OHDSI Collaborator Meeting

Oncology WG Presentation

12/3/2019

Agenda

- 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

Memorial Sloan Kettering Cancer Center

Northwestern University

IMS Health & Quintiles are now IQVIA

ODYSSEUS DATA SERVICES INC

VANDERBILT

Tufts CTSI
Tufts Clinical and Translational Science Institute

COLUMBIA
COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS


Michael Gurley

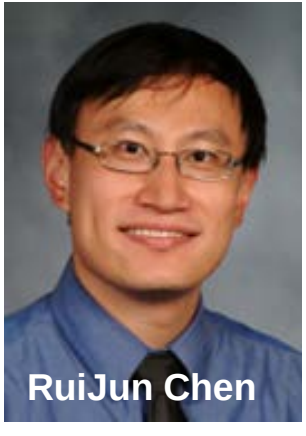

Christian Reich


Dmitry Dymshyts


Robert Miller


Jeremy Warner


Andrew Williams


RuiJun Chen


Rimma Belenkaya

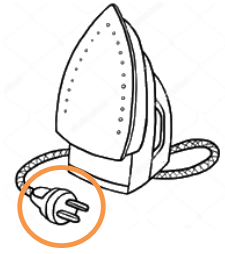
Contributors

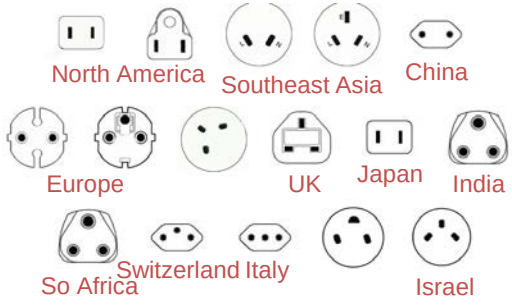
Charles Bailey, Children's Hospital of Philadelphia
Scott Campbell, University of Nebraska
Rachel Chee, IQVIA
Mark Danese, Outcome Insights
Asieh Golozar, Regeneron
George Hripcsak, Columbia University
Ben May, Columbia University
Maxim Moinat, The Hyve
Anna Ostropolets, Columbia University
Meera Patel, 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

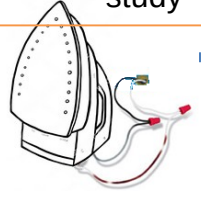
Traditional way

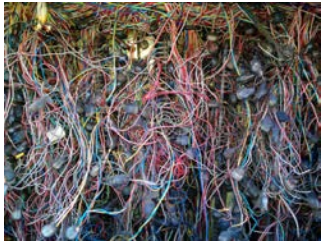
Analytical method: Adherence to Drug





One SAS or R script for each study

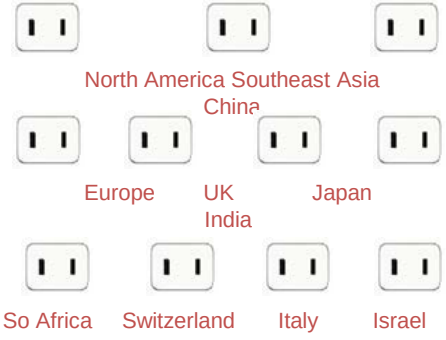




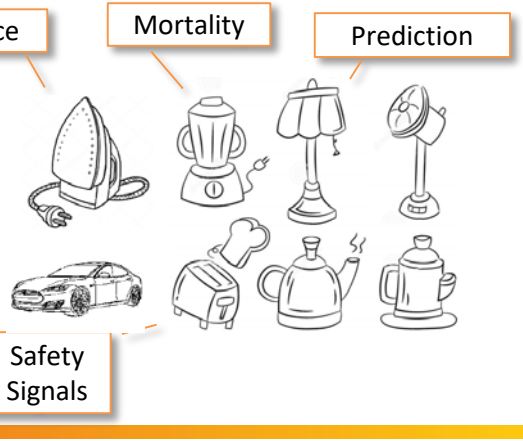
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

OHDSI approach

OMOP CDM



OHDSI Tools



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:

