



CALIFORNIA HEALTHCARE FOUNDATION



Here's Looking at You:
How Personal Health Information
Is Being Tracked and Used

JULY 2014

Contents

About the Author

Jane Sarasohn-Kahn, MA, MHSA, is a health economist and adviser who works with a broad range of stakeholders at the intersection of health and technology. She writes the Health Populi blog.

Acknowledgments

Thanks to these experts who were interviewed for this paper and generously shared their wisdom on the topics of people, data, health and privacy: Ryan Beckland, Validic; Kipp Bradford, KippKits; Mary Cain, HT3; Steven Dean, Prehype; Stephen Downs, Robert Wood Johnson Foundation; David Goldsmith, Dossia; Harry Greenspun, MD, Deloitte; Leslie Kelly Hall, Healthwise; David Harlow Esq., The Harlow Group; Patricia Hyle, Healthwise; Fard Johnmar, Enspektos; Basel Kayyali, McKinsey; Greg Matthews, W20; Deven McGraw, Manatt, Phelps & Phillips LLC; Glen Moy, technology consultant; Shawn Myers, Healthwise; Mikki Nasch, The Activity Exchange; Jody Ranck, health technology advisor; Harry Reynolds, IBM; Robert Rowley MD, FlowHealth; Mary Anne Sterling, Connected Health Resources; Christine Sublett, Sublett Consulting; Fred Trotter, O'Reilly Radar; Jon Wald MD, RTI; and, Paul Wicks, PatientsLikeMe.

Mary Cain of HT3 and Paul Wicks of PatientsLikeMe also reviewed sections of this report.

About the Foundation

The **California HealthCare Foundation** works as a catalyst to fulfill the promise of better health care for all Californians. We support ideas and innovations that improve quality, increase efficiency, and lower the costs of care. For more information, visit www.chcf.org.

© 2014 California HealthCare Foundation

- 4 The Sharing of Personal Data
- 6 Opportunities for Data-Mining
- 7 The Privacy Challenge
- 10 Balancing Benefits and Challenges
- 11 A Preferred Future
- 12 Endnotes

“There is no doubt that big data and the analytics of big data and the discoveries that it will enable are a fundamental pillar of enabling personal health for all.”

— Eric Dishman, Intel

“Electronic health information is like nuclear energy. If it’s harnessed and kept under tight control, it has potential for good. But if it gets out of control, the damage is incalculable.”

— Jim Pyles, Esq., Powers Pyles Sutter & Verville¹

Every day, most of us leave our digital “footprints” in the Internet. We constantly generate these bits of data about ourselves in the course of using cell phones, credit cards, search engines, and websites.² Many of the footprints directly relate to health. They include:

- ▶ Googling health-related terms or using online social networks for health information.³
- ▶ Posing questions to physicians on the Internet or participating in health-related social media sites like PatientsLikeMe, WEGO Health, and Migraine.com.
- ▶ Tracking activity, eating, exercise, clinical data, and other personal health information via mobile phone apps and digital health and medical devices such as weight scales and blood pressure monitors.
- ▶ Purchasing health and medical products through brick-and-mortar and online retailers using credit cards or store coupons.
- ▶ Passively communicating personal information such as location and retail store preferences through GPS sensors integrated into smartphones, and retailers’ growing use of geo-location technologies like iBeacon.⁴

- ▶ Buying fast food, cigarettes, or anything that could have an impact on health.⁵

With the advent of consumer-facing technology companies like Apple, Google, Microsoft, and Samsung expanding offerings in health, people will have more “on ramps” for generating health and other personal data. Aggregated and analyzed, these data flows have the potential to paint a detailed health profile of individuals, as well as to describe whole communities based on location, health conditions, or other factors.

This report provides a broad overview of emerging issues associated with consumer-generated health data. Based on interviews with technology and health care experts, and additional research, it looks at the use of consumer-generated data to enhance both individual and public health. The report also examines the privacy implications that are increasing in importance as health care enters the era of “big data”—in which the proliferation of extremely large databases and new opportunities for combining data challenge regulators’ and society’s ability to assure individuals’ data rights and privacy.⁶

Stephen Downs of the Robert Wood Johnson Foundation observed, “While big data crunching has been going on on some level for a while, the data sources are much more comprehensive now, and analysis techniques are getting better and better.” The data translate into billions of dollars for many types of enterprises, in the form of new markets, new products, and cost-savings. The enormous scale of data collection is hard for many people to envision. (See box, page 4.)

These developments are important to individuals and patients because most people are not aware of the amount of information they are leaving behind that is not covered by HIPAA⁷ or any privacy rules. Without such protection, different kinds of data are being combined and used by third parties in ways that consumers might not anticipate, and some would not want.⁸

Big Data Changes Everything

The definition of “big data” is evolving, but it is often described in terms of the three V’s — volume, velocity, and variety — all of which are fast increasing and have compelling implications for personal and public health according to Gartner research.

