nature > npj digital medicine > articles > article

Article | Open Access | Published: 19 May 2023

Ontologizing health systems data at scale: making translational discovery a reality

Tiffany J. Callahan, Adrianne L. Stefanski, Jordan M. Wyrwa, Chenjie Zeng, Anna Ostropolets, Juan M. Banda, William A. Baumgartner Jr., Richard D. Boyce, Elena Casiraghi, Ben D. Coleman, Janine H. Collins, Sara J. Deakyne Davies, James A. Feinstein, Asiyah Y. Lin, Blake Martin, Nicolas A. Matentzoglu, Daniella Meeker, Justin Reese, Jessica Sinclair, Sanya B. Taneja, Katy E. Trinkley, Nicole A. Vasilevsky, Andrew E. Williams, Xingmin A. Zhang, Joshua C. Denny, Patrick B. Ryan, George Hripcsak, Tellen D. Bennett, Melissa A. Haendel, Peter N. Robinson, Lawrence E. Hunter & Michael G. Kahn

— Show fewer authors

npj Digital Medicine **6**, Article number: 89 (2023) | Cite this article

2081 Accesses | **21** Altmetric | Metrics

OHDSI Community Call - September 25, 2023

@Tiff_callahan

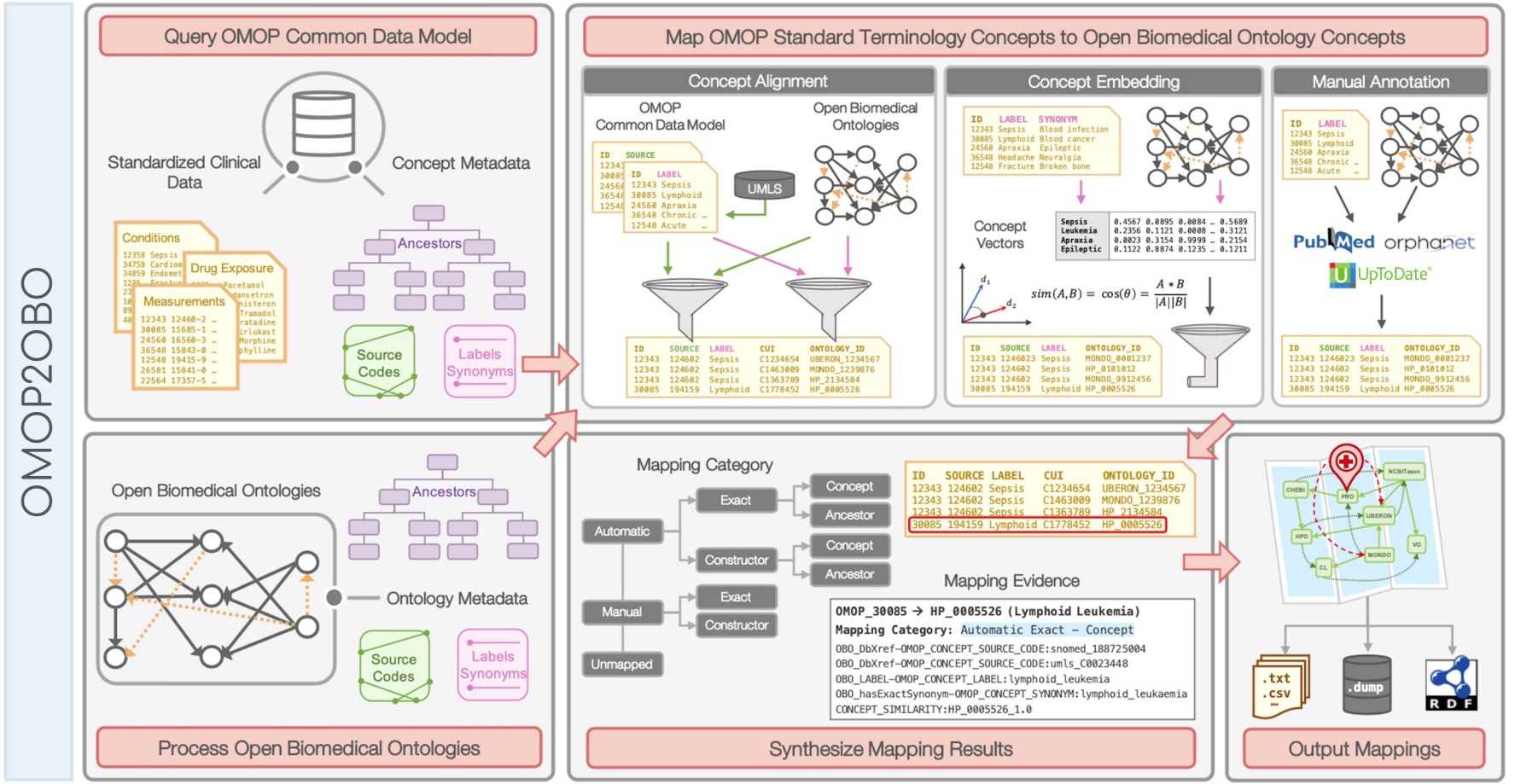
Background

- Deep phenotyping, or "the precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described"[1], requires timely synthesis of multiple types of patient data[2-3]
- Common data models solve many challenges of standardizing electronic health record (EHR) data but are unable to semantically integrate all of the resources needed for deep phenotyping[4]
- Open Biological and Biomedical Ontology (OBO) Foundry ontologies provide computable representations of biological knowledge and enables the integration of heterogeneous data[5]

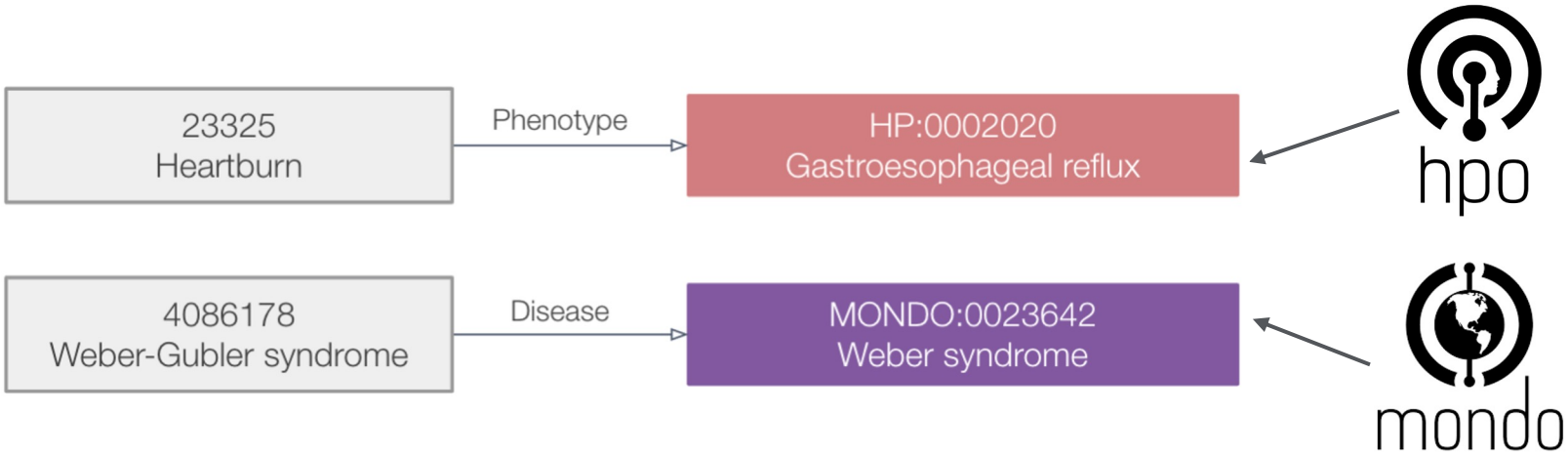
Challenge: Mapping EHR data to OBO ontologies requires significant manual curation and domain expertise

