Could AI Push Sales of Personal Health Data? How to Protect Consumers While Advancing Science

written by Theresa Hush | February 13, 2020

We are just beginning to see the power of Artificial Intelligence (AI) in medicine and management of conditions. AI is being used to enhance and speed diagnostic capabilities in conjunction with wearable devices as well as to identify health care cost issues and high risk patients.

Companies, health care providers, and researchers hoping to move forward with better medical technology—and tools to make health care more affordable and accessible—are eager to use AI-powered data in applications. They are largely invested in the quest to use AI in health care for the good of consumers and their patients. But data is not so simple nor the outcomes so obvious that we can afford to blithely gloss over the privacy risks to health care consumers as AI mines vast and deep personal data resources.

We are on the cusp of an intelligence explosion, thanks to better quality data and more data sources (including genomic), coupled with the growing capability of machine learning and AI.

But the undeniable health care benefits must be balanced by a willingness to address the potential, unintended effects that such a data surge might have on consumers—some of which could be harmful—and how data aggregators and providers should act to mitigate any harm.

AI Will Tap Data Well Beyond Health Care Claims

Fueling health-care AI is a rich reserve of individual consumers' health and close personal data. That data is no longer drawn from medical records alone, but also now includes socioeconomic information and preferences from a variety of sources: credit reporting agencies, social service agencies, public records including criminal history, zip code information, and insurance claims. New sources are wearable devices, which record sleep patterns, exercise, some clinical information, and even personal data, such as use of other software applications, as well as political and social preferences.

The rapid rise of AI-powered data-use should raise an orange flag for protecting consumer privacy in health care. The experiences of credit reporting agencies and face-recognition technology are two examples that show how things might evolve—and what we need to prevent.

TransUnion Data Breach Is a Portent for Risks to Health Care Data

TransUnion describes itself as a "global information and insights company that makes trust possible between businesses and consumers, by ensuring that each consumer is reliably and safely represented in the marketplace." The company states that its "accurate and comprehensive picture of each person" is built on data fusion methodology that integrates public records and credit data. Note that consumers have, in fact, very little power to contest or demand correction of their records. Data is assumed to be correct.

In 2019, Equifax reported a data breach from two years earlier for all three major credit reporting agencies, including TransUnion. It is not yet clear how that identity theft will affect the 150 million consumers involved in that breach, but it reveals the vulnerability of even the largest pools of data from intrusion and theft.

Who aggregates and accesses consumer financial data, and how that data is used, is not a neutral question. There can be ramifications for consumers related to employment tenure, insurance coverage, other employee benefits, retirement, recruitment, selection into career or educational tracks, and so on. The growing use of credit scores—not only to determine creditworthiness, but also as indicators of employability, mental health, stability, and financial

worth—are a prime example of how data could negatively impact consumers. Likewise, a breach of confidential health care data could have catastrophic ramifications for employment, insurance coverage, ability to get loans and mortgages, and more.

Clearview AI's Introduction of Unregulated Face Recognition Technology Foreshadows More Risks to Consumers

The surveillance and investigatory capacity of data has been largely associated with autocracies, with Chinese face-recognition technology, in particular, in the spotlight. Consumers may feel safe in a democracy. But recent revelations about Clearview AI's use of 3 billion photos "scraped" from social media and Internet sites to help law enforcement agencies identify individuals of interest should make health care providers and consumers reconsider the harm from unrestricted use of personal data, to individuals and well as vulnerable groups.

When asked about the risks of Clearview's face-recognition technology app eventually being made available to the public, CEO Hoan Ton-That admitted to *The New York Times*, "There's always going to be a community of bad people who will misuse it." Data aggregated from a variety of health care sources, including those without much privacy protection (for example, wearables), combined with AI and machine learning, creates virtually unlimited capacity for analyzing and using such data to extrapolate conclusions and profiles of individuals. That a company can access personal data without permission and decide on its own how to use such data for profit has serious implications for health care.

Health care data is currently protected only with respect to particular sources, and those consumer protections are focused primarily on the provider entity and, by permission, to a business associate. Personal data, quasi-health care data collected by smart watches, data that is not housed in the EMR—all are vague with respect to ownership and are not protected under HIPAA. Clearview scraped a large tranche of photos from Facebook without the social media giant even knowing, violating its terms of service with impunity. The risks to unprotected personal health care data being collected and mined for entrepreneurial ends cannot be understated.

