Indiana Law Journal

Volume 97 | Issue 4 Article 10

Fall 2022

Big Data, Big Gap: Working Towards a HIPAA Framework that Covers Big Data

Ryan Mueller
Indiana University Maurer School of Law, rpmueller@iu.edu

Big Data, Big Gap: Working Towards a HIPAA Framework that Covers Big Data

RYAN MUELLER*

One lasting impact of the Health Insurance Portability and Accountability Act (HIPAA) is the privacy protections it provides for our sensitive health information. In the era of Big Data, however, much of our health information exists outside the traditional doctor-patient dynamic. From wearable technology, to mobile applications, to social media and internet browsing, Big Data organizations collect swaths of data that shed light on sensitive health information. Big Data organizations largely fall outside of HIPAA's current framework because of the stringent requirements for when the HIPAA protections apply, namely that the data must be held by a covered entity, and it must originate from a select few sources. Thus, the very same sensitive health information is covered by HIPAA when a physician obtains the information while outside of HIPAA's purview when it is in the hands of Big Data organizations. Without HIPAA's protections, Big Data organizations are free to exploit their consumers' sensitive information without their consent and often without their knowledge.

This Note first explores the current HIPAA framework with a goal of identifying the gaps that allow Big Data to fall outside of its reach. This Note identifies two primary requirements that allow Big Data organizations to escape the privacy regulations but could, if amended, force these organizations possessing sensitive health information into compliance with HIPAA. Finally, this Note proposes an amended HIPAA framework to cover Big Data by borrowing solutions employed by the European Union and the state of Texas.

---

* J.D. Candidate, 2022, Indiana University Maurer School of Law; B.S., Indiana University, 2019. I would like to thank Professor Michael Mattioli for his thoughtful feedback and suggestions, the *Indiana Law Journal* associates and editors for their work to get this Note ready for publication, and my parents for their continued support.

INTRODUCTION

Imagine you and your significant other are attempting to have a child. Many choose to keep this deeply personal decision to themselves, as is their right. Mobile applications ("apps") such as Flo can help the couple chart and track the process; so perhaps you and your partner choose to download this app and input the required information. Hours later, while scrolling through Facebook, you start to notice advertisements for prenatal vitamins, parenting books, and other common purchases by pregnant couples. How could Facebook have possibly known to show you these targeted advertisements? After all, you kept this decision to yourself.

This exact scenario is not far-fetched. Rather, a *Wall Street Journal* report in 2019 detailed how apps, such as Flo and many others, share the information users enter into the app.[1] In one case, the authors of the report entered heart rate data into an app and noted the data entered was shared with Facebook immediately.[2] It also made no difference whether the user actually had a Facebook profile or had connected the app to their profile; Facebook created its own profiles of the users if necessary.[3]

---

1. Sam Schechner & Mark Secada, *You Give Apps Sensitive Personal Information. Then They Tell Facebook.*, THE WALL ST. J. (Feb. 22, 2019), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636?mod=e2tw [https://perma.cc/245C-BRAB].

2. *Id.*

3. *Id.* In 2018, Facebook CEO Mark Zuckerberg reluctantly admitted before Congress that the social media giant collects data on individuals that are not on the Facebook platform. David Ingram, *Facebook Fuels Broad Privacy Debate by Tracking Non-Users*, REUTERS (Apr. 15, 2018, 7:04 AM), https://www.reuters.com/article/us-facebook-privacy-tracking/facebook-fuels-broad-privacy-debate-by-tracking-non-users-idUSKBN1HM0DR [https://perma.cc/QU7A-Q7BT]. For a discussion of these so-called "shadow profiles" and how Facebook's insistence on their necessity leaves unanswered questions, see Kurt Wagner,

Congress passed the Health Insurance Portability and Accountability Act (HIPAA) in 1996, and HIPAA contained privacy language preventing the disclosure of personally identifiable information in conjunction with sensitive medical information.[4] In the case of Flo, the app essentially told Facebook the user was trying to get pregnant or was pregnant. Compare this to a situation where a woman learns she is pregnant from her doctor. Imagine the outrage if that doctor then sold her pregnancy information to advertising agencies. In either situation, advertising agencies learned of sensitive health information in conjunction with a particular individual. But, in the case of a doctor, the pregnancy information is protected by HIPAA, and the doctor legally cannot share it with an advertising agency without authorization by the patient.

In the case of the app, however, HIPAA has no teeth to prevent disclosures to advertisers or anyone else because a mobile app is outside the scope of HIPAA's coverage.[5] HIPAA's regulations are narrowly applied to only certain types of entities, such as providers, and only when the health information originates from certain sources.[6] Big Data,[7] through common technology, has become ubiquitous in the field of health data. Fitbits, Apple Watches, mobile apps, social media, and other technology that use Big Data collect vast amounts of sensitive health data: height, weight, pre-existing conditions, heart rate, and even the decision to pursue pregnancy. If doctors disclosed this information for purely economic reasons, we would be outraged, yet we welcome the addition of new technology with open arms—either not considering or not caring about the implications Big Data has for our sensitive health information.

The current HIPAA framework fails to prevent Big Data companies from disclosing sensitive health information they obtain whenever and to whomever they want. Therefore, Congress should amend HIPAA's disclosure regulations to include in its "covered entities" all companies collecting, storing, analyzing, using, and transmitting health information and should further amend the disclosure regulations to remove the source requirement to qualify as health information. These changes will shift the legal framework from narrowly protecting health information originating from certain sources and held by certain custodians, to protecting health information based on the nature of the information itself. This Note proceeds in three parts. Part I briefly details the HIPAA legislation. Part II outlines the many ways Big Data companies collect this sensitive information, focusing on three predominant

---

This Is How Facebook Collects Data on You Even If You Don't Have an Account, VOX (Apr. 20, 2018, 1:02 PM), https://www.vox.com/2018/4/20/17254312/facebook-shadow-profiles-data-collection-non-users-mark-zuckerberg [https://perma.cc/HJ9E-TNPQ].

4. Health Insurance Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

5. *See infra* Part II.D.

6. *See infra* notes 14, 27 and accompanying text.

7. Big Data is often ill-defined, and when it is defined, the definitions vary widely. In this Note, "Big [D]ata refers not only to the collection and storage of extremely large data sets but also the data mining and predictive analytic routines that process the data." Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 77 (2014). For a discussion of the difficulties and limitations of defining Big Data, see Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 394–95 (2014).

methods: wearable technology, mobile health apps, and consumer-generated data. Finally, Part III proposes a solution to fill this gap in protection by broadening the "covered entities" definition and removing the source requirement, a proposal drawn from state and European Union solutions as guidance.

I. AN OVERVIEW OF HIPAA

Congress passed HIPAA in 1996 to address a number of concerns in the U.S. health care system.[8] Initially, Congress intended HIPAA to ease the burden of those with pre-existing medical conditions from "job-lock,"[9] where people remained in certain positions solely out of fear of losing their health insurance while changing jobs. As the legislation progressed, however, the introduction of electronic devices and information sharing sparked a concern for patient privacy, as some of the provisions upended the traditional doctor-patient dynamic.[10] Recognizing the longstanding principle of privacy in the medical setting, HIPAA created privacy protections in the age of information sharing and interaction of multiple players in an individual's medical care.[11] Under HIPAA, the United States Department of Health and Human Services (HHS) is charged with promulgating rules and regulations to carry out its objectives.[12] Although the legislation was passed in 1996, HHS did not publish the Privacy Rule until 2000, and it did not take effect until April 14, 2003.[13]

HIPAA's Privacy Rule is narrowly tailored. The privacy protections only apply to "covered entities" and their "business associates."[14] A covered entity is a health plan, health care clearinghouse, or health care provider.[15] Health plans refer generally to insurance plans, and include various forms of health plans such as traditional private insurance, health maintenance organizations, and government-funded plans.[16] Health care clearinghouses are businesses that compile health information, often for billing or processing purposes.[17] Health care providers include any person or organization that is paid to provide health services, from physicians to hospitals

---

8. *See generally* 110 Stat. at 1936 (discussing the purposes of the HIPAA legislation).

