



HHS Public Access

Author manuscript

Acad Med. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

Acad Med. 2016 April ; 91(4): 469–472. doi:10.1097/ACM.0000000000001022.

Transforming Scientific Inquiry: Tapping Into Digital Data by Building a Culture of Transparency and Consent

Robert J. Smith,

Fourth-year medical student and research fellow, Penn Medicine Social Media and Health Innovation Lab, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

David Grande, MD, MPA, and

Assistant Professor of Medicine, Perelman School of Medicine, and Senior Fellow and Director of Policy, Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania.

Raina M. Merchant, MD, MSHP

Assistant Professor of Emergency Medicine and Director, Penn Medicine Social Media and Health Innovation Lab, Perelman School of Medicine, and Senior Fellow, Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania.

Abstract

With over 1.7 billion individuals engaged in social media, patients and consumers share more about their lives than ever before through wearable devices, smart phone applications, and social media outlets. This cornucopia of data offers significant opportunity for health researchers and clinicians to track and explore how digital presence contributes to patients' health outcomes and use of health care resources. While patients readily share their information with online communities, it is imperative that they maintain a sense of autonomy over who has access to such data. Recent data breaches of major insurance companies and retailers illustrate the challenges and vulnerabilities related to information safety and privacy. Many Websites and mobile apps require users to agree to data policies, but how those data are mined, protected, utilized, and externally shared is frequently non-transparent, resulting in a climate of fear and distrust around all forums of digital information sharing. While such skepticism is perhaps justified, it should not deter health researchers from attempting to collect and analyze these novel data for the purpose of designing unique health interventions. By clarifying intent around digital data acquisition, simplifying consent procedures, and affirming a commitment to privacy, the authors contend that health researchers can partner with patients to transform the boundaries of scientific inquiry.

Correspondence should be addressed to Raina M. Merchant, Penn Medicine Social Media and Health Innovation Lab, 3400 Spruce, Philadelphia, PA 19104; telephone: (215) 746-8681; Raina.merchant@uphs.upenn.edu.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Whatever, in the course of my practice, I may see or hear (even when not invited), whatever I may happen to obtain knowledge of, if it be not proper to repeat it, I will keep sacred and secret within my own breast.

--The Hippocratic Oath¹

The Hippocratic Oath, uttered so earnestly by students as they enter and leave medical school, calls physicians to hold sacrosanct the privacy of every patient in all of its forms. In today's digital era, a patient's health and non-health information is more accessible and readily viewable than ever before, blurring the definition of what we previously called "private". The potential for capitalizing on this electronic data for the purposes of health research and intervention is a tantalizing prospect. Although there will be logistical challenges to protecting individual patient information, we believe the potential benefits are worth the effort.

For example, an individual's digital food and exercise log created in a free smartphone application could integrate with his medical record to demonstrate correlations with changes in body mass index (BMI), blood pressure, and hemoglobin A1c.² The first sign of an impending manic episode or suicidal ideation in a patient with confirmed bipolar disorder could be caught by tracking and processing subtle changes in her language patterns on Twitter or Facebook.³ Similarly, though not yet a reality, we could imagine day when a sudden rise in someone's online comments related to weapons or anger could trigger a digital violence prevention intervention.⁴ These examples illustrate meaningful clinical applications of digital information, but the full potential of these data will only reveal itself if patients become active, consenting partners in the process.

We have an opportunity as health researchers to actively request permission from patients to access their data in a manner that is safe, consensual, and transparent. By establishing a culture of transparency and consent around research and the acquisition of digital data, we harness the reality of the digital era to engage patients meaningfully in the world of health innovation. To establish such a culture, we must address the challenges and risks associated with data sharing, emphasize our commitment to patient privacy, reiterate the importance of patient consent, and call attention to the scope of research opportunity available to us.

