the content of social media data (posts, images, social associations, locations), their format (tone, frequency), and their source (typed in by a user, or passively uploaded from a wearable device) are highly complex. Social media signals don't share a common data structure like a nucleotide code. They are also rapidly changing. Facebook was created in 2004, Twitter in 2006, Snapchat in 2011. If our genes are relatively immutable, social media information is constantly adapting and we need to be able to adapt along with it by constructing databases that accommodate data from existing digital platforms and those of the future.

Second, we need to link the information available from social media sources with the validated clinical data that typically reside in electronic health records or insurance claims databases. Social media clues that a patient with congestive heart failure is at risk of another hospitalization might come in the form of searches about salty foods or posts of swollen legs, but the validity and strength of those associations can be tested only if media samples are linked to evidence of emergency department visits or inpatient admissions.[13]

Third, we need to turn associations and observations into interventions that prevent and cure. There is reason for optimism here, because we are not searching for new molecular targets but instead identifying patients who could benefit from existing approaches to managing conditions like heart disease. The early promise of using social media to advance health is most likely to come from linking patients with existing treatments, just as Amazon helps link customers to existing products.

Fourth, any mining of social media data needs to occur in a way that is transparent to patients' and consistent with their preferences for privacy. Prior work suggests significant variability in how individuals alter their privacy settings and for different types of information shared (e.g. status updates, photos, location).[14] While marketers and others use this data for predicting patient preferences and targeting products, this information is rarely revealed to the social media user. This is important considering the impact that a social media based intervention can have on individuals' actions and emotions-as revealed in a study of emotional contagion through altering Facebook feeds.[15] Needed are clear guidelines for health research with regard to how to best engage patients in a way that is ethical and generative and helps patients in understanding their digital footprints while also accelerating new research.

It is a cliché to talk about how social media are reshaping our world, and it may not be true. Social connections and communications have always been a central part of human history.

What is different now is that an increasing amount of the person to person communication that was once private and fleeting—casual conversations or word-of-mouth advice passed along and never to be heard again by others—is now externally observable, enduringly recorded, and available for analysis. Patients can also be consented to access this information. Progress in human genetics has offered up new targets for discovery and cure and the promise of precision medicine[11] based on the genome. While much of the traffic along social media channels might seem superficial, or lacking the gravitas of our genome, mining the social media data for exposures and behaviors has the promise of radically personalizing the delivery of health care. Just as resources have been marshaled to mine the human genome, with equal zeal we should mine the social mediome in a manner that is transparent to patients and pursue the discoveries and cures that can come from that.

Acknowledgments

none

Funders: none

References

1. Young SD. Trends Microbiol. 2014; 22:601–602. [PubMed: 25438614]
2. Harris JK, et al. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep. 2014; 63:681–685. [PubMed: 25121710]
3. Bardach NS, et al. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ quality & safety. 2013; 22:194–202.
4. Padrez KA, et al. The Patient Diarist in the Digital Age. J Gen Intern Med. 2014
5. Grande D, et al. The importance of purpose: moving beyond consent in the societal use of personal health information. Ann Intern Med. 2014; 161:855–862. [PubMed: 25506854]
6. Deters FG, Mehl MR. Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment. Soc Psychol Pers Sci. 2013; 4:579–586.
7. Moreno MA, et al. A pilot evaluation of associations between displayed depression references on Facebook and self-reported depression using a clinical scale. J Behav Health Serv Res. 2012; 39:295–304. [PubMed: 21863354]
8. Park S, et al. Activities on Facebook reveal the depressive state of users. J Med Internet Res. 2013; 15:e217. [PubMed: 24084314]
9. Tandoc EC, et al. Facebook use, envy, and depression among college students: Is facebooking depressing? Comput Hum Behav. 2015; 43:139–146.
10. Christensen H, et al. E-Health Interventions for Suicide Prevention. Int J Env Res Pub He. 2014; 11:8193–8212.
11. Collins FS, Varmus H. A new initiative on precision medicine. The New England journal of medicine. 2015; 372:793–795. [PubMed: 25635347]
12. Eichstaedt JC, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci. 2015; 26:159–169. [PubMed: 25605707]
13. West R, et al. From cookies to cooks: insights on dietary patterns via analysis of web usage logs. WWW, May 13-17, Rio de Janeiro Brazil. 2013
14. Saeri AK, et al. Predicting Facebook users' online privacy protection: risk, trust, norm focus theory, and the theory of planned behavior. The Journal of social psychology. 2014; 154:352–369. [PubMed: 25154118]
15. Kramer AD, et al. Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:8788–8790. [PubMed: 24889601]