Anonymized EHRs Create Potential Liability for Google and Hospital

By Gail L. Gottehrer, JD, and Ronald J. Hedges, JD

A recently filed class action highlights the risk that the proliferation of health-related smartphone apps, fitness trackers, and the data these programs collect, may enable the re-identification of purportedly de-identified electronic health records (EHRs). The action raises questions about the need to examine existing de-identification requirements and processes to ensure that these keep pace with developments in technology, such as artificial intelligence (AI) and machine learning. It also presents questions of whether those requirements and processes are sufficient to protect patient privacy and confidentiality.

The Allegations

In the class action complaint against Google, the University of Chicago Medical Center, and the University of Chicago (the latter two entities referred to here as the "University"),[1] former University patient Matt Dinerstein alleged that in 2017 the University transferred its EHR consisting of thousands of patient medical records to Google without the express consent of those patients. Dinerstein asserted claims arising out of alleged violation of the Illinois Consumer Fraud and Deceptive Business Practices Act as well as claims for breach of contract, tortious interference with contract, intrusion upon seclusion, and unjust enrichment.

While Google explained that it was partnering with the University to enable it to "research ways to use machine learning to predict medical events," according to the lawsuit, Dinerstein contended that Google's actual motivation was financial—namely, to obtain patient health data that could be used by DeepMind Health, an AI and machine learning entity that Google acquired. This acquisition was allegedly part of Google's decade-long focus on the healthcare market. In recent years Google has worked on creating apps and cloud services such as G Suite for healthcare businesses and Google Cloud for HIPAA-compliant workloads. The company also has launched fitness-tracking products such as Google Fit, and has added features to Google Search to make it easier for consumers to research health-related questions.

In addition to failing to obtain consent from patients to provide these records to Google, Dinerstein alleged that although Google and the University represented that the patient data had been de-identified, the records "were not sufficiently anonymized and put the patients' privacy at grave risk." An article published by Google and the University in 2018[2] revealed that datestamps from University patients' records were maintained, as were "de-identified free-text medical notes." Other data included patient demographics, provider orders, diagnoses, procedures, medications, lab values, vital signs, and flowsheet data.

Underscoring the importance of effective de-identification of medical data, Dinerstein's lawsuit cites two examples of "researchers with limited access to public data sets and supposedly de-identified medical records [who] have been able to re-identify patients." The first was an instance where, using only publicly available data they bought for $50, researchers at Harvard's Data Privacy Lab re-identified 43 percent of de-identified medical discharge records. In the other example, using AI and machine learning tools to analyze data from fitness trackers and information such as age, gender, education level, income, race, and country of birth, researchers were able to re-identify 95 percent of a data set from medical records. One of the authors of that study opined that the results show that "machine learning can successfully re-identify the de-identified physical activity data of a large

percentage of individuals, and this indicates that our current practices for de-identifying physical activity data are insufficient for privacy," according to the lawsuit. He went on to caution that, "[m]ore broadly it suggests that other types of health data that have been thought to be nonidentifying could potentially be matched to individuals by using machine learning and other artificial intelligence technologies."7

Using these examples as guides, Dinerstein concluded that, when medical records that contain datestamps and free-text notes are shared, there is a high probability that the shared data can be re-identified. Moreover, "when the transfer of medical records is made to Google, the ability to re-identify those records becomes a certainty" because "Google is one of the largest and most comprehensive data mining companies in the world."

As the lawsuit alleges, "Google has access to public and nonpublic information that could easily lead to the re-identification of the medical records it received from the University." That information includes geolocation data from Google Maps, and when combined with data from Google email apps, could identify when an individual entered the hospital for treatment, which buildings they entered and exited, and the amount of time spent at each location.

The fact that Google has access to all this information increases the chances that medical records provided by the University could be used to re-identify an individual by name and compromise patient privacy.

The Causes of Action

Dinerstein filed his class action complaint in the US District Court for the Northern District of Ilinois. He alleged that subject matter jurisdiction, which gives the court the authority to adjudicate the claims in issue, arises out of 28 U.S.C. Section 1332 (d)(2). That statute gives district courts subject matter jurisdiction over "any civil action in which the matter in controversy exceeds the sum or value of $5,000,000, exclusive of interest and costs, and is a class action." Dinerstein alleged no claims under federal law. Instead, he asserted claims under Illinois statutory and common law. With that as a very basic background, the first defense that Google and the University might assert is a lack of standing.

Article III of the US Constitution allows federal courts to decide "cases or controversies." To establish Article III standing, a plaintiff must have suffered an "injury in fact," which the US Supreme Court has defined to be "an invasion of a legally protected interest" that is concrete and particularized and actual or imminent, not conjectural or hypothetical, *Patel v. Facebook, Inc.,* No. 18-15982 (9th Cir. Aug. 8, 2019), quoting *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 560 (1992). Standing may be particularly difficult to allege when a plaintiff asserts a claim under a statute—such as the Illinois statute identified in the class action complaint—which sets forth a statutory requirement. The mere failure to comply with that requirement does not establish standing absent a real risk of harm. This has proved to be a stumbling block for plaintiffs who have sought to recover damages for "mere" data breaches without tangible harm since the Supreme Court decided *Spokeo, Inc. v. Robins*, 136 S.Ct. 1540 (2016).

The standing defense is complicated by a decision of the Illinois Supreme Court, *Rosenbach v. Six Flags Entertainment Corp.*, 2019 IL 123186 (Ill. 2019), which interpreted an Illinois statute, the Biometric Information Privacy Act (BIPA), to allow a claim for damages arising out of BIPA to proceed without any allegation of actual damage. The court held that the violation of the statute was enough to state a claim. *Rosenbach* was "featured" in the Ninth Circuit Court of Appeals' *Facebook* decision, cited above. Based on *Rosenbach*, that court held that allegations that Facebook violated BIPA were sufficient to confer standing. In other words, standing can be complicated.

Points to Consider

The Dinerstein case raises many important issues for health information management (HIM) professionals, not covered in the scope of this article. Looking at the lawsuit from the perspective of an HIM professional, some points to consider include:

- When might the duty to preserve relevant health records arise?
- What might the scope of any duty be?
- How might electronically stored information—or physical things such as paper records—be located and in what form should different information be preserved?

These and other topics are certain to be discussed as Google and the University defend against the claims asserted in the class action complaint.

Note

1. *Matt Dinerstein v. The University of Chicago, Google LLC, and The University of Chicago Medical Center.* https://edelson.com/wp-content/uploads/2016/05/Dinerstein-Google-DKT-001-Complaint.pdf.
2. Alvin Rajkomar et al. "Scalable and accurate deep learning with electronic health records." *Digital Medicine*. May 9, 2018. https://www.nature.com/articles/s41746-018-0029-1.

Gail Gottehrer (ggottehrer@outlook.com) is the founder of the Law Office of Gail Gottehrer LLC, a lawfirm focused on emerging technologies. Ron Hedges (r_hedges@live.com) is a former US Magistrate Judge in the District of New Jersey and is a writer, lecturer, and consultant on topics related to electronic information. He is a senior counsel with Dentons US LLP.

Article citation:
Hedges, Ron. "Anonymized EHRs Create Potential Liability for Google and Hospital" *Journal of AHIMA* 90, no.9 (September 2019): 26-27.