The Scope of Data Sharing and Its Associated Risks

Over 1.7 billion individuals now access social media websites such as Twitter, Facebook, Pinterest, Instagram, and Snapchat, sharing more personal information with an exponentially larger network of individuals than ever before.⁵ Individuals regularly relinquish pieces of their "private" information, allowing companies the capability to access their demographics, pictures, and contacts. We are often reminded by news media that every digital move leaves a mark—even Google searches and wearable device updates (e.g., Fitbits) contribute to the individual user's summative digital footprint.⁶ We recognize that the presence of so much digital information in our society has come at the price of some troubling growing pains. In the wake of the recent digital security breach of health insurance giant Anthem that left 80 million patient medical records exposed in the United States (affecting 1 in 9 Americans), the concern for safety of medical information is justified.⁷ In 2014, there were notable high-

profile digital security breaches involving Sony, Target, Home Depot, Android, and Apple, leaving previously “private” consumer information vulnerable to misuse.^{8,9} Similarly, a 2010 report showed that among 101 common smartphone apps, 56 transmitted unique information about the phone and its user, such as age, gender, and location, to external third parties without the user having a mechanism to “opt out” of participation.¹⁰ To consider how we might capitalize on using digital data within health research, it is imperative that we remain mindful of the current and historical risks associated with such availability of information.

Committing to the Protection of Privacy

The efficiency and apparent ease provided by the online exchange of information comes at the price of potentially leaving both consumers and patients exposed. Since the stakes are so high, however, regulations around the development of newer electronic medical records have, in fact, strengthened existing security and privacy requirements under the Health Insurance Portability and Accountability Act (HIPAA), as set forth by the American Recovery and Reinvestment Act of 2009.¹¹ In reaction to anecdotal reports of violations of HIPAA regulations on social media by healthcare providers, such as the case of an emergency responder posting pictures of patient injuries on Facebook, both legal and medical professionals have emphasized the legally protected status of patient information across all platforms, including social media.^{12–14} New “Hippocratic Databases” enable health enterprises to comply with privacy laws without impeding the sharing of data. Encryption and anonymization techniques can allow for data mining across a health database without revealing identities of individual patients.¹⁵ In contrast, regulations around non-health consumer privacy standards are less clear.

Some might posit that, given the risks, patients would be reluctant to allow their personal data to be shared with researchers. For many patients, however, the idea of their online health and non-health information being used for research purposes is acceptable and permissible, depending on the scope of the data shared and the nature of the recipient.^{16,17} Insights from work in the burgeoning field of mobile health (mHealth) technologies suggest that patients are becoming more willing to share information collected through mobile sensors and other devices. Factors that affect willingness to share this information include the type of data shared, whether the shared information is seen by the patient prior to sharing, whether the patient has control over specific pieces of data shared, and whether the patient perceives that there is some inherent benefit to sharing.¹⁸ Patients desire a careful, selective approach in how health information is shared from their medical record, but they are also more willing to share health data with clinicians than with any other entity.¹⁹ The rise of the “quantified self” movement shows that patients are already taking strides to actively track their own diet, exercise, and other health patterns.^{20,21} This movement suggests that if health researchers were transparent about their intent regarding data acquisition and its potential use, patients would be open to “opting in,” sharing their digital information, and learning about what the data within their aggregated digital footprints illustrate.²²

Consent and Transparency Enable Data Sharing

Although people may be aware of how much personal information they make available online, they still blanch at the unanticipated manipulation of their data. A recent study of how emotions can be spread on social media created a significant amount of backlash in the lay press.²³ In this study, Facebook researchers altered the algorithm for showing specific news stories or “posts” to 689,000 individuals (<0.1% of total users), manipulating the percentage of positive- or negative-valence material that appeared for a given user.²⁴ They then examined the content of subsequent posts written by the user, ultimately showing that individuals exposed to less positive content were more likely to post negative material, and vice versa. Similarly, the taxi service Uber created a stir in 2014 by showing that they are able to predict infidelity, job interviews, and chemotherapy treatments of their consumers just by examining deviations from users’ travel trends.²⁵

Why were these social and psychological studies so controversial? The public outcry revolved largely around the issue of *lack* of consent, of social media users and consumers being manipulated as research subjects without knowing they were participating in the experiment.²³ When humans choose to participate in research, they want to make the active decision to do so. When they are otherwise included without providing consent, they feel akin to a lab rat, stripped of agency and control. Even as we share the mundane facets of our lives on social media and through mobile applications, people hold tightly to the concept that information shared still belongs to the individual sharing. Individuals frequently abdicate those rights by consenting to privacy agreements and other Terms of Use policies, but whether or not they have true understanding of the content and implications of those agreements is another matter entirely. To the credit of online developers, it should be noted that many company privacy policies do contain explicit language regarding their rights to collect, share, and target user information for the purposes of advertising. However, agreeing to convoluted legal jargon does not truly equate to informed consent. Some companies, such as Facebook, have recognized this dilemma and have moved to providing simpler, cleaner privacy policies.²⁶

