Crosswalk from Cancer Registry to OMOP Oncology Module
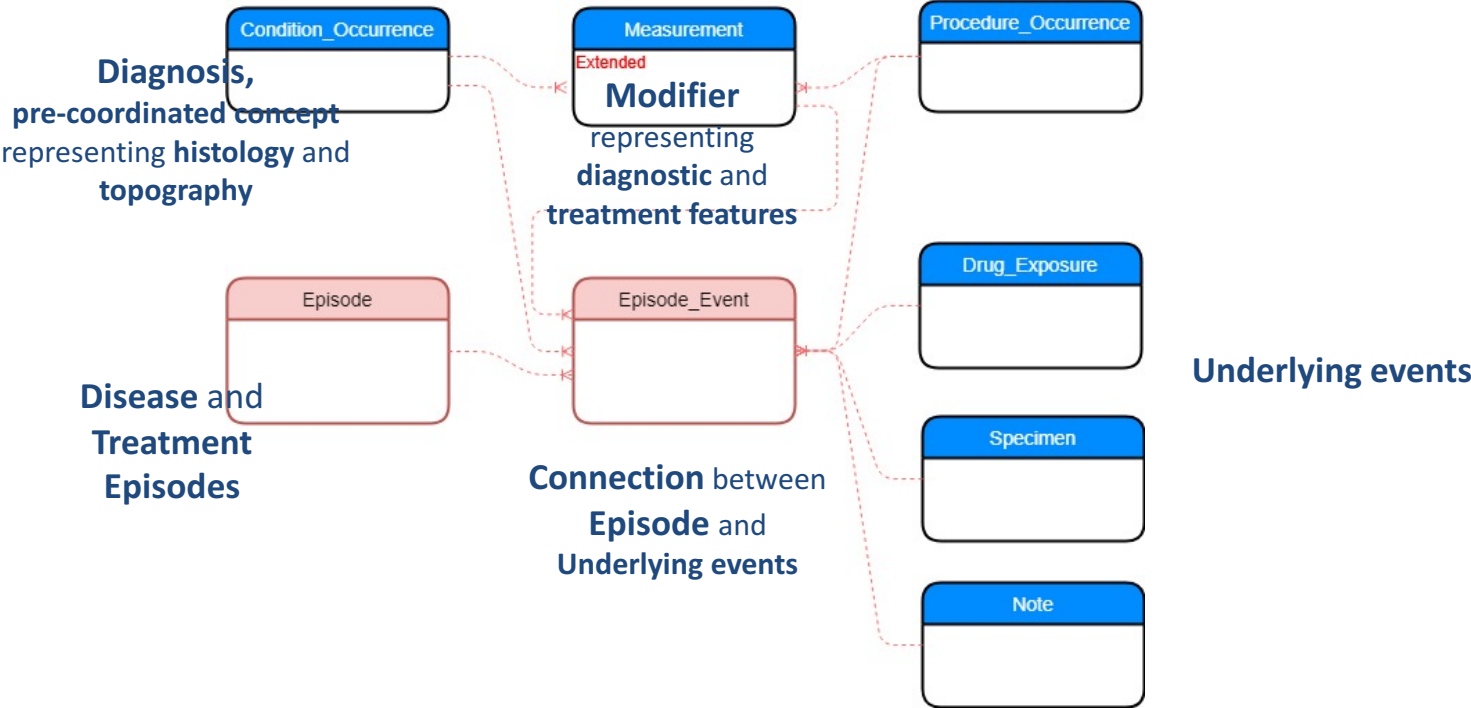
Rimma Belenkaya

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

Cancer Registry Patient Record Sample

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

Source Table Structure

Field Name	NAACCR Item #
variable_id	NAACCR Item #
Patient_id	NAACCR Item Name
value	Value Code
histology_topography	Value Description

Recommendation to add histology_topography. Will Simplify queries.

Source Data Sample

Patient_id	variable_id	value	histology_site
1	400	C50.9	8010/3-C50.9
1	522	8010/3	8010/3-C50.9
1	390	Jun-9-2019	8010/3-C50.9
1	410	1	8010/3-C50.9

1. Site and histology must be converted to the following format:
   C50.9 and 8010/3
2. Adding 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

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

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		

SQL for ETL of Diagnosis

```sql
Insert into condition_occurrence
(person_id,
condition_start_date,
condition_start_date_time,
condition_concept_id,
condition_type_concept_id,
condition_source_value,
condition_source_concept_id)
select s1.patient_id,
s3.value,
s3.value,
c2.concept_id,
35941507,
s1.value || '-' || s2.value,
c1.concept_id
from data_source as s1
join data_source as s2
on s1.patient_id = s2.patient_id and s1.variable_id = '400' and s2.variable_id = '522'
join data_source as s3
on s1.patient_id = s3.patient_id and s3.variable_id = '390'
join concept as c2 on c2.standard_concept ='S'
join concept_relationship as ra on ra.concept_id_2 = c2.concept_id and ra.relationship_id = 'Maps to'
join concept as c1 on c1.concept_id = ra.concept_id_1
where c1.concept_code = s1.value || '-' || s2.value
and c1.vocabulary_id ='ICDO3'
```

