VIEWPOINT

Health Care AI and Patient Privacy—*Dinerstein v Google*

Mindy Nunez Duffourc, JD
Maastricht University Faculty of Law, Maastricht, the Netherlands.

Sara Gerke, Dipl-Jur Univ
Penn State Dickinson Law, Carlisle, Pennsylvania.

A federal appeals court recently rejected a lawsuit that claimed that a hospital violated patients' privacy by sharing electronic health record (EHR) data with Google for medical artificial intelligence (AI) development.[1] This lawsuit provides crucial insight into legal issues hospitals may face if they share EHR data with for-profit companies, even if these data do not explicitly identify individual patients.

EHR Data for AI Development

Google entered a research partnership with the University of Chicago, including its medical center (UC), to develop an AI model that could predict significant medical events and reduce hospital readmissions. UC shared with Google "de-identified" EHR data from adult patients encountered between January 2010 and June 2016.[1] These data still contained "dates of service" and "de-identified, free-text medical notes."[2] The data use agreement (DUA) prohibited Google from reidentifying patients.[1] The DUA also granted UC "a nonexclusive, perpetual license to use the (...) Trained Models and Predictions" developed by Google "for internal non-commercial research purposes."[3]

The lead plaintiff in this case ("MD" herein) was admitted to UC twice during the relevant period; thus, MD's "de-identified" EHR data were shared with Google for the research.[1] On admission, MD received a Notice of Privacy Practices and signed an agreement stating that (1) MD's medical information might be used for research, (2) MD would "not be entitled to any compensation" from such research, (3) UC would make all efforts to preserve MD's privacy, and (4) that UC would comply "with federal and state laws."[1,3]

This lawsuit provides crucial insight into legal issues hospitals may face if they share EHR data with for-profit companies, even if these data do not explicitly identify individual patients.

Patient Initiates Lawsuit in Federal District Court

In 2019, MD filed a class action lawsuit against UC and Google in Illinois federal court.[4] Two of MD's claims are especially notable for hospitals that share EHR data for research purposes.

First, MD claimed that UC breached its contract, wherein it promised to preserve MD's privacy and comply with federal law ("breach of contract claim"). MD asserted that the data UC shared with Google were not properly "de-identified" under the Health Insurance Portability and Accountability Act (HIPAA) since it still contained dates of service and free-text notes.[4]

Second, MD claimed that UC violated patients' privacy because the data could easily be reidentified and UC did not obtain patients' express consent to share these data with Google ("privacy claim").[4] MD argued that Google could combine its geolocation information from Google Maps and Waze with the data received from UC to potentially reidentify MD.[4]

Federal District Court Dismisses the Lawsuit

The district court dismissed MD's lawsuit. It dismissed the breach of contract claim for failure to state a legally cognizable claim—MD could not demonstrate all the elements of a claim recognized by the applicable law.[3] The court first considered that UC's sharing might fall under 2 HIPAA "safe harbor" provisions, which MD did not address in his lawsuit.[3] The first safe harbor permits the use or disclosure of a so-called limited dataset stripped of specific direct identifiers to be shared for research purposes if a DUA with certain conditions is in place (CFR §164.514[e]). A limited dataset can contain elements of dates like the dates of service included in the EHR data shared with Google. Notably, a limited dataset is still considered protected health information (PHI) under HIPAA—ie, generally "individually identifiable health information" (45 CFR §160.103). The second safe harbor allows the use and disclosure of PHI for research with approval from an institutional review board (45 CFR §164.512[i]).

Ultimately, however, the district court found that sharing the EHR data, which included the dates of service, was not properly deidentified (45 CFR §164.514[b]) and violated the HIPAA prohibition of *selling* PHI without written authorization (45 CFR §164.502[a][5][ii] and §164.508[a][4]).[3] The court determined that granting UC a nonexclusive, perpetual license to use Google's Trained Models and Predictions in exchange for EHR data was a "sale" under HIPAA that required prior written authorization by MD, which MD never gave.[3]

While patients do not have a legal right to sue for HIPAA violations under HIPAA itself, in this case, the court found that UC's promises in its agreement with MD went beyond its preexisting duties under HIPAA.[3] However, it still dismissed the breach of contract claim because MD did not sustain economic damages as required by state law.[3]

The district court dismissed the privacy claim for lack of standing—MD did not have a legal right to bring a claim before the court—concluding that a claim for breach of medical confidentiality is unlikely to be recognized in Illinois.[3]

Corresponding Author: Sara Gerke, Dipl-Jur Univ, Penn State Dickinson Law, 150 S College St, 234 Lewis Katz Hall, Carlisle, PA 17013 (sgerke@psu.edu).