Concept	Category
Carcinoma	Histology
Neck of bladder	Anatomical site
Non-metastatic disease	Tumor attribute
Disease in remission	Condition Episode
First line treatment	Treatment Episode
Chemotherapy regimen	Regimen
Gemcitabin	Component of regimen

Five Goals

1. Build standards **on top of OMOP**
   - Vocabularies
   - Module
   - Data model

Oncology

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

Logged in as: Christian Reich (cgreich) Update Profile Log Out

Observational Health Data Sciences and Informatics

Search

Recent Changes Media Manager Sitemap

Trace: • oncology-sg

projects:workgroups:oncology-sg

- Documentation
- Video tutorials
- Development
- Research Studies
- Projects & Workgroups
- Meetings & Events
- Other Resources
  - Community Forums
  - 2018 Data Network
  - 2019 Data Network
  - OHDSI Directory
  - Funding Opportunities
  - Conferences
  - OHDSI Library
  - Mailing Lists
  - Realtime Chat (IRC)
  - Community Publications

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

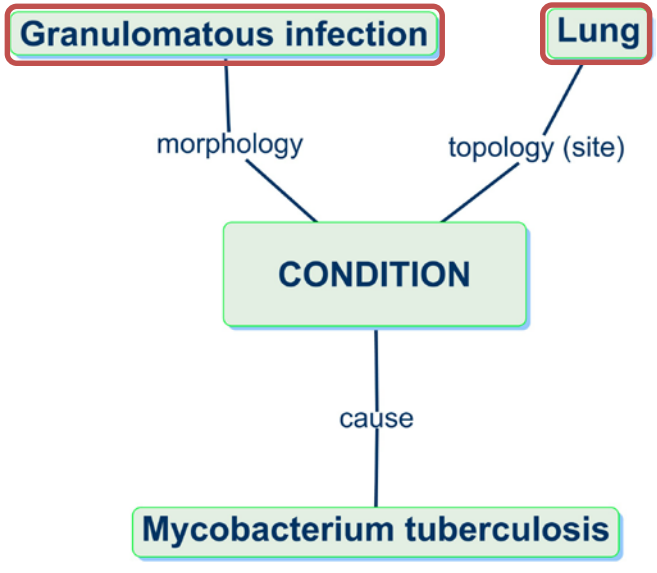
What's Been Accomplished

- 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

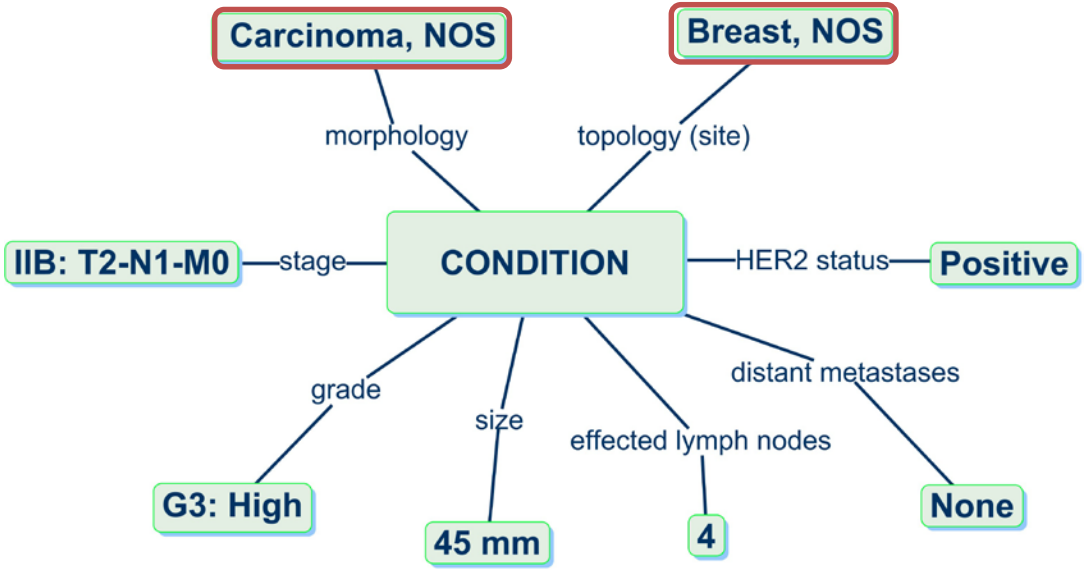
Normal Condition

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



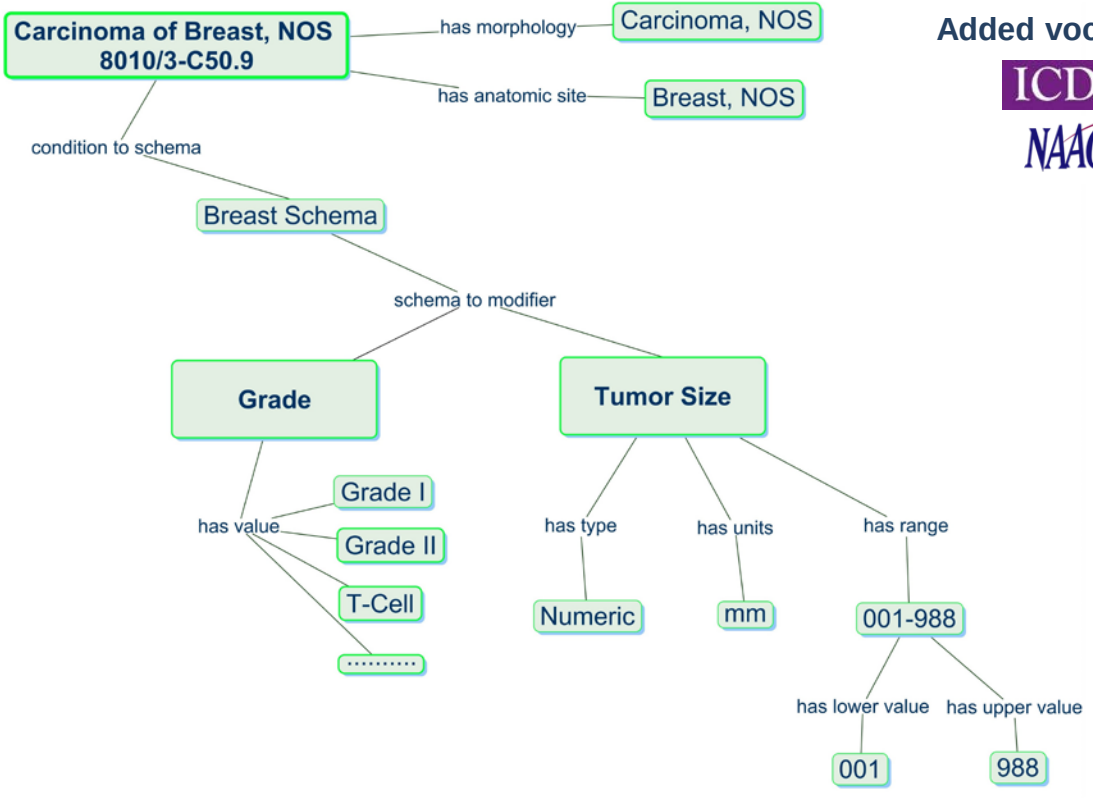
Cancer

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



Solving Granularity Challenge

Cancer Diagnosis Model in the OMOP Vocabulary



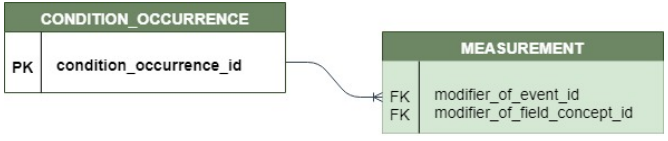
Added vocabularies:

ICD - O

NAACCR

Solving Granularity Challenge

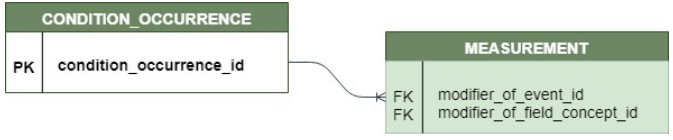
Cancer diagnosis representation in the OMOP CDM



