Journal of the Society for Clinical Data Management

Craven CK, Jackson B, Solomonides T. Introduction to the JSCDM Special Issue on Clinical Data Sharing. *Journal of the Society for Clinical Data Management*. 2023; 3(4): 5, pp. 1–3. DOI: https://doi.org/10.47912/jscdm.315

GUEST EDITORIAL LETTER

Introduction to the JSCDM Special Issue on Clinical Data Sharing

Catherine K. Craven*, Brian Jackson† and Tony Solomonides‡

Keywords: data sharing; EHR to EDC; EHR; EDC

Stores of electronic clinical data have expanded dramatically over the past decade, driven by widespread adoption of electronic health records (EHRs). There is corresponding momentum toward increased sharing of this clinical data for secondary reuse, especially for research purposes, whether those data are generated at healthcare institutions through research visits or regular care. The possibility of access to such truly Big Data for secondary use is tantalizing, including for those who want to deploy machine-learning and artificial intelligence methods on large data sets for hypothesis generation and potential foundational scientific discovery.

However, there is an urgent need for greater understanding and management of clinical data shared and used for secondary purposes. Issues include biases in the data, lack of appropriate representation of populations within data, and potential for misuse (including theft) of private health information, among others. This timely special issue of the Journal of the Society for Clinical Data Management contains a number of thoughtful articles exploring the benefits, risks, and risk-management approaches associated with data sharing.

Increased data sharing for secondary purposes is an explicit international priority. In 2017, following the World Medical Association's Declaration of Helsinki updates for human subjects research studies and data sharing, the International Committee of Medical Journal Editors (ICMJE) began to require that interventional clinical trials post individual participant data (IPD) sharing plans in public registries such as ClinicalTrials.gov prior to participant enrollment. Eligibility for publication in ICMJE member journals is contingent on compliance. In this issue's "Evaluating Individual Participant Data Plans for International Committee of Medical Journal Editors Compliance: A Case Study at University of Michigan," authors report that although technical compliance at their institution is high, actual data sharing is much lower, and they suggest process improvements to bolster adoption for increased data sharing.

From the U.S. perspective, the National Institutes of Health (NIH) has taken multiple steps to drive increased data reuse and sharing for research. Recent NIH initiatives include the Accrual to Clinical Trials (ACT) federated EHR data network, now Evolve to Next-Gen ACT, and the Science and Technology Research Infrastructure for Discovery, Experimentation and Sustainability (STRIDES) to facilitate cloud data storage and use, both launched in 2018. In 2020, the COVID-19 pandemic catalyzed EHR data sharing for research in the cloud. NIH launched the National COVID Cohort Collaborative (N3C), which is a cloud-based secure enclave in which harmonized EHR data, now for almost 20 million people from 78 sites, can be accessed and analyzed but not removed. Three tiers of data are available through various graduated data use agreements: synthetic, de-identified, and limited data sets. In 2023, 10 sites joined a pilot to expand the N3C secure enclave model to include other diseases. Also in 2023, NIH launched the 2023 Data Management and Sharing (DMS) Policy requiring all NIH-funded researchers to prospectively submit plans for data management and sharing.

A wide range of data aggregation and analysis platforms continue to be developed by other types of organizations as well. Authors here from Epic Systems, Inc., discuss the functionality and evolving opportunities within their proprietary platform in "Cosmos: Real-World Data Powered by the Healthcare Community." A different type of data aggregation platform is described in "Big Data Education for Nursing Education," in which the authors describe their institution's efforts to prepare clinicians for a more data-intensive future.

Patients have long expressed concerns about consent, privacy, and security regarding all uses of their clinical data, yet they have also expressed appreciation of the value of sharing such data for research, so that new knowledge can be translated into improved diagnosis, treatment, and health outcomes.[1–7] As technology and scientific approaches discussed in this issue lead to ever-larger data sets, they raise significant data management and governance issues.

* Population Health Science and Policy, University of Texas Health Science Center San Antonio, US

† Departments of Pathology and Biomedical Informatics, University of Utah, US

‡ NorthShore University HealthSystem Research Institute, US

Corresponding author: Catherine K. Craven (catherine.craven@gmail.com)

The author of "Key Considerations When Designing Real World Evidence Research Involving The Secondary Use of U.S.-Based Electronic Health Record Data" describes how global research organizations can navigate US regulation, including HIPAA. Current regulations may no longer be sufficient to protect patient autonomy and privacy, however. The authors of "The HIPAA Deidentification Exception Must Go" explain how data-linking technologies and business practices have evolved in ways that circumvent HIPAA's original intent.

In a similar vein, the authors of "A Privacy Nihilist's Perspective on Clinical Data Sharing: Open Clinical Data Sharing is Dead, Long Live the Walled Garden," probe the present realities of cyberattacks, data breaches, as well as legal yet unethical data aggregation, reidentification, and use by brokers that are made possible because of HIPAA loopholes. They propose the non-open sharing of data used only within secure enclaves, their eponymous "walled gardens," such as N3C, and discuss necessary conditions for walled garden creation and use. The authors also suggest steps that data managers and institutions can take now, to clarify consent for patients, communicate re-identification risk, and safeguard transferred data outside of secure enclaves. They also discuss data access by law enforcement in US states with repressive legislation, e.g., concerning reproductive or gender reassignment medicine, which has the potential to disrupt well-meaning data collection and curation efforts.

