



Memorial Sloan Kettering  
Cancer Center

# Crosswalk from Cancer Registry to OMOP Oncology Module



Rimma Belenkaya

---

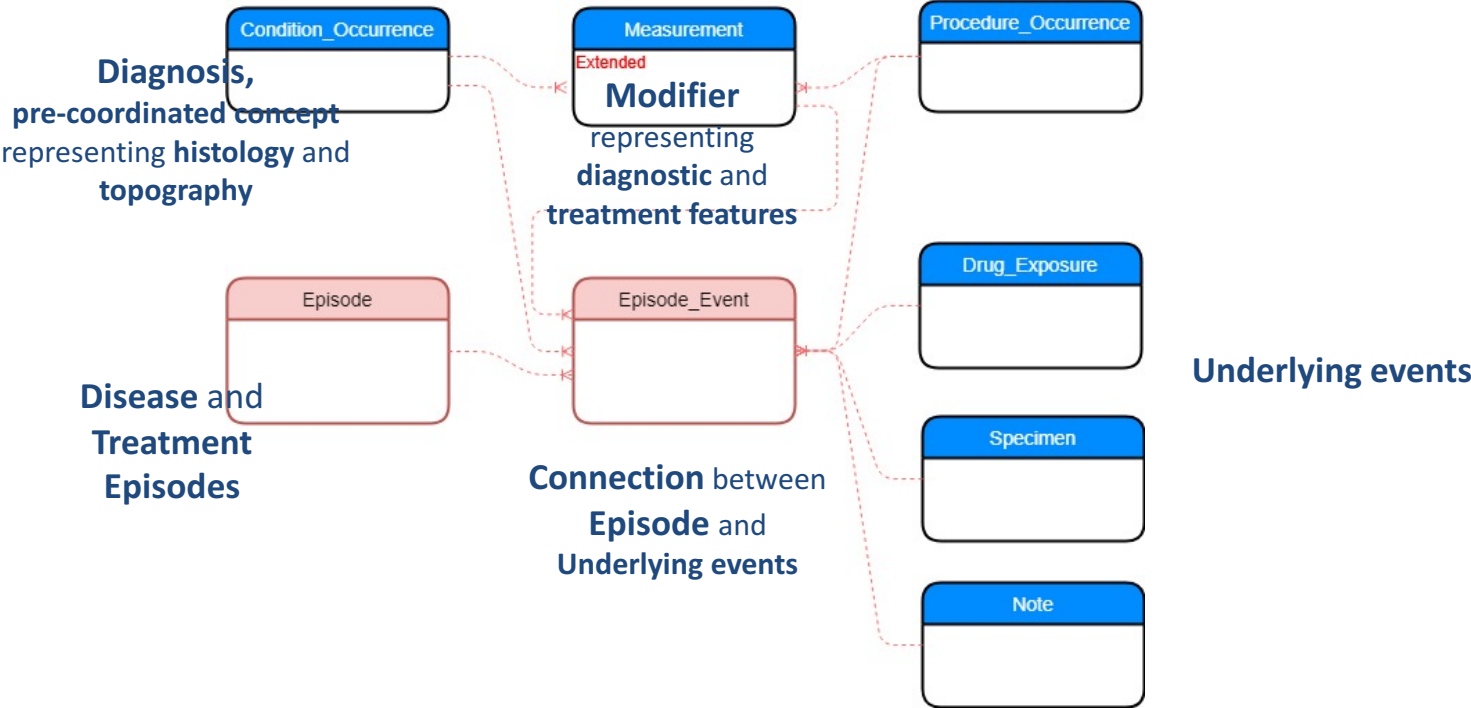


# Outline

- **ETL of Diagnoses**
  - **ETL of Diagnostic Modifiers**
    - Value as concept
    - Value as number
  - **Derivation of disease Episodes**
    - First occurrence
-

