STAT E-BOOK

The health data revolution: promise and pitfalls

SPONSORED BY 

STAT

TABLE OF CONTENTS

In medical parlance, "stat" means important and urgent, and that's what we're all about — quickly and smartly delivering good stories. We take you inside science labs and hospitals, biotech boardrooms, and political backrooms. We dissect crucial discoveries. We examine controversies and puncture hype. We hold individuals and institutions accountable. We introduce you to the power brokers and personalities who are driving a revolution in human health. These are the stories that matter to us all.

BOSTON • WASHINGTON • NEW YORK

SAN FRANCISCO • LOS ANGELES • CLEVELAND

Our team includes talented writers, editors, and producers capable of the kind of explanatory journalism that complicated science issues sometimes demand. And even if you don't work in science, have never stepped foot in a hospital, or hated high school biology, we've got something for you. The world of health, science, and medicine is booming and yielding fascinating stories. We explore how they affect us all. And, with our eBook series, we regularly do deep dives into timely topics to get you the inside scoop you need.

Technology has vastly changed how health data is collected and who can access it.

Electronic data systems now give doctors and other health care providers the means to quickly access data about patients, and allow health systems around the country to share medical records with other institutions securely.

This health data revolution has been beneficial for patients, too. Under federal rules that took effect in 2022, they now have unlimited access to their digital health records. They can get at them quickly, move them around electronically, link accounts in different systems, and choose who else can see the records. It's a development STAT national technology correspondent Casey Ross, in an article in this e-book, calls a "jumping-off point for a patient-mediated data economy that lets consumers in health care benefit from the fluidity they've had for decades in banking."

And, as STAT health tech correspondent Katie Palmer reports in her article included here, the access to "bulk" data will improve under new government rules. "Access to population-level patient data is critical for public health monitoring, health system quality measurements, and research and development," Palmer writes, adding that health informaticians suggest that perhaps the biggest impact of the new rules "is likely the ability to hold providers accountable for the quality and cost of their care."

Yet the technology that underpins this sea change is far from perfect. Health systems, though they promise to protect patient confidentiality, have websites that leak sensitive information.

And problems with what health tech experts refer to as "interoperability" — the ability to exchange and use data from disparate health systems — remain.

The question of how to share data widely but protect patient privacy is a problem that has attracted a slew of companies using a variety of methods and technologies to develop new solutions.

In the articles in this e-book, you will discover the promise and possibilities of this new frontier of health data, and the obstacles that still need to be overcome.

"Let's finish this job," physician Steven Lee writes, in the opinion piece that opens this book.


IMO
Intelligent
Medical Objects
Grounded in data.
Guided by insights.
Intelligent Medical Objects is the foundation and future of health IT. Our terminology and insights solutions ensure the integrity of patient data from the point of care to a range of uses across the healthcare ecosystem.
Learn more at imohealth.com.

STAT

FIRST OPINION

Mr. President, health care has a data problem. 'Let's finish the job'

By Steven Lane | **FEB. 8, 2022**

President Biden's 2023 State of the Union address outlined the administration's plan to reduce health care costs for Americans, including lowering health insurance premiums and expanding the $35-a-month cap on insulin costs to anyone who needs it. He boldly declared that cuts to Medicare and Social Security are off the table, and said he would veto any attempt to repeal the Inflation Reduction Act or institute a national ban on abortion.

As a practicing family physician, I am always pleased to hear about ways to reduce the cost of commonly prescribed medications and increase the chances that patients will receive the high-quality care they deserve. High health care costs disproportionately affect lower-income individuals, including many in racial and ethnic groups. The. U.S. is in dire need of more equity in health and health care, and I believe the president's plan is a step toward ensuring that every American has the opportunity to access high-quality, affordable care regardless of where they live, their race, or socioeconomic status.

As a clinical informaticist, however, I must say that the solutions Biden proposed do not address one of the most significant drivers of administrative waste in the U.S. health care system — health care data.

It's no secret that the U.S. health care system is plagued by inefficiencies, with administrative costs consuming a staggering 25% of total health care spending, representing $1 trillion per year. A major contributor to these costs is the deficiencies of health data interoperability between electronic health records — meaning one system can't talk to or exchange data with another — and other health data systems, which results in a fragmented and siloed view of patients' information. The inability to easily access and exchange data between providers, labs, payers, and public health entities not only leads to increased costs, but also harms patients and clinical outcomes by delaying care, and increasing cost and the risk of medical errors.

The solution won't be simple. Health data are spread across thousands of hospitals, clinics, novel care sites, labs, and pharmacies, most of which use different IT systems and often store data in different formats, making it difficult to piece together an individual's information.

I've spent much of my career trying to address this problem in collaboration with officials at the Department of Health and Human Services, the Office of the National Coordinator for Health Information Technology (ONC), and many national and regional health IT organizations. We've made remarkable progress, but there is still much to be done. As the president said repeatedly in his State of the Union address, "let's finish the job."

Today, using ONC-certified health information technology, the vast majority of health systems have the technical capabilities to exchange health information securely and safely with other systems across the country. It's easier than ever for providers to quickly and securely access extensive data on their patients. Progress has also been made implementing

many of the goals of the 21st Century Cures Act, including the Trusted Exchange Framework and Common Agreement, which establishes a universal floor for interoperability nationwide.

On Feb. 13, HHS will recognize the first set of organizations that are approved for onboarding as Qualified Health Information Networks, which will connect to one another and enable their participants to engage in a new era of health information exchange across the country.

However, patient data now flowing in in ever-larger volumes is creating additional obstacles to overcome. In a recent interoperability report, 60% of health system IT executives said the patient data they retrieved through health information exchanges had quality issues, with problematic amounts of duplicative, incomplete, or "junk" data. In addition, 75% of health systems reported concerns around patient privacy and data security as a result of increases in national health information sharing.

Fortunately, there are many for-profit and nonprofit organizations addressing these issues, with new technologies and data-enhancement techniques, innovative frameworks that allow for secure frictionless exchange of information, and consortia to inform federal and state policy.

Every one of the health and social care programs touted by Biden could be made more efficient and effective through more robust, secure, privacy-protecting health data exchange. While the solutions he proposed in the State of the Union address are positive steps toward improving access to health care, the central issue of data fragmentation lurks in the shadows. Without addressing this problem head on, efforts to lower costs and improve access to safe, high-quality care will be limited in their impact. Fixing the country's data quality and integrity issues must be one of the

nation's top health care priorities to rein in administrative costs, improve secure data access, and, most importantly, allow for an equitable health care system.

Steven Lane is a primary care physician, a clinical informaticist, a clinical professor of family and community medicine at the University of California, San Francisco, and the chief medical officer of Health Gorilla, a California-based health information network and data platform.

STAT

HEALTH TECH

‘Apples to apples’: How new health data rules could hold providers accountable

By Katie Palmer | **DEC. 23, 2022**

Last year, medical records opened up to patients. This year, they’re opening up to the nation.

Before the ball drops on New Year’s Eve, electronic health care record vendors will have to provide tools to easily pull big batches of patient data from their systems. Just as information blocking rules gave individual patients the ability to access their medical records, this next round of federal rules gives a framework for sharing insights — within a health system, or with trusted partners — about groups of patients that reflect different populations.

“[This] really for the first time ever allows you the ability to do true comparisons between providers, with what I would call apples to apples data,” said Don Rucker, the former national coordinator for health information technology who spearheaded the tools. By combining standardized outcomes information — say, on blood glucose levels for patients with diabetes — with claims data in the same format, it will be easier than ever to see the bang a patient gets for their buck at different providers. “I believe that will be, over time, utterly transformative.”

While EHR giant Epic implemented its "bulk data" access updates in a release earlier this year, the existence of these tools — which work with standardized data in a format called FHIR — has largely flown under the radar. Once health systems clue into the new tools, though, advocates think they could unlock the ability of digitized medical records to improve health care quality and affordability.

Access to population-level patient data is critical for public health monitoring, health system quality measurements, and research and development. Providers and other users have been able to extract that information with proprietary APIs — but with so many different systems and formats, it can be a serious slog to share and analyze data between institutions. But starting in 2023, thanks to the 21st Century Cures Act, all certified EHRs will have to provide API technology that taps into a minimum dataset in the standardized FHIR format.

"You can go down the list of the health care economy and ask yourself, where would we not see benefit from more timely, standardized, normalized population-level data?" said Aneesh Chopra, president of CareJourney and former chief technology officer under the Obama administration. That could mean assembling groups by a shared medical condition, insurance status, enrollment in a health plan, and more: "All possible through the Cures Act APIs."

Even with the rules in place, health informatics experts said not to expect change overnight. As usual, the impact of bulk data access will come down to financial incentives — and EHR vendors and health care providers alike have often found ways to push against transparent and interoperable health records.

The promise of standardized bulk data access is what developers have called "push button population health." Imagine a Covid-19 pandemic in which providers large and small could automate nightly federal updates for all their Covid-positive patients. Or an app that taps into the records of cancer patients across facilities to recruit matches for clinical trials, "so we can democratize, expand to underserved populations, reach into new parts of the country that don't normally enroll in clinical trials," said Chopra. Machine learning efforts in medicine, too, could benefit from new sources of relatively clean, standardized data for model training.

But the biggest impact of the bulk data requirements, health informaticians suggested, is likely the ability to hold providers accountable for the quality and cost of their care.

"Now, for the first time, you're going to be able to collect claims and clinical data in the same data format," said Rucker, now chief strategy officer for FHIR platform provider 1upHealth. "Which is ultimately what you need if you're going to search for value."

