OHDSI Collaborator Meeting

Oncology WG Presentation

12/3/2019
• Introduction to the Oncology WG (Christian)

• What’s Been Accomplished (Rimma)

• Next Steps (Michael/Meera/Dima)

• Community Engagement in Development & Research (Andrew)
Oncology WG Core Team

Michael Gurley
Christian Reich
Dmitry Dymshyts
Robert Miller

Jeremy Warner
Andrew Williams
RuiJun Chen
Rimma Belenkaya
Contributors

Charles Bailey, Children’s Hospital of Philadelphia
Scott Campbell, University of Nebraska
Rachel Chee, IQVIA
Mark Danese, Outcome Insights
Asieh Golozar, Regeneron
George Hripcsak, Columbia University
Ben May, Columbia University
Maxim Moinat, The Hyve
Anna Ostropolets, Columbia University
Meera Patel, MSK
Joseph Plasek, Aurora
Gurvaneet Randhawa, NCI
Mitra Rocca, FDA
Anastasios Siapos, IQVIA
Firas Wehbe, Northwestern University
Seng Chan You, Ajou University School of Medicine, Suwon, Korea
Data Standardization to OMOP Enables Systematic Research

**Traditional way**

- Analytical method: Adherence to Drug

One SAS or R script for each study

- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

**OHDSI approach**

- OMOP CDM
- OHDSI Tools

- Adherence
- Mortality
- Prediction
- Safety Signals

North America Southeast Asia China Europe UK Japan India So Africa Switzerland Italy Israel
Cancer Research is different from other diseases

It needs more detail:

“What is the overall survival for patients with non-metastatic carcinoma of the neck of bladder in remission after first line of gemcitabin-containing chemotherapy?“

Concepts in this research question currently not standardized:

<table>
<thead>
<tr>
<th>Concept</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carcinoma</td>
<td>Histology</td>
</tr>
<tr>
<td>Neck of bladder</td>
<td>Anatomical site</td>
</tr>
<tr>
<td>Non-metastatic disease</td>
<td>Tumor attribute</td>
</tr>
<tr>
<td>Disease in remission</td>
<td>Condition Episode</td>
</tr>
<tr>
<td>First line treatment</td>
<td>Treatment Episode</td>
</tr>
<tr>
<td>Chemotherapy regimen</td>
<td>Regimen</td>
</tr>
<tr>
<td>Gemcitabin</td>
<td>Component of regimen</td>
</tr>
</tbody>
</table>
Five Goals

1. Build standards on top of OMOP
   – Vocabularies Oncology Module
   – Data model

2. Create algorithms and heuristics
   – Infer Disease Episodes (automatic abstraction)
   – Infer chemo regimens

3. Build network of data nodes

4. Build network of researchers

5. Do research
## Working Group Detail

### Participants
- OHDSI
- Ajou University
- AstraZeneca
- Center for Surgical Science, Region Sjaelland
- Children’s Hospital of Pennsylvania
- Columbia University
- Digital China Health
- Integraal Kankercentrum Nederland
- IQVIA
- Memorial Sloan Kettering Cancer Center
- Merck
- Montefiore
- Mount Sinai
- Multiple Myeloma Foundation
- NIH
- Northwestern University
- Odysseus
- Oncology Analytics
- Pittsburgh University
- Providence Health
- Vanderbilt

### Subgroups
- Leadership
- Outreach/Research
- Development
- CDM/Vocabulary
- Genomic

### Vocabularies implemented/under Consideration
- ICD-O-3
- NAACCR
- CAP
- IMO
- HemOnc
- OROT
OHDSI Oncology Working Group

MISSION: Extend OMOP CDM/Vocabulary and OHDSI analytic platform to support observational cancer research.

* Documentation
* Participants
* Forum discussions
* Data Repository
* Outreach Repository

Oncology Subgroups

(1) Outreach/Research Subgroup Meeting

* Every 1st and 3rd Tuesday of the month, 10PM EST. Next meeting 12/17.

