Lowering the OMOP ETL Barrier for Clinical Registries

First name	Last name	Affiliation	E-mail address
Smith	Heavner	Critical Path Institute, Tucson, AZ	Sheavner@c-path.org
Trayson	Llano	Prisma Health, Greenville, SC	Trayson.Llano@gmail.com
Zachary	Wang	Johns Hopkins University School of Medicine, Baltimore, MD	zwang380@jh.edu
Marco	Schito	Critical Path Institute, Tucson, AZ	Mschito@c-path.org
Heather	Stone	Food and Drug Administration	Heather.Stone@fda.hhs.gov
Pam	Dasher	Critical Path Institute, Tucson, AZ	Pdasher@c-path.org
Tresha	Russel	Society for Critical Care Medicine, Mount Prospect, IL	Trussel@sccm.org
Vishakha	Kumar	Society for Critical Care Medicine, Mount Prospect, IL	Vkumar@sccm.org
Ben	Saeks	Johns Hopkins University School of Medicine, Baltimore, MD	Bsaeks1@jh.edu
Michael	Cooke	Johns Hopkins University School of Medicine, Baltimore, MD	Michael.cook@jhu.edu
Rahul	Kashyap	Society for Critical Care Medicine, Mount Prospect, IL	Kashyapmd@gmail.com
Matthew	Robinson	Johns Hopkins University School of Medicine, Baltimore, MD	mrobin85@jhmi.edu
Paul	Nagy	Johns Hopkins University School of Medicine, Baltimore, MD	nagy@ohdsi.org

Abstract (701 words; Max 1,000)

Background

Clinical registries of real-world medical records play an important role in biomedical discovery of effective disease treatments. Registries aggregate patient level information across multiple healthcare delivery systems as well as define standardized data collection to answer critical questions. Systematic collection of real-world data through registries may offer early insight into the impact of repurposed drugs for diseases without clear treatment options. The challenge for clinical registries is that data collection requires clinical providers to conduct time intensive chart reviews to manually abstract this data.[1]

The CURE-ID project seeks to demonstrate automated extraction of relevant data from electronic medical records by combining the Observational Medical Outcomes Partnership (OMOP) common data model with registry specific clinical concepts. The OMOP common data model (CDM) enables this automation by providing conventions on bringing in the major components of the medical record such as medications, procedures, conditions, and laboratory measurements. The OMOP CDM also provides the flexibility to add clinical concepts that are typically not coded within the medical records such as custom data forms for respiratory management.

The goals of the project were to lower the costs of a health institution in transforming local medical records into the OMOP CDM as part of a clinical registry. Performance of an extract, transform, load (ETL)

is a time and resource consuming exercise which has traditionally created a high barrier to entry into the Observational Health Data Sciences and Informatics (OHDSI) community. The initial ETL creation at Johns Hopkins required >2000 person-hours of developer time to map the Epic Systems (Verona, Wi) custom clarity data repository into OMOP. We sought to simplify and standardize this ETL process, in order to facilitate participation of a wider-range of healthcare institutions in clinical registries.

Methods

This work builds on experiences with the Society for Critical Care Medicine (SCCM) Discovery VIRUS COVID-19 Registry,[2] specifically, efforts to automate the extraction of data from the electronic health record (EHR). These efforts were heterogeneous with each participating site developing methods and sharing lessons learned with the collaborative.[3] The CURE ID Platform – an FDA and NCATS/NIH collaboration, along with the associated public-private partnership, the CURE Drug Repurposing Collaboratory (CDRC), with funding from the HHS Assistant Secretary for Planning and Evaluation's (ASPE) Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (OS-PCORtf), sought to disseminate harmonized methods developed by Johns Hopkins University. The goal was to expand the CURE ID clinical registry from a clinician-entered case-based system to one that used automated electronic health record (EHR) extraction tools, enabled by conversion of non-common data model EHR systems to OMOP. This was intended to serve as a model for other clinical registries that could be enabled by this data mapping and transformation in the future.

The project began with a modified Delphi process including clinical and scientific experts on COVID-19 to identify minimum viable variables for registry creation and additional variables of interest. In total, the consortium identified 36 high priority variables with an additional 150 lower priority variables to be collected when possible. OMOP experts performed semantic mapping for registry variables to standard terminology. The informatics team adapted the existing ETL codebase for implementation at one alpha test site. The team collected iterative feedback to improve the process.

Results

Specific challenges included respiratory management variables, specifically in anesthesia documentation, which lack standard terminology.[4] Multiple approaches to semantic mapping were evaluated, leading the team to streamline the process for future sites. The alpha test site using the methods developed at Johns Hopkins and were able to complete the transformation into OMOP with high data quality in under 200 hours of effort (a 90% reduction in labor cost).

Conclusion

It is feasible to reduce the ETL implementation time by providing default configuration transformations along with assistance and feedback on the process. Further reduction in the person-hours required to perform an OMOP ETL will be evaluated with the Perseus web based OHDSI ETL project and cloud provider deployments of Atlas and the DQD. Our goal is to increase the adoption of OMOP in sites with fewer resources and enable wider participation in high-quality clinical registries with sufficient patient numbers and data variables to perform appropriate observational research techniques to control for potential confounders (e.g., propensity score matching).

References/Citations

1. Paris N, Lamer A, Parrot A. Transformation and Evaluation of the MIMIC Database in the OMOP Common Data Model: Development and Usability Study. JMIR Med Inform. 2021 Dec 14;9(12):e30970. doi: 10.2196/30970. PMID: 34904958; PMCID: PMC8715361.

2. Walkey AJ, Kumar VK, Harhay MO, Bolesta S, Bansal V, Gajic O, Kashyap R. The Viral Infection and Respiratory Illness Universal Study (VIRUS): An International Registry of Coronavirus 2019-Related Critical Illness. Crit Care Explor. 2020 Apr 29;2(4):e0113. doi: 10.1097/CCE.0000000000000113. PMID: 32426754; PMCID: PMC7188422.
3. Heavner-Sullivan et al. Oral presentation: Practical EHR export patways (PEEP): Rapid implementation of automated data abstraction for a global COVID-19 registry using the dynamic sustainability framework. 13th Annual Conference on the Science of Dissemination and Implementation. Section: The Implementation Journey: From Exploration to Dynamic Adaptation. Virtual. December 16, 2020. https://publons.com/publon/49095562/
4. Lamer A, Abou-Arab O, Bourgeois A, Parrot A, Popoff B, Beuscart JB, Tavernier B, Moussa MD. Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study. J Med Internet Res. 2021 Oct 29;23(10):e29259. doi: 10.2196/29259. PMID: 34714250; PMCID: PMC8590192.