[1]PMID22504886; [2]PMID:26536218; [3]PMID:32335224; [4]PMID:24071798; [5]PMID:25863278

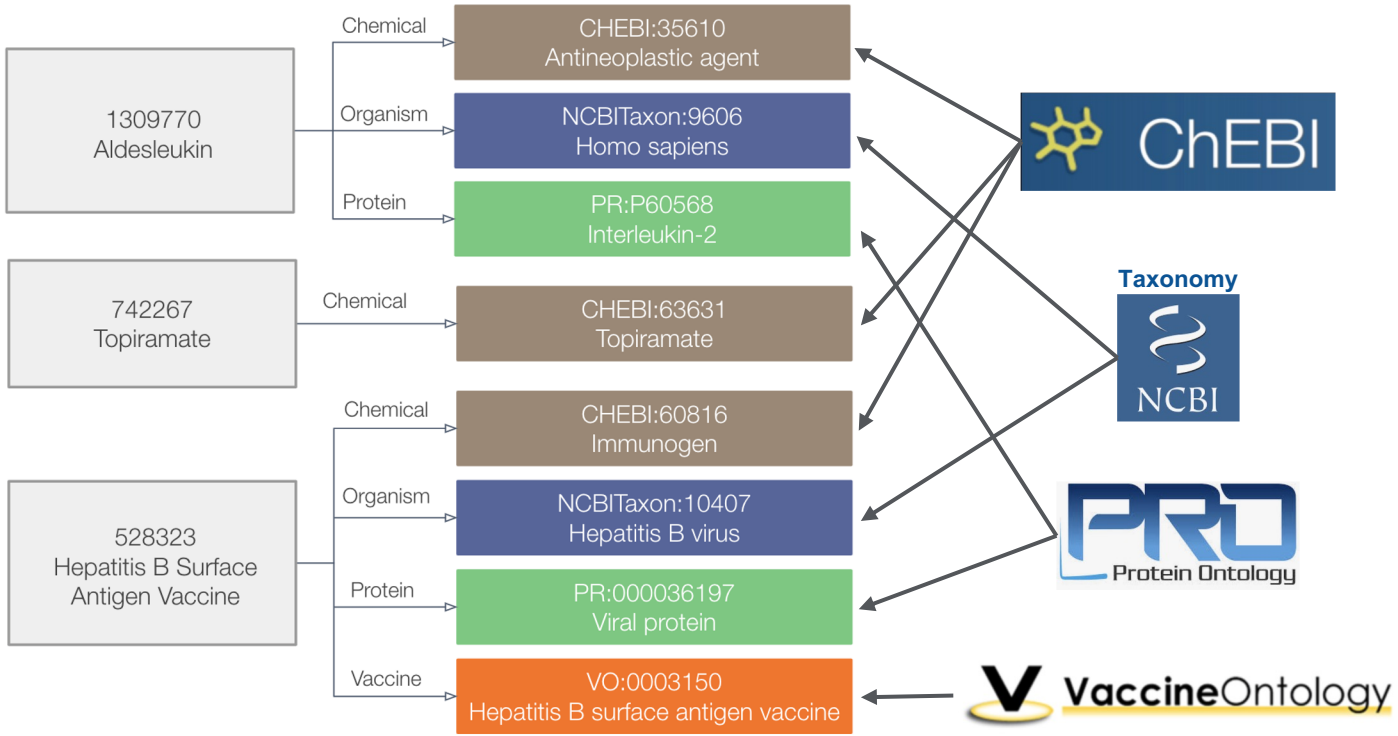
Objective: Develop an algorithm that generates clinically meaningful and biologically relevant mappings between standard OMOP vocabularies and Open Biomedical Ontologies