9. Tamela J. White & Charlie A. Hoffman, *The Privacy Standards Under the Health Insurance Portability and Accountability Act: A Practical Guide to Promote Order and Avoid Potential Chaos*, 106 W. VA. L. REV. 709, 713 (2004).

10. *Id.* ("Individual patient privacy, however, became an increasing concern as healthcare and health insurance reform measures resulted in greater information sharing through electronic information systems, which were accessible by individuals outside the realm of direct health care provider/patient care relationships.").

11. *See id.* at 713–14.

12. Timothy Newman & Jennifer Kreick, *The Impact of HIPAA (and Other Federal Law) on Wearable Technology*, 18 SMU SCI. & TECH. L. REV. 429, 432 (2015).

13. *Id.* The Privacy Rule is codified in Title 45 of the Code of Federal Regulations in section 160 and subparts A and E of section 164. 45 C.F.R. §§ 160.101, 164.104, 164.500 (2019).

14. § 164.500.

15. *Id.* § 160.103.

16. White & Hoffman, *supra* note 9, at 718.

17. *Id.* at 719.

to nursing staff.[18] This Note refers to this requirement as the "custodial requirement" because the Privacy Rule only applies when the information is held by these specific custodians.

A business associate is anyone who, on behalf of a covered entity, "creates, receives, maintains, or transmits protected health information" or "[p]rovides . . . legal, actuarial, accounting, consulting, data aggregation . . . , management, administrative, accreditation, or financial services."[19] Thus, most companies that contract with a covered entity to perform certain functions must also comply with HIPAA regulations.[20] Under President Obama, HHS issued its HIPAA Omnibus Final Rule, which operationalized the changes Congress made to HIPAA when it passed the Health Information Technology Economic and Clinical Health (HITECH) Act of 2009.[21] While the HITECH Act intended to incentivize health care providers and other market operators to adopt electronic health records, it also included an important amendment to the application of HIPAA's privacy and security rules to business associates.[22] Prior to the enforcement of HITECH's amendments through the Omnibus Rule, business associates were not directly responsible for HIPAA regulations; rather, the covered entities were held responsible for ensuring their business associates adequately complied with the HIPAA regulations.[23] After the Omnibus Rule, however, business associates are now directly responsible for complying with the Privacy Rule.[24]

Further, the Privacy Rule only protects certain kinds of information. The Privacy Rule prohibits covered entities and their business associates from using or disclosing "protected health information" (PHI).[25] PHI is any information that is "individually identifiable health information."[26] Thus, there are two separate definitions within PHI that require parsing: the information must (1) be health information and (2) be individually identifiable.

First, to qualify as "health information" under the PHI definition, two components must be satisfied: (1) the information must originate from a "health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse;" and (2) the information must "relate[] to the past, present,

---

18. *Id.*

19. § 160.103.

20. For an example, see the discussion of Optum, *infra* note 38. Optum is a business associate when they partner with covered entities to provide consulting and data analytics services. *See id.*

21. John V. Arnold, *Privacy: What Lawyers Must Do to Comply with HIPAA*, 50 TENN. B.J. 16, 17 (2014).

22. Megan Bradshaw & Benjamin K. Hoover, *Not So Hip? The Expanded Burdens on and Consequences to Law Firms as Business Associates Under HITECH Modifications to HIPAA*, 13 RICH. J.L. & PUB. INT. 313, 326 (2010).

23. Jo-Ellyn Sakowitz Klein, Kelly M. Cleary & Anna R. Dolinsky, *Health Sector Braces for Wide Impact of the New HITECH Omnibus Rule*, 25 INTELL. PROP. & TECH. L.J. 10, 10 (2013). *See also* Carol Stryker, *Two Essentials for HIPAA Omnibus Final Rule Compliance*, PHYSICIANS PRAC. (Sept. 18, 2013), https://www.physicianspractice.com/view/two-essentials-hipaa-omnibus-final-rule-compliance [https://perma.cc/J3K3-E3MD].

24. Klein et al., *supra* note 23.

25. 45 C.F.R. § 164.502 (2019).

26. *Id.* § 160.103.

or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual."[27] Thus, from this definition, health information has both a source requirement and a substance requirement.

Second, health information becomes "individually identifiable" when it can identify an individual.[28] To answer the question of what it means to be able to identify an individual, HHS delineated when health information is *not* individually identifiable and thus can be disclosed, thereby backing into a set of eighteen kinds of information that show when health information is individually identifiable.[29] Health information is *not* individually identifiable health information, and thus can be disclosed by covered entities or their business associates without authorization, when any of the eighteen personal identifiers have been removed.[30] The eighteen identifiers include names, social security numbers, email addresses, fingerprints, and phone numbers.[31] Thus, it is important to note at the outset that the Privacy Rule does not directly protect medical information, such as a diagnosis or prescription. Rather, the Privacy Rule prevents disclosure of such health information only when it is combined with identifiers demonstrating the health information pertains to a specific person.

At the time of its passing, HIPAA's privacy provisions represented a revolutionary codification. But, simply put, the provisions do not apply outside of its narrow drafting. Because HIPAA was intended to apply only to certain health care custodians in the health care system and only to certain kinds of information, much sensitive health information falls outside of HIPAA's framework. The substance requirement[32] and the individually identifiable requirement[33] seem like sensible attempts to limit the privacy protections of HIPAA to the health information arena. Yet other requirements, such as the custodian requirement[34] or the source requirement,[35] cut against the goal of providing privacy protections for sensitive medical information. These two requirements narrowly constrict what constitutes protected information, thus allowing certain actors, such as Big Data organizations discussed below, to escape the Privacy Rule.

---

27. *Id.*

28. *Id.*

29. *Id.* § 164.514(b) ("A covered entity may determine that health information is not individually identifiable health information only if: . . . (2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed[.]").

30. *Id.* The full eighteen categories are as follows: names; location information such as address, zip code, and city; any dates that directly relate to an individual such as birth date, admission date, and death date; phone numbers; fax numbers; email addresses; social security numbers; medical record numbers; health plan numbers; account numbers; license numbers; vehicle identifiers; medical and health-related device numbers; web URLs; biometric data such as finger and voice prints; photographs of the face; and any other identifying number, code, or characteristic. *Id.*

31. *Id.* § 164.514(b)(2).

32. *See supra* text accompanying note 27.

33. *See supra* text accompanying notes 29–30.

34. *See supra* text accompanying notes 14–18.

35. *See supra* text accompanying note 27.

II. HOW BIG DATA FALLS OUTSIDE OF HIPAA'S REGULATIONS

Big Data[36] impacts health information in two main ways. First, Big Data organizations can partner with a player in the health care industry to provide their services.[37] For example, Optum partners with various providers and health plans with the goal of improving patient care and patient value by using its data research and analysis tools.[38] Another example is the rise of telehealth physician visits, where a provider contracts with a telehealth company to provide health services remotely, often via a mobile app.[39] While interesting, this impact of Big Data on health information is not the focus of this Note because the Big Data organizations in both of these situations would likely be subject to HIPAA's privacy regulations by virtue of being a covered entity or a business associate of a covered entity.[40]

The second way Big Data impacts health information, and the subject of this Note, is when Big Data organizations collect health information independent of a covered entity. In this context, Big Data organizations collect vast amounts of sensitive health information that is very much identifiable. Because Big Data organizations are not covered entities or business associates of a covered entity, however, these organizations fall outside the scope of HIPAA's privacy regulations. The following Sections highlight a few of the more prominent ways Big Data collects, stores, and uses health information, followed by an analysis detailing why such uses of health information fall outside of HIPAA's scope.

A. Wearable Technology

Fitbits, Apple Watches, and other smart technology worn by individuals have become increasingly prevalent in society. The Pew Research Center found that

---

36. For a reminder of the definition of Big Data used in this Note, see Terry, *supra* note 7.

37. *See, e.g.*, White & Hoffman, *supra* note 9, at 719–20 (discussing how "interactive information sharing is necessary for the orderly operation of the health care industry" among the patients, covered entities, and business associates).