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

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

Example of cancer diagnosis in the OMOP CDM

Histology+Site diagnosis in **Condition_Occurrence**

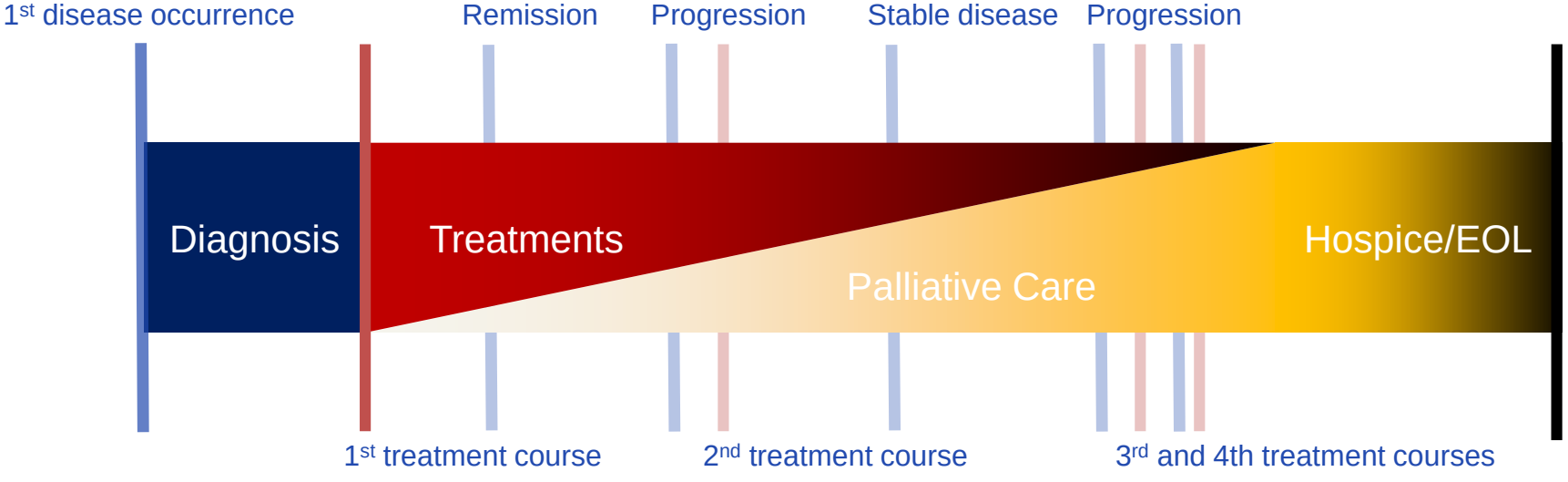
condition_occurrence_id	123456789	
person_id	1	
condition_concept_id	**4116071**	← **SNOMED concept 'Carcinoma of breast'**
condition_start_datetime	June 9, 2019	
condition_type_concept_id	32535	
condition_source_value	**8010/3-C50.9**	← **Precoordinated concept of ICD-O Histology & Site**
condition_source_concept_id	44505310	

Grade modifier in **Measurement**

measurement_id	567890	
person_id	1	
measurement_datetime	June 9, 2019	
measurement_concept_id	**35918640**	← **NAACCR concept 'Grade Pathological'**
measurement_date	June 9, 2019	
value_as_concept_id	**35922509**	← **NAACCR concept 'G3: High combined histologic grade (unfavorable); SBR score of 8-9 points'**
measurement_type_concept_id	32534	← **OMOP concept 'Tumor registry'**
measurement_source_value	3844	← **NAACCR code for 'Grade Pathological'**
measurement_source_concept_id	35918640	← **NAACCR concept 'Grade Pathological'**
value_source_value	breast@3844@3	← **NAACCR code for 'G3: High combined histologic grade (unfavorable); SBR score of 8-9 points'**
modifier_of_event_id	**123456789**	← **Value of the respective condition record condition_occurrence_id**
modifier_field_concept_id	**1147127**	← **Concept for 'condition_occurrence.condition_occurrence_id'**