NOTE JT opinion, UFHealth and GAtorTron may escape this type of scrutiny because it is "internal" and not business contract

The Federal Appeals Court Reviews and Dismisses the Lawsuit

On July 11, 2023, the Seventh Circuit Court of Appeals confirmed the dismissal of MD's case but for reasons different than those given by the district court.[1] The court of appeals dismissed both the breach of contract claim and the privacy claim for lack of standing because MD's alleged harm was neither concrete (not abstract) nor imminent (not speculative).[1]

Regarding the breach of contract claim, the court of appeals found that even if UC breached its contract with MD, MD did not sustain a concrete and imminent harm because MD does not have a monetary property interest in his medical information, and because even if UC benefited from disclosing EHR data, MD did not suffer any loss.[1]

Regarding the privacy claim, while the district court had been more or less silent on the issue of reidentification, the court of appeals noted that Google agreed in the DUA that it would not, and in fact did not, reidentify patients.[1] It also noted that the risk of reidentification in the future (using the dates of service and Google's geolocation information) was too speculative to give MD standing.[1]

Discussion and Key Takeaways for Hospitals and Patients

The court of appeals resolved the case on the procedural issue of standing and did not review the district court's findings regarding whether the EHR data constituted PHI under HIPAA, whether UC's sharing of such data with Google violated HIPAA, or whether a HIPAA violation could support a breach of contract claim under state law. The district court's findings on these issues may nevertheless serve as judicial dicta—an indication of how the court might apply relevant substantive law in future cases similar to MD's case. Additionally, because there is not much precedent addressing issues surrounding patient privacy and EHR data sharing for AI development, other courts may also look to the district court's interpretation of both HIPAA and patient-hospital contract language for guidance in similar cases.

There remains no question that hospitals and physicians participating in medical AI research and development are still responsible for complying with biomedical, ethical, and legal rules, including HIPAA. There is also no doubt that courts will continue to encounter cases involving sharing health data with technology companies like Google. Many will likely face a fate similar to MD's case because patients may not have a property interest in personal information in most states,[5] and because it will be difficult or impossible for them to show "harm" sufficient to take legal action. Others may find some support in state law that recognizes claims specifically related to medical confidentiality.

Hospitals that plan to share EHR data with private parties for AI development can take away the following lessons from the federal courts' decisions in MD's case. First, disclosure and privacy practices should accurately reflect any data-sharing activities that involve patients' EHR data. Second, EHR data should be sufficiently deidentified according to HIPAA before sharing them with third parties for research purposes—ie, either through an expert determination that the reidentification risk is "very small" or by removing 18 specific identifiers (45 CFR §164.514[b]). They should consider using an independent committee that includes experts in ethics, statistics, computer science, and patients to assess certain uses and disclosures of deidentified datasets, including reidentification risks.[6] Third, if hospitals decide to share PHI, including limited datasets, they should get prior written authorization from patients, institutional review board approval, and/or sign a DUA that adequately protects the data and prohibits reidentification. They should also carefully verify whether they really share—vs actually sell—the PHI under HIPAA.

This is where we come in

Last, hospitals and physicians should be aware that the legal landscape surrounding privacy, particularly regarding sensitive personal information like health data, is in flux as concerns about the rapid growth of new technologies and big data mount. Google's increasingly dominant control and influence over information on the internet only adds to these concerns.[7] Additionally, because Google already obtains massive amounts of personal data, including EHR data stemming from research agreements with large medical centers, its status as "the principal purveyor of online health information" remains unchecked for now, which may have serious implications for patient privacy.[7]

ARTICLE INFORMATION

Published Online: February 19, 2024. doi:10.1001/jama.2024.1110

Conflict of Interest Disclosures: Ms Gerke reported receiving grants from the European Union (101057099), National Institute of Biomedical Imaging and Bioengineering and the National Institutes of Health (NIH) Office of the Director (3R01EB027650-03S1, 1R21EB035474-01), and the National Institute on Drug Abuse/NIH (1U54DA058271-01) outside the submitted work.

Funding/Support: This work was supported by grant 101057321 from the European Union.

Role of the Funder/Sponsor: The European Union had no role in the preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

1. *Dinerstein v Google LLC*, No. 20-3134 (7th Cir 2023).

2. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18. doi:10.1038/s41746-018-0029-1

3. *Dinerstein v Google LLC*, 484 F Supp 3d 561 (ND Ill 2020).

4. Amended class action complaint and demand for jury trial, *Dinerstein v Google LLC*, No. 1:19-cv-04311 (October 8, 2019).

5. McGuire AL, Roberts J, Aas S, Evans BJ. Who owns the data in a medical information commons? *J Law Med Ethics*. 2019;47(1):62-69. doi:10.1177/1073110519840485

6. Cohen IG, Mello MM. Big data, big tech, and protecting patient privacy. *JAMA*. 2019;322(12):1141-1142. doi:10.1001/jama.2019.11365

7. Curfman G. *United States v Google*—implications of the antitrust lawsuit for health information. The Source. May 13, 2021. Accessed February 13, 2024. https://sourceonhealthcare.org/united-states-v-google-implications-of-the-antitrust-lawsuit-for-health-information/