In terms of volume, datasets are now being described in petabytes. One petabyte is 1 quadrillion bytes, enough to store the DNA of the entire population of the US, and then clone them twice. According to an article in *The Economist*, Walmart had 2.5 petabytes worth of consumer data in 2010.

The velocity factor (the speed at which data are generated and consumed) and the variety factor (the many types of information) are equally compelling. Because data are far less “siloes” than in the past, they frequently can be processed, analyzed, and viewed virtually in real time.

Much of the data come from non-traditional sources, such as social media and GPS, and can have a variety of structures, such as video, images, and free text, in addition to traditional database formats. Data are being gathered by information-sensing mobile devices (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks.

Sources: Gartner, Big Data, 2014. Data, data everywhere. *The Economist*. 25 February 2010. Parallel Programming in the Age of Big Data. Joe Hellerstein, gigaom, 2008. Segaran, Toby; Hammerbacher, Jeff (2009). *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly Media. p. 257.

The Sharing of Personal Data

Two rapidly growing areas of consumer-generated health data involve social networks and digital health devices combined with mobile apps. Through social networks such as Facebook, Twitter, and YouTube, people share health experiences, side effects of drugs, and participate in research studies.⁹ The purchase of wearable fitness devices, such as smartwatches and digital activity trackers, is also proliferating, tripling among Americans from 3% in 2012 to 9% in 2013.¹⁰ The supply of digital and connected health devices is fast-growing: the Consumer Electronics Show had 40% growth in the convention’s digital health footprint in 2014, following 25% growth in 2013.¹¹ A 2014 Accenture survey found that 54% of adult consumers were interested in buying a health monitor as an application on a phone, laptop, desktop, or as a dedicated device.¹²

Here are a few scenarios that illustrate ways that consumer-generated data are being combined with retail, location, and other information to support better health:

- ▶ David, who is managing type 2 diabetes, uses a mobile nutrition app to buy groceries, a digital glucometer for tracking blood glucose, a Fitbit (digital activity device) to gauge metabolic rate, and a blood pressure monitor that syncs with his computer. His health and retail data are routed to a health management provider that analyzes it in real time and pushes advice back to David: for example, a nudge to take a walk or a reminder to test his blood sugar. A medication adherence program reminds him to take and re-order meds at the right time.
- ▶ Rosa, diagnosed with a rare cancer, participates in a clinical trial through a web-based matching program, into which she enters her data every day. To help manage her illness she also records her mood in a mobile health app, uses a meditation-biofeedback app, and journals her daily food intake. These data are aggregated with those of many other patients for researchers to mine in developing a cure for her type of cancer. The trial is of shorter duration than it would have been because of the large quantity of data made available by patients in a quicker time-frame.
- ▶ Renee uses her mobile phone to check into an online patient social network as she manages her migraines

and depression. She refills medicines online, and purchases over-the-counter pain relievers from a grocery store, collecting loyalty program points and mobile coupon discounts. A mobile app provides virtual counseling (telemental health) with a third-party mental health provider monitoring her via location data from her smartphone. If the provider notices a problem—for example, that she has stayed home for several days in a row — they will check on Renee by phone or text message.

- ▶ Anthony goes to the gym at 5:30 every morning, using a digital fob to record his visit with his employer's wellness program provider. If he meets the program's exercise targets, he receives a discount on the insurance premium. The wellness program has access to his grocery store and other retail receipts. The program provides him with incentives for healthy eating, including mobile coupons for health food stores.
- ▶ Kelly keeps tabs on her children's health on her mobile phone: immunizations, well visits, and food tracking. She often responds to Mom blogs online. She records her exercise through a digital device and participates in a chain pharmacy's health loyalty program, which enables her to accumulate points for store discounts. A third-party data company acquires Kelly's retail data and matches it against her social media activity, then sends coupon offers to her mobile phone tailored to the members of her family.
- ▶ Thomas, who is homebound with congestive heart failure, uses a remote health monitor and a Wi-Fi-enabled scale. His daily weight is routed to a cardiac health management company contracted through his Medicare Advantage program. He buys groceries online. Based on the health and retail data, the company routes personalized health videos to Thomas via his cable TV box and computer, and a cardiac specialty nurse video-chats with him once a week.

"You can buy from any number of data brokers, by malady, the lists of individuals in America who are afflicted with a particular disease or condition."