Modifier for the First Occurrence

- It is added to explicitly indicate condition_occurrence record corresponding to the first occurrence extracted from Cancer Registry

- This modifier is not extracted from the data as other modifiers.

```sql
Insert into measurement
(person_id,
measurement_datetime,
measurement_concept_id,
measurement_type_concept_id,
measurement_source_value,
modifier_of_event_id,
modifier_field_concept_id)
select s1.patient_id,
s3.value,
32528, -- First occurrence
35941507, -- Cancer registry
'First occurrence',
co.condition_occurrence_id,
1147127
from data_source as s1
join data_source as s2
on s1.patient_id = s2.patient_id and s1.variable_id = '400' and s2.variable_id = '522'
join data_source as s3
on s1.patient_id = s3.patient_id and s3.variable_id = '390'
join condition_occurrence as co on s1.patient_id = co.person_id and s3.value=co.condition_start_datetime and
and co.concept_code = s1.value || '-' || s2.value
```

Crosswalk for Diagnostic Modifiers: value as concept

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

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	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

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

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

```
Insert into measurement
(person_id,
measurement_datetime,
measurement_concept_id,
value_as_concept_id,
measurement_type_concept_id,
measurement_source_value,
measurement_concept_id,
value_source_value,
modifier_of_event_id,
modifier_field_concept_id)
```

SQL for ETL of Diagnostic Modifier: value as concept

```sql
select s.patient_id
, sd.value as measurement_date
, c2.concept_id as measurement_concept_id
, c3.concept_id as value_as_concept_id
, 35941507 -- Cancer registry
, c2.concept_code as measurement_source_value
, c2.concept_id as measurement_source_concept_id
, c3.concept_code as value_source_value
FROM data_source AS s
-- Getting schema
JOIN concept d
ON d.vocabulary_id = 'ICDO3' AND d.concept_code = s.histology || '-' || s.site
JOIN concept_relationship cr1
ON d.concept_id = cr1.concept_id_1 AND cr1.relationship_id = 'Variable to Schema'
JOIN concept AS c1
ON cr1.concept_id_2 = c1.concept_id AND c1.vocabulary_id = 'NAACCR'
-- Getting variable
JOIN concept AS c2
ON c2.vocabulary_id = 'NAACCR' AND (c2.concept_code = s.variable_id OR c2.concept_code = c1.concept_code || '-' || s.variable_id)
-- Getting permissible value
JOIN concept AS c3
ON c3.vocabulary_id = 'NAACCR' AND (c3.concept_code = s.variable_id ||  '-' || s.value OR c3.concept_code = c1.concept_code || '-'  ||
s.variable_id  || '-'  || s.value)
-- Getting date
join concept_relationship crd
on c2.concept_id = crd.concept_id_1 and crd.relationship_id = 'Variable has date'
join concept cd on crd.concept_id_2 = cd.concept_id
join data_source as sd
on s.patient_id = sd.patient_id and sd.variable_id = cd.concept_code
join condition_occurrence co
on co.person_id = s.patient_id
and co.condition_start_datetime = sd.value
and co.condition_concept_id = c2.concept_id
And co.condition_type_concept_id = 35941507
```

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 derivation of First Occurrence episode

```sql
Insert into episode
(person_id,
episode_start_datetime,
episode_concept_id,
episode_object_concept_id,
episode_type_concept_id)
select co.person_id,
co.condition_start_datetime,
32528, -- First occurrence
co.condition_concept_id,
35941507 -- Cancer registry
From condition_occurrence as co
join measurement as m on co.condition_occurrence_id = m.modifier_of_event_id
and m.modifier_concept_id = 'condition_occurence.condition_occurrence_id'
and m.measurement_concept_id = 32528
```

SQL for derivation of First Occurrence episode

```sql
Insert into episode_event
(episode_id,
condition_occurrence_id,
episode_event_event_id,
episode_field_concept_id)
select co.person_id,
co.condition_start_datetime,
32528, -- First occurrence
co.condition_concept_id,
35941507 -- Cancer registry
From condition_occurrence as co
join episode as e on e.person_id = co.person_id
and e.episode_concept_id = 32528
and e.episode_type_concept_id = 35941507
join measurement as m on co.condition_occurrence_id = m.modifier_of_event_id
and m.modifier_field_concept_id =
'condition_occurrence.condition_occurrence_id'
and m.measurement_concept_id = 32528
```