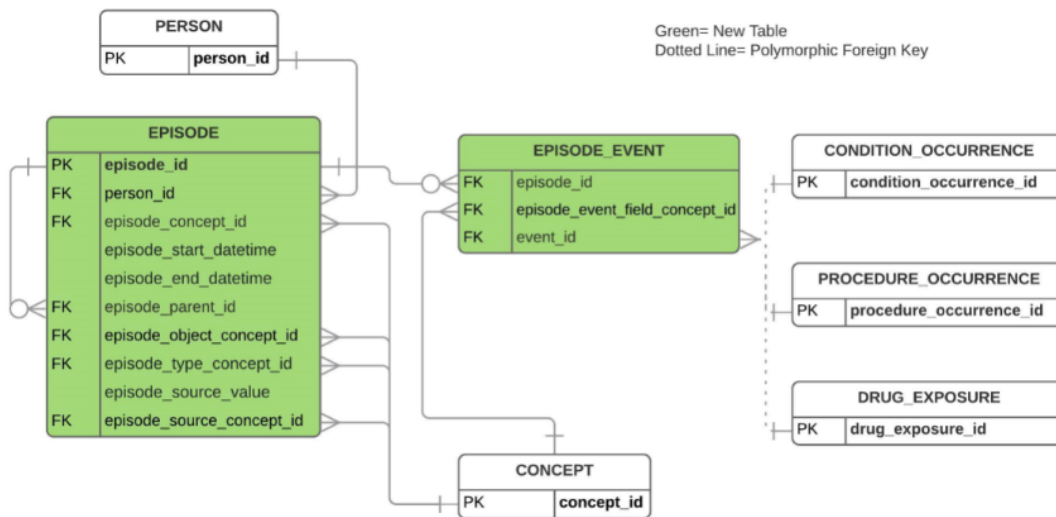


# Oncology Model in OMOP CDM





# EPISODE Model





# Cancer Registry Patient Record Sample

Field#	NAACCR Item #	NAACCR Item Name	NAACCR Value	Value Description
1	20	Patient ID Number	123456	
2	400	Primary Site	C50.9	Breast, NOS
3	522	Histologic Type ICD-O-3	8010/3	Carcinoma, NOS
4	390	Date of Diagnosis	Jun-9-2019	
5	410	Laterality	1	Right: origin of primary
6	440	Grade	4	Grade IV
7	390	Summary Stage 2018	2	Regional by direct extension only



# Source Table Structure

Field Name	Field Type	Required	Field Description
<b>person_id</b>	integer	Yes	Patient ID
<b>record_id</b>	varchar(100)	Yes	Unique primary tumor occurrence ID
<b>naaccr_item_number</b>	integer	Yes	NAACCR Item #
<b>naaccr_item_value</b>	varchar(100)	Yes	NAACCR item value
<b>histology_topography</b>	varchar(20)	Yes	Concatenated histology and topography



# Source Data Sample

person_id	record_id	naaccr_item_number	naaccr_item_value	histology_site
1	<a href="http://mskcc.org/tumorregistry/1_000001">http://mskcc.org/tumorregistry/1_000001</a>	400	C50.9	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000001">http://mskcc.org/tumorregistry/1_000001</a>	522	8010/3	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000001">http://mskcc.org/tumorregistry/1_000001</a>	390	Jun-9-2019	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000001">http://mskcc.org/tumorregistry/1_000001</a>	410	1	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000002">http://mskcc.org/tumorregistry/1_000002</a>	2800	30	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000002">http://mskcc.org/tumorregistry/1_000002</a>	400	C50.9	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000002">http://mskcc.org/tumorregistry/1_000002</a>	522	8010/3	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000002">http://mskcc.org/tumorregistry/1_000002</a>	390	Jun-9-2019	8010/3-C50.9
1	<a href="http://mskcc.org/tumorregistry/1_000002">http://mskcc.org/tumorregistry/1_000002</a>	410	3	8010/3-C50.9

1. Map field names in your source to NAACCR item ids
2. Site and histology must be converted to the following format: C50.9 and 8010/3
3. Add histology\_site field to the source data (will simplify ETL)

# Crosswalk for Diagnosis

## Cancer Registry patient record

Field#	NAACCR Item #	NAACCR Item Name	Value Code	Value Description
1	20	Patient ID Number	123456	
2	400	Primary Site	C50.9	Breast, NOS
3	522	Histologic Type ICD-O-3	8010/3	Carcinoma, NOS
4	390	Date of Diagnosis	Jun-9-2019	
5	410	Laterality	1	Right: origin of primary
6	440	Grade	4	Grade IV
7	390	Summary Stage 2018	2	Regional by direct extension only