Much of the personal data collected from consumers would not be considered "health information," yet is used for a health purpose.¹³ Deven McGraw of Manatt, Phelps & Phillips noted that "the FICO Medication Adherence Score is not based on health data but is instead an algorithm based on other types of demographic data that turn out to be highly predictive of medication adherence." McGraw warned that "Digital dust can have health implications even if the 'dust' is devoid of actual health information. FICO and other 'scores' could have significant implications for consumers — arguably as significant as a score generated using health data," she said.¹⁴

Even consumer footprints that are not expressly about health can be used to help determine a person's physical or mental health. How we shop, the magazines we subscribe to, where we hang out on the weekend—this information is relatively easy to purchase by third parties, as the investigative reporters of "60 Minutes" revealed in their 2014 story, "The Data Brokers: Selling Your Personal Information."¹⁵ In the report, Tim Sparapani said, "You can buy from any number of data brokers, by malady, the lists of individuals in America who are afflicted with a particular disease or condition." Sparapani, once director of public policy at Facebook, currently advises companies as vice president of the Application Developers Alliance.

According to Dr. Harry Greenspun of Deloitte, "changes in your life — such as a trend in your retail behavior that you've stopped shopping at the home improvement store — could indicate your risk of depression is going up. This information can be more predictive of a person's health than anything that's in your EHR or even in your genome."

In fact, research suggests that health care services, such as visits to the doctor, contribute to only about 10% of health outcomes.¹⁶ Health behaviors, on the other hand, determine as much as one-half of health status. People make many choices throughout the course of each day that bolster or diminish their personal health. Thus, behavior change has been called by some the Holy Grail of health care.¹⁷

Healthy behaviors are also key to the effectiveness of health reform, under which providers taking on value-based payment can only flourish when healthy patients stay well and those with chronic illness manage their disease effectively. Payment for health care to doctors and hospitals is shifting from volume-based (such as

visits, tests, and surgeries) to value-based (such as preventing hospital readmissions and effectively managing diabetes).

Opportunities for Data-Mining

There are numerous opportunities for organizations to use consumer-generated data that benefit both public and personal health, including health researchers, health providers, consumer and medical supply companies, technology developers, and public policymakers, among others. Larry Page, Google co-founder and CEO, exhorted people to think about “the tremendous good that could come from sharing the right information with the right people in the right ways.”¹⁸ Following are a few examples of data-mining that are already in practice.

Improving clinical trials. Data-mining offers opportunities to make clinical trials faster and more effective. Working with a biopharmaceutical company to design a recruitment strategy for a Hepatitis C vaccine trial, PA Consulting found 120 million patients that could fit the trial criteria and also learned that these patients were strongly influenced by 10 key individuals who were most active on Twitter in their social networks. The company approached these influencers, who then publicized the clinical trials. Leveraging social networks helped the company reduce the planned recruitment period from six months to five weeks.¹⁹ “Patient-Powered Research Networks” are being developed through PCORnet, the National Patient Centered Research Network, that will bring researchers and patients together – including their self-generated data.²⁰ To encourage data sharing, PatientsLikeMe launched a campaign to promote “Data for Good,” focusing on the value of sharing health information in three respects: to improve one’s personal health, to help other people, and to change medicine.²¹

Managing chronic disease. People with chronic illness have the daily challenge of tracking and managing their health. Propeller Health (formerly known as Asthmapolis) focuses on people with asthma and chronic obstructive pulmonary disease (COPD).²² The company’s FDA-approved GPS-enabled sensor attaches to an inhaler that collects data on the time and precise location that patients experience symptoms. Users receive

feedback from the program via emails, phone calls, or text messages to help understand their asthma triggers. In pilots, users showed an 80% improvement in medication adherence, and over two-thirds got their conditions under control.²³ Propeller can also identify geographic areas with air quality issues and help drive population health in communities.

Forecasting and tracking epidemics. The use of mobile phones with geo-location apps represents a global platform for user-generated data that has been used to anticipate epidemics: cholera in Rwanda, malaria in Kenya, and malnutrition in Tanzania, for example, based on call patterns and mobility patterns.²⁴ Researchers can combine geo-spatial phone data with public health records. In Portugal and Spain, Twitter tweets have been analyzed to understand depression, eating disorders, flu, and pregnancy.²⁵

Harnessing self-tracking data for research. Individuals are often willing to share self-tracking data for research under conditions of anonymity and especially if their personal data would advance knowledge in public health, health care, computer science, and social and behavioral science.²⁶ According to Lauren Solberg of Meharry Medical College, researchers in a number of academic fields are collecting research data through self-tracking.²⁷ Nearly one-half of researchers participating in the Health Data Exploration Project had used self-tracking data, and one-quarter had worked with commercial companies involved in apps, devices, and social media. An alliance struck between PatientsLikeMe, an online patient network, and Genentech will open up the patient community’s data to the biotech field for clinical trials and research projects.²⁸ Much of the data are available in real time.