The distinctive ethical interests of individual patient populations are explored in "Realization of Disability Equity Through Ethical Data Management Practices." The author presents an argument for collection of more detailed patient attributes as a prerequisite to health equity, while simultaneously arguing for more patient control over their data. In "Geographic Information Systems as Data Sharing Infrastructure for Clinical Data Warehouses," the author probes the tensions between regulatory protection of patient addresses and the value of spatial computing when addresses from EHR data in warehouses are geocoded and converted to less sensitive social determinants of health data. The author discusses the value of sharing such data for research and advocates for a more flexible interpretation of legal prohibitions on sharing location data.

As the articles here were entering peer review, a data breach occurred for a Washington, D.C.-based health insurance exchange, DC Health Link, exposing protected health information on "170,000 individuals, although the official notice about the breach says 56,415 people were affected." This data was placed for sale on a dark web site, BreachForums, on which "some of the world's largest hacked databases show up for sale." The owner-administrator of this illegal site, now shuttered by the U.S. Federal Bureau of Investigation, was a 21-year-old who lived with his parents, just two blocks from one of the guest editors of this issue, in Peekskill, New York., a quiet middle- and working-class suburb 40 miles north of NYC. The perpetrator was just 19 when he started the site.[8]

Do we know what are the relative security risks of health systems' secured networks for health data storage v. those of allowed third parties v. a secure enclave for research such as N3C, which is operated on Palantir, the same cloud-based big data analytics platform used by the U.S. Intelligence Community and the U.S. Department of Defense? As important, what are we communicating about this to patients, who hear of health data breaches yet likely do not understand the range of security risks across different technologies and governance structures?

And what about the thorny question of how patient consent (whether broad or study-specific) and its revocation might be managed in large walled-garden repositories?

These are just some of the thoughts that arose as we combed through these varied authors' contributions. It seems that ongoing re-examination of data-sharing for all purposes, from the perspectives of all stakeholders, will be our collective responsibility for the foreseeable future.

We thank the other members of JSCDM Board, the immediate past and current JSCDM Editors-in-Chief, and its managerial staff for this opportunity to serve as Guest Editors for this special issue. We've enjoyed putting together this edition with submissions from long-time colleagues and some new to us. We hope that you will find their work as thought-provoking as we did, and that you will use it as a vehicle for further data-sharing discussions with colleagues in your workplaces and among your professional networks, including members of the Society for Clinical Data Management.

Best regards,
Catherine K. Craven, PhD, FAMIA
Brian Jackson, MD
Anthony Solomonides, PhD FAMIA
JSCDM Special Issue on Clinical Data Sharing Guest Editors

Competing Interests

The author has no competing interests to declare.

References

1. **Luchenski SA, Reed JE, Marston C, Papoutsi C, Majeed A, Bell D.** Patient and public views on electronic health records and their uses in the United kingdom: cross-sectional survey. *J Med Internet Res.* 2013 Aug 23; 15(8): e160. doi: 10.2196/jmir.2701. PMID: 23975239; PMCID: PMC3758045. DOI: https://doi.org/10.2196/jmir.2701
2. **Kim J, Kim H, Bell E, Bath T, Paul P, Pham A, Jiang X, Zheng K, Ohno-Machado L.** Patient Perspectives About Decisions to Share Medical Data and Biospecimens for Research. *JAMA Netw Open.* 2019 Aug 2; 2(8): e199550. PMID: 31433479; PMCID: PMC6707015. DOI: https://doi.org/10.1001/jamanetworkopen.2019.9550
3. **Harle CA, Golembiewski EH, Rahmanian KP, Krieger JL, Hagmajer D, Mainous AG, Moseley RE.** Patient preferences toward an interactive e-consent application for research using electronic health records. *J Am Med Inform Assoc.* 2018 Mar 1; 25(3): 360–368. PMID: 29272408; PMCID: PMC5992814. DOI: https://doi.org/10.1093/jamia/ocx145

4. **Brelsford KM, Spratt SE, Beskow LM.** Research use of electronic health records: patients' perspectives on contact by researchers. *J Am Med Inform Assoc*. 2018 Sep 1; 25(9): 1122–1129. PMID: 29986107; PMCID: PMC6118867. DOI: https://doi.org/10.1093/jamia/ocy087
5. **Kim J, Kim H, Bell E, Bath T, Paul P, Pham A, Jiang X, Zheng K, Ohno-Machado L.** Patient Perspectives About Decisions to Share Medical Data and Biospecimens for Research. *JAMA Netw Open*. 2019 Aug 2; 2(8): e199550. PMID: 31433479; PMCID: PMC6707015. DOI: https://doi.org/10.1001/jamanetworkopen.2019.9550
6. **Petersen C.** User-focused data sharing agreements: a foundation for the genomic future. *JAMIA Open*. December 2019; 2(4): 402–406. DOI: https://doi.org/10.1093/jamiaopen/ooz043
7. **Morse B, others.** Patient and researcher stakeholder preferences for use of electronic health record data: a qualitative study to guide the design and development of a platform to honor patient preferences. *Journal of the American Medical Informatics Association*. June 2023; 30(6): 1137–1149. DOI: https://doi.org/10.1093/jamia/ocad058
8. **Krebs B.** Krebs On Security [Internet]. "Feds Charge NY Man as BreachForums Boss "Pompompurin." Krebs B.; [Published March 17, 2023; Accessed August 7, 2023]. Available from: https://krebsonsecurity.com/2023/03/feds-charge-ny-man-as-breachforums-boss-pompompurin/

How to cite this article: Craven CK, Jackson B, Solomonides T. Introduction to the JSCDM Special Issue on Clinical Data Sharing. *Journal of the Society for Clinical Data Management*. 2023; 3(4): 5, pp. 1–3. DOI: https://doi.org/10.47912/jscdm.315

Submitted: 28 October 2023 **Accepted:** 30 October 2023 **Published:** 06 November 2023

Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS