TheZetemaProject

Privacy, Transparency, and Penalties in the Era of Digital Health

Innovation Outstripping Regulation

Key Takeaways:

As the healthcare sector increasingly adopts digital technology, privacy is becoming a significant challenge.

There is a fundamental tension between rapid innovation and patient privacy protections

The definition of protected health information is changing as companies escaping HIPAA regulations – and perhaps those subject to the same – derive health information and algorithms thereupon that may influence individuals in unclear ways.

Current privacy laws are not written with the era of massive data generation and algorithmic analysis in mind; in particular, crowdsourced information leads to provision of information without end-user consent (such as genomic information screening identifying parents, siblings, and children).

Regulatory standards are largely limited to personal health information identified or provided in care settings. Minimal provisions are tied to derived health information, not how the information may be used on a platform.

Penalties standards hinge principally on data breaches, and not how data may be used to influence user behavior.

Deidentifying users is a unique challenge.

These challenges will be amplified as artificial intelligence algorithms continue to be deployed upon these datasets.

Introduction

"Digital health" describes the intersection of large-scale health information with the exponential growth of digital technology, driving insight into health. It underpins a movement within healthcare toward individually-tailored precision medicine. Yet this technology also threatens patient privacy, and current policy and business practices may not provide adequate safeguards. At the same time, outdated privacy laws - and an overly-conservative interpretation of them due to potential penalties - threatens innovation. How do we resolve this fundamental tension between protecting privacy and allowing innovation to flourish?

Furthermore, the ability to integrate datasets on a large scale has blurred the lines of what constitutes “health information”: phones track physical activity, credit card and online purchases identify food and activity selection, and today genealogical services offer genomic screenings. Integrated, this allows an unprecedented level of insight into an individual, but much of these data are or can be gathered without the end-user’s permission, and algorithms inducing behavioral change may act on users regardless of consent. While expanded health datasets can open the door for improved care, much of this information falls outside the scope of modern health information protections, while behavioral intervention may be wholly opaque to the end user.

Big Data and Precision Medicine: Regulation Impedes Progress

Big data is a modern phenomenon, and opens the door to live-analyzed, individually-tailored medicine. However, regulatory efforts that protect user anonymity may also impede medical progress. Advances in the computational power of personal computers along with decentralization of analytics to technology company data servers have enabled widespread analysis of large-scale databases. Simultaneously, the amount of data being generated – spread across phones, wearables, cameras, texting, etc. – is increasing exponentially. Collectively, this enables an unprecedented level of insight into the early pathogenesis of disease, while potentially enabling individually-tailored treatments to benefit of both individuals and, via reduced costs, the healthcare system. For example, the ability to use continuously generated wearable information could lead to the development of a system for continuous health analysis and feedback. This could open the door for identification of asymptomatic early disease markers uncaptured in current healthcare while also enabling continuous feedback for the purposes of positive behavioral reinforcement. Current privacy standards afford some patient protection. However, many regulations make innovation slower and more expensive, and additional laws being considered could critically hamper entrepreneurship in healthcare.

HIPAA / HITECH

Current privacy laws are not written with today's massive data generation and algorithmic analysis in mind. The key laws in the United States applying to the use of personal health information include the Health Insurance Portability and Accountability Act (HIPAA), which was extended by the 2009 Health Information Technology for Economic and Clinical Health act (HITECH). These provide protections specifically for personally-identifiable health information, with penalties and notification if any breaches occur. Neither law, however, allows individuals to bring a cause of action against a provider. Notably, what data is considered "health information" is more strictly described that information which is used in the diagnosis and treatment of a patient, and so fitness and other wearable-generated information arguably operate outside HIPAA provisions, despite the ability to derive health information from the raw data collected.

Deidentifying users is a unique challenge. Researchers at MIT and Berkeley used a series of national surveys with individual physical activity data for 14,451 individuals, stripped of protected health information. Using machine learning, they were able to reidentify the demographic information of 80% of the children and 95% of the adults [1]. Though HIPAA provides for specific protections, the stream of often-unprotected health activity data generated individually using wearables or phones actively produces a dataset that can be used to individually identify patients, regardless of consent or even demographic obfuscation, while providing deep information about their health status. What constitutes protected health information may be far more broad than initial HIPAA/HITECH verbiage cover. Given how easy it is to reidentify patients, and they are generating more data all the time via wearables, how do we protect privacy - and who is responsible when we don't?