Focusing on individual health goals. Entrepreneurs are providing benefits to people with a specific health objective by comparing their individual consumer-generated data with a larger dataset of people like them. For example, for women who wish to get pregnant, there is a mobile app, Ovuline, with a database of health information on over 70,000 women. The metrics include body temperature, cervical fluids, menstrual cycles, and frequency of sexual intercourse, among other data points. A woman can compare her own data to those of other women to help her predict when she has the greatest opportunity to conceive.²⁹

Learning from retail. The health care industry can better align to consumers' demands by learning from retail's use of data. "If you analyze the Acxiom (retail) and other data in an area, you know the habits of people, when they shop, where they shop, the characteristics of their neighborhood, the bus schedules," said Harry Reynolds of IBM. As patients take on more financial and decision-making responsibility for health services, they will be looking for greater convenience and levels of service, the kind they find in retail businesses. In one example, a hospital learned from retail and other data that many of its outpatients lived in an area far from its clinic. Patients often cancelled appointments because they had to leave work and travel a long way. The hospital responded by opening a clinic in that neighborhood.

The Privacy Challenge

Despite the benefits of consumer-generated data, there are growing concerns. Computers are making decisions based on algorithms, mathematical models, which are built on assumptions and calculations that are not made public. "When the input data explicitly or implicitly encodes for a protected characteristic like gender or race...the resulting algorithm runs the risk of being biased against certain groups, or in the worst case 'redlining' them," a team from the Electronic Frontier Foundation explained.³⁰ In a nutshell, the algorithm is only as fair as the data it analyzes. Because of that, Fred Trotter of O'Reilly Radar warned, "We are outsourcing decisions to a virtual entity which has none of our humanist notions of fairness or a sense of fair play."

"We are outsourcing decisions to a virtual entity which has none of our humanist notions of fairness or a sense of fair play."

Two recent studies raised data privacy and security issues about digital health tools. Research published in August 2013 by the Privacy Rights Clearinghouse identified privacy risks in 43 of the most popular mobile health and fitness applications.³¹ Another analysis, from Evidon, found that the top 20 health and wellness apps were transmitting information to as many as 70 third-party

organizations.³² Thus, consumers cannot assume their personal data generated when using a health app or device is being de-identified and aggregated by the technology provider. And, as noted previously, none of these data are protected by HIPAA until they become part of a HIPAA-covered entity's records.

According to a recent White House report, "Personal health information of various kinds is shared with an array of firms, and even sold by state governments, in ways that might not accord with consumer expectations of the privacy of their medical data."³³

It is important to note that while most US consumers worry about how marketers will use their personal data, two-thirds rarely or never read an online privacy policy, nor do most consumers use "do not track" or "incognito" features that allow them to opt out of website tracking.³⁴

Sharing personal data can trump privacy risks, particularly among patients dealing with serious medical conditions. In a survey of US social media users with a medical condition, 94% agreed with sharing their health data to help doctors improve care or to help other patients like them.

However, most patients willing to share data also believed that their personal health information could be used without their knowing. Specifically, 72% believed their data could be used to deny them health insurance, and 66% believed it could be used to deny them jobs. As Paul Wicks of PatientsLikeMe explained, "Most patients think there's a possibility their personal health data could be used against them, but when they see the benefits, they do it anyway."³⁵

Surveys conducted by Consumer Reports National Testing and Research Center and PatientsLikeMe found that 94% of social media users with medical conditions would be willing to share data about their health (anonymously) to help improve health care.³⁶ Interestingly, most people don't trust commercial private sector companies' use of their health data (e.g., drug and medical device companies) any less than they trust sharing with public sector/government organizations like public health officials.³⁷

When people engage with a particular organization for health, trust is the most important factor.³⁸ Nurses, pharmacists, teachers and medical doctors are the most trusted professions in America, according to the Gallup

Poll.³⁹ When it comes to peoples' levels of trust in organizations to take care of personal data, hospitals rank ahead of banks, government, retailers, mobile technology and internet companies, and insurers, a FICO survey found.⁴⁰

Consumers have a "split personality" when it comes to their open sharing of social data, at the same time expecting others to protect their privacy — say, government or private sector companies. However, most people don't trust these organizations, private or public, to protect their privacy. EMC, a cloud computing and big data company, calls this the Social Sharing Paradox.⁴¹

There are a number of factors that add to the privacy challenges of consumer-generated data, including the following:

Lack of awareness and recourse. There are hundreds of unregulated data brokers that build dossiers on people and sell them to marketers. "Individuals have little recourse if they want to examine their files or correct mistakes," *The Economist* warned in an article titled, "Maintaining your privacy is harder than you think."⁴² People aren't really aware they're "creating digital data exhaust all day long," Mary Cain of HT3 observed. Each click into Facebook and Google represents a little piece of information tied to that person: "Google and Facebook are running multimillion dollar businesses off of these 'free' services which means 'you' are the product," said David Harlow, a health lawyer. "If you go into that with your eyes open, fine. But many people haven't opened their eyes to that."

