OMOP2OBO: Semantic Integration of Standardized Clinical Terminologies to Power Translational Digital Medicine Across Health Systems

PRESENTER: Tiffany J. Callahan

Background
Despite significant progress in biobanking, translational use of electronic health records (EHRs) remains largely aspirational due to its disconnect from biomedical knowledge. Ontological Biomedical Ontologies (OBOs) provide detailed representations of biological domains, are logically verifiable using description logics, and can be easily integrated with basic science data and clinical research (Figure 1).

MAPPING CHALLENGES
- Limitations of existing work in this domain:
  - Focused on specific diseases and biological domains
  - Largely limited to one-to-one mappings
  - Rarely include external validation
- Existing algorithms cannot automatically capture complex biomedical semantics underlying clinical concepts

GOAL: Develop OMOP2OBO, the first health system-wide integration and alignment between OMOP standardized clinical terminologies and OBO ontologies.

Methods
- OMOP-normalized Children’s Hospital Colorado EHR data.
- OBOs were selected by domain experts and included diseases, phenotypes, anatomical entities, cell types, organisms, small molecules, vaccines, and proteins.
- Mappings were performed using the pipeline in Figure 2.
- 20% of the most challenging mappings were verified by a panel of clinical and molecular domain experts.
- Mapping generalizability was assessed by comparing the coverage of mapped concepts to 9 independent EHRs.

Results
- 20,850 condition concepts were mapped to 4,661 phenotypes and 3,614 diseases (Figure 3).
- 1,574 of drug ingredient concepts were mapped to 1,422 chemicals, 91 proteins, 39 organ systems, and 54 vaccines.
- 11,072 measurement results matched to 920 phenotypes, 25 anatomical entities, 27 cell types, 338 chemicals, 194 organisms, and 113 proteins.

VALIDATION
- Domain expert agreement was found for 91.6% of measurements, 75.0% of ingredients, and 73.8% of conditions.
- 90-95% for conditions, 91-96% for ingredients, and 50-55% for measurement concepts on EHR from two independent health systems revealed.

Discussion
OMOP2OBO is the first health system-wide integration of OMOP clinical terminology concepts and OBO biomedical ontologies.

FUTURE WORK
We are currently working on expanding the mapping provenance to include mechanisms of actions and conducting an expanded coverage study, using data from the OHDSI Concept Prevalence Study.