Penalties Without Incentives Complicate Innovation; Loopholes Persist

Current penalties with regard to misuse of data focus on violations of HIPAA as stipulated in HITECH: mandatory penalties for “willful neglect” with civil penalties up to $250,000, and repeated violations extending to $1.5 million. Fear of these penalties stifles innovation, as does HIPAA’s ban on EHR vendors from monetizing data, making data transmission fees uniquely illegal within medicine (though prevalent in other industries, such as finance), disincentivizing interoperable systems. HIPAA and problems relating to compliance with the law have been tied to the impedance of interoperability standards, to the tune of $371 billion per year [2].

While HIPAA bans EHR vendors from monetizing data, HITECH requires free transmission of medical data (interoperability) as an individual right. This complicates incentives. The result has been penalties for health data transmission companies: eClinicalWorks settled with the Department of Justice for $155 million as “eCW’s software failed to satisfy data portability requirements intended to permit healthcare providers to transfer patient data from eCW’s software to the software of other vendors.” The inability to monetize data in order to secure HIPAA compliance also prevents entrepreneurial models involving monetization of patient data and potentially patient data analysis are banned, limiting the ability to innovate on patient data and potentially encouraging companies to circumvent HIPAA as regulation.

Companies such as Facebook may circumvent HIPAA by declaring that information on their platform is not protected health information if voluntarily provided; derived health information, similarly, may be unprotected. Penalties as established by HITECH are applicable to those in violation of HIPAA, and so independent organizations may use user/patient data internally, including experimentation on and/or behavioral manipulation of end-users. This goes without incident if that information does not fall strictly within the purview of HIPAA; outside of EMRs, things can get murky, and a strict definition hinging on “information involved in diagnosis and treatment” may allow companies with information about an individual extending far beyond clinical information to escape HIPAA. It is unclear how companies establishing themselves outside of HIPAA may sell aggregations of data to other players, and it is unclear how either HIPAA or non-HIPAA companies may sell algorithms derived from user data. These algorithms may target specific users or health conditions. Though HITECH specifically stipulated for free transfer of health information with the implication that these data were owned by patients, it remains unclear how recent CMS guidance verifying this may affect the current state of the market.

Case Study – The Golden State Killer

DNA information from a 1976-1979 spree of murders in California existed, but could not be matched to any individual. The alleged Golden State Killer was arrested when a relative uploaded their DNA sequences to the open-source genealogy website GEDmatch, which law-enforcement was able to use to partially match the DNA from the killings. [3, 4] This highlights challenges with the advent of identifiable technologies: Genomic analysis of one individual makes identifiable their entire family tree, without consent by any other individuals. Moreover, the ability of law enforcement to use tools like GEDmatch is not subject to regulation; some identified through such methods have ultimately been shown to be innocent [5]. Healthcare-related information, and medical information derivable from this, is leading to cases in which individuals can have no interaction with healthcare whatsoever, but still have their private health status violated.

Transparency – A Key Area Without Regulation Today

As technology has expanded into healthcare, more individual health and healthcare data are being recorded in different ways. Credit card purchases may capture dietary and activity metrics; watches and phones may derive user metabolic rates; genealogy websites may offer genomic testing, generating information not only on an individual, but on family members. This all occurs generally with user involvement, but without their explicit consent. Applications and companies able to access this information may then use it freely internally, generating insights into users and changing their services accordingly, and often without transparency (or compensation) to the end-user.

While consumers may be ambivalent (or, often, unaware) of the ways in which information is being generated about them, privacy regulations are in place to protect them. However, internal data use by a company is not subject to protection or regulation. For instance, it was reported in 2014 that Facebook had been conducting experiments upon users, ultimately finding that they could be emotionally manipulated by the content they were exposed to [6]. Facebook funded the study independently and so was exempt from Federal Common Rule guidelines, which require informed consent, and further made the claim that user consent to their Data Use policy constituted consent to the experimentation. Though there was significant media coverage of the experiment, no regulatory change was enacted.