As health researchers and clinicians know, a patient’s signature on a written consent document (or an online privacy policy, for that matter) should merely be a tangible proxy for that participant’s understanding of a medical procedure or research protocol. Since the Institute of Medicine’s 2004 publication *Health Literacy: A Prescription to End Confusion*, there has been increased emphasis on the importance of addressing health literacy as a broader entity that encompasses informed consent.²⁷ A report on the progress of this initiative suggests that most people, regardless of literacy levels, are unable to remember or understand what is presented to them in the informed consent process for medical procedures. Many areas of needed improvement in the consent process have been identified, such as simplifying text, encouraging talk-back with patients, shortening consent procedures, and transitioning to visual aids.^{28,29} We can use the challenges experienced within clinical medicine to inform how we approach consent within future, large-scale, digital research.

Patients yearn for transparency with how their data are used.^{16,30} In the world of marketing and advertising, consumer data are mined constantly to target advertising to ideal buyers, largely without an individual consumer's specific knowledge or consent. The Uber and Facebook case studies, among others, raise many questions around data, including what we truly are sharing, how we are consenting, who is monitoring the data collected, and how exactly those data are being used. Instead of staking claim to data that do not belong to us, we have an opportunity as health researchers to actively request permission from patients for access to their data in a manner that is safe and transparent, thereby creating a partnership of trust. A necessary corollary to this idea is that we must only use those data in the manner, time frame, and indication specified within the consent process. We will need to examine the ramifications of having digital data available beyond the specific scope of a given research project. In the current landscape, however, it is still imperative that research groups dispose of data after concluding the terms of their scientific inquiry. Similarly, if novel methodologies emerge for identifying patterns within data that alter the original scope of the consent process, we must have systems in place to inform patients of such changes.

Reasons to Embrace the Opportunity

The reasons for us to engage in this conversation are plentiful. Previously unattainable patient data will inform how regular, routine activities may contribute to the development, maintenance, and prediction of disease. By amassing and refining mobile and social media data from a large group of individuals, we can develop algorithms to predict the development of disease, predict healthcare utilization, track the spread of infectious illness, study health communication, and target public health messaging interventions. "Big data" constructs a platform from which to orchestrate a shift in our typical approach to healthcare research. Traditionally, clinical trials and evidence-based research are hypothesis-driven in their approach, using deductive reasoning to direct the answers to scientific questions. Instead, the depth and breadth of online data allows for an exploratory, hypothesis-generating approach to health research, using inductive reasoning to gather patterns of information about health and health behavior. A recent survey by the PEW Research Center found that 70% of Americans "track" a health indicator for themselves or a loved one in some capacity.³¹ With over 70 million units of fitness tracking devices shipped in 2014, it is evident that individuals are beginning to embrace digital health tracking through mobile phone applications (e.g., Apple iHealth), "smart" watches (e.g., the Charge), and light-weight, portable activity monitors (e.g., the Nike FuelBand).^{21,32} An anticipated criticism of using massive quantities of data for research is will be how to "hone the signal" – how will we determine the difference between white noise and meaningful health-related information? While this is no small feat, the giants in computer science, health informatics, and industry are paving that data-driven road, using algorithms to predict the spread of influenza using Google searches, track earthquake magnitude with Jawbone data, and capture Ebola hysteria on Twitter.^{33–35}

Conclusion

Barring an apocalyptic collapse of the digital era, the online, cloud-based sharing of information will continue to expand. We do not yet know what information will be most

helpful for guiding the advancement of disease diagnosis, health maintenance, and medical intervention. It would be shortsighted of us, however, to ignore the potential wealth of health information contained within this fount of data. It is evident that a vast number of third parties collect information about patients and consumers, and will continue to do so, whether or not individuals truly desire their digital data to be gathered. We can, instead, create a culture of transparency that explicitly engages, enables, and empowers patients to become active participants in the digital era of health research. Clinicians and health scientists alike have taken strides in the past to protect patient privacy, but until now, we have only had to focus on health information. By forging a culture of proactive transparency and consent across data-gathering platforms in the digital age, we preemptively remove the worry that we might be wrongfully eavesdropping on patients and consumers. Instead of viewing privacy as a stumbling block when it comes to gathering data for research, we should see it as an opportunity to engage participants more fully in the types of research that can improve the health of all people in important and lasting ways.