Challenges: Abstraction

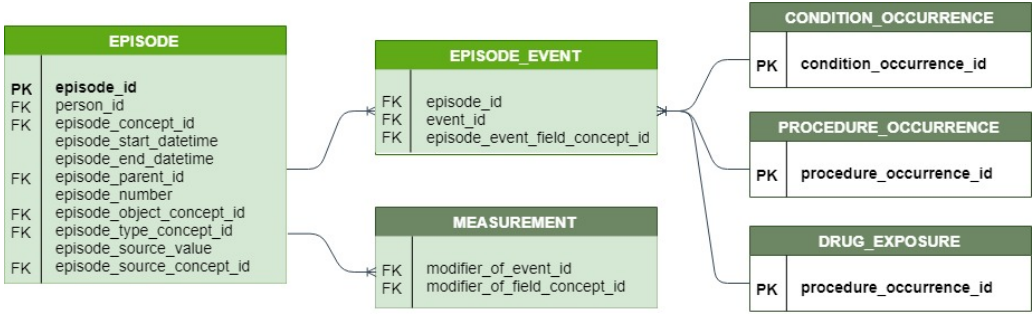
- Clinically and analytically relevant representation of cancer diagnoses, treatments, and outcomes requires data abstraction



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

Solving Abstraction Challenge

Disease and treatment episodes in the OMOP CDM

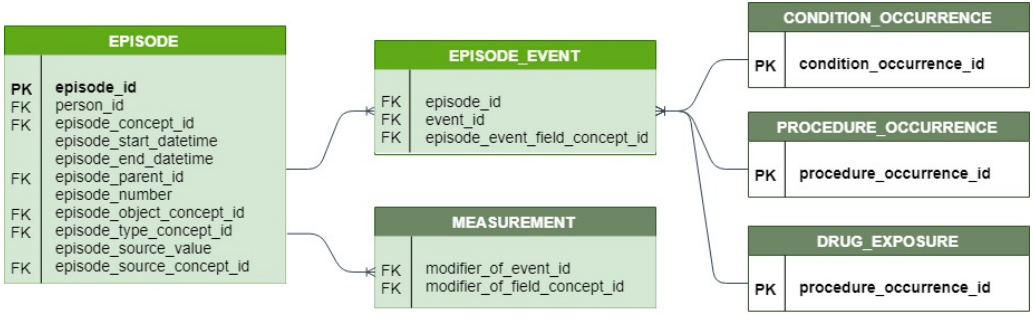


Added vocabularies:

HemOnc.org

Solving Abstraction Challenge

Disease and treatment episodes in the OMOP CDM



Added vocabularies:

HemOnc.org

Example of disease and treatment episodes in the Episode table

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

episode_id	12345
person_id	1
episode_concept_id	**32528**
episode_start_datetime	June 9, 2019
episode_object_concept_id	**4116071**
episode_type_concept_id	32535

OMOP concept 'First disease occurrence'

SNOMED concept 'Carcinoma of breast'

OMOP concept 'Tumor registry'

'Treatment regimen'-of-' Paclitaxel and Bevacizumab'

episode_id	12346
person_id	1
episode_concept_id	**32531**
episode_start_datetime	July 9, 2019
episode_parent_id	12345
episode_object_concept_id	35804255
episode_type_concept_id	32545

OMOP concept 'Treatment Regimen'

Foreign key to the disease Episode record

HemOnc concept Paclitaxel and Bevacizumab'

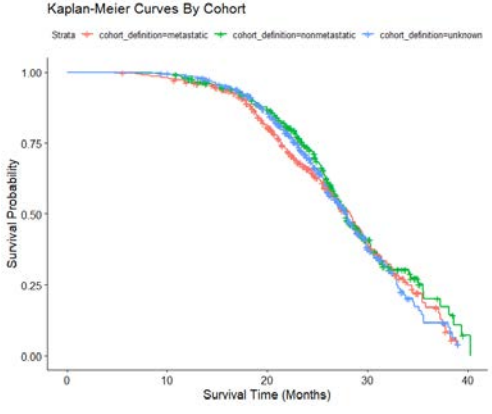
OMOP concept 'Episode algorithmically derived from EHR'

