Memorial Sloan Kettering Cancer Center

OMOP CDM Oncology Module at Work

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

OHDSI Oncology WG

Memorial Sloan Kettering Cancer Center

Northwestern University

IMS Health & Quintiles are now IQVIA™

ODYSSEUS DATA SERVICES INC

VANDERBILT

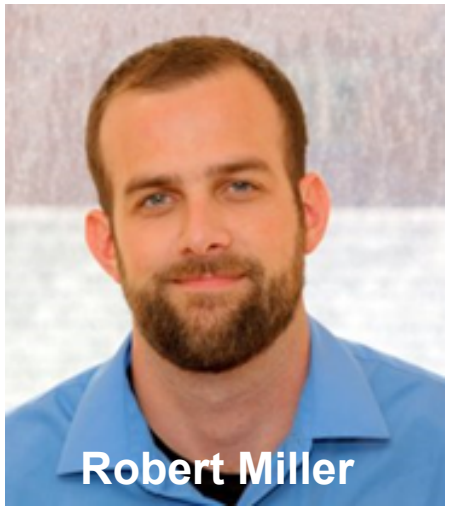
Tufts CTSI
Tufts Clinical and Translational Science Institute

COLUMBIA
COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

Michael Gurley

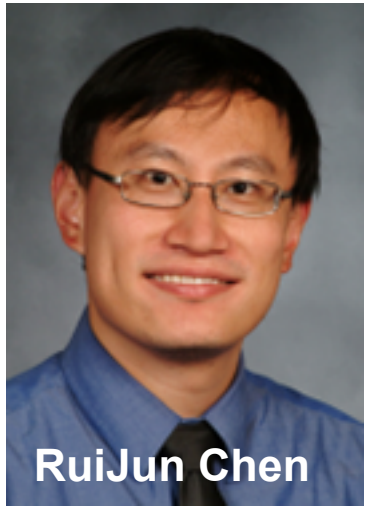
Christian Reich

Dmitry Dymshyts


Robert Miller

Jeremy Warner

Andrew Williams

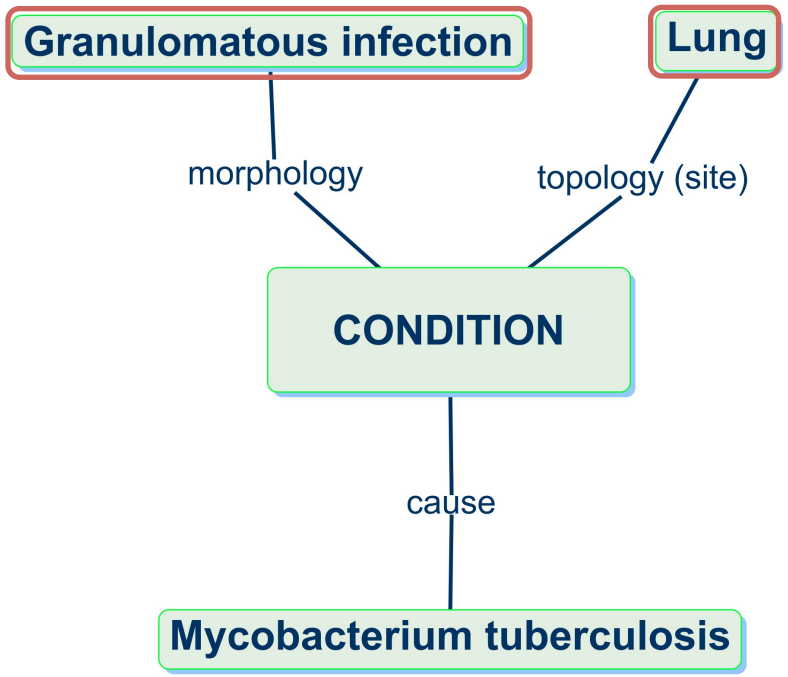

RuiJun Chen

Rimma Belenkaya

Challenges: Granularity

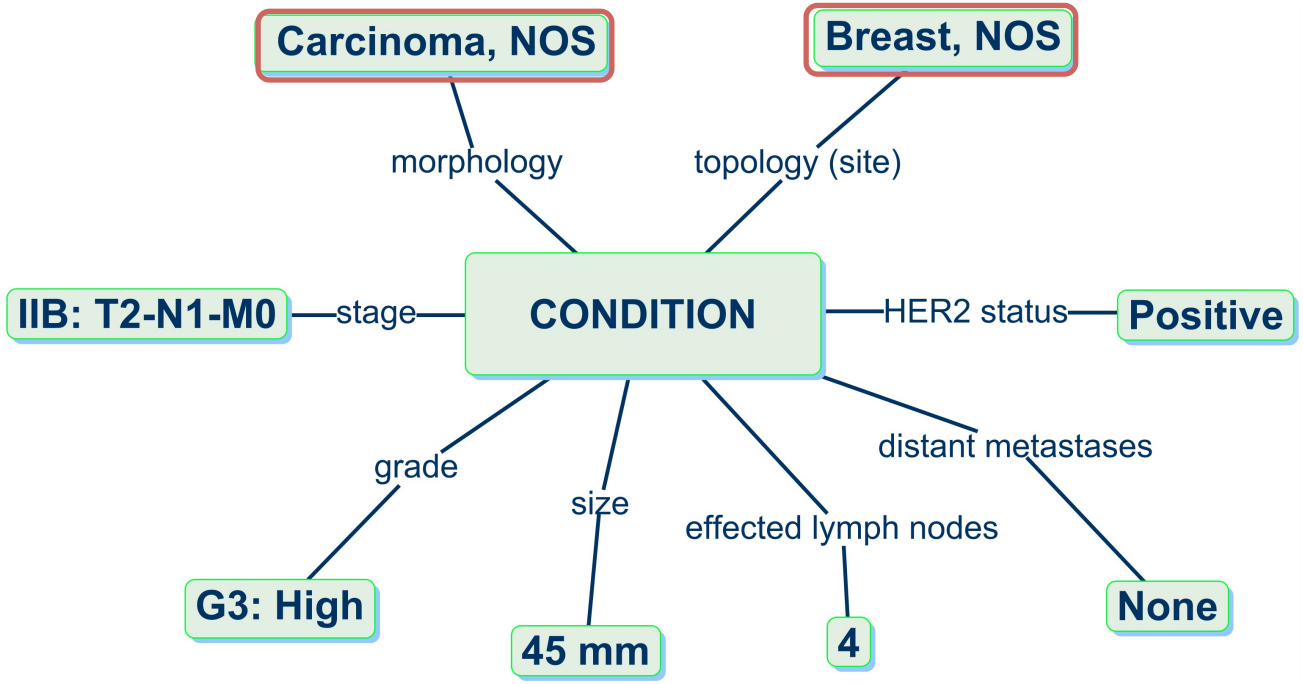
