



Memorial Sloan Kettering  
Cancer Center

# Crosswalk from Cancer Registry to OMOP Oncology Module



Rimma Belenkaya

---

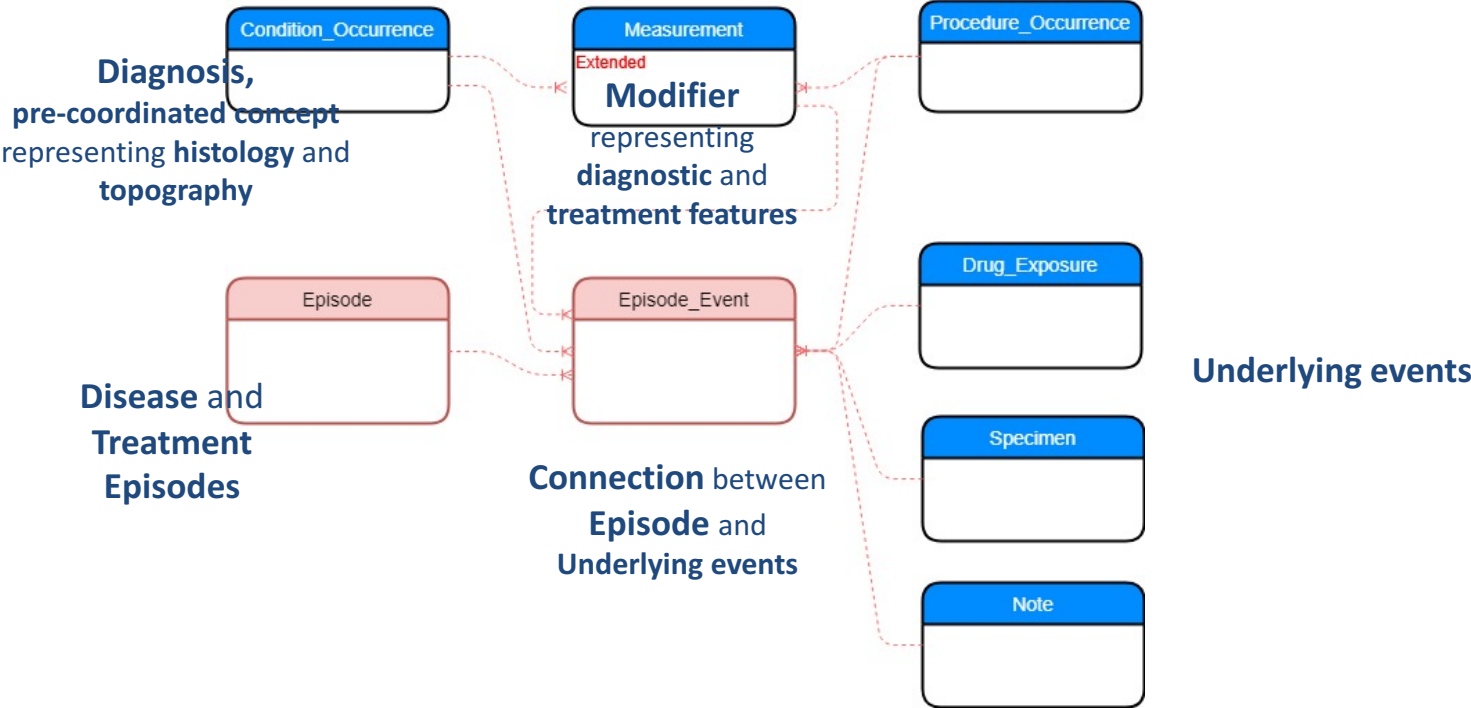


# Outline

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



# Oncology Model in OMOP CDM





# 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              |



# SQL for ETL of Diagnosis

```
Insert into condition_occurrence
(condition_occurrence_id,
person_id,
condition_start_datetime,
condition_concept_id,
condition_type_concept_id,
condition_source_value,
condition_source_concept_id)
select
row_number() over() as condition_occurrence_id,
s.person_id as person_id,
to_date(s.naaccr_item_value,'MM/DD/YYYY') as condition_start_datetime,
c2.concept_id as condition_concept_id,
32534 as condition_type_concept_id, -- 'Tumor registry' concept
s.histology_site as condition_source_value,
c1.concept_id as condition_source_concept_id
from data_source as s
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
and c1.concept_code = s.histology_site
and c1.vocabulary_id = 'ICD03'
where s.naaccr_item_number = '390'
order by s.person_id
```