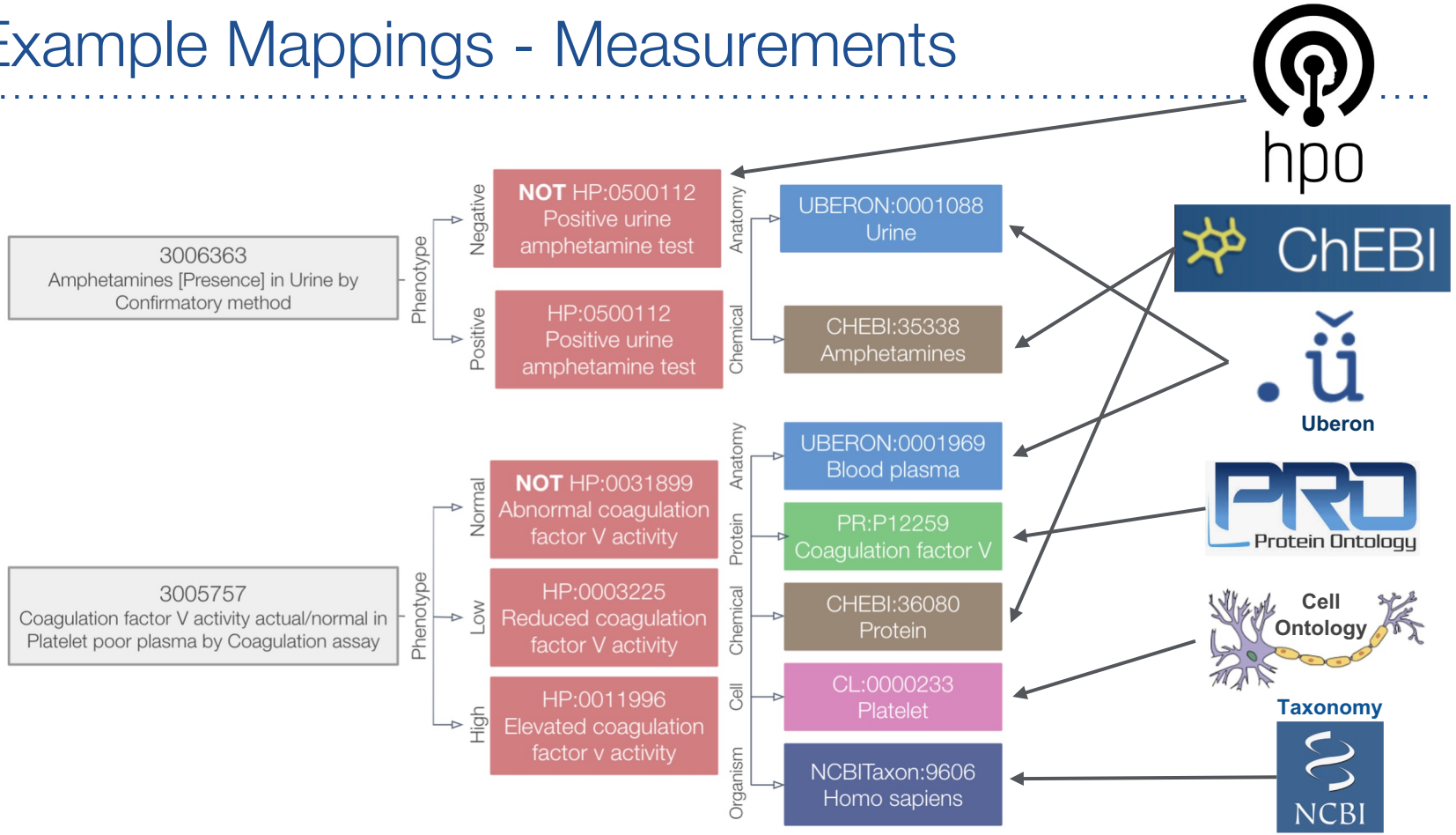
Example Mappings - Condition Occurrences

