



Oncology Workgroup Presentation

October 3, 2017



Outline

- **Use cases**
- **Supporting data questions**
- **Challenges**
 - Source data
 - OMOP CDM and Vocabulary
- **Roadmap**
 - Plan
 - Current activities
- **Call for volunteers**



Use Cases

- **Disease progression**
 - Identifying and predicting disease stages from diagnosis to end of life event
 - Predicting first recurrence, following recurrences, length of remissions, and eventual decline based on phenotype/genotype and treatments
- **Episodes of care**
 - Identifying episodes of care: detecting continuous episodes of care and their characteristics
 - This is closely related to disease progression: the ability to identify treatment periods and characteristics, will help identifying and predicting remissions and recurrences
- **Outcome prediction**
 - Survival
 - Response to treatment
 - Adverse events
 - Readmissions, visits to urgent care, re-surgeries
- **Matching patients to trials**



NCI Feasibility Assessment

Feasibility of OHDSI being used to facilitate cancer care delivery research:

- Understand the sequence of (non-cancer) treatments in cancer patients with diabetes, depression, or high blood pressure.
- Assemble U.S.-based treatment cohorts of cancer patients with diabetes, depression, or high blood pressure; understand the sequence of treatments in these patients, and assess the geographic variations in these pathways, if there are adequate number of patients to do this analysis.
- Understand the feasibility of using existing data infrastructure to conduct cancer treatment and outcomes research.
- Clarify the feasibility of using existing data infrastructure to conduct research on cancer treatment pathways, and the impact of treatment variations on outcomes of cancer patients.



Supporting Data Questions

- **Identify patients with cancer / certain cancer diagnosis**
- **Determine 1st cancer occurrence**
- **Determine primary cancer characteristics**
 - Histology
 - Topography
 - Stage (pathological and clinical): grade: size/spread, lymph nodes, metastases
- **Identify episodes of care**
 - Identify treatment types: Surgery, Radiation Therapy, Chemotherapy, Immunotherapy, Hormone Therapy, Targeted Therapy, Active Surveillance, Palliative care
 - Identify treatment regimens
 - Identify treatment's intent
- **Identify response to treatment**
 - Imaging, pathological, clinical
- **Identify progression of disease**
 - Recurrences
 - Remissions
 - End of life event(s)



Challenges: Source Data

- Several types of data sources
 - Reconciling data granularity between different types of data sources
 - Alignment between well and poorly structured sources
- Lack of terminological standards
- Key data elements are not collected/collected as notes only



Types of Data Sources

	Cancer Registry	EMR/Claims	Cancer Trials
Focus	Epidemiology	Patient care/billing	Research
Granularity	High	Low	High
Structured data	Mostly	Half & Half	All
Data Quality	High	Poor	High
Coverage			
Temporal	1 st occurrence	All occurrences	One occurrence
Cancer types	Reportable only	All	Selected
Domains	Most	Most	Selected
Use of standard Vocabularies	Low	Medium	Minimal
Time lag	6 months	None	Trial-specific



Challenges: OMOP CDM Extensions

- **Conditions**

- Disease episodes/eras: reflect recurrences and remissions
- Add disease attributes (e.g. stage)

- **Treatments**

- Intent: connection between condition and treatment
- Regimens/eras: reflect treatment episodes, combinations of treatments (procedures and medications), temporal relationships between treatments
- Response to treatment
- Add treatment attributes/modifiers (e.g. radiation therapy dose)



Challenges: OMOP Vocabularies

- **Choice of vocabulary**
 - Is granularity of the chosen vocabulary sufficient? Example: ICD-O vs. SNOMED for diagnoses
 - Does a chosen vocabulary has hierarchy/classification that will help identifying treatment intent? Example: CPT vs. SNOMED for procedures
 - Are additional classifications available? Example: NCCN Compendiums to identify medication/radiotherapy regimens and their targeted diagnoses
- **Vocabulary structure vs. CDM extension**
 - Pre-coordinated multi-dimensional concepts vs. additional CDM attributes. Example: SNOMED diagnosis concepts with at least two dimensions, anatomy and morphology
- **Mappings between source and standard vocabularies**
 - Add new: ICD-O to SNOMED
 - Improve existing: CPT to SNOMED



Roadmap

- **Plan**

- Divided into deliverables that will help answer at least one analytical question
- Deliverables prioritized by use case importance and complexity
- Deliverables broken into exploratory, modeling, implementation, and testing phases

- **Current Activities**

- Implementation of ICD-O to SNOMED mapping
 - Exploration of SNOMED for identification of procedure intent
 - Exploration of NCCN Chemotherapy Order Templates and Radiation Therapy Compendium for identification of medication regimens and treatment intent
 - Exploration of CAP Cancer Protocols as a possible target vocabulary for tumor stage and grade
-



Oncology Wokgroup Info

- **Workgroup page, meeting info**

<http://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:oncology-sg>

- **Project plan and volunteer list**

<https://drive.google.com/file/d/0B-aFA5uiBXbzUlppcENINkp1VDA/view?usp=sharing>

- **Face-to-Face meeting at the OHDSI symposium**

- Thursday, October 19th, 6 pm,

- Location: TBD

Join the journey!



Oncology Wokgroup Members

Charlie Bailey

Rimma Belenkaya

Shantha Bethusamy

RuiJun Chen

Melissa Crenshaw

Dima Dymshyts

Krisitn Feeney

Michael Gurley

Rashedul Hasan

George Hripcsak

Iker Huerga-Sanchez

Voijtech Huser

Guoqian Jiang

Shirley Johnson

Gregory Klebanov

Amy Lin

Jin Liu

Robert Miller

Don O'Hara

Anna Ostropolets

Nick Puntikov

Javan Quintela

Christian Reich

Mitra Rocca

Eric Schneider

Chad Smathers

Mark Vance

Andrew Williams



Appendix





Data Sources

- Cancer Registry
 - + Epidemiology focused, granular, well-structured, validated
 - + Includes diagnosis, staging, treatment, cancer-specific attributes
 - + Includes complete first occurrence data
 - Includes only some recurrence data
 - Contains only reportable cancer types
 - Only diagnoses are standard (ICD-O) based
 - Time lag 6 months



Data Sources

- EMR/Claims
 - + Includes data for all cancer types, recurrences and episodes of care
 - + More domains are coded using standards
 - + Data available real-time
 - Billing focused, not granular or structured enough
 - Recurrences, treatment episodes and intent are not delineated
- Clinical Trials
 - + Research focused, granular, well-structured, validated
 - Includes only selected episodes of disease and care
 - Coding is rarely standard based
 - Time lag dependent on data release constraints