38. Tom Davenport & Randy Bean, *Optum Focuses on AI to Improve Administrative Decisions*, FORBES (Oct. 9, 2020, 10:50 AM), https://www.forbes.com/sites/tomdavenport/2020/10/09/optum-focuses-on-ai-to-improve-administrative-decisions/ [https://perma.cc/J6HF-2JBT]. Optum is a business unit within UnitedHealth Group that focuses on improving health care through advanced analytics and data science. *Id.*

39. *See, e.g.*, J. Frazee, M. Finley & JJ Rohack, *MHealth and Unregulated Data: Is this Farewell to Patient Privacy?*, 13 IND. HEALTH L. REV. 384, 393 (2016).

40. For example, Optum is a business associate and thus must comply with HIPAA regulations when it comes into contact with PHI when Optum partners with a covered entity such as a health plan or provider to consult or perform data analysis. *See* 45 C.F.R. § 160.103 (2019).

twenty-one percent of Americans wear some type of wearable fitness tracker on a regular basis,[41] and the trend of wearable technology shows no sign of slowing.[42]

This kind of technology collects vast amounts of health information. The Fitbit might have originated as a step counter, but today's Fitbits can track heart rate and sleeping patterns.[43] Similarly, the Apple Watch can count calories burned and even take electrocardiogram and oxygen saturation readings.[44] Wearable tech also tends to require the user to enter health information such as biological sex, height, and weight.[45]

Wearable technology impacts health information privacy for a number of reasons. First, users continually expect technology to progress, and often a requirement for progress is information sharing between the device and the developer.[46] The shared information may include more than just data about the device's performance; rather, the information sharing gives developers potential access to all of the sensitive health information collected and generated by the device.[47] Many may not see this kind of sharing as a problem; after all, the information sharing improves future generations of this tech.[48] Yet, this kind of sharing certainly implicates sensitive health information.

The second impact on health information privacy comes from users' expectations for the technology. Some users actually want the data collected by their wearable technology to be shared.[49] For example, a user may want their Fitbit or Apple Watch

---

41. Emily A. Vogels, *About One-in-Five Americans Use a Smart Watch or Fitness Tracker*, PEW RSCH. CTR. (Jan. 9, 2020), https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [https://perma.cc/75QA-MJYJ].

42. Alicia Phaneuf, *Latest Trends in Medical Monitoring Devices and Wearable Health Technology*, BUS. INSIDER, https://www.businessinsider.com/wearable-technology-healthcare-medical-devices [https://perma.cc/R5F6-7A2G].

43. Newman & Kreick, *supra* note 12, at 430.

44. *See, e.g.*, Nicole Wetsman, *Why Apple Needed the FDA To Sign Off on Its EKG but Not Its Oxygen Monitor*, THE VERGE (Oct. 7, 2020 12:25 pm), https://www.theverge.com/2020/10/7/21504023/apple-watch-ekg-blood-oxygen-fda-clearance [https://perma.cc/2EED-PU92]. Wetsman also discusses the FDA regulatory challenges of implementing oxygen saturation readings compared to heart rate readings on the Apple Watch. *Id.*

45. *See* Phaneuf, *supra* note 42.

46. Newman & Kreick, *supra* note 12, at 430.

47. *See id.* at 430 (noting that even consumers that want their information shared for the improvement of this technology "should not underestimate the potential for sharing the information that these devices collect").

48. For example, the 2019 Pew Research study previously mentioned found that forty-one percent of wearable technology users feel it is acceptable for these devices to freely share their information with health researchers, compared to thirty-five percent that find this unacceptable. Vogels, *supra* note 41. Although the study notes that those who use fitness trackers tend to be more accepting of data sharing, *id.*, there still appears to be a significant number of Americans who find this kind of data sharing acceptable or even desirable.

49. During the COVID-19 pandemic in 2020, the PGA Tour turned to wearable technology to help identify potential positive cases. One golfer was not experiencing symptoms of the virus, but a wearable device on his wrist called a "Whoop" alerted him that his respiratory rate was increased. This led the golfer to seek a COVID-19 test, and he

data to be communicated to their health care provider so the provider can better track the user's health.[50] Another example is employers wanting access to this kind of information to better track employee health and morale.[51]

In the case where the user wants the information to be shared, the privacy concern is reduced. But because HIPAA's privacy regulations do not apply, these wearable technology companies can share the information with whomever they choose, even without the user's consent.[52] While employers may see information shared from their employees' wearable devices as a way to better gauge employee health,[53] most employees would probably see this as an invasion of privacy. If the employee went to his or her physician and obtained a heart rate reading, the employee has every right to prevent his or her employer from knowing what the heart rate reading was.[54] In fact, that heart rate reading by the doctor is covered by HIPAA, thus the doctor is not allowed to share the reading with the employer unless the employee gives consent.[55] In contrast, there is nothing to stop a wearable device company from sharing the heart rate obtained from the device with the user's employer. Admittedly, this is a bad business practice that would likely cause public outrage. But relying on the marketplace[56] to address these privacy concerns seems contradictory to the HIPAA legislation. Congress just as easily could have decided to allow the market to solve privacy concerns with all health information, where patients choose their doctors in part based on whether the doctor is known to share the personal health information of patients. Congress chose instead, however, to regulate health information so that it cannot be shared in these circumstances. Thus, identical medical information, such as a heart rate reading, is protected in one situation but wholly unprotected in the other.

B. Mobile Applications

Mobile health apps, often referred to as mHealth,[57] collect vast amounts of health information. One estimate predicts the market for mHealth apps will reach $111

---

subsequently tested positive for the virus. Viewing this alert as a potential way to identify positive cases or those that need to be tested, the PGA Tour procured 1000 of the wrist bands to give to golfers, caddies, and other personnel at the events. Jessica Golden, *PGA Tour Procures 1,000 Smart Bands to Help Detect Coronavirus Symptoms in Golfers*, CNBC (June 24, 2020, 1:16 PM), https://www.cnbc.com/2020/06/24/pga-tour-procures-smart-bands-to-detect-coronavirus-symptoms-in-golfers.html [https://perma.cc/CE3L-AEHB].

50. Newman & Kreick, *supra* note 12, at 430.

51. *Id.*

52. *See infra* Part II.D.

53. *See* Phaneuf, *supra* note 42 (discussing how employers can use wearable technology to boost employee health, which lowers turnover).

54. The heart rate reading is PHI because it originates from a provider and relates to a current health condition. 45 C.F.R. § 160.103 (2019). The provider is a covered entity. *Id.* Thus, the provider may not use or disclose the information. *Id.* § 164.502(a).

55. *See, e.g.*, *id.* § 164.508.

56. Newman & Kreick, *supra* note 12, at 430–31

57. Jianyan Fang, *Health Data at your Fingertips: Federal Regulatory Proposals for Consumer-Generated Mobile Health Data*, 4 GEO. L. TECH. REV. 125, 126 (2019).

billion by 2025,[58] showing signs of a field facing tremendous growth. By their nature, these applications collect, store, and use sensitive health information. One way to group these types of applications is based on the kind of health information they collect: (1) applications that directly collect data related to a disease or condition with "inherent medical significance;" (2) applications that collect data related to the user's health in general; and (3) applications that collect other information wholly unrelated to health information, such as birth dates and credit card numbers.[59] The first two categories and their implications will be explored further, as these two relate to the unregulated collection and use of sensitive health information.[60]

Many applications are created for the express purpose of monitoring one's condition. Revisiting the introductory hypothetical, the Flo app was created for the purpose of tracking and monitoring the pregnancy process.[61] With over 300,000 of such applications,[62] users track conditions such as diabetes or depression, and can even monitor alcohol intake.[63]

Other applications, however, collect health information related more generally to an individual's overall health. For example, Under Armour's MyFitnessPal collects a wide array of health, exercise, and diet information.[64] Another example is the Kinsey Reporter mobile app, a product of the Kinsey Institute in Bloomington, Indiana.[65] This app collects information about an individual's sexual health.[66] Of the health information mentioned thus far in this Note, this is perhaps the most sensitive. The app tracks sexual health measures such as contraception use, menstruation, bleeding, and a whole host of other private health information.[67] Although the data is touted as collected anonymously, the application tracks users' location, demographic information, and even the internet protocol (IP) address.[68]

A concern arises when these mobile apps share this sensitive information without the user's knowledge. Some of this information sharing is in furtherance of noble goals. For example, the fertility app, Flo, employs data scientists to analyze the billions of data points it receives from users of the app.[69] Using data obtained, the company hopes to increase understanding of fertility and improve predictions for

---

58. *Id.* at 127.

