OMOP CDM Oncology Module at Work

Rimma Belenkaya, Michael Gurley, Christian Reich, Dmitry Dymshyts, Jeremy Warner, Robert Miller, Andrew Williams, RuiJun Chen

Background

Observational research in cancer requires substantially more detail to represent diagnoses, treatments, and outcomes than most other therapeutic areas. Cancer diagnosis is defined by a constellation of histology, site, stage, grade, genetic biomarkers. Cancer treatments are administered in defined order and cycles, and cannot be fully described by individual medications. At the same time, clinically and analytically relevant representation of cancer diagnoses, treatments, and outcomes requires data abstraction (e.g. recurrence, remission, end-of-life events, chemotherapy regimens, treatment cycles, response to treatments) that is not readily available in the source data and has not been traditionally supported in OMOP CDM. Here, we introduce a new Cancer Module of the OMOP CDM, which allows for both the required granularity and abstraction of cancer data to support transformation from the source data and standardized analytics. We tested the Module in EHR and Cancer Registry data against a number of typical use cases.

Methods

Cancer diagnosis as a complex model

- **Cancer Diagnosis Model in the OMOP Vocabulary**
  -diagnosis: episode_concept_id
  - HTO to EHR → Diagnostic Encounter
  -from
  -OMOP CDM
  -example: Cancer Diagnosis Model in the OMOP Vocabulary

  ![Diagram of Cancer Diagnosis Model in the OMOP Vocabulary](image)

- **Disease and treatment abstraction**
  - example: Disease and treatment episodes in the OMOP CDM
  -New Episode model added to the OMOP CDM
  -Episodes are represented by episode_concept_id and episode_object_concept_id
  -Treatment regimen of Paciﬁcal and Bevacizumab
  -Episodes and lower level events (e.g. conditions, drugs) are linked via Episode_Event table

  ![Diagram of Disease and treatment episodes in the OMOP CDM](image)

- **Database instantiation and ETL**
  - Developed **vocabulary-driven ETL** for data conversion from Tumor Registry
  - Converted EHR and Registry data into OMOP CDM

  ![Diagram of Database instantiation and ETL](image)

- **New vocabularies**
  - ICD-O-3 Site and Site were precoordinated and mapped to SNOMED
  - Unmapped concepts were designated as standard

  ![Diagram of New vocabularies](image)

Results

We converted EHR and Registry data from four participating institutions using uniform vocabulary-driven ETL. We derived First Cancer Occurrence and First Treatment Course using NAACCR and HemOnc vocabularies. We achieved 95% of coverage for the diagnoses reported in the source data by the Standardized Vocabularies, the remaining 5% representing rare cancers. We tested the following clinical characterization use cases:

1. Survival from initial diagnosis
2. Time from diagnosis to treatment
3. High-level treatment courses for 1st cancer occurrence
4. Derivation of chemotherapy regimens from atomic drugs

Conclusions

We incorporated foundational structural and semantic support into the OMOP CDM to represent clinical cancer disease and treatment data. This significantly improves specificity of cancer cohort definitions. Introduction of disease and treatment abstractions supports key clinical characterization use cases.

Future work on the Oncology Module will include:

- Adding domains for genomics, imaging and outcomes
- Improving ICD-O-3 to SNOMED mapping precision
- Mapping of NAACCR data dictionary concepts to SNOMED, using Nebraska Lexicon
- Improving precision of chemo regimen identification.