Phenotype modelling tools utilizing standardized EHR data in a Common Data Model format

Ravi Teja Bhupatiraju, MBBS PhD	extsuperscript{a}, Vojtech Huser, MD PhD	extsuperscript{a}, Kin-Wah Fung MD MS	extsuperscript{a}

	extsuperscript{a}Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD

Introduction
As EHR systems are widely deployed, we have access to large amounts of patient data for research purposes. Addressing the challenges of data representation, the Common Data Model, defined by Observational Medical Outcomes Partnership (OMOP), allows for the development of tools and analyses that work on multiple datasets of Electronic Health Record (EHR) data or administrative claims data. It allows for phenotype modelling across institutions and has been adopted by NIH All-Of-Us research project.

Methods
Our goal was to replicate the current technology stack of OMOP-based tools for visualization and phenotype modelling LHNCBC, namely, the data profiling tool Achilles and the data visualization tool Atlas.

As sample EHR data, we used NIH Biomedical Translational Research Information System (BTRIS) data from deceased subjects obtained via BTRIS limited data set application (with 19,195 diagnoses, 638,024 medications, and 532,184 lab results)

Results
We created an Extract-Transform-Load pipeline that transforms EHR data in phases into the OMOP CDM. Our phased approach can be used on other input data and extends existing OMOP tools for ETL, taking advantage of OMOP vocabularies to translate source data into OMOP targeted terminologies (e.g., RxNorm, LOINC and SNOMED CT). With the converted data in place, we also implemented the data validation tool and were able to execute one OMOP-based network study.

Conclusion
In conclusion, the fully implemented OMOP based technology stack allows us to pilot clinical analyses using OHDSI statistical packages and tools. In addition to BTRIS data, we hope to re-use this experience with two additional datasets (e.g., MIMIC-III data).

References