Third parties' use of dark data. When consumers are not consciously aware of their data being used, Mikki Nasch of The Activity Exchange refers to this as dark data. She offered an example: "If you are someone who checks in with Foursquare for a week when you're in a bar, you've just told people you've been on a bender for five days.... You are socially bragging about all the things you never say to your insurance company." In real time, she added, people check into social networks in a personal context. However, when these data are used for statistical modeling (for health or financial risk profiling, for example), the information can be taken out of context. An example might be someone chatting on social networks about missing work due to migraines and falling behind with mortgage payments as a consequence: These facts could

be used as dark data for scoring consumers — negatively and out of context — in a variety of ways.

Scoring personal health. The phenomenon of predictive consumer scoring has emerged, largely unregulated by the Fair Credit Reporting Act or the Equal Credit Opportunity Act.⁴³ Consumer scores are now ubiquitous across peoples' activities: financial and credit, energy use, law enforcement, environmental, social clout, tax returns, environmental "green-ness," and health. In 2014, there were at least a dozen health scores available in the marketplace, including the Affordable Care Act (ACA) Individual Health Risk Score, FICO Medication Adherence Score, several frailty scores, personal health scores (e.g., WebMD, One Health Score), and medical complexity scores (e.g., Aristotle for scoring of surgery for congenital health conditions). Consumers are largely unaware of the existence and use of these scores and the algorithms that create them.

Under the ACA, each health plan enrollee is subject to risk adjustment under an Individual Health Risk Score. This score was developed to help health plans manage the risks of adverse selection and predict health costs for an individual enrollee.⁴⁴ The ACA says nothing about individuals having access to their Individual Health Risk Score. The Activity Exchange's Nasch pointed to lessons consumers can learn from the history of the FICO score for consumer credit rating. Although FICO began in the 1950s, "It took 50 years to get that data into consumers' hands," she said. "Now consumers know how to repair their lagging credit scores. The same is not true in health. We have to be aware that the same math used to ascertain risk by a car insurance company and mortgage lender is being used in health, too."

Risk management digs deeper. Consumers' health scores could be useful for providers and payers as they move to value-based payment. A McKinsey & Company analysis points to a trillion-dollar savings over the next decade by tying payments more aggressively to outcomes. "Who is on the hook for the risk of that patient?" asked Basel Kayyali, partner at McKinsey. "That's the stakeholder who will want to embed user-generated data into its data warehouse." In the risk-based era for health care, those who bear the financial risk must manage that risk at a "much more granular level," Kayyali advised. Consumers' behavior will be modeled based on both historical claims data and peoples' user-generated

data. Those data are already being purchased by hedge funds, hospitals, large provider networks, payers, pharmaceutical companies, and others involved in health care, according to Nasch.

HIPAA limitations. While personal health information held by health care providers and insurers is protected under HIPAA, many other sources of consumer data are not covered and can be disclosed to third parties. Unless and until a patient shares user-generated data with a HIPAA-covered entity, that information is not covered by HIPAA. User-generated data that could be used in health profiling are held by gyms, websites, banks, credit card companies, cosmetic medicine groups, fitness clubs, home testing labs, massage therapists, nutrition counselors, alternative medicine practitioners, disease advocacy groups, and marketers of non-prescription health products and foods.⁴⁵ Exhibit 1 lists some of the types of consumer data available for purchase and use in analytics. Latanya Sweeney, chief technology officer at the Federal Trade Commission, found in 2010 that at least one-half of consumer data flows fall outside of HIPAA protection.⁴⁶

De-Identified data may be re-identified. Sweeney also found that anonymous data may not remain anonymous. From a database of 1,130 de-identified participants of a genomic surveillance study, 241 participants were re-identified based on three pieces of information — date of birth, gender, and ZIP code — that were then combined with public records, such as voter registration.⁴⁷ Further, health information may not be secure. HHS has been tracking data breaches since September 2009, when the HIPAA breach notification rule went into effect. The agency reported health information breaches affecting more than 500 individuals.⁴⁸

Listening in for profit. Another form of consumer-generated data is personal check-ins on social networks. “Savvy health care companies have begun launching innovative social media listening programs to aggregate data and to capture critical patient insights,” according to an analyst report.⁴⁹ Hundreds of for-profit companies have developed successful business models to “scrape” social networks. However, the growth and adoption of social networking sites have outpaced public policies designed to protect the interests of consumers.⁵⁰ Data shared on a social networking site can become a “digital tattoo” for a consumer, and it is almost impossible to

Exhibit 1. Consumer Data Available for Purchase and Use in Analytics (partial list)

