Claiming 'Certainty' of Re-Identification, HIPAA Violations, Patient Sues Google, U. of Chicago

By Theresa Defino

A patient who was twice hospitalized at the University of Chicago Medical Center four years ago has filed a proposed class action suit against UC and Google, alleging UC gave his protected health information (PHI) to Google for use in a research study utilizing electronic health records (EHRs) without adequately de-identifying it. The suit terms UC's sharing of PHI with Google a "massive medical data grab."

UC and Google have called the suit baseless and said they would mount a defense.

The suit, filed June 26 in the U.S. District Court for the Northern District of Illinois, Eastern District, would have to be certified as class action litigation. It is being brought by Matt Dinerstein, who was admitted to UC medical center for a total of six days in June 2015. UC and Google began a research collaboration in May 2017.

UC, the suit claims, was "happy to turn over [to Google] the confidential, highly sensitive and HIPAA-protected records of every patient who walked through its doors between 2009 and 2016."

Both Google and UC "violated HIPAA by sharing and receiving medical records that included sufficient information for Google to re-identify the patients," the suit alleges. "Both were aware at the time of the transfer that the medical records contained information outside of HIPAA's safe harbor provisions, that a competent expert determination was not made, and that the thousands of patients had not given proper consent to allow Google to take possession of the records for the purpose of creating a commercial product."

Little information is available about how data were scrubbed of PHI. Just one paper appears to have resulted from the UC-Google project so far. Titled "Scalable and accurate deep learning with electronic health records," the paper was published last year in the Nature Partner Journals: *Digital Medicine.*

Using "de-identified EHR data from two U.S. academic medical centers with 216,221 adult patients hospitalized for at least 24 hours," authors from UC, Google, Stanford Medicine and the University of California San Francisco (UCSF) wrote that they "demonstrate[d] that deep learning methods using this representation are capable of accurately predicting multiple medical events from multiple centers without site-specific data harmonization."

According to the paper, "All electronic health records were de-identified, except that dates of service were maintained in the UCM [University of Chicago Medicine] dataset. Both datasets contained patient demographics, provider orders, diagnoses, procedures, medications, laboratory values, vital signs, and flowsheet data, which represents all other structured data elements (e.g., nursing flowsheets), from all inpatient and outpatient encounters. The UCM dataset additionally contained de-identified, free-text medical notes. Each dataset was kept in an encrypted, access-controlled, and audited sandbox. Ethics review and institutional review boards approved the study with waiver of informed consent or exemption at each institution."

Purchase Login