When payers — commercial insurers and federal plans alike — negotiate rates with providers, standardized cost and outcome data could give them more leverage. And while many providers will still have a strong incentive to gatekeep that information, when every major EHR has a standardized API available at the end of this year, "your excuse for not sharing data … becomes way less tenable," said Rucker.

The same kinds of analysis could also drive health systems' internal efforts to improve care. "It lowers the burden of developing decision support and other analytic products that can be delivered and acted upon by the frontline clinical workforce," said Chopra. An early use has been under

development at the Centers for Medicare and Medicaid Services, whose Beneficiary Claims Data API allows accountable care organizations to access Medicare claims data for their beneficiaries — from within and outside of the ACO.

Getting people on board with this form of data sharing, though, could take some salesmanship. Just because bulk FHIR APIs must now be made available to health providers doesn't mean they have to use them (though they do have to start using the updated EHRs by the last quarter of 2023.) However kluged-together they may be, existing bulk data-sharing systems have been painstakingly built and integrated over time, and health systems may not be enthusiastic about abandoning those sunk costs — especially when legacy tools may initially offer more functionality than a health system can get with the new public option.

"There's no question that there's going to be some tension between a public API and proprietary APIs," said Kenneth Mandl, director of the computational health informatics program at Boston Children's Hospital and co-developer of SMART on FHIR. Vendor-built systems may initially be more functional within a single health care system and allow for more uses, as they enable access to more than the "core" interoperability dataset defined by the government. And it's unclear how much EHR companies will charge their customers to use the bulk FHIR tools, and how those costs compare to their proprietary access points.

Those tradeoffs may look even less appealing early on, as new systems run into inevitable technical hurdles. "When these APIs are first turned on there will be data quality issues to address," said Mandl. Just because patient data can be easily extracted in the FHIR format doesn't mean that providers collect it consistently, and in a way that makes transfers most

accurate and useful. There's also a possibility that data requests could overload EHR systems, slowing their performance for patient care.

For bulk FHIR to have its greatest impact, organizations at every level of the health care system will have to pay what Chopra calls a "technology implementation tax"— from federal agencies and huge hospital systems to state public health departments and low-resourced nonprofits. And each of those data-sharing partners will have to spend significant time making sure the records they're sharing don't jeopardize patient privacy or a health system's strategic advantages.

But if standardized bulk data sharing is widely adopted, it has the potential to create a virtuous feedback loop for interoperability efforts. Typically, said Mandl, EHRs don't collect particularly high-fidelity data — except in the case of billing, where errors and omissions cost the provider. If payers like CMS lean into standardized APIs to calculate quality measures, providers will have just as good a reason to pay attention to the reliability of their clinical data — whether that means leaning into best practices for collection of race and ethnicity or more complete immunization records.

"Ultimately, when data are being used, people care much more about those data. And so that has a strong potential to reverberate back in the chain," said Mandl. "We might care more about how those data are collected in the first place."

STAT

HEALTH TECH

Call it data liberation day: Patients can now access all their health records digitally

By Casey Ross | **OCT. 6, 2022**

The American Revolution had July 4. The allies had D-Day. And now U.S. patients, held down for decades by information hoarders, can rally around a new turning point, October 6, 2022 — the day they got their health data back.

Under federal rules taking effect Thursday, health care organizations must give patients unfettered access to their full health records in digital format. No more long delays. No more fax machines. No more exorbitant charges for printed pages.

Just the data, please — now.

"My great hope is that this will turn the tide on the culture of information blocking," said Lisa Bari, CEO of Civitas Networks for Health, a nonprofit that supports medical data sharing. "It's a ground level thing to me: We need to make sure information flows the way patients want it to."

That's the opposite of the situation now in place. Health systems, data networks, and the companies that sell electronic medical records determine

how much data patients can access, when, and under what circumstances. Meanwhile, private data brokers make huge profits by amassing hundreds of millions of de-identified medical records and selling insights to drug companies, device makers, and insurers without patients' knowledge or consent.

The new federal rules — passed under the 21st Century Cures Act — are designed to shift the balance of power to ensure that patients can not only get their data, but also choose who else to share it with. It is the jumping-off point for a patient-mediated data economy that lets consumers in health care benefit from the fluidity they've had for decades in banking: they can move their information easily and electronically, and link their accounts to new services and software applications.

"To think that we actually have greater transparency about our personal finances than about our own health is quite an indictment," said Isaac Kohane, a professor of biomedical informatics at Harvard Medical School. "This will go some distance toward reversing that."

Even with the rules now in place, health data experts said change will not be fast or easy. Providers and other data holders — who have dug in their heels at every step — can still withhold information under certain exceptions. And many questions remain about protocols for sharing digital records, how to verify access rights, and even what it means to give patients all their data. Does that extend to every measurement in the ICU? Every log entry? Every email? And how will it all get standardized?

For months, patients have been able to obtain a minimum data set specified under federal law, and applications such as Apple Health Records have already dramatically expanded access. But the new rules throw open

the floodgates to a much wider swath of information, including medical images, doctors' notes, genetic data and other details normally kept under lock and key.

"It's really simple — I have access to all my data, and people need to make that available to me digitally at my request," said Harlan Krumholz, a cardiologist at Yale University and founder of Hugo Health, a company that helps patients collect and organize their health data.

He said it will take time for providers and other data holders to fully comply, especially since enforcement remains spotty and unclear under the new rules. But patients' ability to get their data means they can better understand their care, shop for services, and participate in research without waiting for a clinician, or drug company, to present them with an opportunity.

"I hope it will become clear that we need to switch from a paternalistic system where a lot of data is moving behind peoples' backs and without their permission or knowledge, to one where people have more control and agency over their data," Krumholz said. Now, Krumholz said, patients can have their own personal repository of data that they can build on and ferry from one health care setting to another.

A growing number of data companies are popping up to help them in that quest, seeking to act as fiduciaries for consumers who want access to their records, but don't have the time or technical savvy to wade through the bureaucracy.

The new environment is a radical departure from the status quo. For decades, it's been all but impossible for patients to quickly and easily access

their records. Hospitals and other organizations are loath to relinquish that information for a couple of business reasons. It makes it easier to retain patients in their care, and it keeps them in control of information with high commercial and research value.

The federal law known as HIPAA requires that providers turn over records when patients ask for them. But such requests are often met with delays, fees, and sometimes requests that they fetch them via fax. "HIPAA's been in place for a long time," Bari said. "But it's simply not respected and used in that way."

The roadblocks made it harder for patients like Liz Salmi, who has brain cancer, to get the care they need. For the first eight years of her cancer treatment, she sought care at Kaiser Permanente in California. But a change in insurance coverage meant she eventually had to switch to new providers.

"I naively thought that because they were all on Epic, they could easily share my records," Salmi said. But she found out that their computer systems didn't talk to each other, so one hand didn't know what the other was doing. She ended up going to the medical records office at Kaiser in person to request copies of her records.

"They said, 'OK, what parts of your record do you want? Lab results? Visit summaries?'" Salmi recalled. "I said, 'No, I want everything.'"

She was told her full record — comprising eight years of care at Kaiser — was 4,823 pages. If she wanted it printed, she would have to pay 15 cents per page, for a total of $723.45. Salmi said she opted to pay $45 for three DVDs instead. It was 2017: well into the era of streaming services,

smartphones, and same-day delivery. But to read her medical records, Salmi had to buy an external hard drive to load the disks into her computer.

When she finally opened them, it was like experiencing health care for the first time.

"I had no idea there was a whole other narrative going on behind the scenes," Salmi said. "I could see all the emails back and forth with my doctors. I could see my progress notes. There was so much I had forgotten as a patient and here it was in black and white, because somebody had taken the time to write it down. I was blown away."

Salmi, who is now undergoing treatment for a recurrence of her cancer, joined OpenNotes, an organization that promotes data sharing, where she is director of communications and patient initiatives. She said the effective date for the new data rules marks an important milestone. But to have a real impact it must be accompanied by an education campaign to help patients understand their rights and the benefits of getting their data.

Too many patients, she said, are unaware of the volume of information recorded about them or its value in an environment with new opportunities to participate in clinical studies and digital health services, without leaving their communities, or even their homes. Many also may be unaware of the security risks and how to separate reputable data users from swindlers.

"You need to see and experience your health records to even know what you'd be sharing," she said. "But for years that decision was made for us — 'you can see this, but you can't see that.' When I peeled back the curtain and actually could see what was in there I said, 'Oh my gosh, this is incredible. I want to keep reading. I wish I knew this sooner.'"

STAT

STAT is tracking the effects of a new federal law that requires health care organizations to give patients access to their full health records in digital format. Are you willing to discuss your experience requesting your health records? We will not share your name or story without your permission.

SPONSOR CONTENT

Labs, meds, and data quality: Taming complexity through normalization

By Amy Loriaux, P.h. D., health industry and standards writer

The advent of the electronic health record (EHR) created exciting possibilities for using patient-level medication and laboratory data in new ways. With EHRs, such information could now not only improve patient care, but also help drive clinical research, inform public health policy, and ultimately progress the field of healthcare.

Yet realizing this potential remains difficult. In this white paper, we outline some of the current challenges related to capturing, sharing, and using medication and laboratory data, along with how terminology-driven data normalization can help unlock its potential for downstream use

Labs and meds data — not only for providers

While it's essential that providers are able to access information about a patient's lab results or current medications, this data is also critical to support a range of uses beyond the point of care, including life sciences and public health.