59. *Id.* at 134.

60. The third category is outside the scope of this Note because it relates to data collected by apps that have no medical or health significance. While privacy and data protection are just as salient in the third category, this Note addresses health information specifically, which is lacking in this third category.

61. *See supra* text accompanying note 1.

62. Remy Franklin, *11 Surprising Mobile Health Statistics*, MOBIUS MD (Oct. 25, 2021) https://www.mobius.md/blog/2019/03/11-mobile-health-statistics/ [https://perma.cc/6Z8U-9HBT].

63. *See* Frazee, Finley & Rohack, *supra* note 39, at 393.

64. Fang, *supra* note 57, at 126.

65. Stacey A. Tovino, *Going Rogue: Mobile Research Applications and the Right to Privacy*, 95 NOTRE DAME L. REV. 155, 163 (2019).

66. *Id.*

67. *Id.*

68. *Id.* at 163–64.

69. Leah R. Fowler & Stephanie R. Morain, *Schrödinger's App*, 46 AM. J.L. & MED. 203, 208 (2020).

pregnancy.[70] As previously noted, however, some mobile apps share the sensitive information not to improve the app or for research purposes but rather for financial gain.[71] For example, Facebook was caught obtaining this type of information from third-party apps.[72] Advertisers flock to Facebook to purchase this kind of data because, as some advertising buyers explain, "Facebook's insights into users' behavior" allow these advertising buyers to "offer marketers better return on their investment."[73] One study found that the top twenty mobile health apps disclosed sensitive health information to as many as seventy third-party companies for financial gain.[74]

Another way to divide health information from mHealth apps is into two broad categories: active data that the user inputs manually and passive data that is automatically collected.[75] Some passive data can be further classified as inferred data, which is data that analytical models infer from other inputs by a user.[76] One example of inferred data is using data analytics tools to infer one might have diabetes based on that user's entry of dietary information.[77] Even users that are aware that their dietary information entered into the application might be shared rarely understand the conclusions these data analytics tools can draw.[78] Additionally, inferred data is often constant, providing even less privacy than the same information taken as a snapshot at a physician's office.[79]

The data collected through these apps are often used to create user profiles, and often without the user's knowledge the profile exists.[80] Some of these profiles include descriptive elements such as tobacco user, allergy sufferer, or dieter.[81] Other profiles, however, include more general labels such as expecting mother or diabetic.[82] Collectors of this information may sell it to third parties, known as data brokers.[83]

---

70. *Id.*

71. *See supra* notes 1–3 and accompanying text.

72. Fang, *supra* note 57, at 127–128. The data Facebook shared "included diet information, exercise activities, ovulation cycle, and intention to get pregnant." *Id.*

73. Schechner & Secada, *supra* note 1.

74. Frazee, Finley & Rohack, *supra* note 39, at 394. Another study found that twenty out of forty-three top wellness apps disclosed information about its users to third parties. *Id.*

75. *Id.* at 396. Active data is "voluntarily revealed to the service provider by the user," such as a user entering dietary information. *Id.* Passive data is "automatically revealed to the service provider and does not require active participation by the user," such as location data. *Id.*

76. *Id.*

77. *Id.* at 396–97. Thus, in the situation of someone entering dietary information into a diet-tracking app, the active data is the actual dietary input the user enters, the passive data may be the location data from where the user entered the dietary information, and the inferred data will come from the analytical tools predicting this user may have or develop diabetes. *Id.*

78. *Id.*

79. *Id.* at 396 ("A physician is only able to test a finite number of variables during a patient visit, whereas mHealth apps continuously monitor patients' habits.").

80. *See id.* at 398.

81. *Id.*

82. *Id.*

83. *Id.* at 397. Data brokers are simply buyers of consumer information. *Id.* One data broker in particular, Acxiom, reportedly has data segments for almost all U.S. consumers. *Id.*

While some uses of this data may seem more benign, such as targeting a diabetic with sugar-free advertisements, nefarious uses such as selling this information to an insurance company are also possible absent regulation in this context.[84]

C. Non-Health Consumer-Generated Data

Consumers generate significant amounts of data, whether through online interactions, search queries, or spending habits.[85] This data exists largely outside of the health information realm. Yet, as this Part argues, this subset of Big Data use is equally as concerning as other impacts of Big Data on sensitive health information. Nicolas Terry, a leading scholar in the health information privacy field, refers to this kind of data as "medically inflected data."[86] The hallmark characteristic of this kind of data is that it is not generated directly for health or medical purposes.[87]

Online interactions, including social media and blogs, should be considered within the realm of health data.[88] There are countless Facebook groups, Twitter feeds, and blogs that relate to specific conditions.[89] Users that participate in these interactions signal a connection to the specific condition.[90] One example, PatientsLikeMe, asked users to share their experience with various conditions, and these shares were linked with Twitter and Facebook, thus increasing the likelihood this data would be aggregated by commercial entities.[91] Other online activities, such as search queries, allow companies to classify a user based on his or her interests, allowing the companies to sell these labels to advertisers and others.[92]

Once data is collected and aggregated, analytical tools allow Big Data companies to use the consumer-generated data in a number of ways: purchasing data from a store can show when someone is pregnant, online shopping coupled with high cable

---

84. *Id.* at 398.

85. Terry, *supra* note 7, at 85–86. Terry identifies a number of online activities that create this kind of data, including web browsing, online transactions, social media, and smartphone usage. *Id.*

86. *Id.* at 85.

87. *Id.* Big Data is often defined by three key characteristics: volume, velocity, and variety. *Id.* at 78. Terry refers to medically inflected data as "quintessential *high-variety* big data." *Id.* at 85.

88. *See* Janine S. Hiller, *Health Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L.J. 251, 271 (2016).

89. For example, one such group is for women that have a mutation in the BRCA gene, which leads to an increased risk of breast cancer. Kate Fazzini, *Facebook Recently Closed a Loophole That Allowed Third Parties to Discover the Names of People in Private, 'Closed' Facebook Groups*, CNBC (Jul. 12, 2018), https://www.cnbc.com/2018/07/11/facebook-private-groups-breast-cancer-privacy-loophole.html [https://perma.cc/FWN6-LTCX]. In 2018, Facebook found a loophole in its system that allowed third parties to uncover the names of users who had joined what they thought were private, closed groups, such as the breast cancer group. *Id.*

90. Hiller, *supra* note 88, at 272.

91. *Id.*

92. *See, e.g.*, Terry, *supra* note 7, at 85 (discussing Datalogix which classifies consumers as "allergy sufferers" or "dieters" and Acxiom which sells "online search propensity" for diseases or medicines).

TV bills and a van as a vehicle can show a man is overweight, plus-size clothing orders can show someone is depressed, and much more.[93] Much of these analyses lend to targeted advertising; common utilizers of targeted advertising include insurance and credit card companies.[94] Yet, this sensitive health information gleaned from consumer-generated data can be used for more nefarious purposes.[95] One such purpose is discriminatory pricing.[96] With these advanced analytical tools utilized by Big Data companies, consumers could face higher prices for insurance or face negative consequences with employers and financial and educational institutions.[97]

All of these uses occur with both health-related information and non-health information. Targeted advertising occurs with all sorts of products, regardless of the domain, and most readers are likely familiar with suddenly seeing advertisements after a recent internet search.[98] But these kinds of tools become problematic when the information obtained is personal health information. While targeted advertising can be benign, albeit a nuisance, often the advertising itself is not thought to be too much of an invasion of privacy.[99] This reasoning, however, fails for two reasons. First, discriminatory pricing can result when the advertisers know sensitive details about their targets.[100] Second, regardless of whether the advertising is benign, the disclosure itself is a violation of privacy. Returning to the analogy of a brick-and-mortar physician visit, patients do not want their physicians selling information such as their seasonal allergies to a pharmaceutical company so the company can send targeted advertising. Absent express consent, this would breach HIPAA,[101] and the patient would be understandably upset about the occurrence. Yet, the very same information can be shared with the very same pharmaceutical company without any consent by the consumer simply because it was collected and held by a company to which HIPAA does not apply.[102]

---

93. Fang, *supra* note 57, at 135–36.

94. *Id.*

95. One egregious example is where a list of "Suffering Seniors" was created based on data about Alzheimer's and cancer, and the list was sold to nefarious actors that tricked the seniors into revealing financial information. *Id.* at 140.

