Conversion of a Myositis Precision Medicine Center into a Common Data Model: A Case Study

Zachary Wang MS[1], Will Kelly[1], Paul Nagy PhD[1], Christopher A Mecoli MD MHS[1]

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore MD

Background

Idiopathic Inflammatory Myopathies (IIM) comprise a group of rare chronic autoimmune diseases affecting multiple organ systems which can lead to substantial morbidity and mortality. Single center studies are often inadequate to robustly study rare subgroups within rare diseases. Therefore, there is a large need for multi-center observational studies. However, individual centers often have center-specific data dictionaries, codebooks, and structures that impede data aggregation and validation. A solution to this problem would be to combine deep phenotyping along with a common data model (CDM), which enables data standardization and thus pooling of databases. The Observational Health Data Sciences and Informatics (OHDSI) organization maintains standards and tools for the Observational Medical Outcomes Partnership (OMOP) CDM. However, a roadmap for executing and trouble-shooting OHDSI's open-source tools in a Center of Excellence (COE) within a large academic institution is not currently available, and individual efforts to "OMOP-ify" their data may be prohibitively difficult or time-intensive.

Methods

We developed and documented a multi-step process that successfully led to our adoption of OMOP and OHDSI open-source tools that may be adopted by other CoEs. This Community Development process involves the setup, integration, and troubleshooting of the CohortDiagnostics R package within an academic institution's constricted computational environment (CrunchR), as well as interaction with administration and OHDSI community. We hope that our documentation may serve as a guideline for future implementors who wish to reproduce our efforts in creating their own OMOP instances. For our process, we used the Hopkins Myositis Clinical Registry as source data, that we then deidentified and converted into the OMOP CDM format.

Results

[1] Assemble the Team

Before we began, we recognized that several skillsets are needed to be successful in this project. Content expertise is needed in the form of clinicians caring for IIM patients, as well as expertise in bioinformatics and data science. We also were fortunate enough to identify institutional leadership/mentorship within our home institution who were familiar with CDMs, understood their potential value, and could serve as a guidepost throughout the process.

[2] Identify all concepts in our clinical and research databases that need to be mapped to OMOP standard concepts (Content coverage analysis)

We started by extracting all concepts in both our clinical and research databases. The sources for these concepts included our electronic medical record (EPIC), SmartForms contained within EPIC, a historical database contained within Microsoft Access, and a research database in SQL. A list of each concept was recorded in Microsoft Excel and annotated by subject matter experts.

[3] Map all concepts to OMOP as able via Extract-Transform-Load (ETL) and create list of all concepts that do not have an appropriate mapping.

Our home institution developed an ETL for the majority of concepts contained within EPIC to be mapped to OMOP. However, for our IIM-specific concepts, we explored open-source tools available to researchers through Observational Health Data Sciences and Informatics (OHDSI). Several of these tools, including Athena, USAGI, and Perseus, help determine the optimal mappings; that is, they facilitate mapping source concepts to standard OMOP concepts. In our experience, a large number of concepts were determined to be specific to IIM and did not have standard mappings. In addition to EMR data already mapped by our institution, 461 elements specific to our COE were reviewed and 325 were mapped to OMOP concepts. The source to concept mapping required approximately 45 hours by non-clinical personnel and adjudication by a physician took approximately 5 hours.

[4] Engage Stakeholders in the IIM Field

In parallel with steps 2-4, we engaged both OHDSI and the International Myositis Assessment & Clinical Studies Group (IMACS) groups. IMACS is a coalition of health care providers and researchers with experience and interest in the myositis syndromes. Several of our team members attended in-person OHDSI conferences and contributed to forums and Microsoft Teams standing meetings. We also engaged the IMACS leadership on ways to discuss the potential benefits and limitations of OMOP. We presented what OMOP is and what it could offer to the international IIM field, and have developed a Data Sharing Special Interest Group within IMACS with a goal of OMOP-ifying as many myositis cohorts internationally as possible as well as approaching existing research Registries.

[5] Navigate OHDSI Open-Source tools in the confines of our academic environment

Of the many OHDSI open-source tools available, we prioritized accessing CohortDiagnostics for its prevalence in the OHDSI community and quantitative feedback capabilities. We found that the CrunchR environment within our institution was not compatible with the assumed workflow of CohortDiagnostics, nor many of the OHDSI libraries CohortDiagnostics is built upon. Numerous obstacles were encountered, including a package dependency on Java, lack of support for Linux-based systems, incompatible authentication mechanisms for both MS SQL Server and WebAuth as well as many auxiliary functions that do not work within a secure, academic workspace. To address these issues of installation, authentication, and publication, we developed technical workarounds for each, which required significant time investment and collaboration with programming and security subject matter experts.