## Modifier for the First Occurrence

- It is temporarily 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 for ETL of First Occurrence Modifier

## **Insert into measurement**

```
(measurement_id,  
person_id,  
measurement_datetime,  
measurement_concept_id,  
measurement_type_concept_id,  
measurement_source_value,  
modifier_of_event_id,  
modifier_field_concept_id)  
select distinct  
row_number() over() as measurement_id, fco.*  
from  
(select distinct  
co.person_id as person_id,  
co.condition_start_datetime as measurement_datetime,  
32528 as measurement_concept_id, -- 'Disease First Occurrence' concept  
32534 as measurement_type_concept_id, -- 'Tumor registry' concept  
'First disease occurrence' as measurement_source_value,  
co.condition_occurrence_id as modifier_of_event_id,  
1147127 as modifier_field_concept_id -- 'condition_occurrence.condition_occurrence_id' concept  
from data_source as s  
join condition_occurrence as co on s.person_id = co.person_id  
and to_date(s.naaccr_item_value,'MM/DD/YYYY') = co.condition_start_datetime  
and s.naaccr_item_number = '390'  
and co.condition_source_value = s.histology_site) as fco
```

# 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        |
| vocabulary_id | NAACCR     |

## 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 |



## SQL for ETL of Diagnostic Modifier: value as concept INSERT clause

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



## SQL for ETL of Diagnostic Modifier: value as concept SELECT clause

```
select row_number() over() as measurement_id, cm.*  
from (  
select distinct  
s.person_id as person_id  
, to_date(sd.naaccr_item_value,'MM/DD/YYYY') as measurement_datetime  
, c2.concept_id as measurement_concept_id  
, c3.concept_id as value_as_concept_id  
, 32534 as measurement_type_concept_id -- 'Tumor registry' concept  
, c2.concept_code as measurement_source_value  
, c2.concept_id as measurement_source_concept_id  
, c3.concept_code as value_source_value  
, co.condition_occurrence_id  
, 1147127 as modifier_field_concept_id --  
    'condition_occurrence.condition_occurrence_id' concept
```



# SQL for ETL of Diagnostic Modifier: value as concept FROM clause

```
FROM data_source AS s
-- Getting schema
JOIN concept d
ON d.vocabulary_id = 'ICD03' AND d.concept_code = s.histology_site
JOIN concept_relationship cr1
ON d.concept_id = cr1.concept_id_1 AND cr1.relationship_id = 'ICDO to Schema'
JOIN concept AS c1
ON cr1.concept_id_2 = c1.concept_id AND c1.vocabulary_id = 'NAACCR'
---- Getting variables
JOIN concept AS c2
ON c2.vocabulary_id = 'NAACCR' AND (c2.concept_code = s.naacrr_item_number OR c2.concept_code = c1.concept_code || '@' || s.naacrr_item_number)
-- Identify numeric type variables
left join concept_relationship cn
on c2.concept_id = cn.concept_id_1 and cn.relationship_id = 'Has type' and cn.concept_id_2 = 32676
---- Getting permissible value
JOIN concept AS c3
ON c3.vocabulary_id = 'NAACCR' AND (c3.concept_code = s.naacrr_item_number || '@' || s.naacrr_item_value OR c3.concept_code = c1.concept_code || '@' || s.naacrr_item_number || '@' || s.naacrr_item_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.person_id = sd.person_id and sd.naacrr_item_number = cd.concept_code
---- Getting condition_occurrence record

join
(select
co.person_id
, co.condition_occurrence_id
, s.record_id
from
(select
co.person_id
, co.condition_occurrence_id
, date(co.condition_start_datetime) as condition_start_date
, co.condition_source_value
, rank () OVER (PARTITION BY co.person_id ORDER BY co.person_id, co.condition_occurrence_id) AS occurrence_number
FROM condition_occurrence co
where co.condition_type_concept_id = 32534) as co
join
(select
s.person_id
, s.record_id
, to_date(s.naacrr_item_value, 'MM/DD/YYYY') as condition_start_date
, s.histology_site
, rank () OVER (PARTITION BY s.person_id ORDER BY s.person_id, s.record_id) AS occurrence_number
FROM data_source AS s
where s.naacrr_item_number = '390'
) as s
on co.person_id = s.person_id
and co.occurrence_number = s.occurrence_number
and co.condition_source_value = s.histology_site) as co
on co.person_id = s.person_id
and co.record_id = s.record_id
where cn.concept_id_1 is null -- excluding numeric types
) as cm
```