Acknowledgments

Funding/Support: R.M. Merchant is funded by a K23 Grant from the National Institutes of Health, #10714038, and by research grant R01 HL122457.

References

1. Edelstein, L. *The Hippocratic Oath: Text, Translation, and Interpretation*. Baltimore, MD: The Johns Hopkins Press; 1943.
2. Sen AP, Sewell TB, Riley EB, et al. Financial incentives for home-based health monitoring: a randomized controlled trial. *J Gen Intern Med*. 2014; 29(5):770–777. [PubMed: 24522623]
3. Jashinsky J, Burton SH, Hanson CL, et al. Tracking suicide risk factors through Twitter in the US. *Crisis*. 2014; 35(1):51–59. [PubMed: 24121153]
4. Hawkins JW, Pearce CW, Skeith J, Dimitruk B, Roche R. Using Technology to Expedite Screening and Intervention for Domestic Abuse and Neglect. *Public Health Nurs*. 2009; 26(1):58–69. [PubMed: 19154193]
5. [Accessed October 9, 2015] Social Networking Reaches Nearly One in Four Around the World. eMarketer Report. 2013. <http://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976>
6. Dwoskin, E. [Accessed October 9, 2015] Where Were You 3 Minutes Ago? Your Apps Know. *The Wall Street Journal*. <http://blogs.wsj.com/digits/2015/03/23/where-were-you-3-minutes-ago-your-apps-know/?mod=LS1>. Published March 23, 2015
7. Barbash, F.; Phillip, A. [Accessed October 9, 2015] Massive data hack of health insurer Anthem potentially exposes millions. *The Washington Post*. <http://www.washingtonpost.com/news/morning-mix/wp/2015/02/05/massive-data-hack-of-health-insurer-anthem-exposes-millions/>. Published February 5, 2015
8. Cieply, M.; Barnes, B. [Accessed October 9, 2015] Sony Cyberattack, First a Nuisance, Swiftly Grew Into a Firestorm. *The New York Times*. <http://www.nytimes.com/2014/12/31/business/media/sony-attack-first-a-nuisance-swiftly-grew-into-a-firestorm-.html>. Published December 30, 2014
9. Perlroth, N. [Accessed October 9, 2015] Home Depot Data Breach Could Be the Largest Yet. *The New York Times*. <http://bits.blogs.nytimes.com/2014/09/08/home-depot-confirms-that-it-was-hacked/>. Published September 8, 2013
10. Thurm, S.; Kane, IT. [Accessed October 9, 2015] Your Apps Are Watching You. *The Wall Street Journal*. <http://online.wsj.com/articles/SB10001424052748704694004576020083703574602>. Published December 17, 2014

11. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011; 4:47–55. [PubMed: 22312227]
12. Rorer, SS. [Accessed October 9, 2015] Social Media Compliance Challenges?: From HIPAA to the NLRA. 2013. https://www.healthlawyers.org/Events/Programs/Materials/Documents/HHS13/Z_rorer.pdf.
13. Spector N, Kappel DM. Guidelines for using electronic and social media: the regulatory perspective. *Online J Issues Nurs*. 2012; 17(3):1. [PubMed: 23036057]
14. [Accessed October 9, 2015] New York paramedics post helpless patients' photos online. RT USA. <http://rt.com/usa/new-york-paramedics-photos-179/>. Published April 1, 2013
15. Agrawal R, Johnson C. Securing electronic health records without impeding the flow of information. *Int J Med Inform*. 2007; 76(5–6):471–479. [PubMed: 17204451]
16. Harris Interactive Inc. Many U.S. Adults Are Satisfied with Use of Their Personal Health Information. Rochester NY: <http://www.harrisinteractive.com/vault/Harris-Interactive-Poll-Research-Health-Privacy-2007-03.pdf>. Published March 27, 2007 [Accessed October 9, 2015]
17. Cutrona SL, Roblin DW, Wagner JL, et al. Adult Willingness to Use Email and Social Media for Peer-to-Peer Cancer Screening Communication: Quantitative Interview Study. *JMIR Res Protoc*. 2013; 2(2):e52. [PubMed: 24287495]
18. Prasad, A.; Sorber, J.; Stablein, T.; Anthony, D.; Kotz, D. Understanding User Privacy Preferences for mHealth Data Sharing. In: Adibi, S., editor. *mHealth Multidisciplinary Verticals*. Boca Raton, FL: CRC Press; 2014. p. 545-570.
19. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. *J Am Med Inform Assoc*. 2013; 20(1):7–15. [PubMed: 23184192]
20. Patel MS, Asch DA, Volpp KG. Wearable Devices as Facilitators, Not Drivers, of Health Behavior Change. *JAMA*. 2015; 313(5):459–460. [PubMed: 25569175]
21. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data. *JAMA*. 2015; 313(6):625. [PubMed: 25668268]
22. Kumar S, Nilsen WJ, Abernethy A, et al. Mobile health technology evaluation: the mHealth evidence workshop. *Am J Prev Med*. 2013; 45(2):228–236. [PubMed: 23867031]
23. Goel, V. [Accessed October 9, 2015] Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. *The New York Times*. <http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html>. Published June 29, 2014
24. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci*. 2014; 111(24):8788–8790. [PubMed: 24889601]
25. Lowrey, A. [Accessed October 9, 2015] Uber Promises It's Not Looking at Your Data, Except When It Wants To. *New York Magazine*. <http://nymag.com/daily/intelligencer/2014/11/uber-were-not-looking-at-your-data-except.html>. Published December 19, 2014
26. [Accessed October 9, 2015] Facebook. Data Policy. 2015. <https://www.facebook.com/policy.php>.
27. Nielsen-Bohlman, L.; Panzer, AM.; Kindig, D. *Health Literacy: A Prescription to End Confusion*. Washington, DC: Institute of Medicine Committee on Health Literacy. National Academies Press; 2004. <http://www.ncbi.nlm.nih.gov/books/NBK216032/>. [Accessed October 9, 2015]
28. Schenker Y, Meisel A. Informed consent in clinical care: practical considerations in the effort to achieve ethical goals. *JAMA*. 2011; 305(11):1130–1131. [PubMed: 21406651]
29. Aldoory, L.; Horowitz, H.; Rouhani, A. [Accessed October 9, 2015] Best Practices and New Models of Health Literacy for Informed Consent: Review of the Impact of Informed Consent Regulations on Health Literate Communications. 2014. Available at: http://www.iom.edu/Activities/PublicHealth/HealthLiteracy/~media/Files/ActivityFiles/PublicHealth/HealthLiteracy/Commissioned-Papers/Informed_Consent_HealthLit.pdf.
30. Stone, J. [Accessed October 9, 2015] Why Transparency And Data Sharing In Clinical Trials Matters. *Forbes Online*. 2015. <http://www.forbes.com/sites/judystone/2015/01/15/why-transparency-and-data-sharing-in-clinical-trials-matters/>. Published January 15, 2015

31. Fox, S.; Duggan, M. [Accessed October 9, 2015] Tracking for Health. 2013. <http://www.pewinternet.org/2013/01/28/tracking-for-health/>. Published January 28, 2013
32. Agomuo, F. [Accessed October 9, 2015] Fitbit Dominates The Wearables Market, But Can It Survive The Coming Onslaught Of Smart Watches?. International Business Times. <http://www.ibtimes.com/fitbit-dominates-wearables-market-can-it-survive-coming-onslaught-smart-watches-1725780>. Published November 19, 2014
33. Hotz, RL. [Accessed October 9, 2015] Decoding Our Chatter. The Wall Street Journal. <http://online.wsj.com/news/articles/SB10001424052970204138204576598942105167646>. Published October 1, 2011
34. Issa NT, Byers SW, Dakshanamurthy S. Big data: the next frontier for innovation in therapeutics and healthcare. *Expert Rev Clin Pharmacol*. 2014; 7(3):293–298. [PubMed: 24702684]
35. Pentland A, Lazer D, Brewer D, Heibeck T. Using reality mining to improve public health and medicine. *Stud Health Technol Inform*. 2009; 149:93–102. [PubMed: 19745474]