Life sciences

One of the most important uses of medication and lab data in the life sciences is applied research — specifically for clinical trials. Patient recruitment for these trials is often a long and arduous process, particularly for research about rare diseases. However, information about labs and meds can help identify potential candidates more easily.

But while many registries house patient data that can be used to find individuals with specific conditions, this information is not always complete. Critically, data about labs and meds is often sparse, inconsistent, or absent. And even when it is present, it may be improperly coded or poorly mapped. Indeed, in a recent study fewer than 15% of hospital providers used standardized code sets like LOINC® to code their patient lab data.[1]

Public Health

Public health agencies (PHAs) use patient data to track the spread of communicable diseases, monitor patient safety, and allocate community resources to vulnerable populations. But this data lives in a variety of places — EHRs, prescription drug monitoring programs (PDMPs), and laboratory information systems (LIS'), to name a few.

For PHA initiatives aimed at increasing patient safety around substances, accurate information from PDMPs — state-level databases that monitor the dispensing of controlled substances like opioids — is crucial. But PDMPs can't achieve their goals of deterring opioid over-prescribing, identifying drug-seeking behavior, and informing clinical decision-making without the exchange of medication data. Accurate and complete information about the medications that a patient has previously been prescribed — including whether it has been dispensed — is critical for

these initiatives to successfully prevent medication abuse and overdoses.

Data needs beyond the point of care

Consider the following situations: A scientist wants to know if menopause impacts the efficacy of adjuvant hormone treatment in women with estrogen-positive (ER+), early-stage breast cancer. Separately, a PHA is trying to track the spread of a sudden outbreak, like COVID-19.

In the first scenario, finding the right patients to study or track relies on the detail available within each patient record. The scientist needs information on age, menopausal status, ER+ status, and disease stage in order to successfully recruit for her trial. And, if she's testing a novel treatment for ER+ breast cancer patients, she will also need to rule out anyone on medications or undergoing treatments that may interfere with the investigational drug.

As for the epidemiologists working at the PHA, granular data about labs and meds is key to tracking an outbreak. Without metrics such as the type of test used to diagnose a case, the types of medications used to treat it, and the related outcomes data, they cannot recommend appropriate public health interventions. And without that detailed data, tracking efforts and the resulting insights would be inaccurate and potentially harmful.

The role of standardized clinical terminologies

Maintaining semantic interoperability — or ensuring a clinical term and standardized code preserve the meaning of patient information when it is transferred between systems — is challenging when it comes to highly detailed concepts like labs and meds. That's because a standardized code

has to convey more than just a test's name. It must also communicate many more specific pieces of information — such as dosage or unit of measurement — in a way that makes sense and can be made machine computable.

To this end, two systems — LOINC and RxNORM® — are freely available and preferred by the federal government for preserving semantic interoperability with labs and meds. However, they are still less widely used by providers compared to systems like CPT®, ICD-10-CM and SNOMED CT®.

LOINC

LOINC codes represent laboratory orders and test results. Unlike ICD-10-CM or SNOMED CT, which represent problems or diagnoses with one code, LOINC represents test results or observations across six dimensions or parts,[2] as follows:

1. **Component –** The substance or entity being measured or observed
2. **Property –** The characteristic or attribute of the analyte
3. **Time –** The interval of time over which an observation was made
4. **System (specimen) –** The specimen or thing upon which the observation was made
5. **Scale –** How the observation value is quantified or expressed: quantitative, ordinal, nominal
6. **Method –** A high-level classification of how the observation was made

RxNORM

Although National Drug Codes (NDCs) are assigned by the Food and Drug Administration (FDA) for all drugs manufactured, prepared, propagated, compounded, or processed for sale in the US,[3] these codes are used primarily for prescription billing. But since many NDC codes can exist for one product, they are problematic for research or epidemiology use cases.

Therefore, RxNORM is the recommended nomenclature for medications by the federal government[4] and is used in many payment programs run by CMS. The system contains both the branded and generic names for drugs available by prescription and over the counter.

RxNORM works by taking drug concepts from other sources or drug terminologies and groups them under a single, normalized name, or code. Therefore, RxNORM functions as a database of vocabularies whose goal is to link synonymous drug terms from a variety of sources together under one umbrella concept or term.

Problems inherent in the coding systems

RxNORM and LOINC were designed to be universally adopted standards with an eye towards increasing interoperability. Either via their overall structure or comprehensiveness, these terminologies are meant to capture and convey highly detailed information that can be useful for both providers and secondary users. However, there are still problems inherent with these systems — including their lack of universal adoptability, the practice of reusing old codes for new terms, and frequent changes and updates.

What's more, these standards are not as widely adopted as other code systems. In fact, much of the medication and lab data in the EHR goes uncoded. And when it is coded, providers often use CPT codes to document lab data, and NDC or other proprietary formular service vendor codes for medication documentation.

This is ultimately the greatest roadblock to effective use of lab and medication data for secondary applications. When most of this data goes uncoded, it becomes incredibly difficult to extract from the EHR, aggregate, and use for meaningful secondary purposes.

Normalization of labs and meds data

What's needed to assist secondary users of labs and meds data are tools that can be employed to preserve its clinical meaning when it is transferred between systems. This typically means that data coded with terminologies such as ICD-10-CM or NDC need to be cross-mapped to more granular code systems line LOINC and RxNORM.

However, it is important to note that labs and meds data is particularly complex and must include factors like dose, formulation, analyte, or

units of measurement. Communicating all of this information requires a standardized terminology that is large enough to encapsulate all the possible combinations of terms and values used to represent a medication or lab result. Additionally, since the mappings to code sets like LOINC and RxNORM change frequently, there is additional need for continual maintenance to keep mappings up to date.

Fortunately, terminology vendors with a log track record of providing services are uniquely positioned to provide these solutions for medication and lab data.

By leveraging a highly granular, regularly maintained, foundational terminology, normalization engines can preserve the precise meaning of lab and medication data through comprehensive code maps. By linking meds and labs to standardized code sets like LOINC and RxNORM, this data can be effectively aggregated with very little information loss.

These tools serve not only point-of-care providers, but also secondary use cases like research and public health. With these solutions available, researchers, PHAs, and other can more fully realize the value in real-word data from the EHR to make novel insights, ensure public safety, and ultimately contribute to higher-quality healthcare for all.

To learn how normalization solution grounded in clinical terminology can help solve the struggle with labs and meds data, visit imohealth.com/imo-precision-normalize.

Works Cited:

1. Nelson, H. Health IT Vendors Partner on EHR-Based Clinical Research Network. EHR Intelligence. Accessed via: https://ehrintelligence.com/news/health-it-vendors-partner-on-ehr-based-clinical-research-network

2. No author given. Major parts of a LOINC term. The Regenstrief Institute. Accessed via: https://loinc.org/kb/users-guide/major-parts-of-a-loinc-term/

3. No author given. National Drug Code (NDC). Codonics. Accessed via: https://www.codonics.com/products/our-solutions/national-drug-code-ndc/

4. No author given. Advancing Interoperability and Improving Prior Authorization Processes Proposed Rule CMS-0057-P: Fact Sheet. Centers for Medicare & Medicaid Services. Accessed via: https://www.cms.gov/newsroom/fact-sheets/advancing-interoperability-and-improving-prior-authorization-processes-proposed-rule-cms-0057-p-fact.

STAT

FIRST OPINION

Unlocking the promise of learning from everyone with cancer

By Jay J. Schnitzer | **FEB. 1, 2023**

Locked behind the firewalls of proprietary systems sits a treasure trove of data that could help diagnose heart disease, diabetes, cancer, and other conditions faster and more accurately and better treat people with them. But there it sits, largely untapped, because the electronic health record infrastructure was never designed to let organizations easily share data.

Electronic health records were first developed in the 1960s but didn't become mainstream until about 12 years ago when the federal government provided incentives for their use. At the time, expectations were high that they would be the solution for seamlessly and securely collecting and sharing valuable patient data. EHRs would reduce the need for faxing records from one doctor's office to another and end the practice of manually inputting the same information into multiple databases.

The country isn't there yet. The rush to develop electronic health records produced proprietary and competing data systems that are customized for each health provider organization, like hospitals and medical offices. In addition, it's taken years to develop software standards to enable the sharing of data across systems. Thanks to organizations such as Health

Level Seven International and the Office of the National Coordinator for Health Information Technology, some progress has been made.

The chaos of Covid emphasized the lack of standards and what electronic health records can't easily do, like quickly make shareable patient data available so doctors can evaluate which treatments worked for which patients.

At the height of the pandemic, as physicians at the Mayo Clinic were battling to keep patients alive, the medical staff frequently had to hit pause to fill out lengthy REDCap surveys to inform Minnesota state health officials about the number of patients they were seeing with Covid-19. The state-mandated surveys lived outside Mayo's regular EHR system, and so required painstaking manual work to record Covid case information on giant Excel spreadsheets. "When we were going through our surges, we were drowning and burdened" by all the paperwork, Priya Sampathkumar, an infectious disease and critical care specialist at Mayo, told my team at MITRE, the nonprofit research and development organization I work for. System A couldn't talk to System B. "Filling out these pieces of paper just added insult to injury."

Given the ways electronic health records are currently structured, it's difficult, if not impossible, to share and analyze high-quality data from millions of patients in disparate health systems to drive research, improve current treatments, and inform meaningful discussions between patients and their providers.