* Meeting Information

(2) Development Subgroup Meeting

* Every Wednesday, 10 am ET. Next meeting 12/27.

* Meeting Information

(3) CDM/Vocabulary Subgroup Meeting

* Every Thursday, 10 am ET. Next meeting 12/5.

* Meeting Information

(4) Genomic Subgroup Meeting

* Every Friday, 9 am ET. Next meeting 12/6.

* Meeting Information

(5) Leadership Subgroup Meeting
Use Cases

- **Survival**
  - Overall
  - Disease-free
  - Symptom-free
  - From diagnosis
  - From treatment

- **Time**
  - From diagnosis to treatment
  - From screening to diagnosis
  - From symptoms/initial primary care visit to diagnosis

- **Variations in outcomes of bladder cancer with and w/o liver metastases**
- **Define uptake of genomic test**

- **Identify treatment regimens**
- **Compare tumor registry chemo with identified chemo regimens**
- **Validate identified chemo regimens against Beacon**
- **Compare uptake of newer medications vs. older medications**
- **Number of medications taken daily by a cancer patient**
- **Speed of drug administrations and the risk of allergic reaction/rejection**
- **Time of administration**
- **Comparative effectiveness of adhering to the administration rules vs deviations**
- **Metastatic hormone–sensitive prostate cancer and non-metastatic castration-resistant pros**
What’s Been Accomplished

• Extension of CDM and Vocabulary to support required granularity of cancer representation
  – Incorporation of ICD-O into vocabulary
  – Incorporation of NAACCR into vocabulary
  – CDM support for cancer modifiers

• Extension CDM and Vocabulary to support abstractions required for cancer representation
  – Incorporation of HemOnc into vocabulary
  – Development of the Episode CDM module

• Development of ETL from US Tumor Registries to OMOP

• Testing typical use cases
Challenges: Granularity

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

- Granulomatous infection
- Lung

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

CONDITION
- Mycobacterium tuberculosis
- IIB: T2-N1-M0
- Carcinoma, NOS
- Breast, NOS
- G3: High
- 45 mm
- 4
- None
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

NAACCR

Grade

Tumor Size

Grade I

Grade II

T-Cell

has type

Numeric

has units

mm

has range

001-988

has lower value

001

has upper value

988
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

<table>
<thead>
<tr>
<th>condition occurrence id</th>
<th>123456789</th>
</tr>
</thead>
<tbody>
<tr>
<td>person_id</td>
<td>1</td>
</tr>
<tr>
<td>condition_concept_id</td>
<td>4116071</td>
</tr>
<tr>
<td>condition_start_datetime</td>
<td>June 9, 2019</td>
</tr>
<tr>
<td>condition_type_concept_id</td>
<td>32535</td>
</tr>
<tr>
<td>condition_source_value</td>
<td>8010/3-C50.9</td>
</tr>
<tr>
<td>condition_source_concept_id</td>
<td>44505310</td>
</tr>
</tbody>
</table>

**Grade** modifier in Measurement

<table>
<thead>
<tr>
<th>measurement_concept_id</th>
<th>35918640</th>
</tr>
</thead>
<tbody>
<tr>
<td>measurement_datetime</td>
<td>June 9, 2019</td>
</tr>
<tr>
<td>value_concept_id</td>
<td>35922509</td>
</tr>
<tr>
<td>measurement_concept_id</td>
<td>32624</td>
</tr>
<tr>
<td>measurement_source_concept_id</td>
<td>35918640</td>
</tr>
<tr>
<td>measurement_date</td>
<td>June 9, 2019</td>
</tr>
<tr>
<td>value_source_concept_id</td>
<td>35922509</td>
</tr>
<tr>
<td>modifier_of_event_id</td>
<td>123456789</td>
</tr>
<tr>
<td>modifier_field_concept_id</td>
<td>1147127</td>
</tr>
</tbody>
</table>