## CONDITION\_OCCURRENCE record

Source Field #	Field	Value	Value Description
	condition_occurrence_id	987654	
1	person_id	123456	
	condition_concept_id	4116071	Carcinoma of breast
4	condition_start_date	June 9, 2019	
4	condition_start_datetime	June 9, 2019	
	condition_end_date		
	condition_end_datetime		
	condition_type_concept_id	35941507	Hospital cancer registry
	condition_status_concept_id		
	stop_reason		
	provider_id		
	visit_occurrence_id		
	visit_detail_id		
2+3	condition_source_value	8010/3-C50.9	
	condition_source_concept_id	44505310	Carcinoma of breast
	condition_status_source_value		

## CONCEPT: source concept for diagnosis

Source Field #	Field	Value
	concept_id	44505310
	concept_name	Carcinoma of breast
2+3	concept_code	8010/3-C50.9
	vocabulary_id	NAACCR

## CONCEPT\_RELATIONSHIP

Field	Value
concept_id_1	44505310
concept_id_2	4116071
relationship_id	Maps to

## CONCEPT: standard concept for diagnosis

Field	Value
concept_id	4116071
concept_name	Carcinoma of breast
concept_code	254838004
vocabulary_id	SNOMED



# Crosswalk for Diagnostic Modifiers: value as concept

## CONCEPT: source concept for diagnosis

Source Field #	Field	Value
	concept_id	44505310
	concept_name	Carcinoma of breast
2+3	concept_code	8010/3-C50.9
	vocabulary_id	NAACCR

## CONCEPT\_RELATIONSHIP

Field	Value
concept_id_1	44505310
concept_id_2	35909870
relationship_id	ICDO to Schema

## CONCEPT: concept for diagnostic schema

Field	Value
concept_id	35909870
concept_name	Breast
concept_code	breast
vocabulary_id	NAACCR

## CONCEPT\_RELATIONSHIP

Field	Value
concept_id_1	35918306
concept_id_2	35909870
relationship_id	Variable to Schema

## CONCEPT: standard concept for laterality

Field	Value
concept_id	35918306
concept_name	Laterality
concept_code	410
vocabulary_id	NAACCR

## Cancer Registry patient record

NAACCR Item #	NAACCR Item Name	Value Code	Value Description
20	Patient ID Number	123456	
400	Primary Site	C50.9	Breast, NOS
522	Histologic Type ICD-O-3	8010/3	Carcinoma, NOS
390	Date of Diagnosis	Jun-9-2019	
410	Laterality	1	Right: origin of primary
440	Grade	4	Grade IV
390	Summary Stage 2018	2	Regional by direct extension only

## CONCEPT\_RELATIONSHIP

Field	Value
concept_id_1	35918306
concept_id_2	35940926
relationship_id	Has Answer

## CONCEPT: standard concept for value

Field	Value
concept_id	35940926
concept_name	Right: origin of primary
concept_code	410@1
vocabulary_id	NAACCR

## CONCEPT\_RELATIONSHIP

Field	Value
concept_id_1	35918306
concept_id_2	35918536
relationship_id	Variable has date

## CONCEPT: standard concept for date

Field	Value
concept_id	35918536
concept_name	Date of Diagnosis

## CONDITION\_OCCURRENCE record

Source Field #	Field	Value	Value Description
	condition_occurrence_id	987654	
1	person_id	123456	
	condition_concept_id	4116071	Carcinoma of breast
4	condition_start_date	June 9, 2019	
4	condition_start_datetime	June 9, 2019	
	condition_end_date		
	condition_end_datetime		
	condition_type_concept_id	35941507	Hospital cancer registry
	condition_status_concept_id		
	stop_reason		
	provider_id		
	visit_occurrence_id		
	visit_detail_id		
2+3	condition_source_value	8010/3-C50.9	
	condition_source_concept_id	44505310	Carcinoma of breast
	condition_status_source_value		