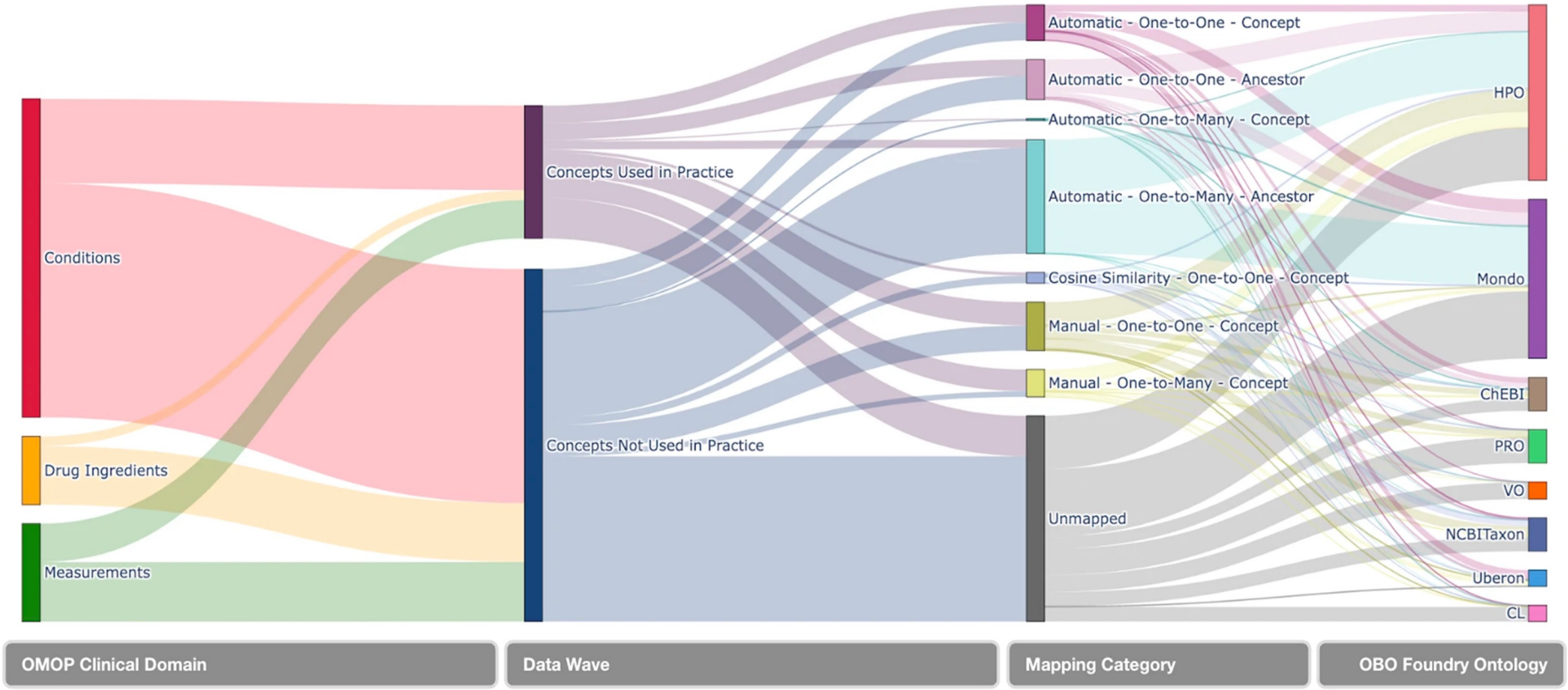


Example Mappings - Drug Exposure Ingredients



Example Mappings - Measurements





Using OMOP2OBO, we built mappings for 92,367 conditions, 8615 drug ingredients, and 10,673 measurement results to ontology concepts representing **9636** diseases, **6309** phenotypes, **83** anatomical entities, **2704** organisms, **4261** chemicals, **132** vaccines, and **272** proteins

Evaluation

Accuracy

10 domain experts manually reviewed 20% of the most challenging manual OMOP2OBO mappings

- Correct Mappings: 73.9% (condition), 70.7% (drug ingredient), and 92.9% (measurements)

