JOCHEN LENNERZ, NICK SCHNEIDER, AND KARL LAUTERBACH

How Health Data Integrity Can Earn Trust and Advance Health

Efforts to share health data across borders snag on legal and regulatory barriers. Before detangling the fine print, let's agree on overarching principles.

Imagine a scenario in which Mary, an individual with a rare disease, has agreed to share her medical records for a research project aimed at finding better treatments for genetic disorders. Mary's consent is grounded in trust that her data will be handled with the utmost care, protected from unauthorized access, and used according to her wishes.

It may sound simple, but meeting these standards comes with myriad complications. Whose job is it to weigh the risk that Mary might be reidentified, even if her information is de-identified and stored securely? How should that assessment be done? How can data from Mary's records be aggregated with patients from health systems in other countries, each with their own requirements for data protection and formats for record keeping? How can Mary's wishes be respected, both in terms of what research is conducted and in returning relevant results to her?

From electronic medical records to genomic sequencing, health care providers and researchers now have an unprecedented wealth of information that could help tailor treatments to individual needs, revolutionize understanding of disease, and enhance the overall quality of health care. Data protection, privacy safeguards, and cybersecurity are all paramount for safeguarding sensitive medical information, but much of the potential that lies in this abundance of data is being lost because well-intentioned regulations have not been set up to allow for data sharing and collaboration. This stymies efforts to study rare diseases, map disease patterns, improve public health surveillance, and advance evidence-based policymaking (for instance, by comparing effectiveness of interventions across regions and demographics). Projects that could excel with enough data get bogged down in bureaucracy and uncertainty. For example, Germany now has strict data protection laws—with heavy punishment for violations—that should allow de-identified health insurance claims to be used for research within secure processing environments, but the legality of such use has been challenged.

What will help is to step back from focusing on the minutia and embrace a larger principle: health data integrity. We see this term as encompassing both technical safeguards (accuracy, security, access) and ethical values (protecting patients, respecting their wishes, and advancing equitable, high-quality health care). As the Belmont Report and the Declarations of Helsinki and Taipei did for clinical research on human subjects, we believe that an international, multistakeholder effort to define and commit to health data integrity can help facilitate frameworks and cultural norms that justify Mary's trust that her data will not be altered or misused, that her privacy will be respected, and that her contribution to medical science will be meaningful and secure. In other words, health data integrity can serve as a guiding principle that embodies the collective conscience of health care.

Integrity in a big data era

When Nick Schneider was a boy living in Argentina in 1988, he was hit by a car. After a brain scan resulted in an (incorrect) incidental finding of early dementia, his parents were able to mail medical data to experts in Germany and the United States for helpful second opinions. Today, such cross-country consulting could put health care providers in legal limbo. Numerous other practices that could potentially benefit individual patients are often challenging to navigate. For example, one of us (Lennerz) regularly encounters situations where he needs to identify patients similar to Mary in national or international databases. However, finding patients with identical or related genetic alterations and obtaining dependable medical information is a challenging—if not impossible—task, not because patients have opted out, but because health systems aren't set up to enable it.

The regulatory and legislative frameworks governing health care data have, in many cases, struggled to keep pace with the requirements for collaborative research and innovation. The late Robert Eiss, who helped coordinate international projects at the US National Institutes of Health (NIH), highlighted several significant consequences of data sharing restrictions: almost 50 clinical research sites in the European Union were prevented from participating in NIH-sponsored COVID-19 trials, and dozens of EU projects assessing genetic and environmental factors for cancer risk were stalled. Prohibitions against exporting data prevent EU-run trials from submitting evidence to non-EU regulators, including the US Food and Drug Administration.

Different specifications around essential ethical practices—such as protecting sensitive data and obtaining informed consent—can also prevent collaborations. And the practical realities of working with real-world data, such as the heterogeneity of electronic medical records, often undercut efforts to put data to use. As health data science advances, the need for coordinated, internationally standardized, and reliable frameworks grows more apparent.

Effective frameworks for establishing health data integrity need to accomplish many aims simultaneously. They should honor informed consent and balance privacy needs with the benefits of sharing data—while also encouraging collection of the broadly representative data required to inform equitable health care practices. Frameworks should provide overarching requirements that ensure ethical data handling, responsible data use, and the transparent operation of language models to prevent fraud and abuse; and they need to enforce strict authentication protocols. International data sharing might seem to add to the complexity of these tasks, but we think it could actually ease them. These multifaceted ethical, regulatory, and practical challenges are best tackled via collaboration across countries and functions.