Normal Condition

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



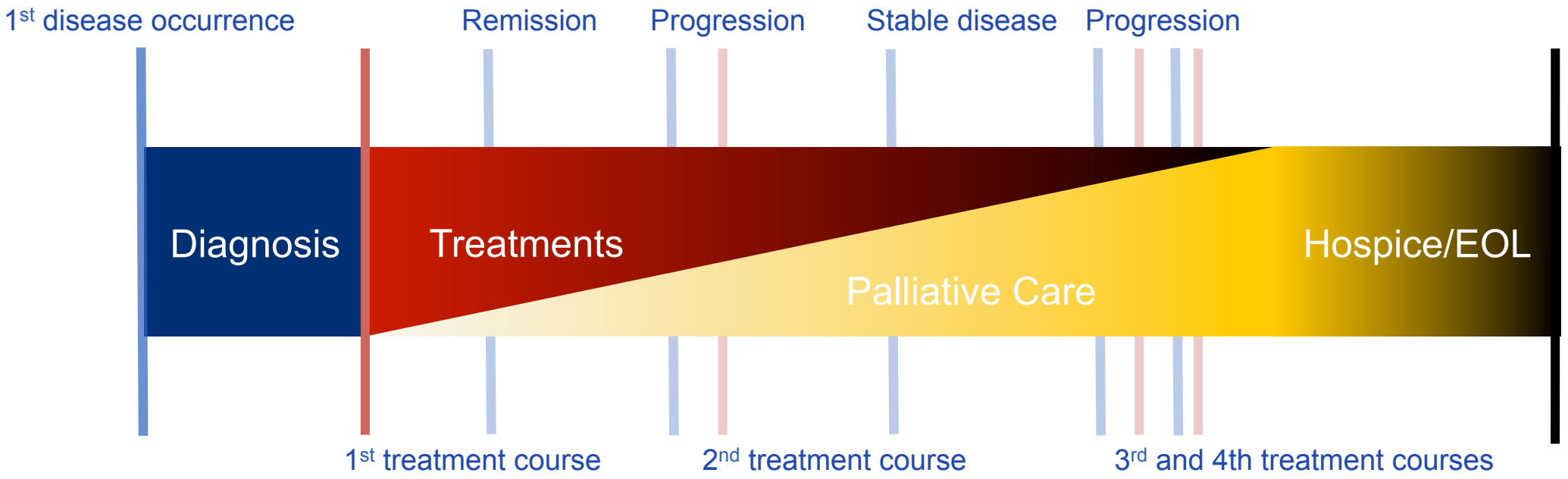
Cancer

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



Challenges: Abstraction

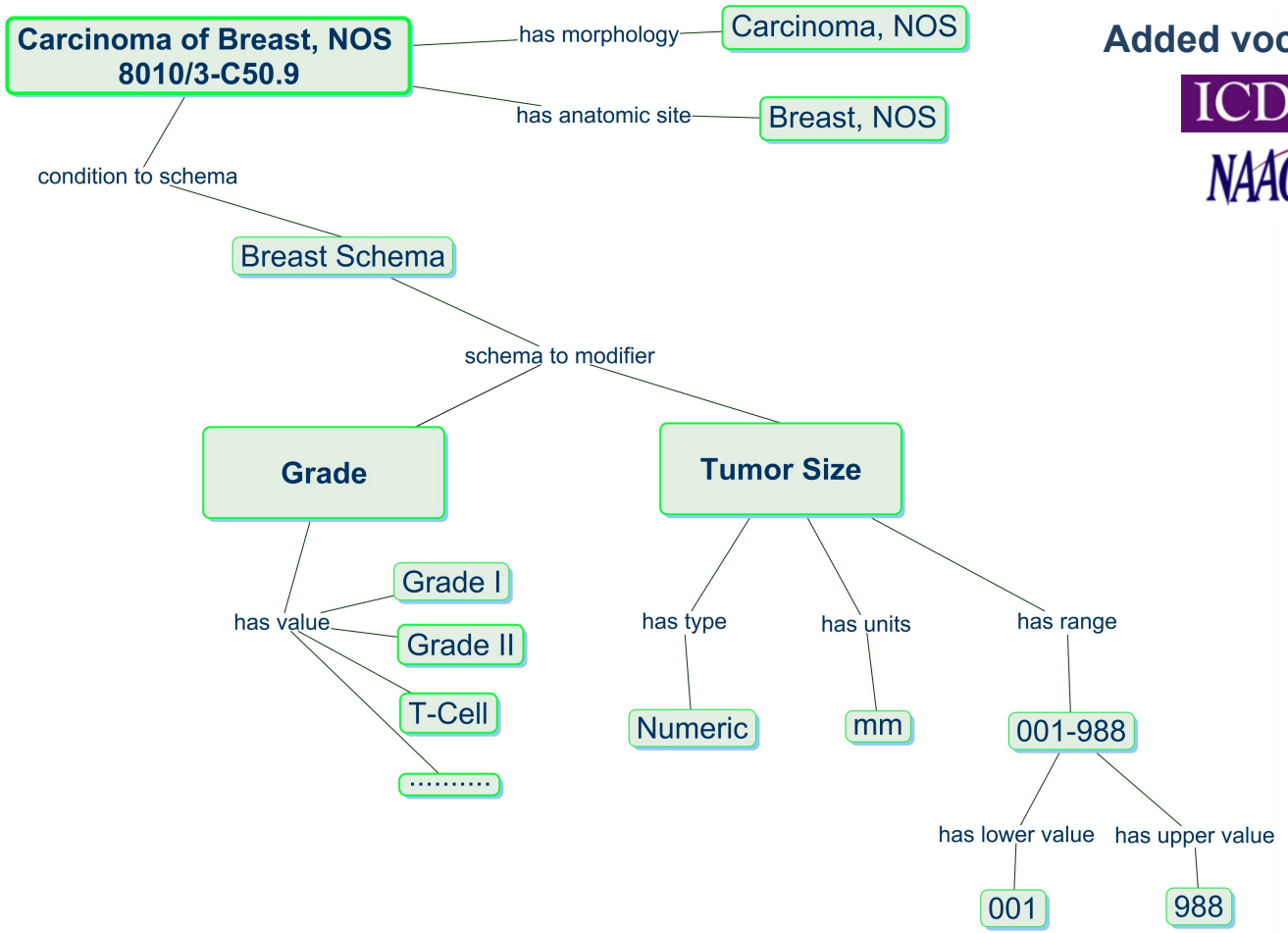
- Clinically and analytically relevant representation of cancer diagnoses, treatments, and outcomes requires data abstraction