Generalizability

OMOP2OBO mapping concept coverage in 24 Independent hospitals (OHDSI Concept Prevalence Study)

- Coverage: 99.5% of conditions, 99.9% of drug ingredients, and 68% of measurement results

Clinical Utility

Compare OMOP2OBO mappings to validated manual mappings when used to identify patients with rare genetic diseases using AoU data

- Query Performance: OMOP2OBO mappings identified 99.3% of the patients identified by the validated manual mappings using fewer codes and one-third of the query time

Acknowledgements

Tri-Advisory


Dr. Michael G. Kahn


Dr. Lawrence E. Hunter


Dr. Tellen D. Bennett

Translational Research Experts


Adrianne L. Stefanski, PhD


Peter Robinson, MD


Nicole Vasilevsky, PhD


Melissa Haendel, PhD

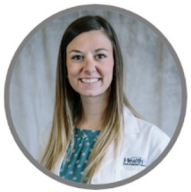
Clinical Experts


James Feinstein, MD


Blake Martin, MD


Jordan Wyrwa, DO


Katy Trinkley, PharmD


Jessica Sinclair, PharmD

OHDSI / OMOP Experts


Patrick Ryan, PhD


Anna Ostropolets, MD


Andrew Williams, PhD

Organizations

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Children's Hospital Colorado

HEALTH DATA Compass

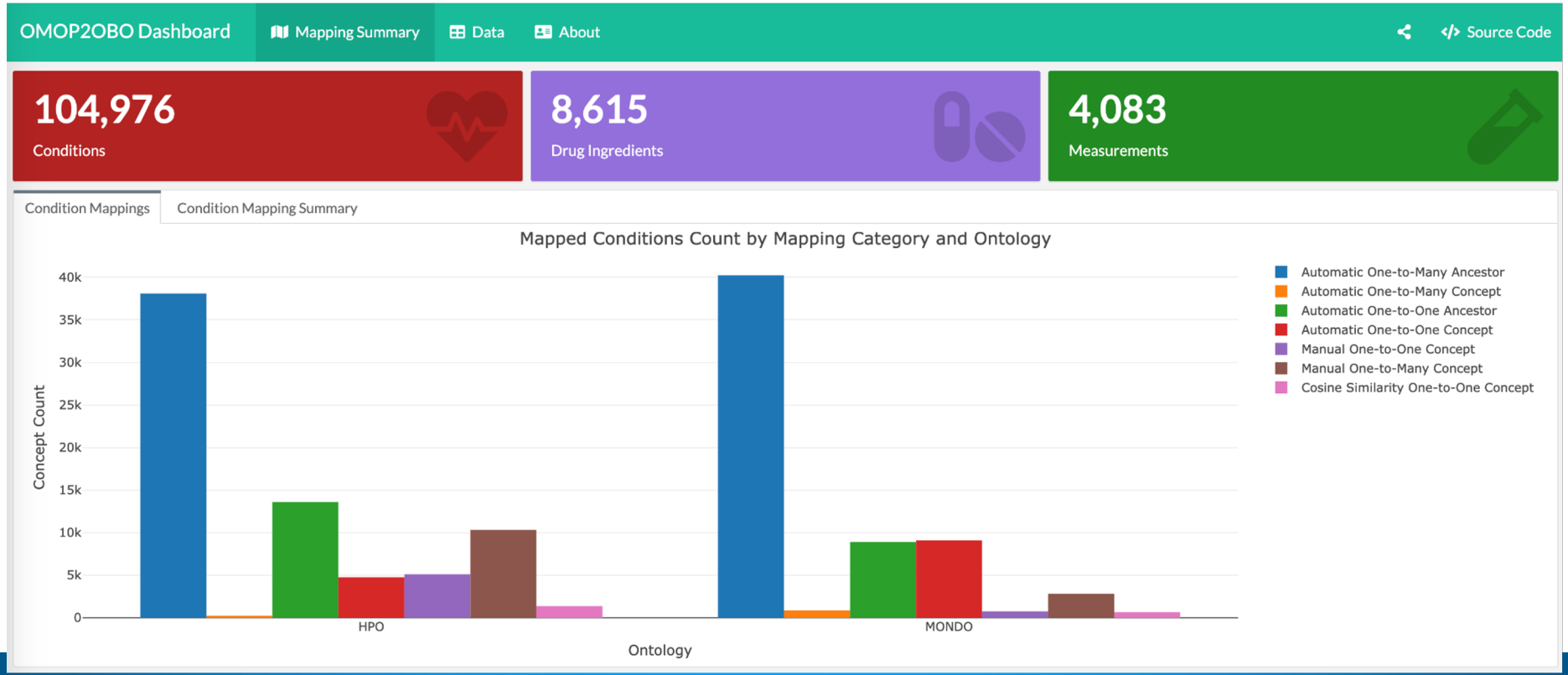
PEDSnet
A Pediatric Learning Health System