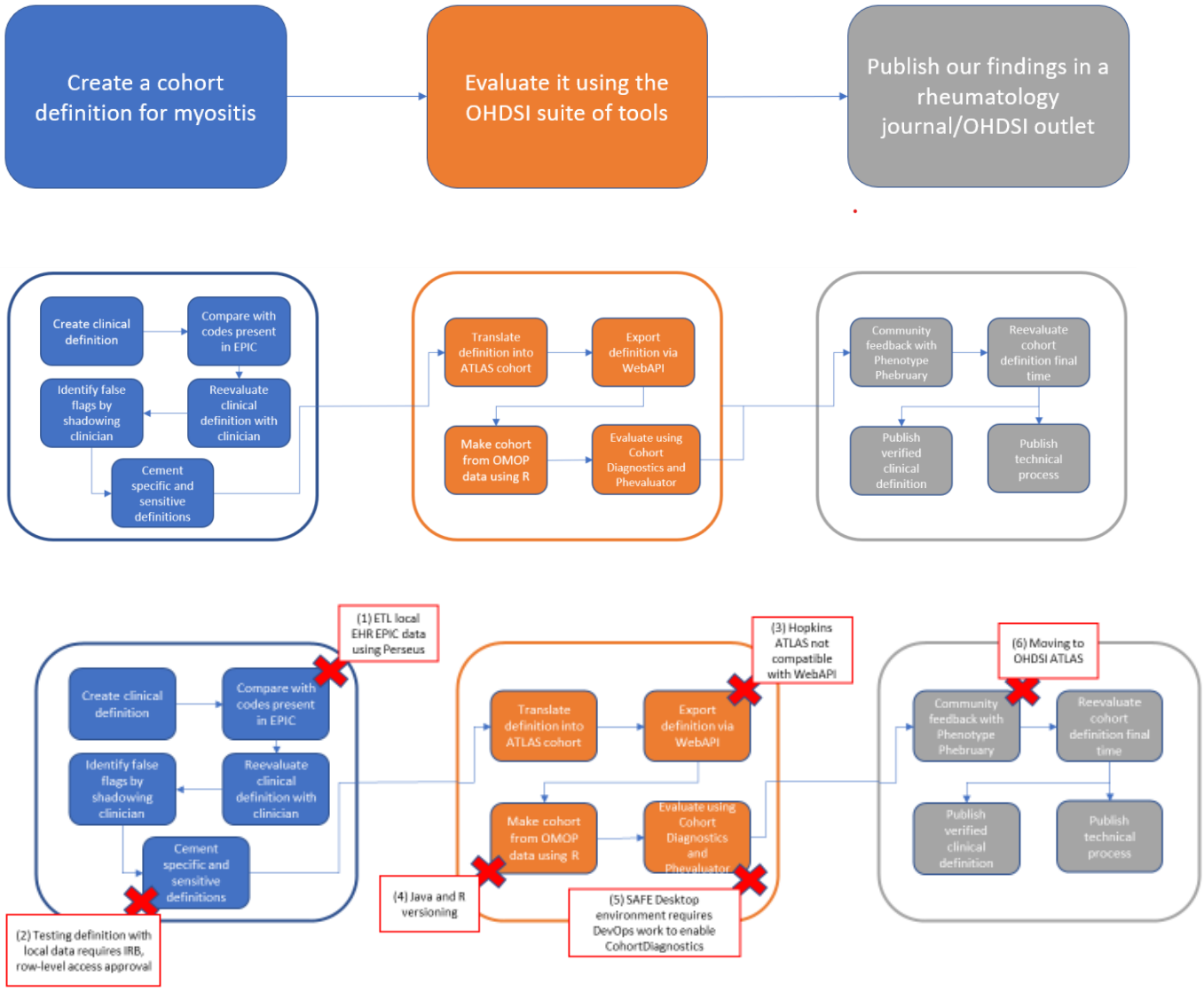
[6] Institutional Change and Iteration

After developing our solution, we engaged with leadership in our academic administration and the broader OHDSI community, resulting in several systemic improvements. CrunchR was modernized to align with the Java and R specifications of CohortDiagnostics, and the package itself was updated to support Linux environments. Our efforts have also established a direct connection between the maintainers and users of the package, leading to increased activity within the Phenotype workgroup and support from our administration on future development of the project.

Conclusion

We developed a process that successfully led to our adoption of OMOP and OHDSI open-source tools that

may be adopted by other institutions. Variations of this process and the obstacles we encountered may be applicable to other academic institutions with disease-specific registries.



Figures 1-3. As a visual demonstration of our process throughout our work, we've created three figures on the simplified, detailed, and full workflow involved in establishing our OMOP-ifying procedure. Obstacles are numerically marked: (1) Mapping IIM-specific fields required creation of Hopkins PERSEUS instance; (2) Initial myositis cohort definitions yielded zero results, despite clinicians knowing these patients existed. IRB approval needed to be sought to deidentify data and enable row-level access to troubleshoot cohort definition code; (3&6) Hopkins ATLAS instance could not be accessed by OHDSI community and was not compatible with WebAPI; JSON cohort definitions required copy/pasting from Hopkins ATLAS instance to OHDSI ATLAS instance; (4) Dependency on Java and latest version of R required appealing to institution to download; (5) Institution-specific changes were required to our CrunchR/SAFE

Desktop environment.