For example, most of the data that lead to novel cancer treatments today come from clinical trials. That's a problem, because these trials involve less than 6% of adult Americans living with cancer. The percentage for

children is even smaller. That means there is limited information about what treatments work for which patients. Clinicians and researchers don't have ready access to data from the vast majority of cancer patients, data that could potentially identify what treatments have worked well for, say, a 52-year-old Hispanic woman who has diabetes as well as breast cancer or a 70-year-old man with Stage 3 lung cancer.

The quest to expand standards and interoperability across electronic health record systems to improve the quality, safety, and effectiveness of patient care is, fortunately, not starting from scratch. The Fast Healthcare Interoperability Resources (FHIR) standard makes it easier to share data by defining how health information can be exchanged among different computer networks, regardless of how it is stored in those systems.

In 2019, MITRE and several other nonprofits launched an effort to develop a common standard and language for cancer care that could be incorporated into EHRs and used to capture the characteristics, treatments, and outcomes of every person with cancer. We built our standard on existing best practices, such as HL7's experience developing the Fast Healthcare Interoperability Resources standard.

The result, mCODE (short for minimal Common Oncology Data Elements), is now being tested by more than 60 health organizations and other stakeholders — including EHR vendors — who see the potential of learning from millions of patients' experiences. We chose cancer to test the hypothesis that only a minimal amount of critical information is necessary to produce valuable results comparable to those found in clinical trial reports.

We also learned the necessity of involving the community to build

consensus around new standards and drive them forward. mCODE was developed by a multidisciplinary group of subject matter experts, including cancer clinicians, informaticists, health services researchers, experts in data standards and interoperability, people living with cancer, and others under the auspices of MITRE and the American Society of Clinical Oncology.

Rather than focusing on standards for exchanging data, mCODE aims to standardize health records so diverse stakeholders can share information in a meaningful way to achieve large-scale outcomes, such as more efficient research, faster trial matching, and more personalized medicine. Using mCODE's common data language and open-source, nonproprietary model, organizations can access and analyze data from various EHR systems, including essential data that can be hard to find today, such as patients' cancer stages or the outcome of specific treatments.

In its first pilot project, the mCODE team collaborated with a clinical trial group that is testing a new use of an existing drug to treat breast cancer. Initial results on preliminary data, not yet published, indicate that the accuracy of mCODE's results match those of the clinical trial team's 95% of the time. Since then, mCODE has been incorporated into several other clinical trials, and is being tested for other uses, such as cancer registries and prior authorization of treatments. The team is also freely sharing its expertise and open-source technology with organizations involved in cardiac diseases, genomics, and dementia who want to use the mCODE approach to develop and test standards for their specialties.

With a standards-based approach like mCODE, every doctor would have valuable information about a patient's disease and possible treatments at their fingertips at the point of care. The insights have the potential to improve patient care and shared decision-making, drive innovation, and set

the foundation for a national cancer health learning system.

The possibilities to improve patient care and research through sharable data are endless. But it will take the whole community to make it happen: electronic health record vendors, health systems, payers, researchers, and patients. It also will take a change of incentives. Many of today's players, such as EHR vendors and health systems, have made their patient data proprietary, believing it gives them a competitive advantage. That's not the case anymore, since no one organization can ever have enough data on its own to solve big problems.

As doctors, scientists, and patients desperate for effective treatments see the obvious benefits of gaining access to data at scale, they will drive this approach forward. Unlocking the potential of these proprietary systems could not be more important.

Jay J. Schnitzer is a pediatric surgeon and the senior vice president, chief medical officer, and chief technology officer at MITRE.

STAT

STAT INVESTIGATION

‘Out of control’: Dozens of telehealth startups sent sensitive health information to big tech companies

By Katie Palmer – STAT and Todd Feathers and Simon Fondrie-Tieitler – The Markup | **DEC. 13, 2022**

Open the website of Workit Health, and the path to treatment starts with a simple intake form: Are you in danger of harming yourself or others? If not, what’s your current opioid and alcohol use? How much methadone do you use?

Within minutes, patients looking for online treatment for opioid use and other addictions can complete the assessment and book a video visit with a provider licensed to prescribe suboxone and other drugs.

But what patients probably don’t know is that Workit was sending their delicate, even intimate, answers about drug use and self-harm to Facebook.

A joint investigation by STAT and The Markup of 50 direct-to-consumer telehealth companies like Workit found that quick, online access to medications often comes with a hidden cost for patients: Virtual care

websites were leaking sensitive medical information they collect to the world's largest advertising platforms.

Tech companies received people's sensitive information from telehealth sites

Of the 50 telehealth websites analyzed, tech companies were found to send:

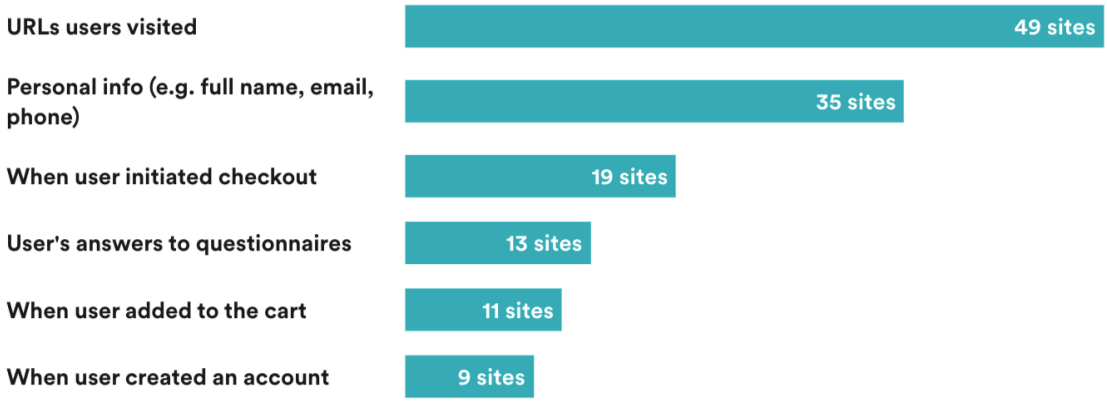


Chart: Joel Eastwood/The Markup • Source: STAT and The Markup analysis

On 13 of the 50 websites, STAT and The Markup documented at least one tracker — from Meta, Google, TikTok, Bing, Snap, Twitter, LinkedIn, or Pinterest — that collected patients' answers to medical intake questions. Trackers on 25 sites, including those run by industry leaders Hims & Hers, Ro, and Thirty Madison, told at least one big tech platform that the user had added an item like a prescription medication to their cart, or checked out with a subscription for a treatment plan.

The trackers that STAT and The Markup were able to detect, and what information they sent, is a floor, not a ceiling. Companies choose where to install trackers on their websites and how to configure them. Different pages of a company's website can have different trackers, and this analysis did not test every page on each company's site.

All but one website examined sent URLs users visited on the site and their IP addresses — akin to a mailing address for a computer, which can be used to link information to a specific patient or household — to at least one tech company. The only telehealth platform that the analysis did not find sharing data with outside tech giants was Amazon Clinic, a platform recently launched by Amazon.

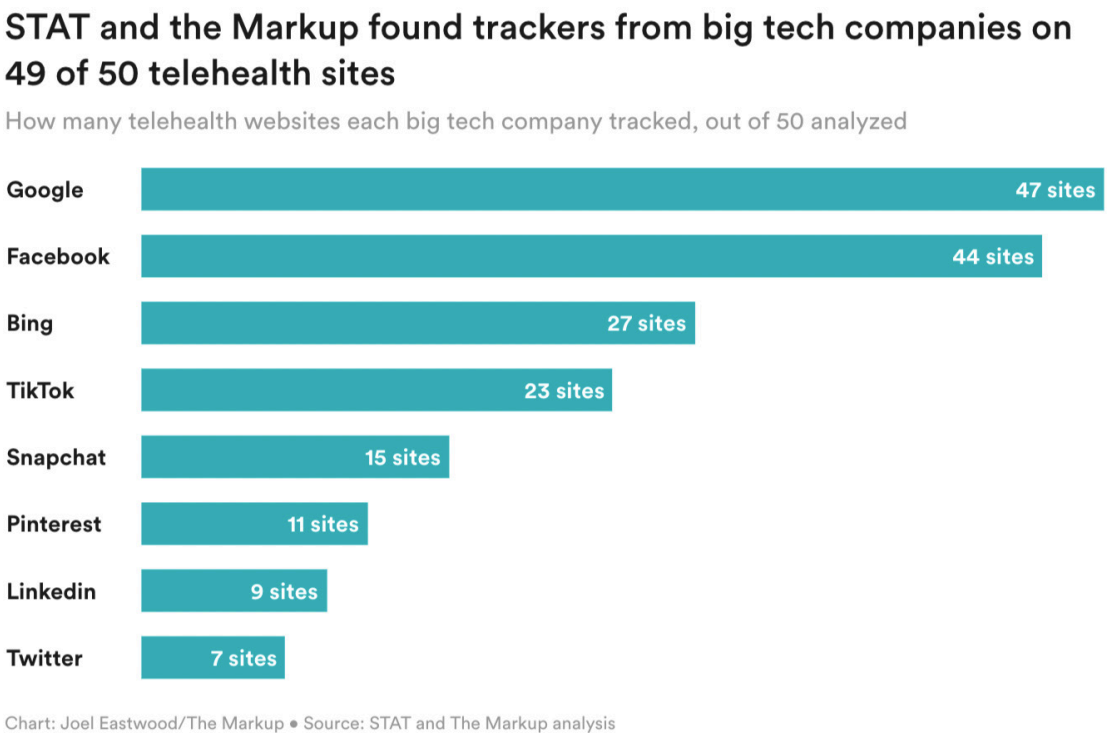
Health privacy experts and former regulators said sharing such sensitive medical information with the world's largest advertising platforms threatens patient privacy and trust and could run afoul of unfair business practices laws. They also emphasized that privacy regulations like the Health Insurance Portability and Accountability Act (HIPAA) were not built for telehealth. That leaves "ethical and moral gray areas" that allow for the legal sharing of health-related data, said Andrew Mahler, a former investigator at the U.S. Department of Health and Human Services' Office for Civil Rights.