NIH U.S. National Library of Medicine

Funding: T15LM009451 and T15LM007079

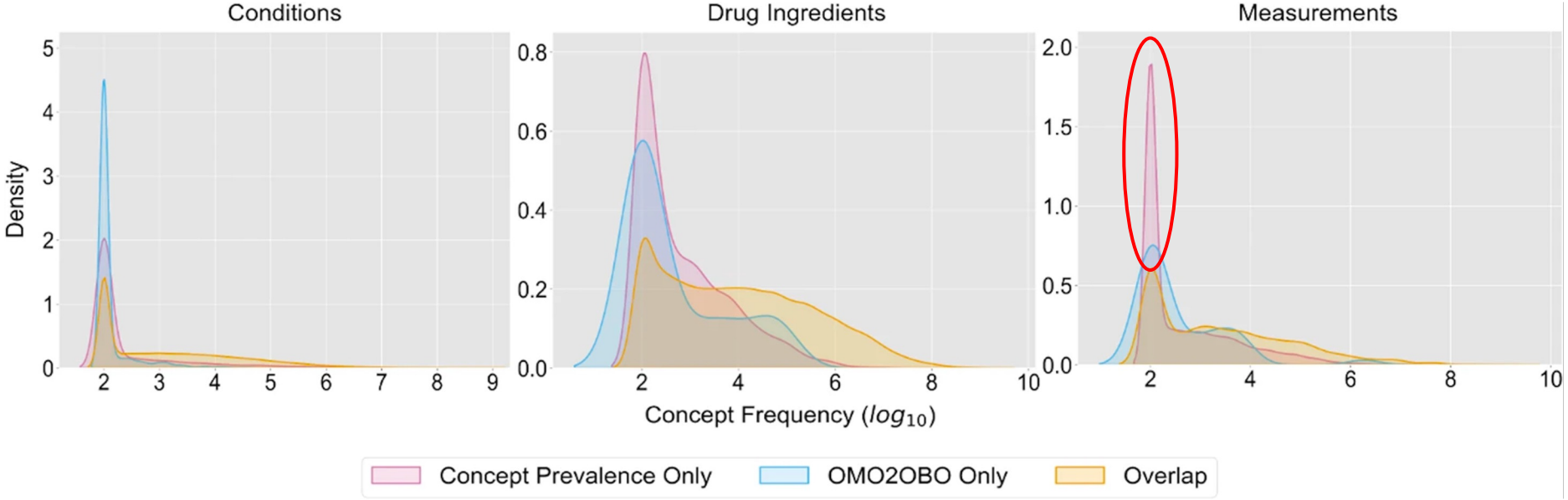
Dashboard

Conditions: https://doi.org/10.5281/zenodo.6774363
Drug Ingredients: https://doi.org/10.5281/zenodo.6774401
Measurements: https://doi.org/10.5281/zenodo.6774443



http://tiffanycallahan.com/OMOP2OBO_Dashboard

Evaluation


Conditions
Drug Ingredients
Measurements
Density
Concept Frequency (log_{10})
Concept Prevalence Only
OMO2OBO Only
Overlap