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

Solving Granularity Challenge

Cancer Diagnosis Model in the OMOP Vocabulary



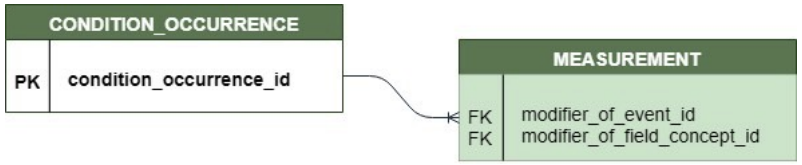
Added vocabularies:

ICD - O

NAACCR

Solving Granularity Challenge

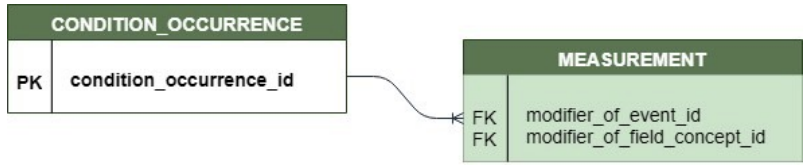
Cancer diagnosis representation in the OMOP CDM



- **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**

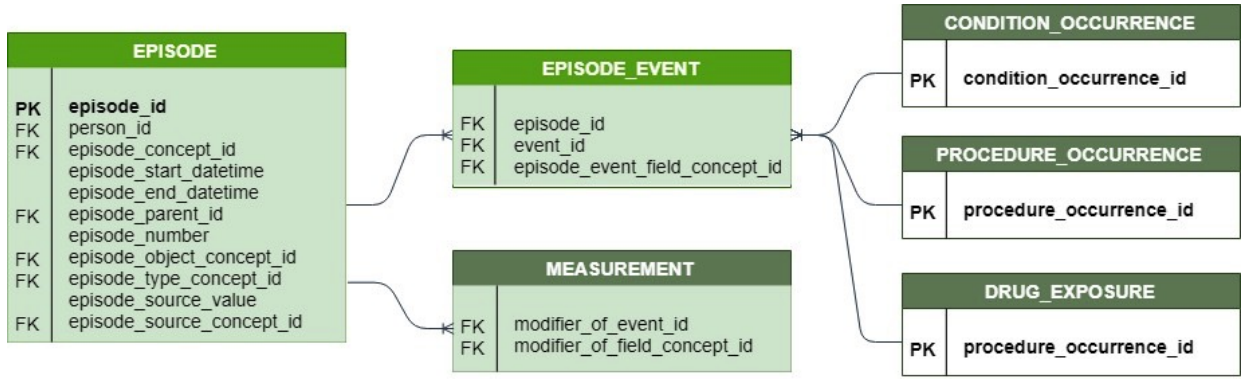
condition_occurrence_id	123456789	
person_id	1	
condition_concept_id	**4116071**	← **SNOMED concept 'Carcinoma of breast'**
condition_start_datetime	June 9, 2019	
condition_type_concept_id	32535	
condition_source_value	**8010/3-C50.9**	← **Precoordinated concept of ICD-O Histology & Site**
condition_source_concept_id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