"I thought I was at this point hard to shock," said Ari Friedman, an emergency medicine physician at the University of Pennsylvania who researches digital health privacy. "And I find this particularly shocking."

In October and November, STAT and The Markup signed up for accounts and completed onboarding forms on 50 telehealth sites using a fictional identity with dummy email and social media accounts. To determine what data was being shared by the telehealth sites as users completed their forms, reporters examined the network traffic between trackers using Chrome DevTools, a tool built into Google's Chrome browser.

On Workit's site, for example, STAT and The Markup found that a piece of code Meta calls a pixel sent responses about self-harm, drug and alcohol

use, and personal information — including first name, email address, and phone number — to Facebook.



The investigation found trackers collecting information on websites that sell everything from addiction treatments and antidepressants to pills for weight loss and migraines. Despite efforts to trace the data using the tech companies' own transparency tools, STAT and The Markup couldn't independently confirm how or whether Meta and the other tech companies used the data they collected.

After STAT and The Markup shared detailed findings with all 50 companies, Workit said it had changed its use of trackers. When reporters tested the website again on Dec. 7, they found no evidence of tech platform trackers during the company's intake or checkout process.

"Workit Health takes the privacy of our members seriously," Kali Lux, a spokesperson for the company, wrote in an email. "Out of an abundance of caution, we elected to adjust the usage of a number of pixels for now as we continue to evaluate the issue."

"Advertisers should not send sensitive information about people through our Business Tools," Dale Hogan, a spokesperson for Meta, wrote in an email.

Patients may assume that health-related data is always protected by privacy regulations including HIPAA. Workit, for one, begins its intake form with a promise that "all of the information you share is kept private and is protected by our HIPAA-compliant software."

"The very reason why people pursue some of these services online is that they're seeking privacy," said David Grande, a digital health privacy researcher at the University of Pennsylvania.

But the reality online is more complex, making it all but impossible for the average user to know whether the company they're entrusting with their data is obligated to protect it. "Individually, we have a sense that this information should be protected," said Mahler, who is now vice president of privacy and compliance at CynergisTek, a health care risk auditing company. "But then from a legal and a regulatory perspective, you have organizations saying … technically, we don't have to."

Rather than providing care themselves, telehealth companies often act as middlemen connecting patients to affiliated providers covered by HIPAA. As a result, information collected during a telehealth company's intake

may not be protected by HIPAA, while the same information given to the provider would be.

"All the privacy risks are there, with the mistaken but entirely reasonable illusion of security," said Matthew McCoy, a medical ethics and health policy researcher at the University of Pennsylvania. "That's a really dangerous combination of things to force the average consumer to deal with."

In response to questions for this story, representatives of Meta, Google, TikTok, Bing, Snap, and Pinterest said advertisers are responsible for ensuring they aren't sending sensitive information via the tools. Twitter did not respond to requests for comment.

"Doing so is against our policies and we educate advertisers on properly setting up Business tools to prevent this from occurring," wrote Meta's Hogan. "Our system is designed to filter out potentially sensitive data it is able to detect."

LinkedIn's tracker "collects URL information which we immediately encrypt when it reaches our servers, delete within 7 days and do not add to a profile," Leonna Spilman, a spokesperson for the company, wrote in an email.

Nevertheless, three of the seven big tech companies also said they had taken action to investigate or stop the data sharing.

Google is "currently investigating the accounts" in question, spokesperson Elijah Lawal wrote in an email.

"In response to this new information, we have paused data collection from these advertisers' sites while we investigate," Snap spokesperson Peter Boogaard wrote in an email.

Pinterest "offboarded the companies in question," spokesperson Crystal Espinosa wrote in an email.

A boom industry on the edge of the law

Together, the companies in this analysis reflect an increasingly competitive — and lucrative — direct-to-consumer health care market. The promise of a streamlined, private prescription process has helped telehealth startups raise billions as they seek to capitalize on a pandemic-driven boom in virtual care.

Hims & Hers, one of the largest players in the space, is now a publicly traded company valued at more than $1 billion; competitor Ro has raised $1 billion since its founding in 2017, with investors valuing the company at $7 billion. Thirty Madison, which operates several telehealth companies focused on different medical needs, is valued at more than $1 billion.

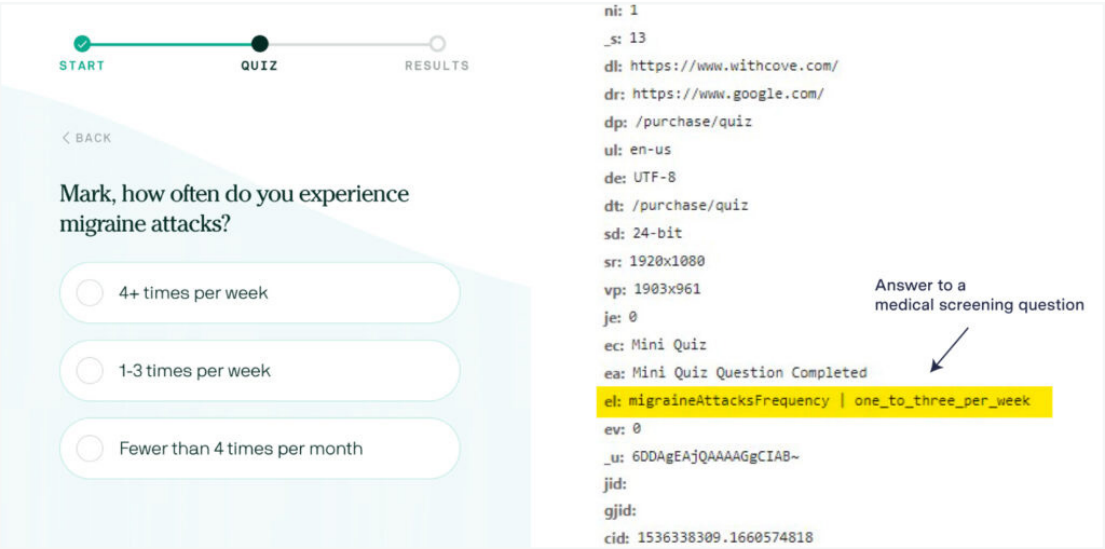
The industry's rapid growth has been enhanced by its ability to use data from tools like pixels to target advertisements to increasingly specific patient populations and to put ads in front of users who have visited their site before. The companies we analyzed mostly provide care and prescriptions for conditions like migraines, sexual health, or mental health disorders rather than comprehensive primary or urgent care — making browsing their websites inherently sensitive.

In the same way visiting an opioid use disorder treatment center can identify an individual as an addiction patient, data about someone visiting a telehealth

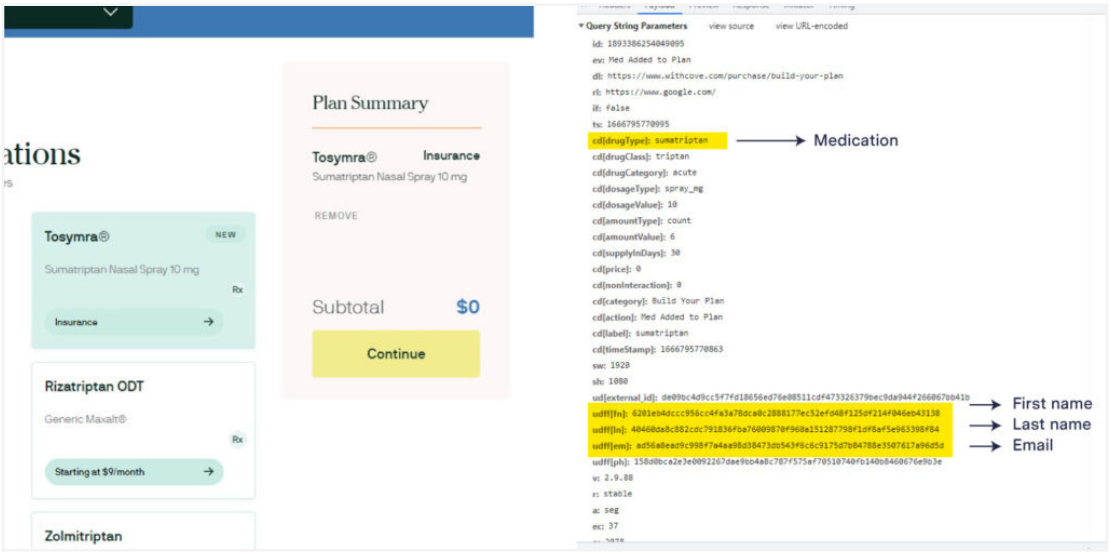
site that treats only one condition or provides only one medication can give advertisers a clear window into that person's health. Direct answers to onboarding forms could be even more valuable because they're more detailed and specific, said McCoy. "And it's more insidious because I think it would be all that much more surprising to the average person that information that you put in a form wouldn't be protected. It's both worse and more unexpected."