# 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

- NAACCR Values

| Scenario | NAACCR Values | Description   |
|----------|---------------|---|
| 1        | 001-988       | 001 - 988 millimeters (mm)<br>(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 Diagnostic Modifier: value as number

## INSERT clause

**Insert into** measurement

```
(measurement_id,  
person_id,  
measurement_datetime,  
measurement_concept_id,  
value_as_concept_id,  
value_as_number,  
unit_concept_id,  
operator_concept_id,  
measurement_type_concept_id,  
measurement_source_value,  
measurement_source_concept_id,  
value_source_value,  
modifier_of_event_id,  
modifier_field_concept_id)
```



# SQL for ETL of Diagnostic Modifier: value as number SELECT clause

```
select row_number() over() as measurement_id, cm.*  
from (  
select distinct  
  s.person_id as person_id  
  , to_date(sd.naaccr_item_value,'MM/DD/YYYY') as measurement_datetime  
  , c2.concept_id as measurement_concept_id  
  , c3.concept_id as value_as_concept_id  
  , case when c3.concept_id is null then cast(s.naaccr_item_value as integer) else cn.value_as_number  
    end as value_as_number  
  , case when c3.concept_id is null then cru.concept_id_2 else cn.unit_concept_id end as  
    measurement_unit_id  
  , case when c3.concept_id is null then null else cn.operator_concept_id end as operator_concept_id  
  , 32534 as measurement_type_concept_id -- 'Tumor registry' concept  
  , c2.concept_code as measurement_source_value  
  , c2.concept_id as measurement_source_concept_id  
  , c3.concept_code as value_source_value  
  , co.condition_occurrence_id  
  , 1147127 as modifier_field_concept_id -- 'condition_occurrence.condition_occurrence_id' concept
```

# SQL for ETL of Diagnostic Modifier: value as number FROM clause

```
FROM data_source AS s
-- Getting schema
JOIN concept d
ON d.vocabulary_id = 'ICDO3' AND d.concept_code = s.histology_site
JOIN concept_relationship cr1
ON d.concept_id = cr1.concept_id_1 AND cr1.relationship_id = 'ICDO 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.naacr_item_number OR c2.concept_code = c1.concept_code || '@' || s.naacr_item_number)
-- Constraining to numeric concepts only
JOIN concept_relationship AS crn
on c2.concept_id = crn.concept_id_1 and crn.relationship_id = 'Has type'
and crn.concept_id_2 = 32676 -- numeric
-- Getting units if exist
left JOIN concept_relationship AS cru
on c2.concept_id = cru.concept_id_1 and cru.relationship_id = 'Has unit'
-- Getting permissible value for ranges
left JOIN concept AS c3
ON c3.vocabulary_id = 'NAACCR' AND (c3.concept_code = s.naacr_item_number || '@' || s.naacr_item_value OR c3.concept_code = c1.concept_code || '@' || s.naacr_item_number || '@' || s.naacr_item_value)
left JOIN concept_numeric AS cn
ON c3.concept_id = cn.concept_id
---- 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.person_id = sd.person_id and sd.naacr_item_number = cd.concept_code
---- Getting condition_occurrence record
join
(select
co.person_id
, co.condition_occurrence_id
, s.record_id
from
(select
co.person_id
, co.condition_occurrence_id
, date(co.condition_start_datetime) as condition_start_date
, co.condition_source_value
, rank () OVER (PARTITION BY co.person_id ORDER BY co.person_id, co.condition_occurrence_id) AS occurrence_number
FROM condition_occurrence co
where co.condition_type_concept_id = 32534) as co
join
(select
s.person_id
, s.record_id
, to_date(s.naacr_item_value,'MM/DD/YYYY') as condition_start_date
, s.histology_site
, rank () OVER (PARTITION BY s.person_id ORDER BY s.person_id, s.record_id) AS occurrence_number
FROM data_source AS s
where s.naacr_item_number = '390'
) as s
on co.person_id = s.person_id
and co.occurrence_number = s.occurrence_number
and co.condition_source_value = s.histology_site) as co
on co.person_id = s.person_id
and co.record_id = s.record_id
) as cm
```