Solving Abstraction Challenge

Disease and treatment episodes in the OMOP CDM

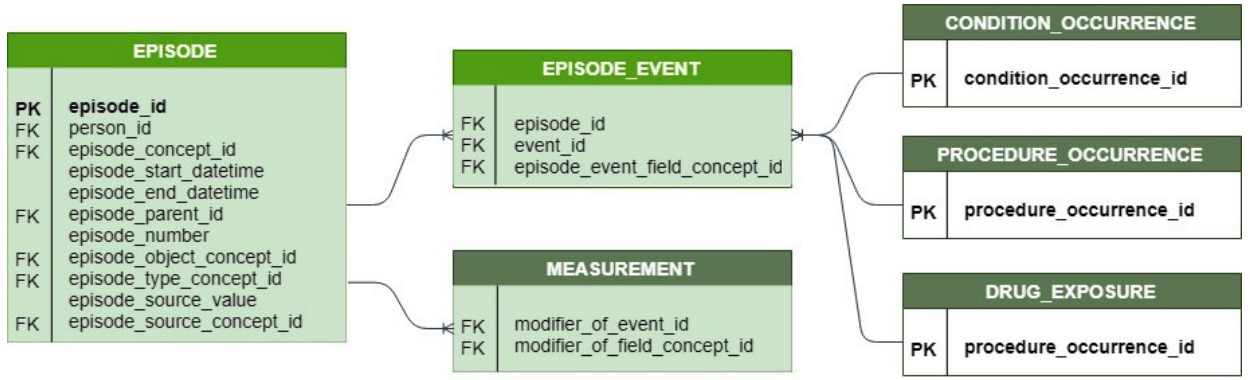


Added vocabularies:

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

- episode_concept_id: **OMOP concept 'First disease occurrence'**
- episode_object_concept_id: **SNOMED concept 'Carcinoma of breast'**
- episode_type_concept_id: **OMOP concept 'Tumor registry'**

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

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

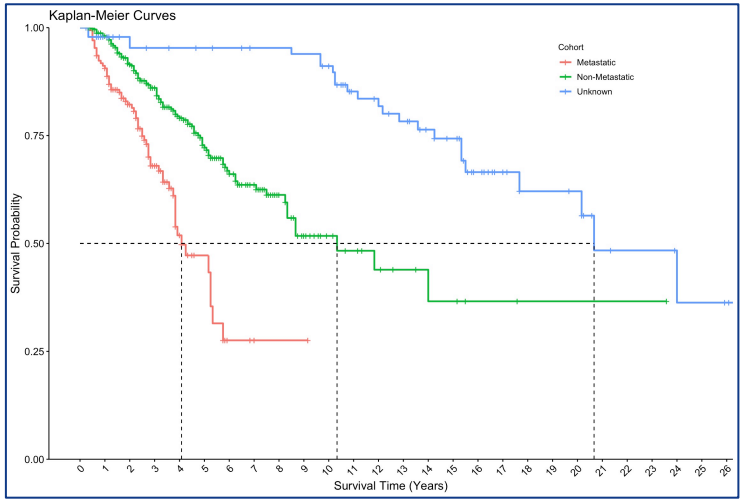
- episode_concept_id: **OMOP concept 'Treatment Regimen'**
- episode_start_datetime: **Foreign key to the disease Episode record**
- episode_parent_id: **HemOnc concept 'Carboplatin and Paclitaxel'**
- episode_object_concept_id: **OMOP concept 'Episode algorithmically derived from EHR'**