Testing

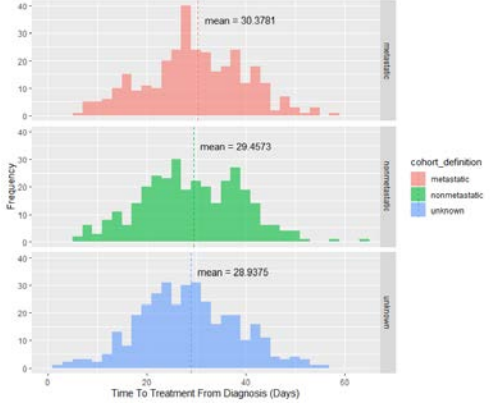
- Developed **ontology-driven ETL** for data conversion from Tumor Registry
- **Converted EHR** and **Registry data** from four participating institutions
- Tested **clinical characterization use cases**
  - Survival from initial diagnosis
  - Time from diagnosis to treatment
  - High-level treatment 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 of 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

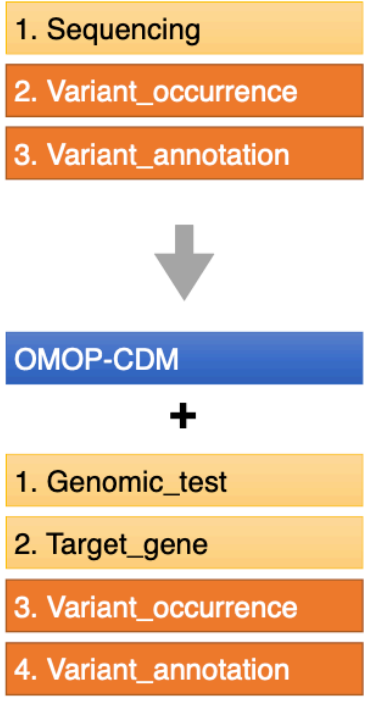
Next Steps – Vocabulary Subgroup

- 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

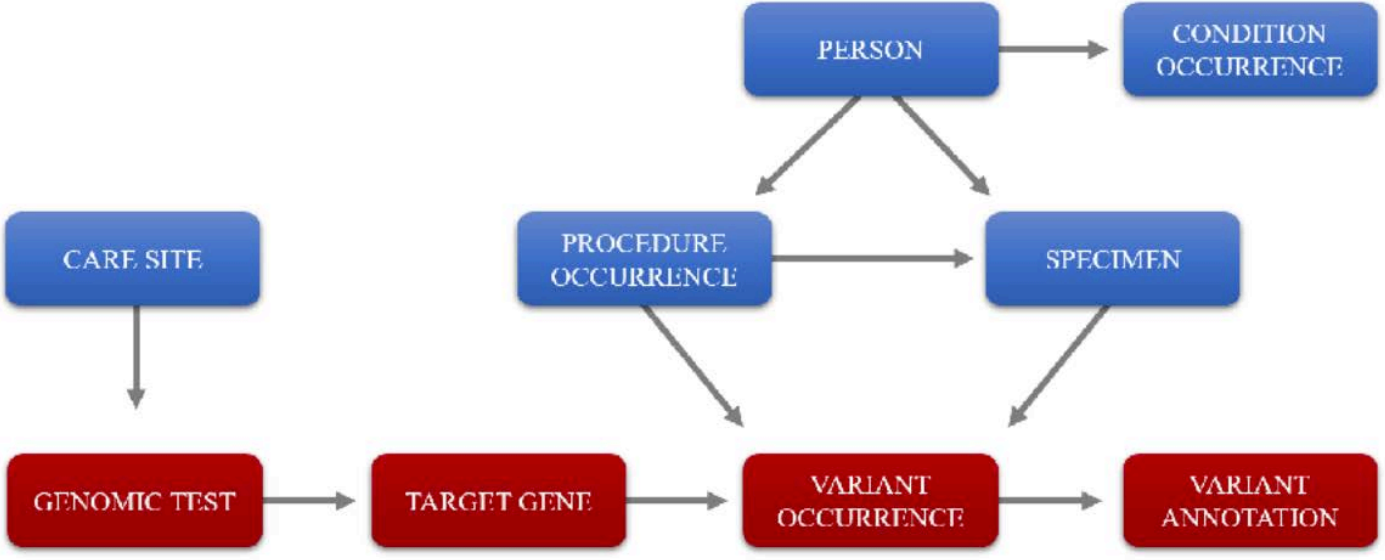
G-CDM Structure

- 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



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?