Moreover, the deployment of these technologies both in and out of healthcare can be problematic: facial recognition has notable biases depending on skin color, while audio information gathered through services such as Siri, Google Home, or Alexa may offer key insights into end-user health to the company at hand, including diagnostics, without user input or even knowledge. Companies are not required to divulge algorithmically-derived diagnoses, and company awareness of a user's health condition may enable directed targeting ranging from display of ads to behavioral change. Collectively, these datasets enable unprecedented levels of information about human health and may accelerate diagnosis and garner personalized treatment in ways previously impossible. They may also represent a potential for abuse. Should companies be restricted from developing this information or required to reveal this derived metadata to end-users?

The guidelines in place for managing healthcare data within the healthcare system, particularly with regard to end-user experimentation, are therefore unclear if not absent for corporations. The definition of protected health information is changing as companies avoiding HIPAA regulations – and perhaps those subject to the same – derive health information and algorithms thereupon that may influence individuals in unclear ways. Moreover, freemium and other business models may allow generation and use of health information to drive human behavior, without violating HITECH regulations through the explicit sale of data. The landscape is notably murky outside the use-case the regulations were designed for, EMRs, and as in-lab and at-home data science grows increasingly prevalent, regulations ensuring transparent use of health information and/or compensation to patients may be needed.

Conclusion

The ability to harness individual-level information at scale offers dramatic opportunity for improving human health. It is difficult to overstate the potential of early detection and treatment of illness, as well as personalized care and potential positive behavioral change that may already be possible through information technology. Current regulations stifle entrepreneurship in the healthcare space and slow the ability for existing entities to enact positive research or change at scale. However, the growth of digital health has notable implications on patient privacy and the transparency with which data are used. Technology enables personal identification with minimal information, companies are able to experiment with behavioral change opaquely upon their platforms, and all with minimal user consent or knowledge. Penalties that currently exist focus principally on data breaches or negligence on the part of information carriers, but not on how that data may be used.

Discussion Questions

Is regulation of digital health feasible? Desirable?

How should digital health be regulated, while preserving entrepreneurial considerations?

Regulations currently focus on patient privacy, though the capacity to deidentify patients is becoming increasingly difficult. How can we maintain privacy, or must we shift our model to one of transparency?

What are the implications and decision points of this debate for many of the key stakeholders: providers, payers, pharma, etc.?

What opportunities and risks are posed to the healthcare sector by innovations in digital health?

References

[1] Na L, Yang C, Lo C, Zhao F, Fukuoka Y, Aswani A. “Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning.” JAMA. 2018-12-21. doi:10.1001/jamanetworkopen.2018.6040

[2] Yaraghi, Niam. “To Foster Information Exchange, Revise HIPAA and HITECH” Health Affairs. 2019-09-19.

[3] Winton, Richard, Serna, J., St. John, P., Oreskes, B. “Police used consumer genealogical websites to identify Golden State Killer suspect.” Los Angeles Times, 2018-04-26.

[4] Wade, Christian M. “Massachusetts Considers Bill to Limit Facial Recognition." Government Technology, 2019-02-11.

[5] Mustian, Jim. “New Orleans filmmaker cleared in cold-case murder; false positive highlights limitations of familial DNA searching.” The New Orleans Advocate. 2015-03-12.

[6] Waldman, Katy. “Facebook’s Unethical Experiment.” Slate. 2014-06-28.

Chan, Tara Francis. “A school in China is monitoring students with facial-recognition technology that scans the classroom every 30 seconds.” Business Insider. 2018-05-20

Sauer, Gerald. “A murder case tests Alexa’s devotion to your privacy.” Wired. 2017-02-28.

“Artificial Intelligence.” Privacy International. 2019-02.

Tebani A, Afonso C, Marret S, Bekri S. “Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations.” Int J Mol Sci. 2016;17(9):1555. Published 2016 Sep 14. doi:10.3390/ijms17091555

O’Shea, Dan. “RBC: About 41% of Americans now own smart speakers.” Retail Dive. 2019-01-03.

“GDPR Right to be Forgotten” Intersoft Consulting.

Smith, Chris. “Facebook tracks you even if you’re not a user, and you can’t really do anything about it”. BGR 2018.