Solutions to these disparate problems share a common prerequisite: health care depends upon trust. Trust in the context of health data science encompasses trust between researchers, between patients and their health care providers, between humans and the technology they apply, and between nations in transnational collaborations. Health care workers must also trust that, say, blood samples and biopsies are analyzed in ways that enable good decisionmaking and patient care. And trust is earned through integrity.

Dedication to integrity

The need for integrity as a larger principle was brought home to us several years ago in a fortuitous encounter between two of us, Karl Lauterbach, Germany's health minister, and Jochen Lennerz, who, at the time, ran a technology assessment laboratory at Massachusetts General Hospital. Lauterbach, an epidemiologist, faces national and supranational barriers to enabling health data research that improves care while simultaneously addressing privacy concerns in Germany and Europe. Lennerz faces practical and regulatory challenges to introducing cutting-edge diagnostics into cancer and other clinical care. For both, proper, effective handling of highly complex data is of paramount concern.

We joined forces with our third author, Nick Schneider, who negotiated the European Union's General Data Protection Regulation (GDPR) on behalf of the German Federal Ministry of Health and led both the taskforce to adapt German federal health laws to the GDPR and the current negotiations toward a European Health Data Space, an infrastructure and framework set up to empower patients, protect their data, and foster health data science. Together we organized a high-level brainstorming meeting in Berlin, hoping to set the stage for strategic alignment.

This Data for Health conference brought together over 400 international stakeholders in the summer of 2023—representatives from industry, academia, law, biomedical sciences, and civil society. There were patient advocates, legislators, health policy advisors, consultants, students, ethicists, creative commons legal experts, data protection and cybersecurity experts, government and tech industry representatives, as well as private citizens and patients.

Instead of the usual conference setup with lectures and posters, we had panel discussions and participant-driven conversations, following the BarCamp format. This let us delve into some of the most pressing questions in health data science: assuring adequate levels of data protection and consent, assessing current data transfer practices, identifying legal bases for transfers, implementing additional safeguards within a legal vacuum, and creating mechanisms that enable health data to be treated differently from consumer data. We made much of the content available online for anyone who wants to follow the conversation and perhaps join in the future.

In these sessions, we also learned the depth of the conundrum this effort faces: discrepant regulatory and legislative frameworks on either side of the Atlantic lack any appropriate, practical working guidelines for enabling collaborative research. The threats to inadequately secured data are very real, as made clear by the list of breaches maintained at the US Department of Health. A better focus on the most pressing risks could improve data security and health data science. One theme that came up at the conference was that data protection officers within hospitals and health agencies often see their roles as solely protecting data against, say, an abstract risk of reidentification or unauthorized disclosures—rather than considering how data could be used to advance health care or how patients wish their data to be used.

The concept of health data integrity emerged as a guiding principle that resonated throughout the gathering in Berlin, surprising even the most seasoned participants. Integrity extends beyond the realm of data accuracy or security; it encompasses a commitment to fairness, honesty, and respect throughout the entire health data life cycle. It includes enabling appropriate use of data to advance health care, drive innovation, and enhance the well-being of diverse populations.

It also intertwines with the pursuit of equitable health care. Health data integrity is essential in any effort to share sensitive data, and sharing diverse, representative datasets is the only way to gain insights across a spectrum of patients and so enable a more comprehensive understanding of health patterns, treatment efficacy, and various health influences on different demographic groups. In this context, health record vendors, health care providers, or any other stakeholders that interfere with permitted access, exchange,

or use of health data violate integrity by hindering research and patient care. Recent laws in both the United States and Europe already ban this kind of interference as "information blocking," but it still happens in practice. The development of common patient information and consent forms, as well as collaboratively written codes of conduct, can serve as practical means to ensure transparency, shareability, and accountability across the system. This was proposed by the Council of the European Union in its conclusions on COVID-19 lessons learned in health and confirmed by conference participants in Berlin.

We continued to address these topics in a follow-up workshop in Boston last fall. The initiative demonstrated that commitment to integrity is an essential enabler; without that assurance, the medical field will not be able to move on from outdated, contradictory frameworks and embrace overarching ones to protect patients *and* advance research. The situation demands a concerted, comprehensive effort to produce an effective regulatory landscape. By embracing integrity, health care professionals, vendors, researchers, and policymakers can establish a financially sustainable health data science ecosystem that honors data subjects and drives improved patient care.

Creating a cultural imperative

At this point, instead of getting bogged down in detailed discussions about the numerous complex regulations that complicate the landscape, it might be more effective and efficient to simply commit to a clear declaration: it is not enough to merely share data; it must be done with integrity.

This approach has helped before. The declarations of Helsinki (established in 1964 and updated several times) and Taipei (established in 2002) have long served as beacons of

WORKING DEFINITION OF INTEGRITY

For citizen advocates, researchers, health care providers, vendors, regulators, and other stakeholders to coalesce around the concept of health data integrity, the concept must be defined. Here is our definition based on multiple inputs so far.