# 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

## **Insert into episode**

```
(episode_id,  
person_id,  
episode_start_datetime,  
episode_end_datetime,  
episode_concept_id,  
episode_object_concept_id,  
episode_type_concept_id)  
select row_number() over() as episode_id  
, co.person_id as person_id  
, co.condition_start_datetime as episode_start_datetime  
, now()  
, 32528 as episode_concept_id -- 'Disease First Occurrence' concept  
, co.condition_concept_id as episode_object_concept_id  
, 32534 as episode_type_concept_id -- 'Tumor registry' concept  
From condition_occurrence as co  
join measurement as m on co.condition_occurrence_id = m.modifier_of_event_id  
and m.modifier_field_concept_id = 1147127 -- condition_occurrence.condition_occurrence_id  
and m.measurement_concept_id = 32528 -- 'First disease occurrence' concept
```



## SQL for derivation of First Occurrence: Episode\_Event

```
Insert into episode_event
(episode_id,
event_id,
event_field_concept_id)
select e.episode_id as episode_id
, co.condition_occurrence_id as episode_event_id
, 1147127 as episode_event_field_concept_id --
'condition_occurrence.condition_occurrence_id' concept
From condition_occurrence as co
join episode as e on e.person_id = co.person_id
and e.episode_concept_id = 32528 -- 'Disease first occurrence'
and e.episode_type_concept_id = 32534 -- 'Tumor registry'
join measurement as m on co.condition_occurrence_id = m.modifier_of_event_id
and m.modifier_field_concept_id = 1147127 --
condition_occurrence.condition_occurrence_id
and m.measurement_concept_id = 32528
```



# SQL for derivation of First Occurrence modifiers

## INSERT clause

**Insert into** measurement

```
(measurement_id,  
person_id,  
measurement_datetime,  
measurement_concept_id,  
value_as_concept_id,  
value_as_number,  
unit_concept_id,  
operator_concept_id,  
measurement_type_concept_id,  
measurement_source_value,  
measurement_source_concept_id,  
value_source_value,  
modifier_of_event_id,  
modifier_field_concept_id)
```



# SQL for derivation of First Occurrence modifiers

## INSERT clause

```
select row_number() over() as measurement_id,  
m.person_id,  
m.measurement_datetime,  
m.measurement_concept_id,  
m.value_as_concept_id,  
m.value_as_number,  
m.unit_concept_id,  
m.operator_concept_id,  
m.measurement_type_concept_id,  
m.measurement_source_value,  
m.measurement_concept_id,  
m.value_source_value,  
e.episode_id ,  
0 as episode_event_field_concept_id -- 'condition_occurrence.condition_occurrence_id' concept  
From episode as e  
join episode_event as ee on e.episode_id = ee.episode_id  
join condition_occurrence as co on ee.event_id = co.condition_occurrence_id  
and ee.event_field_concept_id = 1147127  
join measurement as m on co.condition_occurrence_id = m.modifier_of_event_id  
and m.modifier_field_concept_id = 1147127
```