96. *Id.* at 139.

97. *Id.*

98. A Pew Research study in 2019 found that 77% of Americans are aware of targeted advertising, and 64% have seen targeted advertisements based on their data. Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner, *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*, PEW RSCH. CTR. (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/ [https://perma.cc/D7LW-85NJ].

99. The aforementioned Pew Research study found that despite broad concerns about data collection and monitoring, only 39% of Americans worried about the information advertisers collected. *Id.*

100. Fang, *supra* note 57, at 139.

101. *See supra* note 54.

102. For an analysis of why HIPAA does not apply, see *infra* Part II.D.

D. Why HIPAA's Privacy Protections Do Not Apply

In the context of the health information collected by Big Data companies, such health information falls outside of HIPAA's regulations. To best frame the inquiry into HIPAA's protections, the first thing to note is that the Privacy Rule focuses on downstream data protection.[103] The implication of downstream data protection is that HIPAA does not restrict the collection of health information; rather, it only regulates how those who have collected the health information may handle the information, which essentially limits HIPAA's protections to confidentiality.[104]

Thus, the proper inquiry is not whether HIPAA prevents the collection of the information obtained and aggregated by Big Data organizations, but rather whether HIPAA prevents the disclosure of this information. The answer is only rarely.[105] For HIPAA's privacy provisions to apply, the Big Data organization must be either a covered entity or a business associate.[106] Under the current framework, though, a covered entity is only a health care provider, health plan, or health care clearinghouse.[107] Granted, some mobile apps are covered entities and thus subject to HIPAA's regulations.[108] On the whole, however, Big Data organizations are rarely considered to be one of the three types of covered entities. Consider the vast majority of mobile health apps as an example. Apps that collect general health and wellness information, such as MyFitnessPal, are not covered entities because they are not a health care plan, provider, or clearinghouse.[109] Apps that collect and track health information related to specific conditions likewise are not covered entities, again, because these applications are not health plans, providers, or clearinghouses.

Unable to qualify as a covered entity, a Big Data organization would otherwise be subject to HIPAA's regulations only if it is a business associate of a covered entity, which requires the Big Data organization to create, maintain, or receive the protected health information on behalf of a covered entity.[110] Business associates, therefore, typically arise only out of contractual relationships between covered entities and Big Data organizations. For example, when considering whether HIPAA applies to wearable technology, we must determine "who the users of the wearable technology will be and who will have access to the information collected by a device or application."[111] If the user inputs health information into a device, or if the user generates the health information, HIPAA does not apply because the custodian is not

---

103. Terry, *supra* note 7, at 68.

104. *Id.*

105. *See, e.g.*, Tovino, *supra* note 65, at 175–76 (discussing how mobile apps that collect sensitive health information, such as PatientsLikeMe and MyFitnessPal, fall outside of HIPAA's regulation).

106. *See, e.g.*, Fowler & Morain, *supra* note 69, at 210.

107. 45 C.F.R. § 160.103 (2019).

108. For example, telehealth apps, which a patient and a physician use to communicate with each other remotely, are covered entities because it is likened to a physically present health care provider. Frazee, Finley & Rohack, *supra* note 39, at 393.

109. Tovino, *supra* note 65, at 175.

110. *See* 45 C.F.R. § 160.103 (2019).

111. Newman & Kreick, *supra* note 12, at 448.

a covered entity, there is no contractual relationship between the wearable technology company and a covered entity, and the information came from the user.[112]

Mobile apps are likewise often not business associates of covered entities. This is because in a case where a user downloads the app and populates the app with sensitive health information either directly or indirectly, the app is not receiving or maintaining the data specifically on behalf of a covered entity or another business associate; rather, the individual is using the app independently of a covered entity.[113] Mobile apps, like other Big Data companies, are business associates only when they contract with providers to offer the providers' patients services such as health counseling, electronic health records, or other health monitoring services.[114] Other consumer-generated data such as search queries and social media fall outside of HIPAA's protections for the same reasons: health information obtained by Big Data through the vast data analytics tools is not held by a covered entity or a business associate absent a contractual agreement.[115] Thus, sensitive health information collected, stored, analyzed, or used by Big Data organizations through media such as wearable technology, mobile apps, and other consumer-generated means fall outside the scope of HIPAA because the data fails the custodial requirement.

Further complicating the picture is the source requirement. Even if the custodial requirement ceased to exist, the information obtained comes from the users themselves or from analytical tools.[116] HIPAA currently requires that, in addition to the custodial requirement, the health information must also be collected or received by a covered entity or another included source such as an employer or school.[117] While much of the focus of this Section has been on how Big Data falls outside the covered entity-business associate paradigm, and thus falls outside HIPAA's purview, most of Big Data's health information also falls outside of the source requirement.[118] If the Big Data organization is not a covered entity or business associate, then the health information collected will also fail to meet the source requirement.[119] A heart rate reading collected by wearable technology or inputted into a mobile app falls outside of HIPAA not only because the wearable tech or app is not a covered entity or business associate but also because the data was collected from the user and thus was not collected or received by one of the required sources for HIPAA to apply.

The result of HIPAA's narrow coverage is that none of the aforementioned personal health information is protected, even though this exact information would

---

112. *Id.*

113. OFFICE FOR CIVIL RIGHTS U.S. DEP'T OF HEALTH & HUMAN SERVS., HEALTH APP USE SCENARIOS & HIPAA 2 (2016), https://www.hhs.gov/sites/default/files/ocr-health-app-developer-scenarios-2-2016.pdf [https://perma.cc/2CF3-G4E8].

114. *Id.* at 3.

115. *See* Fang, *supra* note 57, at 146–48.

116. Newman & Kreick, *supra* note 12, at 449.

117. *See* 45 C.F.R. § 160.103; *see also supra* text accompanying note 27.

118. The information must originate from a "health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse" to be health information. 45 C.F.R. § 160.103 (2019).

119. *See id.* Mobile apps, wearable technology, and other forms of Big Data do not meet the source requirement because the Big Data company is not a provider, health plan, public health authority, employer, life insurer, educational institution, or clearinghouse.

be protected from disclosure if it were obtained by a provider or insurance plan. The profiles created through wearable technology, mobile applications, and consumer-generated data that show either general health or specific conditions are not subject to HIPAA regulation. Take an application that tracks depression as an example; the application is free to link the mood state to a personal identifier for the user and disclose this information to anyone and everyone, subject only to the language the application puts in its privacy policy.[120] Even the Kinsey Institute's Kinsey Reporter app, which it advertises as "anonymous," collects sensitive information that would be protected if the very same information were collected by a provider.[121] Though touted as "anonymous," personal identifiers such as geographic location means the data is not sufficiently de-identified by HIPAA standards to allow for disclosure without consent.[122] Thus, the Kinsey Institute is free to disclose this sensitive information that would otherwise be protected under HIPAA if the same was obtained in a traditional doctor-patient interaction.

As this Part demonstrates, Big Data organizations compile sensitive health information in a multitude of ways, such as wearable technology, mobile apps, and other consumer activities. This sensitive information, however, falls outside of HIPAA's protections. The next Part proposes amending HIPAA to include this health information that currently falls victim to Big Data.