Integrity of data in health refers to the overall accuracy, completeness, and consistency of health-related data, ensuring that it can be trusted for critical decisionmaking processes in the health care domain. It encompasses the principles of accuracy, completeness, and consistency as per the ALCOA (attributable, legible, contemporaneously recorded, original or a true copy, and accurate) principle developed for integrity in the life sciences. In the context of health care, data integrity is essential to maintain the quality and trustworthiness of patient records, medical research, and health care operations. Health data integrity serves as a cultural imperative, one that demands all stakeholders actively safeguard patient confidentiality, respect individual autonomy, and provide high-quality care based on evidence and patient-specific needs.

This definition should also encompass integrity in the broader sense—the practice of being honest, reliable, and adhering to strong moral and ethical principles and values. It entails a commitment to dependability, loyalty, honesty, good judgment, and respect, and is a fundamental principle for promoting trust, fairness, and reliability in various aspects of life and professional practice.

ethical conduct in medical research. The first declaration sets rules for medical research involving human subjects, and the second specifies research on health databases, big data, and biobanks. The infamous Tuskegee syphilis study in the United States also brought substantial changes in ethical guidelines for medical research. This study, conducted by the US Public Health Service from 1932 to 1972, withheld treatment from African American men with syphilis without their knowledge or consent. The resulting outrage led to the US National Research Act of 1974 and Belmont Report of 1979, which mandated the creation of institutional review boards and established basic bioethical principles, such as respect for persons, fair treatment, and an expectation that research subjects will benefit from participation. Together these declarations built up a cultural imperative to uphold ethical research on human subjects.

Now it's time to extend the conversation to a new ethical and moral code for the use of data technologies in medicine. The medical profession, research communities, patient organizations, and civil societies need to set clear

used around the world. When she signs the forms, she is provided with realistic options to opt out in the context of a conversation with her trusted provider. In this future, trans-agency coordination fosters health data integrity across health care institutions and regulatory bodies; seamless collaboration and information sharing are designed to benefit patients while upholding ethical standards. Additionally, trans-Atlantic consent mechanisms are established, integrating the requirements of both sides to foster cross-border health care data exchange that respects individual privacy and security needs.

Through the Data for Health Initiative, we have uncovered more than a dozen forms of integrity across various technical, professional, and other contexts. To move forward, several concepts must converge and harmonize with secure data practices to enable the power of large language and other artificial intelligence models. For instance, interoperability can enhance data sharing and collaboration, tokenization can provide a secure way to handle sensitive information, and blockchain can ensure the transparency and integrity

> Integrity extends beyond the realm of data accuracy or security; it encompasses a commitment to fairness, honesty, and respect throughout the entire health data life cycle.

ethical and moral boundaries to underpin technical and legal requirements. The cultural imperative of health data integrity should be made strong enough to prevent health care providers, researchers, or vendors from violating the spirit of integrity, with appropriate legal implications.

Reasons why people wouldn't want their data shared should be proactively assessed and honored, and everyone within health data science should be frank with how data might be used, including that it may not be feasible to retrain models if patients opt out and that there can be no guaranteed protections against hacking, resale of data, or nefarious unanticipated uses of data.

Regulatory and legislative governance structures ensure that ethical standards are upheld, patient rights are protected, and data privacy is maintained. We argue that elevating health data integrity to a cultural imperative can achieve a kind of commitment that frameworks alone cannot. A cultural imperative compels people in the field to focus on more than meeting requirements and avoiding liability; they will be expected to do right by their patients and to enable data practices that produce better health care.

Imagine a future where the consent that Mary gives within her health care setting is compatible with processes

of data—all of which are essential to unleash the potential of these technologies to transform health care while safeguarding patient privacy and security. We cannot imagine accomplishing these huge, important tasks without the concept of health data integrity to unite and motivate efforts.

We implore the medical profession, research communities, patient organizations, and civil society at large to take proactive steps in shaping the future of health data science. By embracing integrity as a cultural imperative, stakeholders can navigate the complexities of health data science with ethics, transparency, responsibility, and improved care as guiding stars. This will help overcome the challenges of interdisciplinary miscommunication and other barriers to drive meaningful advancements in health care. A culture of health data integrity can ensure that patients have less to risk when they share data, and more to gain.

Karl Lauterbach is the federal health minister of Germany and adjunct professor at the Harvard T. H. Chan School of Public Health, in the Department of Health Policy and Management. Nick Schneider is the head of the division on new technologies and data use at the German Federal Ministry of Health. Jochen Lennerz is the chief scientific officer at BostonGene.