- | | |
|--|----------------------------------|
| DEMOGRAPHIC INFORMATION | ▶ Fitness devices |
| ▶ Date of birth | ▶ Dieting/weight loss |
| ▶ Education | ▶ Gun ownership |
| ▶ Gender | ▶ Health and medical |
| ▶ Marital status | ▶ Political affiliation |
| ▶ Own or rent home | ▶ Sexual orientation |
| ▶ Ethnicity | ▶ Social media sites |
| ▶ Number and age of children | ▶ Exercise and fitness |
| ▶ Language preference | ▶ Vehicle make, model, year |
| ▶ Religion | |
| ▶ Occupation | FINANCIAL AND CREDIT DATA |
| ▶ Military history | ▶ Income/household income |
| | ▶ Home value |
| CONTACT INFORMATION | ▶ Length of residence |
| ▶ Name | ▶ Payment data: mortgage lates |
| ▶ Home address | ▶ Bankruptcy |
| ▶ Email address | ▶ Credit score |
| ▶ City, state, ZIP | ▶ IRAs |
| ▶ Type of phone | ▶ Life insurance |
| ▶ Carrier | ▶ Investments |
| | ▶ Tax liens |
| LIFESTYLE, INTERESTS, ACTIVITIES | ▶ Card data by brand |
| ▶ Average direct mail purchasing amounts | ▶ Stocks/bonds |
| ▶ Books/audiobooks purchased | ▶ Average online purchase |
| ▶ Charitable giving | ▶ Average offline purchase |
| ▶ Consumer electronics | |

Source: The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future, World Privacy Forum, April 2, 2014, www.worldprivacyforum.org.

remove.⁵¹ “It’s one thing to know you’re on a statin. It’s another thing to know that you eat fast food three times a week. What is more predictive?” asked Harry Greenspun of Deloitte. The prescription for that statin would be covered as privacy-protected personal health information under HIPAA. Social media sites, on the other hand, are not bound by HIPAA.

Balancing Benefits and Challenges

How can third parties deliver on the benefits of consumer-generated data for health while balancing the real risks of individuals' rights to privacy? Following are a few strategies offered by stakeholders.

Help people gain control. Consumers are worried about privacy, research shows, but they feel relatively powerless to take action.⁵² Brian Loew of Inspire, an organization that connects patient communities, observed, "When it comes to information about you, you want to be in control of when it's shared, and how, and with whom."⁵³ Organizations that want to use consumer-generated data in health should recognize that empowerment is an aspect of design. According to Shawn Myers of Healthwise, "The really well-designed applications give the consumer underlying efficiency and an ability to bring order out of the mess — an application that gives one a sense of control and a sense of peace."

Health lawyer David Harlow recommends that the framework for notices of privacy practices of HIPAA-covered entities include check-the-box provisions that give consumers clear, structured choices such as "Yes, please text me" or "I don't care if you email me data as open text, not encrypted. I just want my test results."

Stephen Downs of Robert Wood Johnson Foundation believes that many startups would like a simplified privacy statement for users, and would find value in a template for (privacy) terms that passed muster and that didn't harm their core business models. Downs' sense of the prevailing attitude among smaller start-up companies is that "they'd prefer not to have to spend a lot of money on lawyers."

However, Deven McGraw cautions that obtaining meaningful consent in an environment where the data are shared ubiquitously over and over is very hard. Two recent reports on big data — one from the White House and one from PCAST — both criticize consent as a means of providing meaningful protections for consumers.^{54,55} However, the PCAST report does recommend strengthening choice through the use of privacy templates or frameworks.

Simplify the fragmented regulatory environment. It has become clear that existing laws and policy frameworks have not kept pace with technology. Further, there is no over-arching national law that addresses citizens' privacy. Instead, user-generated data and health information relate to a patchwork of laws and regulations for which responsibility falls into many federal agencies, along with individual state regulations for specific health and privacy issues.⁵⁶ An IOM report on social networks and health research said, "Data protection policies in the near future must evaluate the delicate balance of a free and democratic society that supports the sales of goods and services (based on a user's data) while at the same time protecting people from discrimination."⁵⁷ The Federal Trade Commission called for Congress to consider legislation that would require data brokers to offer consumers tools to view, suppress, and fix their personal information at the source.⁵⁸

Consider personal health data lockers and clouds. Several projects and companies are developing tools for consumers to control their user-generated data. The personal data locker is one such concept. Dr. Robert Rowley, former chief medical officer of Practice Fusion, is developing FlowHealth, a locker that would enable consumers to collect and consolidate data from different online sources. This project, along with health records products Gliimpse and DrChrono, among others, incorporate a "permissions model" for consumers that lets people determine with whom and how they will share their data.

