Challenges and Opportunities for Secondary Use of Observational Data Following an EHR Transition

Michael E. Matheny, MD, MS, MPH[1,2,3], Hannah M. Gelman, PhD[4,5], Maria Souden, MSI, PhD[4,5], Zhenyu Lu, PhD, MSIS[6], Scott L. DuVall, PhD[6,7], and Margaret E. Gonsoulin, PhD[8]

[1]VA Informatics and Computing Infrastructure, Tennessee Valley Healthcare System VA, Nashville, TN, USA; [2]Geriatrics Research Education and Clinical Care Service, TVHS VA, Nashville, TN, USA; [3]Department of Biomedical Informatics, Biostatistics, and Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [4]VA Information Resource Center, VA Office of Research & Development (ORD), Hines, IL, USA; [5]VA ORD Strategic Initiative for Research and EHR Synergy (OSIRES), Hines, IL, USA; [6]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA; [7]Department of Internal Medicine Division of Epidemiology, University of Utah School of Medicine, Salt Lake City, UT, USA; [8]VA Electronic Health Record Modernization- Integration Office, Washington, DC, USA

INTRODUCTION

The Meaningful Use program and the Health Information Technology for Economic and Clinical Health (HITECH) Act generated strong incentives resulting in most US healthcare systems transitioning from paper records to electronic health record systems (EHR) in the last 15 years.[1] Data from the Office of the National Coordinator for Health Information Technology (ONC) from 2021 showed that 96% of non-federal acute care hospitals and 78% of office-based practices had a certified HER.[2] In addition, many healthcare systems have switched vendors or shifted from a homegrown system to a vendor EHR, a trend that continues nationally.[3] For example, the Department of Veteran Affairs (VA), a large integrated healthcare system with 171 medical centers and 1113 outpatient clinics, is currently transitioning from a self-developed EHR to a commercial ONC-certified vendor EHR developed by Oracle Cerner.

EHR transitions are extremely costly and challenging, creating substantial impacts on organizational culture, workflow, and processes. Overall, the literature surrounding various challenges, needs, and opportunities of EHR transitions is sparse.[4] Most of the studies of EHR transitions have been around data migration or providing the historical data necessary for care continuity following a transition.[4] In addition to clinical care concerns, healthcare organizations need to continue using routinely collected EHR data to conduct quality improvement, research, and operational analyses, often leveraging data extracted from the legacy and new EHRs, separately, and seeking to align these data for reuse.

There is a large literature about the considerations for the use of routinely collected EHR data for operational and research purposes, including data inaccuracy, bias in data collection, information "locked" in clinical text, and multiple conflicting source data provenances, among others.[5] In addition, there is a body of work on evaluation criteria for observational data quality and data assessment best practices.[6,7] Within this larger domain, we highlight key considerations, opportunities, and challenges for assessing the utility of routinely collected data across an EHR transition for analytic purposes.

ALIGNMENT OF DATA DOMAINS BETWEEN A LEGACY AND NEW EHR

As summarized in Fig. 1, the data structures, formats, elements, and terminologies of two EHR systems are generally different, and there are some data domains that will exist in one but not both EHRs. In some cases, these differences mean that analyses and uses of the data can only be executed in one of the data sources. It is critically important to characterize, assess, and document where data do not align between the two EHRs. This has implications for reuse (i.e., the secondary use of the data for research and operational analyses) that can continue after the new EHR goes live, and provides opportunities to ask new questions and develop new methods and designs that capitalize on data domains available only in the new EHR. Also, request to capitalize EHR all letters for all use in the document (section headers, etc.).