Consider the form for Thirty Madison's Cove, which offers migraine medications. It prompts visitors to share details about their migraines, past diagnoses, and family history — and during our testing sent the answers to Facebook and Google. If a user added a medication to the cart, detailed information about the purchase, including the drug's name, dose, and price, were also sent to Facebook, along with the user's hashed full name, email, and phone number.

While hashing obscures those details into a string of letters and numbers, it does not prevent tech platforms from linking them to a specific person's profile, which Facebook explicitly says it does before discarding the hashed data.



A Google tracker collects answers to medical screening questions on Cove's website.



A tracker tells Facebook when a user adds a medication to the cart. It also sends the user's hashed name, email, and phone number.

"It's a pure monetization play," said Eric Perakslis, chief science and digital officer at the Duke Clinical Research Institute. "And yes, everybody else is doing it, it's the way the internet works. … But I think that it's out of step with medical ethics, clearly."

In particular, experts worry that health data could be used to target patients in need with ads for services and therapies that are unnecessary or even harmful.

The big tech platforms that responded for this story say they do not allow targeted advertising based on specific health conditions, and some telehealth companies said they only use the data collected to measure the success of their advertising. However, as The Markup has previously reported, advertisers may still be able to target ads on Facebook using terms that are close proxies for health conditions.

On 35 of the 50 websites, STAT and The Markup found trackers sending individually identifying information to at least one tech company, including

names, email addresses, and phone numbers.

That presents patients with a Catch-22. "It requires anyone that wants to take advantage of telehealth … to expose a lot of the same information that they would reveal within a protected health care relationship," said Woodrow Hartzog, a privacy and technology law professor at Boston University — but without the same protections.

In recent months, regulators have begun cracking down on the indiscriminate collection and sale of personal health data.

After issuing a warning to businesses about selling health information in July, the Federal Trade Commission sued data broker Kochava, alleging that the company put consumers at risk by failing to protect location data that could reveal sensitive details about people's health, such as a visit to a reproductive health clinic or addiction recovery center. Kochava has asked for the case to be dismissed and countersued the FTC.
Meta has also come under significant scrutiny, including congressional questioning, following a Markup investigation that found its pixels sending patient data from hospitals' websites. Meta is also facing a large class-action lawsuit over the breaches.

The increased attention reflects growing fears about how health data may be used once it enters the black boxes of corporate data warehouses — whether it originates from a hospital, a location tracker, or a telehealth website.

"The health data market just continues to kind of spiral out of control, as you're seeing here," said Perakslis.

But thanks to their business structures, many of the companies behind telehealth websites appear to be operating on the outskirts of health privacy regulations.

'It does seem deceptive'

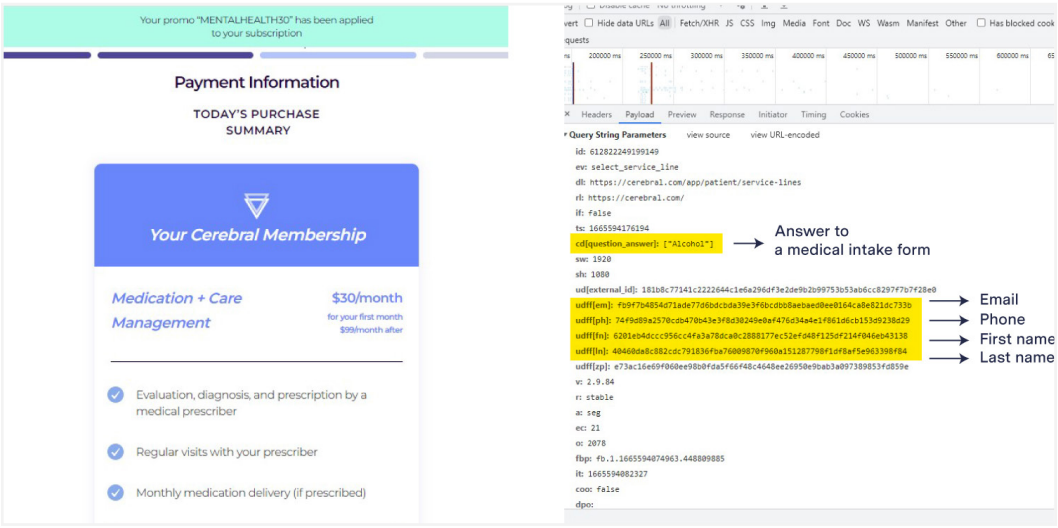
When users visit Cerebral, a mental health company whose prescribing and business practices came under federal investigation this year, they are required to answer a series of "clinically tested questions" that can cover a wide range of conditions, including depression, anxiety, bipolar disorder, and insomnia. During testing, with every response — such as clicking a button to indicate feeling depressed "more than half the days" over the last two weeks — a pixel sent Facebook the text of the answer button, the specific URL the user was visiting when clicking the button, and the user's hashed name, email address, phone number.

At a doctor's office, that kind of detail collected on an intake form would likely be subject to HIPAA. But as with most of the telehealth companies in this analysis, Cerebral Inc. itself doesn't provide care; its website connects patients with providers like those employed by Cerebral Medical Group, P.A. and others. While those medical groups are HIPAA-covered entities that cannot share protected health information with third parties except under narrow circumstances, Cerebral claims in its privacy policy to be a go-between that is not covered by HIPAA — except in limited cases when it acts as a business associate of a medical group, pharmacy, or lab.

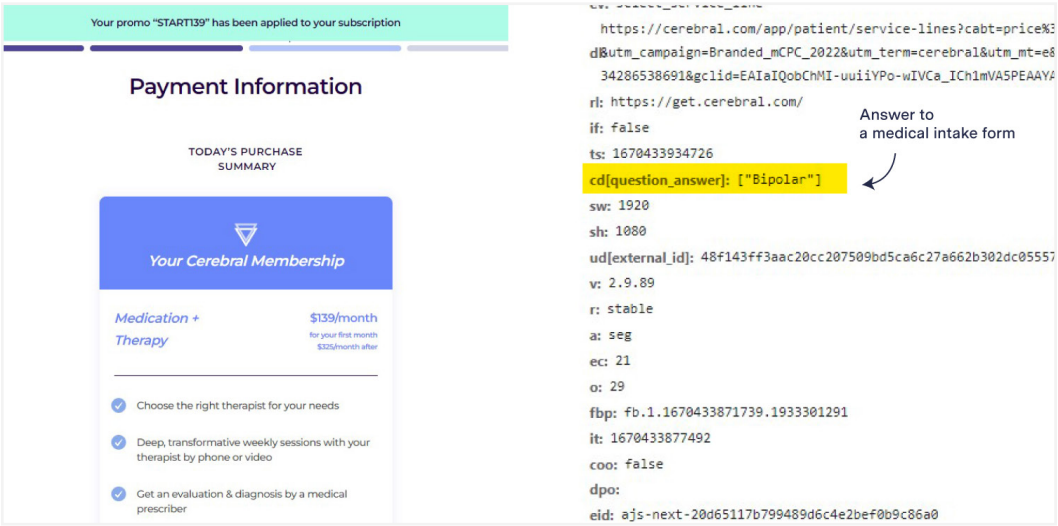
Cerebral did not answer detailed questions that would clarify what these cases might be. But in a Nov. 30 email, spokesperson Chris Savarese said the company would adjust its use of tracking tools. "We are removing any personally identifiable information, including name, date of birth, and zip

code from being collected by the Meta Pixel," he wrote.

However, when STAT and The Markup tested Cerebral's website again on Dec. 7, reporters found that a Meta Pixel was still sending answers to some intake questions and hashed names to Facebook, and trackers from Snap and Pinterest were also collecting hashed email addresses.



A Facebook tracker collected answers from a Cerebral intake form during an October test by STAT and The Markup.



During a December test, a Facebook tracker was still collecting Cerebral's intake form answers.

The telehealth companies that responded to detailed queries said their data-sharing practices adhered to their privacy policies. Those kinds of policies commonly include notice that some — but not all — health data shared with the site is subject to HIPAA. Many companies responded that they were careful to ensure that data shared via third-party tools was not considered protected health information.

But the structure of the companies' businesses — and the inscrutable language in their privacy policies and terms of use — make it difficult for consumers to know what data would qualify as protected, and when.

"There is so much intransparency, and that makes it complex and maybe even deceptive for consumers," said Sara Gerke, a professor of health law and policy at Penn State Dickinson Law.

Several telehealth companies claimed that the information collected from their websites was not personally identifiable because it was hashed. HIPAA allows health information to be shared when it has been de-identified. However, hashing does not anonymize data for the tech platforms that receive it and match it to user profiles. And every data packet sent by a tech company's tracker includes the user's IP address, which is one of several unique identifiers that explicitly qualify health data for protection under HIPAA.

Further complicating decisions for patients, at least 12 of the direct-to-consumer companies examined in this investigation promise on their websites that they are "HIPAA-compliant." That could encourage users to think all the data they share is protected and lead them to divulge more, said Hartzog. Yet the regulations apply to the websites' data use only in limited cases.

Monument, a site that offers alcohol treatment, starts its intake form by saying, "Any information you enter with Monument is 100% confidential, secure, and HIPAA compliant." Yet in its responses to STAT and The Markup, it said that it does not consider information transmitted to third parties from that form — including answers to questions like "In the past year, have you continued to drink even though it was making you feel depressed or anxious or adding to another health problem? or after having had a memory blackout?" — to be protected health information under HIPAA.