III. AMENDING HIPAA'S REGULATIONS TO FOCUS ON THE HEALTH INFORMATION ITSELF

Since its inception, HIPAA has received criticism for its narrowness.[123] This Note is far from the first to suggest structural changes to the privacy framework. Many commentators have proposed various solutions to address the aforementioned gap in HIPAA protections. For example, some have proposed amending the covered entities definition to include certain additional categories such as wearable technology or mHealth apps.[124] Others have addressed the issue from the business associate angle, where these proposals amend the definition of a business associate to include, for example, wearable technology companies.[125] Yet, these proposals struggle from two common problems: (1) the proposals are limited to certain additional contexts that ignore other substantial uses of health information by Big Data organizations, and

120. Eric Rakestraw, *One Size Doesn't Fit All Why HIPAA Should Not Be Extended to Cover PHRs*, 30 J. LEGAL MED. 269, 283 (2009) (discussing how a company's violation of its privacy policy is a "deceptive business practice" subject to FTC action).

121. *See* Tovino, *supra* note 65, at 163.

122. *Id.*

123. Terry, *supra* note 7, at 67.

124. *See, e.g.*, Fang, *supra* note 57, at 171–78 (proposing including mHealth application companies as covered entities along with co-regulation with the FTC); Paige Papandrea, Note, *Addressing the HIPAA-potamus Sized Gap in Wearable Technology Regulation*, 104 MINN. L. REV. 1095, 1121–25 (2019) (proposing amending the covered entities definition to include wearable technology companies).

125. *See, e.g.*, Grant Arnow, Note, *Apple Watch-ing You: Why Wearable Technology Should Be Federally Regulated*, 49 LOY. L.A. L. REV. 607, 632–33 (2016) (proposing amending the business associates definition to include wearable technology).

(2) these proposals retain the custodial nature of HIPAA protections. By proposing amending the covered entity or business associate definitions, the commentators focus on including only a subset of Big Data companies, such as wearable technology[126] or mHealth applications.[127] While these plans address health information privacy in an additional setting, they ignore the broader impacts of Big Data on private health information in other settings. Furthermore, expanding the business associate definition fails because many of these Big Data companies exploit health information independently of covered entities;[128] thus, an expanded definition of business associates would not apply in those situations.[129] Narrowly amending the definition also has a secondary consequence: the regulatory framework still depends on the custodian of the health information rather than the health information itself.

This Note provides a framework that encompasses all potential impacts of Big Data on personal health information and shifts the focus from the custodian and source of the information to protecting the health information itself. This Part explores a state's solution to this problem as well as the European Union's solution to this problem, using these approaches as a guide to a proposal for a uniform, national standard for health information disclosure regulation.

A. Texas's State Definition of Covered Entities

Under the current system, only a handful of states have health information privacy provisions that apply privacy protections more broadly than the current HIPAA framework. This is due to several reasons.[130] First, some states have adopted HIPAA regulations in a wholesale manner by conforming identically to the federal regulations.[131] Another reason is that previous administrations have characterized diverging state laws as slowing the implementation of HIPAA's other goals,[132] such as increasing the use of technology in the traditional health care sector. One state, however, has taken an interesting approach that contributes to this Note's analysis: Texas.

Texas has codified a much broader definition of covered entity.[133] The Texas statute provides the following: "'Covered entity' means any person who: (A) for

126. *See* Papandrea, *supra* note 124, at 1122–23.

127. *See* Fang, *supra* note 57, at 171–78.

128. OFF. FOR CIV. RTS. U.S. DEP'T OF HEALTH & HUM. SERVS., HEALTH APP USE SCENARIOS & HIPAA 2 (2016), https://www.hhs.gov/sites/default/files/ocr-health-app-developer-scenarios-2-2016.pdf [https://perma.cc/CM37-TGN4].

129. The business associate definition requires the business associate to perform services specifically for a covered entity. 45 C.F.R. § 160.103. Thus, if a Big Data company collects and uses health information independently of a covered entity, such as with mobile apps, an expanded business associate definition is insufficient because there is no corresponding covered entity.

130. Terry, *supra* note 7, at 90.

131. *Id.*

132. *Id.* The Bush Administration said diverging state laws hindered implementation of electronic health records. *Id.*

133. *See* TEX. HEALTH & SAFETY CODE ANN. § 181.001(b)(2)(A) (West, Westlaw through 2019).

commercial, financial, or professional gain, monetary fees, or dues, or on a cooperative, nonprofit, or pro bono basis, engages . . . in the practice of assembling, collecting, analyzing, using, evaluating, storing, or transmitting protected health information."[134] Thus, Texas has rejected the three narrow categories that make up the HIPAA definition of a covered entity[135] in favor of including virtually any organization that comes into contact with health information. It is important to note, however, that while this solution eases the narrowness of the custodial requirements for health information, it does not address the source of health information. Rather, Texas's broader regulations include organizations that come into contact with health information only after the health information originated from a few select sources.[136]

B. The European Union's GDPR

The European Union recently passed the General Data Protection Regulation (GDPR), and as one commenter notes, it is "quickly becoming known as the global standard in all industries."[137] The GDPR regulates not just health data but all data, regardless of the industry or manner in which it is obtained or stored, and it focuses on regulating data controllers and data processors.[138] The GDPR does, however, have provisions targeted specifically to the regulation of health data, specifying that this health data includes "all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject."[139] Furthermore, the default consent for disclosure of health information under the GDPR requires explicit consent.[140] Thus, these provisions demonstrate how the GDPR's approach to health data includes all health information without restricting the protections to a particular class of custodians or sources.[141]

---

134. *Id.*

135. *See supra* note 15 and accompanying text. Texas does not require a covered entity to be a provider, health plan, or health care clearinghouse; rather, Texas defines a covered entity as almost any person that possesses health information. *See* HEALTH & SAFETY § 181.001(b)(2)(A).

136. *See* HEALTH & SAFETY CODE ANN. § 181.001(a) (stating that any definition not enumerated in this section, such as the definition of health information, adopts the meaning in HIPAA's privacy regulations). Because Texas adopts the HIPAA definition of health information, the particular health information must satisfy the substance and the source requirement. *See supra* note 27 and accompanying text.

137. Tovino, *supra* note 65, at 174.

138. *Id.* at 177. A data controller under the GDPR is a "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data." Commission Regulation 2016/679, art. 4(7), 2016 O.J. (L 119) 1, 33 [hereinafter GDPR]. A data processor under the GDPR is a "natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." *Id.* art. 4(8).

139. GDPR, *supra* note 138, prmbl., para 35.

140. *Id.* art. 9(1)–(9)(2)(a).

141. Tovino, *supra* note 65, at 178–79.

C. Moving Towards a Framework that Includes all Big Data Uses of Health Information

This Note proposes two changes to the current privacy framework of HIPAA which would protect the disclosure of sensitive health information in the context of Big Data. First, Congress should broaden the covered entities definition,[142] whereby the amended definition would include all organizations or persons collecting, storing, analyzing, using, or transmitting health information. This addition is analogous to Texas's broad definition of covered entities. Second, Congress should amend the definition of health information to remove the requirement that the information must originate from a provider, health plan, or other limited source, thus leaving the definition of health information as "any information, including genetic information, whether oral or recorded in any form or medium, that: . . . (2) [r]elates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual."[143] This change reflects the definition of health information employed by the GDPR.

These changes have two important implications. First, applying the restrictions of covered entities more broadly to any person or organization possessing sensitive health information mitigates the custodial nature of the current HIPAA framework. Additionally, removing the requirement that health information must originate from a particular source, such as a provider or health plan, abrogates the source requirement. Therefore, the new framework, after just minor adjustments, regulates the privacy of health information based on the nature and substance of the information itself, and not merely by who is currently holding the information and from where the information came.

Reviewing the many examples of Big Data impacts on health information shows how this new definition provides a HIPAA framework that protects health information from unwanted disclosure by Big Data organizations. By altering covered entities to include any organization collecting, analyzing, or using sensitive health information, no longer will Big Data escape HIPAA's regulations by arguing they are not a health provider, health plan, or health care clearinghouse. Mobile applications or Big Data analytical tools that obtain information about a user's condition will be in possession of sensitive health information and thus, under the proposed framework, must comply with the disclosure requirements.