IMPACT OF EHR CLINICAL WORKFLOW ON ROUTINELY COLLECTED DATA

As noted in this special issue's editorials and articles and in the literature, deployment of a new EHR is a complex socio-technical process. Within this process, the organizational culture and business needs drive the customization and configuration of the EHR. At the same time, the core features and workflows of the new EHR change and reshape clinical workflow to conform to what is possible. These unfamiliar processes can result in increased time to complete tasks, dependence on workarounds, and propensity for error.[8]
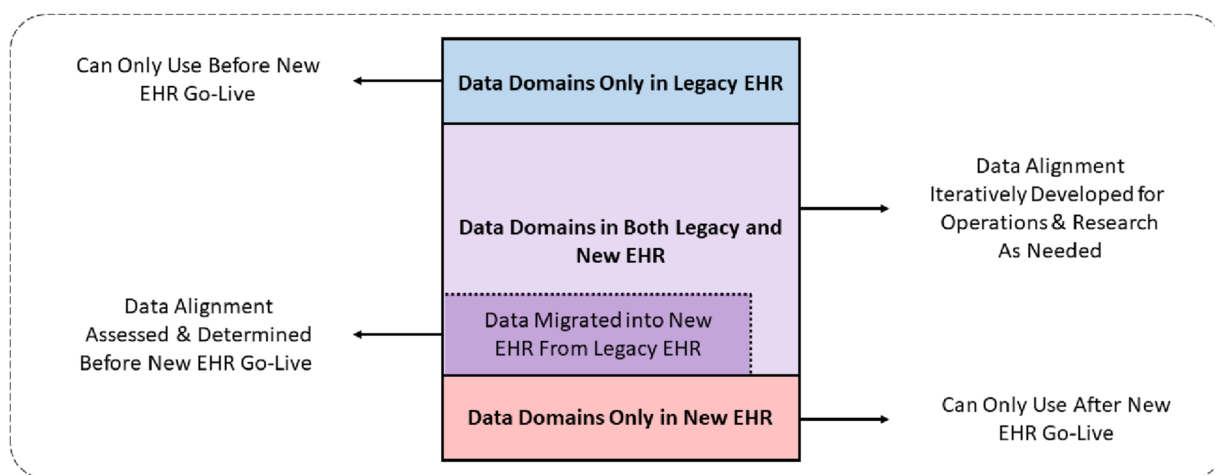
Fig. 1 **Overview of data domain alignment between a legacy and a new EHR**

In addition to usability challenges, workflow changes also result in data interpretation and reuse challenges. In some cases, the configuration of the system itself could lead to or contribute to errors in the entry or recording of data, which would be reflected in the data extracted from the system and used for analytics, quality improvement, or research. This has been observed in VA's current implementation of the Oracle Cerner EHR, where patient safety issues related to "lost" orders have been linked to order entry interfaces that made it difficult for providers to select the appropriate receiving location, resulting in unanticipated (i.e., different from the legacy system's) system behavior when a data entry error was made.[9]

Data entered in error or incorrectly routed through the system will, of course, also impact data reuse, as analyses of healthcare use patterns may not reflect reality or provider intentions. In other cases, workflows that differ between the new and old systems may maintain functionality for clinical users but pose challenges to the interoperability of data generated across the transition. For example, new workflows may change how providers coordinate care and may result in changes to when orders or referrals are placed. Data reuse will need to account for these changes when examining patterns in healthcare utilization across the transition.

It is important to evaluate data collection in the context of the end-user workflow, and to assess the data in both the legacy and new EHR workflows, to understand some of the potential limitations and caveats in the use of the data. As noted in Fig. 1, some of these challenges may be obvious when data are no longer collected or begin to be collected; others may be subtle and challenging when the elements to be collected are the same before and after the transition, but the workflow creates systematic differences in the volume and veracity of that data collection.[10] Though this issue is ubiquitous, a comprehensive mapping between new and old data elements is rarely available and may be impractical to create due to time and cost considerations for reconciling semantic interoperability.[4]

ALIGNMENT BETWEEN DATA FROM TWO DIFFERENT EHRS WITHIN A DATA DOMAIN

While any EHR transition will stop collecting some data domains and begin collecting others, many core data domains are consistent across implementations. This is because the basic purpose of an EHR, to conduct documentation of care and support billing, is retained in both cases.

However, there are several challenges and difficulties in data integration, interpretation, and analysis from using data from the same domain in two EHRs together. Semantic representational variations (differences in the underlying meaning) among data collected for the same purposes by different EHR systems are very common. There is a relative lack of literature regarding the reconciliation of semantic meaning, but most of it focuses on initial data migration during an EHR-to-EHR transition, i.e., the direct ingestion of legacy data into the new EHR to provide access to historical patient data needed for clinical care delivery.[4] Automated conversions can migrate structured data relatively reliably, but there are still notable instances of discrepant data structures, data inconsistencies, and patient safety events.[4]

Using VA's migration to the Oracle Cerner EHR as an example, encounter location data was lost because the new EHR primarily uses the encounter to capture location; while the legacy EHR, while not always connecting clinical observations to encounters, does reliably assign a clinic location. Other examples include the decision to migrate data as text or documents instead of as structured data; in the VA, this was done for the migration of radiology and microbiology data. These decisions appropriately focus on supporting ease of clinical use, severely limiting the secondary use of these transformed data, and creating the need to use historic source

data in parallel with data from the new EHR in use cases (specific analytic task or query) requiring both.