A Preferred Future

Paul Wicks of PatientsLikeMe imagines “a system continuously monitoring your health and tweaking your medication, your behavior, your activity, to maximize your health. It’s a bit like how technology is used to control the performance of a Formula 1 race car,” he said. “So why can’t we do this in medicine for, say, a grandma with Parkinson’s disease? A connected system of sensors, algorithms, and scientists could develop software that realizes it’s Friday, which is when we need more medication on board to get her safely to her weekly game of bingo so the command center suggests to grandma that she ups her dose of medication, and sensors in her smartphone can monitor her journey and alert caregivers if something goes wrong, like a fall.”

Wicks’ vision has similarities to that of Kipp Bradford, a biomedical engineer. “People have two check-ups a year that measure ten to 20 variables,” said Bradford. “I could develop something that says you’re going to get a cold in three days in the same way an airplane diagnostic says you need to change the compressor blades on your jet engine in two weeks.”

However, Bradford acknowledged that having one’s own biological black box “is a big leap from using and tracking with a Fitbit.” The irony, he pointed out, is that it takes a lot of data from a lot of people to help health care get more personalized for the individual. “When you do a clinical trial with 1,000 people you’re taking 1,000 peoples’ physiologic measurements and generalizing that across a billion people. But if you can have a trial with 7 billion people, then you can understand the nuances of the effects on one person,” Bradford explained.

David Goldsmith of Dossia calls this notion “population health made personal.” He anticipates the patient-centered *digital* home (versus the current model of the patient-centered *medical* home) where the growing phenomenon of web-enabled devices meets people in daily life generating data for analyses that will be useful for personalizing health to the individual based on their preferences.

In that future of *extreme personalization* in health, it is the analytical process that calls for *extreme transparency* between the data collectors, data analysts, and the people whose data are being accessed and manipulated. In big data, there are three players: those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have expertise to analyze it.⁵⁹ That third group, the data analysts, is the small group that builds the statistical models that can eventually determine health access, insurance coverage, and therapeutic options for the consumers who knowingly or unknowingly create the data. “Being conscious about how you are generating data about yourself will quantify into behavioral attributes,” Mikki Nasch of The Activity Exchange advised. “This will be your good driving record in the future.”

Fard Johnmar of Enspektos, a digital health communications company, believes that the time is ripe to address digital privacy issues, in part because the recent Edward Snowden revelations about the National Security Agency’s activities raised the issue of digital privacy to a high level of interest. Within eight months of these revelations going public, one in two US adults said they changed their online behavior and think more carefully about where they go, what they say, and what they do online.⁶⁰

“We have financial literacy, we have health literacy, and we need digital privacy literacy, Johnmar said. “We need to educate people to make informed decisions about what they choose to share and not choose to share.”

Endnotes

1. Robertson J. "Your Medical Records Are For Sale." *Bloomberg Business Week*, August 8, 2013.
2. Big Think Editors. Big Data. February 2, 2014. Accessed May 4, 2014 at: www.bigthink.com.
3. Fox S, Duggan M. Health Online 2013. Pew Research Center, January 15, 2013.
4. Panzarino M. iBeacons Usher In The Contextual Shopping Era. *TechCrunch*, June 16, 2014.
5. Pettypiece S, Robertson J. Your Doctor Knows You're Killing Yourself. The Data Brokers Told Her. *Bloomberg*, June 26, 2014, www.bloomberg.com.
6. Gartner. Big Data. Accessed May 4, 2014 at: www.gartner.com.
7. The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, enacted in 1996, is the first comprehensive federal protection for the privacy of personal health information.
8. Fernandez-Luque LN, Elahi N, Grajales F. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. *Studies in Health Technology and Informatics*, 150:292–296, 2009.
9. Ramo D, Prochaska J. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. *Journal of Medical Internet Research*, 14(1):e28, 2012.
10. Consumer Electronics Association. Understanding the Market for Wearable Health and Fitness Devices, December 2013.
11. Consumer Electronics Association. Digital Health Exhibits See 40 Percent Growth at 2014 International CES. Press release, December 2, 2013.
12. Accenture. Racing Toward a Complete Digital Lifestyle: Digital Consumers Crave More. Accenture Digital Consumer Tech Survey 2014, January 3, 2014.
13. New WPF Report — The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future, www.worldprivacyforum.org.
14. Parker-Pope T. Keeping Score on How You Take Your Medicine. *Well*, *New York Times*, June 20, 2011.
15. CBS News, 60 Minutes. The Data Brokers: Selling Your Personal Information. Aired on March 9, 2014. Transcript accessed June 6 2014 at: www.cbsnews.com.
16. Institute for the Future. Health and Health Care 2010: The Forecast, The Challenge. January 2003. Prepared for the Robert Wood Johnson Foundation.
17. Rissel C. Empowerment: The holy grail of health promotion? *Health Promotion International*, 9(1):39–47, 1994.
18. Ha Thu-Huong. Computing is still too clunky: Charlie Rose and Larry Page in conversation, *TED Blog*, March 19, 2014.
19. PA Consulting. The Rise and Rise of Social Influencers – the Most Powerful New Voice in Life Sciences, January, 2014.
20. PCORnet: the National Patient-Centered Clinical Research Network. Accessed June 22, 2014, at: www.pcori.org.
21. PatientsLikeMe press release. PatientsLikeMe Launches "Data for Good" Campaign to Encourage Health Data Sharing to Advance Medicine, March 10, 2014. Accessed April 9, 2014 at news.patientslikeme.com.
22. The California HealthCare Foundation has contributed funding to this organization.
23. Propeller Health press release. Asthmapolis Relaunches as Propeller Health to Advance Broader Respiratory Mission, September 10, 2013. Accessed April 8, 2014, at: propellerhealth.com.
24. Dougherty E. Mobilizing a revolution: how cellphones are transforming public health. *Harvard School of Public Health News*, Winter 2012.
25. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: A Good Place to Detect Health Conditions. *PLoS ONE* 9(1): e86191, 2014.
26. Health Data Exploration Project and the Robert Wood Johnson Foundation. Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health, March 2014.
27. Solberg LB. Regulating Human Subjects Research in the Information Age: Data Mining on Social Networking Sites. *Northern Kentucky Law Review*, Vol. 39, No. 2, 2012.
28. Patients Like Me press release. Genentech and PatientsLikeMe Enter Patient-Centric Research Collaboration, April 7, 2014. Accessed April 8, 2014, at blog.patientslikeme.com.
29. Bertolucci J. Big Data Knows When You're Fertile. *Information Week*, July 12, 2013.
30. Gillula J, Opsahl K, Reitman R. The White House Big Data Report: The Good, The Bad, and The Missing. *Electronic Frontier Foundation*, May 4, 2014.
31. Njje CML. Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications. *Privacy Rights Clearinghouse*, August 12, 2013.
32. Steel E, Dembosky A. Worried well online have new symptom to fear. *Financial Times*, September 1, 2013.
33. The White House, Executive Office of the President. Big Data: Seizing Opportunities, Preserving Values. May 2014.
34. SDL. Marketing Data and Consumer Privacy: What Your Customers REALLY Think, February 26, 2014.
35. Zamosky L. Social Media Users Willing To Share Health Data Despite Concerns About Privacy. *iHealthBeat*, March 10, 2014
36. Grajales F, Clifford D, Loupos P, Okun S, Quattrone S, Simon M, Wicks P, Henderson D. Social Networking Sites and the Continuously Learning Health System: A Survey. *Institute of Medicine*, January 23, 2014.
37. Ibid.

38. Edelman. Health Engagement Barometer: Health Influence in the Era of Public Engagement, October 2008.
39. Gallup. Honest and Ethics in Professions, December 2013.
40. FICO. Mobile Thought Leadership — US, June 2014.
41. Churchward G. On Internet Privacy, Consumers Show Split Personalities. *Forbes*, June 12, 2014.
42. *The Economist*. Watching the watchers: maintaining your privacy is harder than you think, March 1, 2014.
43. Dixon P, Gellman R. The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future, World Privacy Forum, April 2, 2014.
44. Ibid.
45. Ibid.
46. Testimony of Latanya Sweeney, PhD, before the 21st Century Healthcare Caucus Roundtable. April 22, 2010.
47. Dolan, P. Revealed names expose flaw in de-identified patient data, www.amednews.com.
48. Conn J, Major Medical Records Breaches Pass 1,000 Milestone as Enforcement Ramps Up, www.modernhealthcare.com.
49. Best Practices LLC. Social Media Listening: Capturing Critical Pharmaceutical Market Insights. February 22, 2014.
50. Solberg.
51. Enriquez J. Your Online Life, Permanent as a Tattoo. TED2013, February 2013.
52. International Institute of Communications. Personal Data Management: The User's Perspective, Appendix B, Detailed Research Survey Results, March-August 2012. Survey sponsored by Microsoft.
53. Zamosky.
54. Big Data: Seizing Opportunities, Preserving Values, www.whitehouse.gov.
55. Report to the President — Big Data and Privacy: A Technological Perspective, www.whitehouse.gov.
56. The California HealthCare Foundation will be supporting new work by the Center for Democracy and Technology to examine privacy issues in the health realm and propose some solutions.
57. Grajales F, et. al.
58. Federal Trade Commission. Data Brokers: A Call for Transparency and Accountability. May 2014.
59. Manovich L. The Promises and Challenges of Big Social Data. Accessed April 9, 2014 at www.manovich.net.
60. Cobb S. New Harris poll shows NSA revelations impact online shopping, banking, and more. *We Live Security*, ESET newsletter. Accessed June 6 2014 at: www.welivesecurity.com.