Equally important is the removal of the source requirement. For too long, Big Data has been able to claim that they received the information directly from the consumers, and thus the health information did not originate from a health provider or the like. Under the proposed framework, these organizations will now possess protected health information. In the case of Big Data organizations that obtain information about an individual's condition, whether directly from the individual or through analytical tools, the information will certainly relate to the individual's current or future health condition and thus fall under HIPAA's regulations. In the

---

142. 45 C.F.R. § 160.103 (2019).

143. *Id.* The proposed definition of health information retains only subsection (2) of the current definition.

situations of wearable technology that measure heart rate, applications that track a user's diabetes, or analytical tools that label individuals as diabetic, all three sets of data will reflect an individual's health condition and thus be subject to HIPAA's protections.

This Note's proposed changes came after the consideration of other alternatives. One such consideration was to remove the covered entities language entirely and apply HIPAA's privacy regulations universally to all persons and organizations, a solution that other commentators have previously suggested.[144] Although this solution would have resulted in greater privacy to consumers and users of wearable technology, mobile health applications, and other Big Data analytical tools, this solution ultimately proved unwieldy. First, this is a much larger change to the legislation and regulation, thus increasing the likelihood of strong opposition. Second, this solution would apply too broadly. For example, neighborly conversations and familial interactions could suddenly become subject to HIPAA's regulations if health information was discussed. Thus, the solution on which this Note settled strikes a balance between protecting health information privacy from disclosure by Big Data companies while not overzealously extending the privacy provisions beyond the Big Data context.

While this Note's proposed framework will certainly increase privacy protections of sensitive medical information, the framework will not be without criticism. For starters, this proposal requires a legislative solution, which can prove difficult in polarized government. This Note's proposed solution, however, includes the language adopted by the state of Texas in an overwhelmingly bipartisan, and nearly unanimous, legislative vote.[145] The result in Texas provides hope that a federal solution employing substantially the same language could garner bipartisan support. There are some critiques in addition to whether the proposal is politically feasible, however, that will be addressed in the following sections.

1. Reliance on the Competitive Market

Many favor relying on market forces to create consumer privacy, even in the context of health information.[146] Recently, a few notable companies have pledged privacy protections of sensitive health information, including Fitbit and Apple.[147] These companies have extensive privacy policies, many of which include similar protections of health data.[148] The reliance on the competitive market theory is as

---

144. *See, e.g.*, Alexis Guadarrama, Comment, *Mind the Gap: Addressing Gaps in HIPAA Coverage in the Mobile Health Apps Industry*, 55 HOUS. L. REV. 999, 1020 (2018).

145. *See, e.g.*, H.R. JOURNAL, 82d Leg., Reg. Sess. 6700 (Tex. 2011), https://journals.house.texas.gov/hjrnl/82r/pdf/82RDAY89FINAL.PDF#page=240 [https://perma.cc/U3AZ-BTYX] (145 Yeas, 0 Nays, with some not voting in the Texas House); S. JOURNAL, 82d Leg., Reg. Sess. 5046 (Tex. 2011), https://journals.senate.texas.gov/sjrnl/82r/pdf/82RSJ05-29-F.PDF#page=6 [https://perma.cc/2XDQ-VBVN] (31 Yeas, 0 Nays in the Texas Senate).

146. Rakestraw, *supra* note 120, at 284.

147. Newman & Kreick, *supra* note 12, at 430–31.

148. *See, e.g.,* Frazee, Finley & Rohack, *supra* note 39, at 394 (discussing a study by Privacy Rights Clearinghouse that analyzed forty-three of the top wellness apps and found

follows: the market will ensure a proper level of privacy protection because consumers will dictate their desired privacy level based on purchasing and usage habits.[149] Thus, companies that fall below the consumer-dictated level of privacy would suffer economically as the companies lose business to those with more desirable levels of privacy.[150] Additionally, these companies are bound to follow their privacy policies or face reprimand by the Federal Trade Commission (FTC).[151]

This argument fails, however for several reasons. First, the market is often unable to properly correct when consumers face information asymmetry.[152] For example, consumers may be aware that applications that track diet or certain conditions may share this information with advertisers.[153] This conclusion is not unfounded in modern society. But consumers may not be aware that inferred information passively collected is used to label users for advertisers.[154] Furthermore, consumers may be unaware of the depths of information Big Data can glean from the variety of data sources ubiquitous in society. The Flo example in the Introduction illustrates this point.[155] It noted that Facebook used the third-party information regardless of whether the user actually had a Facebook profile; if the user did not have a profile, then Facebook simply made its own profile of the user for its purposes.[156] In this situation, a consumer without a Facebook profile may incorrectly believe he or she is not at risk of the health information being shared, and this is due to information asymmetry. Relying on consumer preferences to correct markets can work in situations in which the consumer has enough information to adequately make informed decisions, but that is often not true when it comes to Big Data's collection of sensitive health information.

Second, the reliance on the enforcement of privacy provisions requires two things: there must actually be a privacy provision that includes these kinds of protections, and the consumer must read and understand the scope of the protection.[157] While good business practices would support companies including a privacy policy,[158] without a regulatory framework for Big Data companies, companies can exploit the lack of a privacy policy requirement to carefully draft or omit provisions of a privacy

---

only half the apps complied with their own privacy policy).

149. Rakestraw, *supra* note 120.

150. *Id.* at 284–85.

151. *Id.*

152. *See* Shmuel I. Becher, *Asymmetric Information in Consumer Contracts: The Challenge That Is Yet to Be Met*, 45 AM. BUS. L.J. 723, 734 (2008) ("Where imperfect information exists, the ability of parties to maximize utility via open market transactions will inevitably decrease."). Information asymmetry is defined as "situations where parties are differently informed, with one party having access to better or more information that the other." *Id.* at 733.

153. *See supra* note 98 and accompanying text (finding that a majority of Americans are aware of targeted advertising).

154. Frazee, Finley & Rohack, *supra* note 39, at 397–98.

155. *See supra* note 1 and accompanying text.

156. *See supra* note 1.

157. The Pew Research Center found that, per a 2019 study, only 9% of adults in America always read the privacy policy, and only 13% often read the privacy policy. *See* PEW RSCH. CTR., *supra* note 98.

158. Rakestraw, *supra* note 120.

policy for information that would otherwise be protected under HIPAA regulations. The reliance on the FTC to protect consumers is misplaced in this setting because the FTC will only take action against an app or wearable technology company when it misleads the consumer.[159] The FTC will not step in merely because a company chose to provide little to no privacy so long as there are no false or misleading statements.[160] Additionally, a mere nine percent of Americans actually read every terms and conditions request before agreeing.[161] Further complicating the picture is that even if consumers do take the time to search out and read these policies, many such privacy policies allow the organizations to unilaterally alter the terms.[162] While consumer apathy towards reading privacy policies carefully does not control, it cuts in favor of imposing a federal regulatory scheme when coupled with everchanging privacy polies and information asymmetry between the consumer and the Big Data organizations.

2. Stifling Innovation

Imposing privacy burdens on Big Data companies, as some argue, would have the effect of stifling innovation in this area.[163] One relevant concern is that these companies must profit in order to survive and provide the services such as tracking a particular condition.[164] Supporters argue that physicians and other covered entities under the current framework have other ways to profit, such as billing for their services, so allowing these covered entities to profit from the information they obtain from patients should not be allowed.[165] On the other hand, because the Big Data organizations do not necessarily have a stream of revenue outside of using the health information for profit, it is acceptable, or even desirable, to allow these companies to profit from users' sensitive health information in the name of continuing the utility and innovation of these devices, applications, and analytical tools.[166]

This argument is disconcerting for several reasons. First, this kind of argument ignores the substance of the information and rather relies on the custodial view of HIPAA. Information that one is a diabetic cannot be shared if a physician obtains that information, but it can be shared for an easy profit to an advertising company if it is a mobile application that obtains the information. Thus, the argument focuses not on whether the information is sensitive and thus worthy of privacy but rather whether the custodian of the information deserves privacy regulations.