"If they're not covered by HIPAA and they have a HIPAA-compliant badge, that seems like a case the FTC could bring," said Justin Brookman, the director of technology policy for Consumer Reports and former policy director with the FTC, which has previously charged companies for deceptive use of HIPAA-compliant badges. "There's an implication there that you're regulated in certain ways, that your data is protected, and so it does seem deceptive."

Such data sharing could be particularly damaging to patients seeking care for substance use disorders, said Jacqueline Seitz, senior staff attorney for health privacy at the Legal Action Center — especially if it enters opaque data brokerages where it can be resold and repurposed indefinitely.

Several companies in this analysis are capitalizing on federal waivers activated during the pandemic that allow controlled substances like suboxone, which is used to treat opioid use disorder, to be prescribed virtually. Under federal law, qualifying addiction treatment providers — including those that prescribe suboxone — are held to patient privacy standards even stricter than HIPAA. For example, Workit's physician group states it is forbidden from acknowledging "to anyone outside of the

program that you are a patient or disclos[ing] any information identifying you as a substance use disorder patient" except in narrow situations.

Nonetheless, STAT and The Markup found that Workit and other telehealth companies — in their role connecting patients to providers — share information that identifies a user as someone seeking addiction treatment. On Boulder Care's website, a pixel sent Facebook our name and email when we joined a suboxone treatment program waitlist. And trackers on the website of Bicycle Health, another online suboxone provider, notified Google and Bing that our email address had been entered on an "enrollment confirmation" URL.

Boulder Care chief operating officer Rose Bromka said the company had started improving its "website hygiene" before being contacted for this article, and restricted the information sent by the Meta pixel after reviewing our findings.

However, Bromka added that Boulder still tracks some information about website visitors to guide its advertising.

"We are always looking to balance ensuring we are able to get the word out about options with holding to our value set," she said.

Big tech's black boxes

Meta, Google, TikTok, Bing, LinkedIn, Snap, and Pinterest say they have policies against using sensitive health data to help target advertisements.

"We clearly instruct advertisers not to share certain data with us and we continuously work with our partners to avoid inadvertent transmission of

such data," TikTok spokesperson Kate Amery wrote in an email, adding, "[W]e also have a policy against targeting users based on their individual health status."

Meta and Google claim to have algorithmic filters that identify and block sensitive health information from entering their advertising systems. But the companies did not explain how those systems work or their effectiveness. By Facebook's own admission to investigators from the New York Department of Financial Services in 2021, its system was "not yet operating with complete accuracy."

To trace what happened to data collected by trackers, STAT and The Markup created dummy accounts logged into Facebook, TikTok, and Twitter while testing the telehealth websites. Reporters then used the platforms' "download your data" tools in an attempt to determine whether any health information the trackers collected was added to our profiles.

The information provided by those tools was so limited, however, that STAT and The Markup couldn't confirm how or whether the sensitive health information was used.

For example, a Meta Pixel on RexMD, which prescribes erectile dysfunction drugs, collected the name of the medication in our cart, our email, gender, and date of birth. Facebook's transparency tool, however, only showed 10 "interactions" on RexMD's website, with generic descriptions like "ADD_TO_CART." It did not provide details about the specific data Facebook ingested during those interactions. A TikTok pixel collected some of that same information from RexMD, but TikTok's report on our "usage data from third-party apps and websites" had just one line: "You have no data in this section."

Our Twitter data showed that the company knew the dummy account user had selected a product on RexMD's website and the exact URL on which that product was selected.

On some websites, users' data was also being collected by "custom events," meaning that a website owner deliberately created a custom tracking label that could have a phrase such as "checkout" in it but wouldn't necessarily show up in the tech platforms' transparency tools.

Only four companies answered whether they had ever been notified by Facebook of potentially sensitive health information. Monument and Favor had data flagged but said they determined it wasn't sensitive. Lemonaid received a notification in error related to a promotional code, and Boulder Care had received none.

Telehealth websites should be held accountable for the trackers they install, said Hartzog, the Boston University law professor. But "big platforms that are deploying these surveillance technologies also need to be held accountable, because they're able to vacuum up every ounce of personal data on the internet in the absence of a rule that tells them not to." The companies in this investigation said their services fill an important need. "The makeup of the traditional health care system has in many cases prevented people from accessing treatment for conditions that should be easy to treat," Scott Coriell, a spokesperson for Hims & Hers, wrote in an email. Companies that serve patients with mental health or substance use disorders emphasized that long wait times to see in-person providers, and the stigma associated with seeking care, made virtual services especially valuable.

Marketing supported by third-party tracking is part of making that care

accessible, some argued. "Monument uses online advertising platforms to raise awareness of our evidence-based treatment for alcohol use disorder, and get people the support and relief they deserve," wrote CEO Michael Russell. "We transmit the minimum amount of data required to allow us to track the effectiveness of our advertising campaigns." Favor spokesperson Sarah Abboud argued that calling standard industry practices into question could threaten trust in those services.

But health privacy and policy experts see a disconnect between the industry's stated emphasis on privacy and its data-sharing practices. "Telemedicine providers should have realized from the get-go that if their entire business model is to seamlessly move people from marketing to care and the care will be online, then there's going to be more personal identifiable information submitted and thus more privacy risk and thus more privacy liability," said Christopher Robertson, a health law and policy professor at Boston University.

One problem may be that marketing teams don't fully understand privacy regulations, and legal teams don't have a handle on how the marketing tools work.

Sara Juster, privacy officer for the weight-loss telehealth company Calibrate, wrote in an email that the company doesn't "send any health information collected in our eligibility flow back to platforms." But a Meta Pixel on its site sent data including height, weight, BMI, and other diagnoses, like diabetes, to Facebook. Juster then clarified the pixel was a duplicate that should have been removed in a tracking audit earlier this year.

However, as of Dec. 7, a Meta Pixel was still present on the site and

sharing hashed identifiers and checkout events with Facebook. The pixel appeared to have been reconfigured, though, to send less information than it had during our original testing.

Without updated laws and regulations, experts said patients are left to the whims of rapidly evolving telehealth companies and tech platforms, who may choose to change their privacy policies or alter their trackers at any time.

"It doesn't make any sense that right now, we only have protections for sensitive health information generated in certain settings," said McCoy, "but not what can be equally sensitive health information generated in your navigation of a website, or your filling out of a very detailed form about your history and your prescription use."

This article was co-reported with The Markup, a nonprofit newsroom that investigates how powerful institutions are using technology to change our society. Sign up for its newsletters here.

STAT

HEALTH TECH

A new crop of companies is reshaping the health data economy

By Casey Ross | **APR. 7, 2022**

It is one of health care's most vexing quandaries: Patient data must be shared to develop more effective medicines and artificial intelligence tools, but there's no way to share it without violating privacy and basic data rights.

Or is there?

A fresh crop of companies is building a new data economy that enables the shared use of personal health information while enforcing ironclad privacy protections. They are not a monolithic group: each one uses its own methods and technologies, serves different customers, and is motivated by distinct problems and personal philosophies.

But they share a belief that the nation's system of exchanging health data — which relies on buying and selling personal information without patients' knowledge or explicit consent — is fundamentally broken. Some spoke of health data as an extension of personhood, or a digital self — not a commodity to be traded for profit.

"We don't like to think about personal data as being owned, because we don't think people can be owned," said Heather Flannery, founder and chief executive of Washington, D.C.-based Equideum Health. "We think analogies that conceive of data more like a person's labor than property are more ethically appropriate."

To ensure data are kept private and secure, these companies are erecting firewalls around health information that allow it to be used for clinical research and testing without exposing patients' identities or allowing their data to be traded between third-party owners.

Here's a closer look at the entrepreneurs and technologies behind the effort.

BeeKeeperAI

Spun out of the University of California, San Francisco, BeeKeeperAI is focused on supporting the development of more ethical and effective AI products. Its technology connects novel algorithms and data in a kind of secure computing container where the details of both remain confidential.

"No one sees anything," said Michael Blum, a physician and co-founder of the company. "The health care organization can't see the algorithm. The algorithm owner can't see the data."

By enforcing those two-sided protections, Blum hopes to unleash a freer flow of health data that would allow AI developers, including private companies and health systems, to test novel algorithms on more diverse populations of patients. Currently, algorithm developers must go on a yearslong fishing expedition to convince hospitals or health entities to grant

access to their data for training and validation. Because so few are willing to do so, though, many algorithms aren't tested enough to ensure they will perform effectively in the real world.

"What you really need are high-quality, annotated data to improve the models so they're generalizable," Blum said. "That data isn't the kind you can buy from organizations or countries that are willing to sell it. It exists in organizations that curate their data very carefully — and those are the organizations that are concerned about sharing it."

Once the data and algorithms are connected in BeeKeeper's computing container, the only thing that leaves is a report detailing how the algorithm performed in the validation and the basic characteristics of the dataset that was used to conduct the test. The company has established partnerships with Intel, Microsoft, and Fortanix, a maker of confidential computing software. It is now honing the design and usability features of its product while preparing for its commercial launch and a fundraising push in the coming months.

Equideum

Formerly known as ConsenSys Health, Equideum relies on a constellation of technologies, including blockchain, cryptography, and decentralized artificial intelligence, to create a suite of privacy-preserving products to serve consumers, pharmaceutical companies, algorithm developers, and other clients.

Its direct-to-consumer product, to be released later this year, allows patients to create a "personal data estate" so they can control the use of their health data, provide consent for specific projects, and receive compensation. They

can choose to be paid or make a tax-deductible donation.