Solving Granularity Challenge
Challenges: Abstraction

- 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 Abstraction Challenge

Disease and treatment episodes in the OMOP CDM

- **EPISODE**
  - episode_id
  - person_id
  - episode_concept_id
  - episode_start_datetime
  - episode_end_datetime
  - episode_parent_id
  - episode_number
  - episode_object_concept_id
  - episode_type_concept_id
  - episode_source_value
  - episode_source_concept_id

- **EPISODE_EVENT**
  - episode_id
  - event_id
  - episode_event_field_concept_id

- **CONDITION_OCCURRENCE**
  - condition_occurrence_id

- **PROCEDURE_OCCURRENCE**
  - procedure_occurrence_id

- **MEASUREMENT**
  - modifier_of_event_id
  - modifier_of_field_concept_id

- **DRUG_EXPOSURE**
  - procedure_occurrence_id

Added vocabularies:
Example of disease and treatment episodes in the Episode table

### 'First occurrence'-of-'Carcinoma of breast'

<table>
<thead>
<tr>
<th>episode_id</th>
<th>12345</th>
</tr>
</thead>
<tbody>
<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>

### ‘Treatment regimen’-of-‘Paclitaxel and Bevacizumab’

<table>
<thead>
<tr>
<th>episode_id</th>
<th>12346</th>
</tr>
</thead>
<tbody>
<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>

Added vocabularies:

- Foreign key to the disease Episode record
- OMOP concept 'Treatment Regimen'
- OMOP concept 'Episode algorithmically derived from EHR'
- SNOMED concept 'Carcinoma of breast'
- OMOP concept 'Tumor registry'
- HemOnc concept 'Paclitaxel and Bevacizumab'

Disease and treatment episodes in the OMOP CDM
Testing

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

• **Converted EHR** and **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 1st cancer occurrence
  – Derivation of chemotherapy regimens from atomic drugs
Results

Survival from diagnosis

Time from diagnosis to treatment
What You Can Do Now

• Represent most granular cancer diagnosis based on ICD-O
• Ingest Tumor Registry data using standardized ETL
• Identify cancer patient cohorts based on multiple diagnostic features
• Ingest or derive chemotherapy regimens
• Ingest or derive cancer disease and treatment episodes
• Test existing use cases and implement your own
Next Steps – Development Subgroup

- Drug Regimen Algorithm and the challenge we plan to organize at the Hackathon
- Data quality checks for NAACCR ETL
- Robust NAACCR ETL including different dialects
- Analytical package and expansion with additional use cases
- Algorithm for the identification of disease progression and other episodes
Next Steps – Vocabulary Subgroup

- De-duplicate NAACCR variables and values and map duplicates to a selected primary code
- Ingest CAP
- Compare CAP variable-value pairs to NAACCR variable-value pairs
- Map NAACCR items (variables) and values to equivalent LOINC and SNOMED concepts
- Map CAP items (variables) and values to LOINC and SNOMED concepts.
- Align this effort with the ongoing Nebraska Lexicon and CAP standardization efforts and with the evolving mCODE standard
Next Steps – Genomic Subgroup

G-CDM Structure

- **Beginning version**
  - In the OHDSI Symposium in May, 2018

- **Upgrade version**
  - Take full utilize of the existing OMOP-CDM tables
  - Adapt a standard vocabulary system

1. Sequencing
2. Variant_occurrence
3. Variant_annotation

OMOP-CDM

1. Genomic_test
2. Target_gene
3. Variant_occurrence
4. Variant_annotation
Next Steps – Genomic Subgroup

G-CDM Structure

Schematic diagram of the relationship between the tables that make up the GCDM.
Community Engagement in Development & Research

- Data: US tumor registry, non-US tumor registry, EHR, Claims, trial (Future)
- Research questions: High impact use cases
- Domain modelers and vocab developers: Radiology, surgery, precision medicine
- ETL developers
- Methodologists: Support of best practices
Questions?