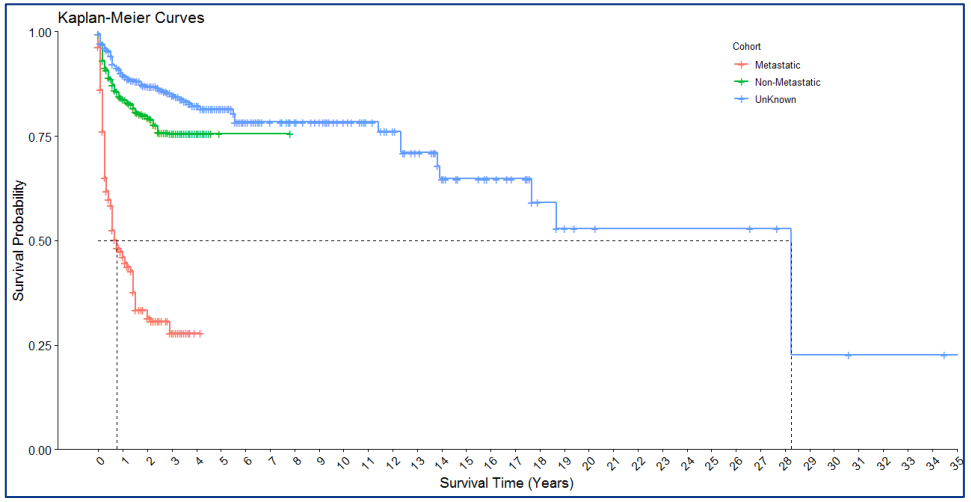
Testing

- 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 1st cancer occurrence
  - Derivation of chemotherapy regimens from atomic drugs

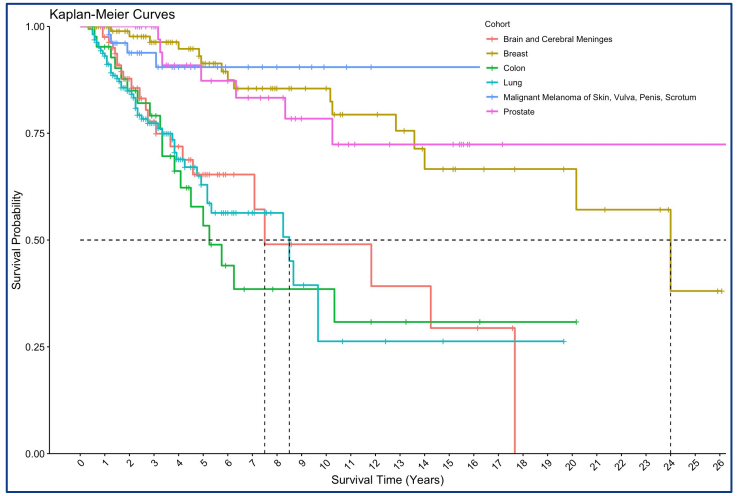
Survival from diagnosis

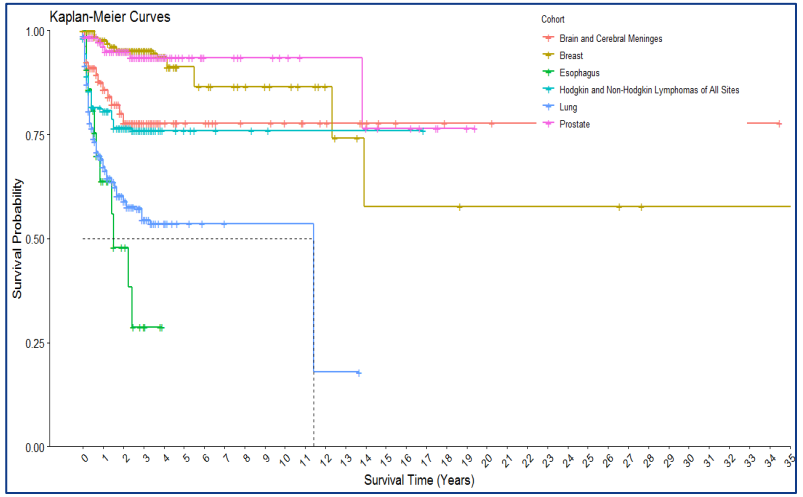
Metastatic vs. Non-metastatic cancers





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!

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank you!