"Data is arguably the single most valuable asset class in the human economy today," Flannery said, noting that its monetization has created some of the world's most valuable companies. "There is a huge data economy that is capital inefficient, that is opaque, that has not benefited from market economics."

Equideum is not just seeking to empower consumers in this market, but also allow health entities to collaborate on research and product development in secure data networks. The networks would enable health organizations to keep their data in their own IT systems, rather than combining it in ways that could undermine privacy.

"We will not participate in the brokerage of today's standard de-identified data economy," Flannery said, noting that such methods do not protect privacy or allow for patient consent. She said many health care organizations, including pharmaceutical companies, are looking for alternatives to that approach. One of the company's first product offerings is a privacy-preserving tool that helps match patients to clinical trials.

"I think we're going to get substantial traction with pharmaceutical companies who are going to be willing to tell the world that buyers want this," Flannery said.

Secure AI Labs

Founded in 2017, Secure AI Labs was created by researchers at the Massachusetts Institute of Technology to enable the development of more robust AI tools and build a data ecosystem useful for scientists studying

everything from cancer genomics to rare disease.

Its technology, dubbed the Uniform Patient Registry, allows AI tools to be connected with encrypted datasets that never leave the information silos of data custodians such as hospitals, patient advocacy organizations, and health researchers. Its containerized approach also allows organizations to amass enough data to tackle questions that no institution could answer on its own.

"When you pool together many researchers and many datasets, then you're able to really make a difference for discovery," said Manolis Kellis, co-founder of the company and a professor of computer science in machine learning and genomics at MIT.

Secure AI Labs (SAIL) is not just focused on creating more fluid data exchange, but harnessing the data to advance health equity. One of its founders, Anne Kim, became interested in the technology after examining pediatric asthma research. Most datasets from clinical trials had poor representation of minority patients. That made differences in their response to treatments difficult to detect — but in one use case, combining datasets with SAIL's technology gave a "very clear signal" that the drug, albuterol, was less effective for Black patients and those of Puerto Rican descent.

That was the jumping off point for a foray into multiple clinical domains. A key proof point for the company will come later this month when it discusses the use of its technology in a collaboration with hospitals and the Kidney Cancer Association to zero in on the best treatments approaches. One of the largest hospitals in the partnership sees more than 100,000 patients a year, but only 250 kidney cancer patients. "So even though the

hospital is huge and has a lot of power academically, it can't fight against the law of small (sample sizes)," said Kim. "They need to collaborate. If you're going to have anything generalizable, you need to have diversity in those patients."

Nference

Based in Cambridge, Mass., Nference has struck up a partnership with Mayo Clinic to build a massive de-identified database that allows its caregivers to answer specific questions about patient care, including treatment approaches and how well different drugs work.

The company's "data under glass" approach to privacy means that the data do not travel beyond the walls of the institution. Its technology can also de-identify information kept in unstructured formats, such as doctors' notes, to enable the secure use of granular data on a wider scale.

The company is set up to help both hospitals and pharmaceutical companies find answers to questions hidden in their data. Because the information is de-identified, that work can be carried out privately and swiftly, without need for lengthy legal and institutional review processes. In many cases, that type of bureaucracy impedes efforts to determine how to best treat patients with pressing medical problems.

"To get anecdotal knowledge, you can talk with any individual physician at Mayo who may have some subset of this knowledge captured, but if you want to know what is the institution's brain or collective knowledge — that is what this technology does," said Venky Soundararajan, the company's co-founder and chief scientific officer.

Nference has used the technology to help answer questions throughout the Covid-19 pandemic, which has often left physicians scrambling to understand new variants and tailor their treatment approaches. The company has raised about $150 million from investors, including Mayo Clinic Ventures and Matrix Capital Management. It expects to announce new health system customers later this year.

Triple Blind

After launching its data privacy technology in November 2020, Kansas City-based TripleBlind received a flood of interest from health care businesses interested in the opportunity to exchange data with global partners, including those operating under Europe's more stringent data regulations.

The company emphasizes that its technology supports the secure exchange of all data types, including images, clinical trial data, and genomic sequences. Built on a simple application programming interface, or API, it requires strict permissions for data use and allows multiple parties to participate in research, or validate an AI algorithm, without patient information leaving their institutions. Instead, the algorithm travels to them, with the results of the training shared by the entities involved.

In an emailed response to STAT, CEO Riddhiman Das said that simple de-identification of data not only fails to protect privacy, but removes demographic details that provide crucial context for research and product development.

"From a quality perspective, stripping datasets of information in order to provide a higher standard of privacy inherently reduces the utility and

precision of the data," he wrote.

TripleBlind's goal is to keep those details intact without compromising privacy. "There is tremendous intellectual property value that remains currently trapped in private data stores and proprietary algorithms," Das said.

In health care, TripleBlind said its technology can be used by customers for an array of uses, including to rapidly select clinical trials sites, generate data to validate AI tools, and monitor outcomes and adverse events related to approved drugs. It expects to announce new customers later this year.

STAT

HEALTH TECH

Hospitals pledge to protect patient privacy. Almost all their websites leak visitor data like a sieve

By Casey Ross | **APR. 3, 2023**

Every hospital in America promises to protect the privacy of its patients and the details of their medical care. And almost every one of them uses sophisticated data tools to track and share the personal information of visitors as soon as they start clicking on their websites.

A new study found that 99% of U.S. hospitals employed online data trackers in 2021 that transmitted visitors' information to a broad network of outside parties, including major technology companies, data brokers, and private equity firms.

The data captured included visits to pages on specific conditions such as depression, breast cancer, and Alzheimer's disease. The ubiquitous use of the tracking tools may clash with the privacy expectations — if not the legal protections — that consumers take for granted as they browse online in search of medical care and information.

"The scale and scope of this continues to shock me even as I work on this research," said Matthew McCoy, a co-author of the study and assistant professor of medical ethics and health policy at the University of

Pennsylvania. "One cannot really access a hospital website in this country without being exposed to really significant levels of tracking."

The study found that hospitals were not only commonly sharing visitor information with the online advertising giants Meta and Alphabet, but also with companies such as AT&T, Verizon, Amazon, the media giant Nielsen, and Golden Gate Capital, a San Francisco-based private equity company.

The data trade forms the backbone of a multi-billion dollar economy that quietly compiles information on consumers to target advertisements and help make decisions about how to recruit employees and distribute products such as prescription drugs and medical devices. Because such decisions are made behind corporate walls, it remains unclear how much personal information these companies gather, and exactly how they use it.

The federal privacy rules created under HIPAA, which governs the sharing of personal information collected on patients, prohibits the disclosure of certain pieces of information that could identify patients. In December 2022, the federal Department of Health and Human Services clarified that those rules apply to hospital websites that use tracking codes to collect and share information such as patients' IP addresses, health conditions, and symptoms.

That doesn't necessarily mean that the information scraping spotlighted in the study, published Monday in Health Affairs, constitutes a HIPAA violation, said Brad Malin, director of the health information privacy lab at Vanderbilt University. That's because it involved data transmitted on the hospital home pages and public-facing areas, not portals where patients share specific information about their conditions and health needs with their doctors.

"If the user had logged in to these sites, such that the trackers were on pages associated with their diagnosis…then it would be a violation of HIPAA without a doubt," Malin said.

To conduct the study, researchers at the University of Pennsylvania used an open-source tool known as webXray to record third-party tracking tools present on hospital websites during a three-day period in August 2021. The researchers also recorded the presence of "cookies," or snippets of data stored on a user's web browser that allow them to be tracked across multiple sites. They used a webXray database to link the tracking domains to their parent companies so they could see where the data were being routed.

Hospitals use tracking tools supplied by technology companies for the same reason many other businesses do: They want data on the use of their web pages as consumers interact with them online.

"Companies have become hyper-specialized in providing this type of support, such that the health care organizations are going to take it because it's cheap and it's useful for them," Malin said. "But it ends up creating a view into an individual's life that the (hospitals) probably were not really considering" when they created their websites.

The study found that the home pages of more than 3,700 hospitals initiated a median of 16 data transfers to third parties. It also found that the tracking tools were equally present on pages used by patients to research specific medical conditions. Malin said that it is difficult to know what other information the companies receiving the data already have about a person, such as consumer data on shopping or personal interests.

Although the study found nearly all hospitals used such tools, it also revealed that nonprofit hospitals with medical school affiliations and those serving urban areas tended to expose patients to higher levels of third-party tracking.

The issue of health data tracking extends beyond hospitals: In December, an investigation by STAT and The Markup found that dozens of direct-to-consumer telehealth companies were collecting sensitive information from users and sharing it with the world's largest advertising platforms. The Federal Trade Commission has started to crack down on that type of data sharing, and has reached settlements with both BetterHelp and GoodRx for health data leaks this year.

But ultimately, the burden still largely falls on consumers to protect themselves as they seek out health care services online — even if their ability to do so is significantly constrained by the volume of information now floating around about them. Those data may be used to shape both the information and opportunities that surround them on a daily basis.

"It might also be that you don't get shown an ad for a particular job because of things that are found out from your health-related tracking," said Ari Friedman, a co-author of the study and physician at the University of Pennsylvania. "The remedy there is hard because the details are so obscure, and so difficult to access."

This story is part of a series examining the use of artificial intelligence in health care and practices for exchanging and analyzing patient data. It is supported with funding from the Gordon and Betty Moore Foundation.