## Measurement record

Field	Value	Value Description
measurement_id	567890	
person_id	123456	
measurement_concept_id	35909870	Laterality
measurement_date	June 9, 2019	
measurement_datetime	June 9, 2019	
measurement_type_concept_id	46940927	Measurement
operator_concept_id		
value_as_number		
value_as_concept_id	35940926	Right: origin of primary
unit_concept_id		
range_low		
range_high		
provider_id		
visit_occurrence_id		
visit_detail_id		
measurement_source_value	410	
measurement_source_concept_id	35909870	Laterality
unit_source_value		
value_source_value	1	
modifier_of_event_id	987654	
modifier_field_concept_id	1147127	condition_occurrence.condition_occurrence_id



# NAACCR Numeric Items in Vocabulary

- NAACCR Item #2800 ‘CS Tumor Size’

## Concept

Field	Value
concept_id	4325345
concept_name	CS Tumor Size
concept_code	June 9, 2019
vocabulary_id	NAACCR
domain_id	Measurement

## Concept Relationship

Field	Value	Description
concept_id_1	4325345	CS Tumor Size
concept_id_2	999991	Numeric
relationship_id	has type	

Field	Value	Description
concept_id_1	4325345	CS Tumor Size
concept_id_2	999992	mm
relationship_id	has units	



# NAACCR Numeric Items in Vocabulary

Scenario	NAACCR Values	Description
1	001-988	001 - 988 millimeters (mm) (Exact size to nearest mm)
2	989	989 mm or larger
3	991	Described as "less than 1 centimeter (cm)"
4	992	Described as "less than 2 cm," or "greater than 1 cm," or "between 1 cm and 2 cm"

## Concept

Field	Value	Value	Value
concept_id	4325346	4325347	4325348
concept_name	989 mm or larger	less than 1 cm	Between 1 and 2 cm
concept_code	2800@989	2800@991	2800@992
vocabulary_id	NAACCR	NAACCR	NAACCR
domain_id	Meas Value	Meas Value	Meas Value

## Concept\_Numeric

Field	Value	Value	Value	Value
concept_id	4325346	4325347	4325348	4325348
concept_as_value	989	1	1	2
concept_unit_id	cm	cm	cm	cm
operator_concept_id	Equal of greater	Equal of less	Equal of greater	Equal of less



# SQL for ETL of Diagnosis and Modifiers

1. Insert a `CONDITION_OCCURRENCE` record
2. Insert the first disease occurrence modifier into `MEASUREMENT`
3. Insert diagnostic modifiers with values as concepts into `MEASUREMENT`
4. Ingest diagnostic modifiers with values as number into `MEASUREMENT`



# Derivation of First Occurrence Episode

## CONDITION\_OCCURRENCE record

Source Field #	Field	Value	Value Description
	condition_occurrence_id	987654	
1	person_id	123456	
	condition_concept_id	4116071	Carcinoma of breast
4	condition_start_date	June 9, 2019	
4	condition_start_datetime	June 9, 2019	
	condition_end_date		
	condition_end_datetime		
	condition_type_concept_id	35941507	Hospital cancer registry
	condition_status_concept_id		
	stop_reason		
	provider_id		
	visit_occurrence_id		
	visit_detail_id		
2+3	condition_source_value	8010/3-C50.9	
	condition_source_concept_id	44505310	Carcinoma of breast
	condition_status_source_value		

## EPISODE record

Field	Value	Value Description
episode_id	4325345	
person_id	123456	
episode_concept_id	32528	Disease First Occurrence
episode_start_datetime	June 9, 2019	
episode_end_datetime		
episode_object_concept_id	4116071	Carcinoma of breast
episode_type_concept_id	999888777	Algorithm #1

## EPISODE\_EVENT record

Field	Value	Value Description
episode_id	4325345	
episode_event_event_id	987654	value of condition_occurrence_id
episode_event_field_concept_id	1147127	condition_occurrence.condition_occurrence_id



## SQL for the First Occurrence Episode

1. Insert an EPISODE record  
from CONDITION\_OCCURRENCE based on the First Occurrence modifier
2. Insert EPISODE\_EVENT record  
to connect EPISODE and CONDITION\_OCCURRENCE records
3. Insert a set of “clean” EPISODE modifiers into MEASUREMENT  
based on the CONDITION modifiers