The literature around the parallel use of data from different active EHRs is drawn mostly from large observational consortia, such as OHDSI, I2B2, Sentinel, and PCORNet. This is because transforming the data into an observational consortium model forces the determination of aligning data collected for the same purposes across EHRs together. These consortia also conduct data quality checks for many different processes, some of which are relevant to data alignment for concurrent use across an EHR transition.[7] In particular, the terminology mapping portion checks for coverage of a clinical vocabulary across domains.

However, controlled vocabularies used in the historic and new EHRs may be different. While the National Library of Medicine maintains an extensive library of vocabularies and crosswalks between vocabularies that are assembled from decades of work by clinical societies, library science, and informatics communities, many gaps and errors remain. For example, a challenge in evaluating medication use across the VA EHR transition was that the VA used the National Drug Code and VA Product vocabularies, and Oracle Cerner used proprietary Multum and product ID vocabularies. In the historical VA EHR, there were also mapping challenges because each facility was responsible for mapping medications to a national reference, with variable results. This resulted in a large harmonization effort needed to align data using these two vocabularies.

When using observational data across an EHR transition, it is important to leverage both technical EHR expertise and clinical and healthcare system user expertise to develop conventions in aligning the data for analytic use. Transparency in design promotes reusability and consistency for analyses using new and legacy data, but also forces specific conventions on the data. It is also important to understand the key design decisions in aligning data from two EHRs for a use case and to determine "fit for use."

CONCLUSION

In conclusion, there are several challenges and opportunities inherent in a transition between EHRs. Understanding, characterizing, and assessing the limitations and capacities of both data sources following such a transition is important. In some cases, reuse will no longer be possible because certain elements are no longer collected in the new EHR. However, opportunities and utility in data use will be possible for new data domains collected in the new EHR, although data volume and longitudinal experience may have to accumulate to make them fully viable. Perhaps most importantly, careful assessment of data utility and validation for use in specific analytical questions is important for those data that are collected both before and after an EHR transition to understand the feasibility of including both sets of data.

Corresponding Author: *Michael E. Matheny, MD, MS, MPH; , VA Informatics and Computing Infrastructure, Tennessee Valley Healthcare System VA, Nashville, TN, USA (e-mail: Michael.Matheny@va.gov).*

REFERENCES

1. **Nordo AH, Levaux HP, Becnel LB, et al.** Use of EHRs data for clinical research: Historical progress and current applications. *Learning Health Systems.* 2019;3(1):e10076.
2. National Trends in Hospital and Physician Adoption of Electronic Health Records. *Health IT Quick Stat #61* 2022; https://www.healthit.gov/data/quickstats/national-trends-hospital-and-physician-adoption-electronic-health-records. Accessed 05/01/2023, 2023.
3. **Sittig DF, Lakhani P, Singh H.** Applying requisite imagination to safeguard electronic health record transitions. *Journal of the American Medical Informatics Association : JAMIA.* 2022;29(5):1014-1018.
4. **Huang C, Koppel R, McGreevey JD, 3rd, Craven CK, Schreiber R.** Transitions from One Electronic Health Record to Another: Challenges, Pitfalls, and Recommendations. *Applied clinical informatics.* 2020;11(5):742-754.
5. **Hersh WR, Weiner MG, Embi PJ, et al.** Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care.* 2013;51(8 Suppl 3):S30-37.
6. **Kahn MG, Callahan TJ, Barnard J, et al.** A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC).* 2016;4(1):1244.
7. **Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR.** Increasing trust in real-world evidence through evaluation of observational data quality. *Journal of the American Medical Informatics Association.* 2021;28(10):2251-2257.
8. **Moy AJ, Hobensack M, Marshall K, et al.** Understanding the perceived role of electronic health records and workflow fragmentation on clinician documentation burden in emergency departments. *Journal of the American Medical Informatics Association.* 2023;30(5):797-808.
9. VA Office of the Inspector General. The New Electronic Health Record's Unknown Queue Caused Multiple Events of Patient Harm. 2022; https://www.va.gov/oig/pubs/VAOIG-22-01137-204.pdf.
10. **Diaz-Garelli F, Strowd R, Lawson VL, et al.** Workflow Differences Affect Data Accuracy in Oncologic EHRs: A First Step Toward Detangling the Diagnosis Data Babel. *JCO Clinical Cancer Informatics.* 2020(4):529-538.