Wearables Have the Power to Influence Employer-Based Insurance Coverage

As AI taps into behavior and clinical data from wearable devices, we must also confront serious ethical questions about how we validate and use data that could cause potential harm to

consumers.

AI capabilities are most powerful and distinct when they harness granular, patient-specific data such as individual attributes, symptoms, conditions, and behaviors. The aggregation of that granular data is how we learn how to predict risk. The appeal of wearable devices is that they passively collect consumer behavioral and clinical data for determination of risk and cost—and thus point to solutions that mitigate those concerns.

As more companies strive to lower their cost of health care coverage, they are turning to tools that lower their costs—like tiered provider networks, centers of excellence, and hiring or contracting for their own providers—and also to methods that engage consumers.

As a result, consumers are being drawn into incentive programs involving wearables, even if they have reservations about the deal they may be making. Yet they have little ability to change the conditions or data use after the fact. Federal HIPAA and HITECH provisions offer insufficient protections.

Patient data, valid or not, is already being tied to incentives in employer wellness programs; in the future it will likely feed into differential copays, premium sharing, or other benefit plan components based on behavior and underlying risk.
While, so far, companies are avoiding measures that could result in employee backlash, that does not mean that they (or health plans) will not turn to additional data to empower better control of patient risk factors, or patient engagement in lowering costs. Case-in-point: U-Haul recently announced its decision to quit hiring smokers, effective this month.

Aggregation Efforts Should Stipulate Protections for Consumer Health Data

It's no surprise that tech giants are enthusiastically investing in collecting and analyzing wearable data. Apple's smart watch and health record projects and its partnership with Aetna/CVS show both how data can be monetized to help consumers as well as potentially complicate coverage benefits; Apple has also established that it is free to decide how to manage its wearable data assets. Google's planned acquisition of Fitbit will, no doubt, further Google's expertise in profiling consumers and creating analytics around personal data. Again, the use of that data is not regulated, even as it is entirely personal.

Protection and privacy of data must be extended to all data that is aggregated, not just identified data. Lest we assume that aggregate, de-identified data already protects consumers, one study on this topic demonstrated 95 percent accuracy in re-aidentifying adults from a large

de-identified data set through machine learning.

Policies and rules must be developed that stipulate how consumer personal data can be used. Such policies will necessitate a major update of federal privacy legislation and rules, an inevitably lengthy process that will not be able to contain the first wave of harmful effects as health-care AI gains momentum.

At the least, health care providers and insurers, with consumers and other stakeholders, can create the groundwork and code of conduct for their part in generating, aggregating, and analyzing data, by meeting these four goals:

> Ensure transparency and voice for consumers whose data is being collected. Data that is used in relationship to health—including lifestyle behaviors, preferences, and all personal behaviors—should require some limitations of use, protection and privacy, and consent as to their use, especially when the data can be re-identified.
> Validate data by multiple sources. Even clinical data is filled with problematic codes, errors, and misidentifications of individuals. Wearables data is too new and already proven insufficient, in and of itself, to be independently used in incentive programs. Use of every data source should require sampling or other validation methods to verify the data source.
> Never make health insurance benefits "contingent" on good consumer behavior. Because of barriers and social determinants, not everyone can afford time off, transportation, or child-care to adhere to a treatment plan. If that plan had been constructed around the patient, it would have been modified to meet the circumstances as well as health needs.
> Never share or sell patient health or personal information, under any circumstances, or permit the sale of such data under HIPAA Business Associate Contracts. Consumer personal data should be used only to improve and provide better direct care to the patient.

None of this analysis is intended to decry the pioneering use of Artificial Intelligence in exciting and promising efforts to provide smarter and more effective health care to individuals. But in our enthusiastic embrace of technological data mining and algorithms, we must not forget that quality of life is at stake. AI's potential to radically improve health care can only be realized if efforts are carefully designed to ensure fair and rigorous protection of data accuracy and use.

Founded in 2002, Roji Health Intelligence guides health care systems, providers and patients on the path to better health through Solutions that help providers improve their value and succeed in Risk.

Image: ev