OMOP CDM Oncology Module at Work

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

Normal Condition
Most normal conditions are defined by three main dimensions implicitly, plus some extra attributes

- Granulomatous infection
- Morphology
- Topology (site)
- Cause
- Mycobacterium tuberculosis

Cancer
- Cause is not known, but morphology and topology are detailed and explicit
- The many tumor attributes (modifiers) are also explicit and well defined

- Carcinoma, NOS
  - Morphology
  - Topology (site)
  - Stage: IIB, T2-N1-M0
  - Grade: High
  - Size: 45 mm
  - Effected lymph nodes: None
  - HER2 status: Positive

- Breast, NOS
  - Morphology
  - Topology (site)
  - Stage: IIB, T2-N1-M0
  - Grade: High
  - Size: 45 mm
  - Effected lymph nodes: None
  - HER2 status: Positive
Clinically and analytically relevant representation of cancer diagnoses, treatments, and outcomes requires data abstraction.

- Not readily available in the source data
- Traditionally not supported in OMOP CDM
Solving Granularity Challenge

Cancer Diagnosis Model in the OMOP Vocabulary

Added vocabularies:

- Carcinoma of Breast, NOS 8010/3-C50.9
- Carcinoma, NOS
- Breast, NOS

ICD-O

Grade

Grade I

Grade II

T-Cell

Tumor Size

Numeric

mm

001-988

has value

has type

has units

has range

has lower value

has upper value

schema to modifier

condition to schema

has morphology

has anatomic site
Solving Granularity Challenge

Cancer diagnosis representation in the OMOP CDM

- **Precoordinated concept** of cancer Morphology + Site is stored in Condition_Occurrence
- **Diagnostic modifiers** are stored in Measurement and linked to the Condition_Occurrence record
Cancer diagnosis representation in the OMOP CDM

- Precoordinated concept of cancer Morphology + Site is stored in Condition_Occurrence
- Diagnostic modifiers are stored in Measurement and linked to the Condition_Occurrence record

Example of cancer diagnosis in the OMOP CDM

**Histology+Site** diagnosis in Condition_Occurrence

- condition_occurrence_id: 123456789
- person_id: 1
- condition_concept_id: 4116071
- condition_start_datetime: June 9, 2019
- condition_type_concept_id: 32535
- condition_source_value: 8010/3-C50.9
- condition_source_concept_id: 44505310

**Grade** modifier in Measurement

- measurement_id: 567890
- person_id: 1
- measurement_datetime: June 9, 2019
- measurement_concept_id: 35918640
- value_as_concept_id: 35922509
- measurement_type_concept_id: 32534
- measurement_source_value: 3844
- measurement_source_concept_id: 35918640
- value_source_value: breast@3844@3
- modifier_of_event_id: 123456789
- modifier_field_concept_id: 1147127

SNOMED concept ‘Carcinoma of breast’

Precoordinated concept of ICD-O Histology & Site

NAACCR concept ‘Grade Pathological’

NAACCR concept ‘G3: High combined histologic grade (unfavorable); SBR score of 8-9 points’

OMOP concept ‘Tumor registry’

NAACCR code for ‘Grade Pathological’

NAACCR concept ‘Grade Pathological’

NAACCR code for ‘G3: High combined histologic grade (unfavorable); SBR score of 8-9 points’

Value of the respective condition record condition_occurrence_id

Concept for ‘condition_occurrence,condition_occurrence_id’
Solving Abstraction Challenge

Disease and treatment episodes in the OMOP CDM

Added vocabularies:
Disease and treatment episodes in the OMOP CDM

Example of disease and treatment episodes in the Episode table

**'First occurrence'**-of-**'Carcinoma of breast'**

<table>
<thead>
<tr>
<th>Episode</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>episode_id</td>
<td>12345</td>
</tr>
<tr>
<td>person_id</td>
<td>1</td>
</tr>
<tr>
<td>episode_concept_id</td>
<td>32528</td>
</tr>
<tr>
<td>episode_start_datetime</td>
<td>June 9, 2019</td>
</tr>
<tr>
<td>episode_object_concept_id</td>
<td>4116071</td>
</tr>
<tr>
<td>episode_type_concept_id</td>
<td>32535</td>
</tr>
</tbody>
</table>

**OMOP concept**
- First disease occurrence
- SNOMED concept
- Carcinoma of breast

**OMOP concept** Tumor registry

**Added vocabularies:**

**'Treatment regimen'**-of-**'Carboplatin and Paclitaxel'**

<table>
<thead>
<tr>
<th>Episode</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>episode_id</td>
<td>12346</td>
</tr>
<tr>
<td>person_id</td>
<td>1</td>
</tr>
<tr>
<td>episode_concept_id</td>
<td>32531</td>
</tr>
<tr>
<td>episode_start_datetime</td>
<td>July 9, 2019</td>
</tr>
<tr>
<td>episode_parent_id</td>
<td>12345</td>
</tr>
<tr>
<td>episode_object_concept_id</td>
<td>35804255</td>
</tr>
<tr>
<td>episode_type_concept_id</td>
<td>32545</td>
</tr>
</tbody>
</table>

**OMOP concept**
- Treatment Regimen
- Foreign key to the disease Episode record

**HemOnC concept**
- Carboplatin and Paclitaxel

**OMOP concept**
- Episode algorithmically derived from EHR
Testing

• Developed **vocabulary-driven ETL** for data conversion from Tumor Registry

• **Converted EHR** and **Tumor Registry data** from four participating institutions

• Tested **clinical characterization use cases**
  – Survival from initial diagnosis
  – Time from diagnosis to treatment
  – High-level treatment course for 1\textsuperscript{st} cancer occurrence
  – Derivation of chemotherapy regimens from atomic drugs
Survival from diagnosis

Metastatic vs. Non-metastatic cancers

Six most prevalent cancers
Join the Effort!

- **CDM and Vocabulary Work**
  - Adding domains for genomics, imaging and outcomes
  - Improving ICD-O-3 to SNOMED mapping precision
  - Mapping of NAACCR data dictionary to SNOMED

- **Oncology-specific THEMIS conventions**

- **ETL**
  - Validation
  - Conventions and algorithms for fusing tumor registry and EHR data on the same patient

- **Use-case-driven algorithms for**
  - identifying & characterizing cancer populations
  - identifying treatment pathways and disease progression
  - predicting disease progression

belenkar@mskcc.org
m-gurley@northwestern.edu
Many thanks to

Charles Bailey, Children’s Hospital of Philadelphia, US
Scott Campbell, University of Nebraska, US
Rachel Chee, IQVIA, Great Britain
Mark Danese, Outcome Insights, US
Asieh Golozar, Regeneron, US
George Hripcsak, Columbia University, US
Ben May, Columbia University, US
Maxim Moinat, The Hyve, Netherlands
Anna Ostropolets, Columbia University, US
Meera Patel, MSK, US
Joseph Plasek, Aurora, US
Gurvaneet Randhawa, NCI, US
Donna Rivera, NIH, US
Mitra Rocca, FDA, US
Anastasios Siapos, IQVIA, Great Britain
Firas Wehbe, Northwestern University, US
Seng Chan You, Ajou University School of Medicine, Korea
Thank you!
Thank you!