---

159. Fowler & Morain, *supra* note 69, at 211.

160. *Id.*

161. Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner, *Americans' Attitudes and Experiences with Privacy Policies*, PEW RSCH. CTR. (Nov. 15, 2019) https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/ [https://perma.cc/Z7AH-PZ2P].

162. Fowler & Morain, *supra* note 69, at 211.

163. Rakestraw, *supra* note 120.

164. *Id.*

165. *Id.* at 284 ("Regulations would have little or no effect on the overall viability of most covered entities.").

166. *See id.*

This leads to the second concern stemming from this argument: it accepts that sensitive health information should just be sold for a profit. As the quid pro quo of providing these applications and offering these devices, the companies and organizations are free to exploit the users' information, regardless of how sensitive the information is.[167] Yet, this draws a false dichotomy. The decision is not binary, with these devices and applications but no privacy on one end and privacy but none of these devices and applications on the other. Rather, there can be a middle ground—for example, HIPAA allows disclosure of health information with the patient's or user's consent.[168] Even if these organizations fall under HIPAA's purview, the organizations are still able to sell the information they obtain; the organizations just have to obtain the user's or individual's authorization before sale. Thus, innovation in Big Data and privacy of sensitive health information do not have to be mutually exclusive.

3. First Amendment Concerns

Because this Note advocates for a shift in HIPAA's coverage towards regulating sensitive health information due to the information itself rather than based on the custodian or source of the information, opponents may raise First Amendment constitutional concerns. Because information is speech,[169] this Note's proposal to regulate and prevent the unauthorized disclosure of health information provides, at the very least, a colorable argument of a First Amendment violation against free speech. Though these concerns largely exist in HIPAA's current form, as even with the custodian and source requirement the current framework still regulates the information based on the content and speaker of the information,[170] the proposed framework makes this kind of regulation more explicit and thus opens the door to additional challenges.

Opponents could argue that the proposed framework is a content-based restriction on speech, as the proposal regulates sensitive health information due to the content

---

167. *Id.* at 284 ("Simply put, if the ability to profit from these services were removed, Google and its competitors would have no reason to offer the services in the first place.").

168. The issue of consent, or authorization, is complex under HIPAA. There are certain situations in which a covered entity must disclose the information regardless of authorization. *See, e.g.*, 45 C.F.R. § 164.502(a)(2) (2019). There are other situations in which a covered entity is permitted to disclose regardless of authorization. *See, e.g.*, *id.*§ 164.502(a)(1). The situations envisioned here, however, are those in which authorization is required for disclosure. One instance is sale of the health information. Sale of this information is prohibited, *id.* § 164.502(a)(5)(2), but a covered entity may sell the information so long as it first informs the patient and obtains sufficient authorization. *See id.* § 164.508(a).

169. *See* Bartnicki v. Vopper, 532 U.S. 514, 527 (2001) (quoting the lower court's opinion that "[i]f the acts of 'disclosing' and 'publishing' information do not constitute speech, it is hard to image what does fall within that category").

170. HIPAA currently regulates based on content through its definition of health information, which requires the information must relate to a past, present, or future medical condition. 45 C.F.R. § 160.103 (2019). HIPAA currently regulates based on speaker through its application of the privacy provisions to covered entities only, which consists of three groups of speakers. *Id.*

of the information.[171] Content-based speech restrictions must pass strict scrutiny,[172] a high bar that invalidates most legislation falling into this category. On the other hand, if this proposal is deemed content-neutral, it need only pass intermediate scrutiny,[173] a lower bar that is much easier to satisfy. Although this Note's focus is not on the Court's First Amendment jurisprudence, a brief look at a particular case can shed light on how constitutional concerns about the proposed framework may be addressed even if a court determines the proposal is a content-based regulation and thus subject to strict scrutiny.

In *Sorrell*, the United States Supreme Court examined a state law that prevented the sale and disclosure of prescriber-identifying information to pharmaceutical companies for marketing purposes.[174] Though not directly analogous to the proposed framework, the Court confronted content- and speaker-based restrictions on speech in the health care field.[175] Although the Court ultimately held the state law violated the First Amendment,[176] dicta in Justice Kennedy's opinion provides insight into how a court may view a First Amendment challenge to the proposed framework.[177] Noting that one of the goals of the state law was to protect the privacy of prescriber information, Justice Kennedy was troubled by how the state law "allow[ed] the information to be studied and used by all but a narrow class of disfavored speakers."[178] Justice Kennedy even pointed to HIPAA as an example of a "more coherent policy" for privacy, where information is allowed to be shared in only a few "well-justified circumstances."[179] Justice Kennedy wrote that the state would have a "stronger position" had the statute prevented the sale and disclosure except in limited situations.[180]

While not directly analogous, Justice Kennedy's opinion in *Sorrell* sheds light on this Note's proposed framework, which follows very closely with his suggestion of a stronger argument for constitutionality by preventing the sale and disclosure of sensitive health information except in limited circumstances (e.g., authorization by the individual). While constitutional concerns over the proposed framework are

---

171. A law or regulation is content-based if it "applies to particular speech because of the topic discussed or the idea or message expressed." Reed v. Town of Gilbert, 576 U.S. 155, 163 (2015). There may be an argument that this Note's proposal is not content-based because it is not discriminatory based on the message or idea expressed but rather prohibits a narrowly tailored category of speech absent an individual's authorization. However, for a cursory discussion of the constitutionality of the proposed framework, this Note will address the strongest opposing argument, which is that the proposed framework is content-based and thus must pass strict scrutiny.

172. *See, e.g.*, *id.* at 163–64 (2015). Strict scrutiny requires the regulation or law to serve a compelling state interest and be narrowly tailored in the least restrictive way possible. *Id.*

173. *See, e.g.*, Turner Broad. Sys. v. FCC, 520 U.S. 180, 189 (1994). Intermediate scrutiny requires the regulation serve an important state interest and "not burden substantially more speech than necessary to further those interests." *Id.*

174. Sorrell v. IMS Health Inc., 564 U.S. 552, 557 (2011).

175. *Id.* at 563–64.

176. *See id.*

177. *See id.* at 573–80.

178. *Id.* at 573.

179. *Id.*

180. *Id.* at 580.

certainly valid, these concerns should not outweigh the benefits of moving towards protecting sensitive health information in the Big Data context.

Although these three critiques of this Note's proposal, prioritizing the competitive marketplace, stifling innovation, and First Amendment values, are certainly valid concerns, on the whole, these concerns should not outweigh the advantages of placing these privacy restrictions on Big Data companies. Furthermore, these companies will still be able to compete in the marketplace and profit, just within the HIPAA framework.

CONCLUSION

From wearable technology, to mobile applications, to consumer-generated data, Big Data is becoming ubiquitous in our lives. As Big Data collects more and more information about individuals, privacy concerns arise. This Note has focused on one particular privacy concern: health information. Big Data is able to collect and use vast amounts of sensitive health information, the very same kind of information that would be protected from disclosure if it were obtained by a provider or health plan and then held by a provider, health plan, or health care clearinghouse. Because HIPAA narrowly focuses on applying its regulations to health information only from these sources and only when held by these custodians, sensitive health information obtained by Big Data organizations falls outside of the current framework of HIPAA. By removing the source requirement and amending the custodian categories, or covered entities, to include all persons or companies interacting with health information, HIPAA's focus will shift towards protecting the health information itself. Thus, under the proposed framework, Big Data companies that obtain sensitive health information by any means will be subject to HIPAA, and they will be unable to disclose or sell the information without the individual's authorization.

This Note merely adds to the literature of proposed solutions to fix the gaps in HIPAA's regulations. But this Note's proposal strikes a balance between regulating all instances of Big Data, as opposed to being focused solely on mobile health applications or wearable technology, while also not applying these protections overbroadly so as to discourage innovation or become infeasibly difficult to introduce and administer. With this Note's proposal, the couple who has made the decision to pursue pregnancy and enlisted a mobile application to help track and chart the progress will not have to worry that the application will disclose this private information, absent the couple's authorization.