1. **ETL of diagnosis and diagnostic modifiers into CONDITION\_OCCURRENCE& MEASUREMENT.** 
   1. **Ingest condition occurrence**

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, -- use your sequence generator instead

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 NAACCR\_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 ='ICDO3'

where s.naaccr\_item\_number = '390'

order by s.person\_id

* 1. **Ingest first disease 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, -- use your sequence generator instead

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 NAACCR\_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

* 1. **Ingest diagnostic modifiers with values as concepts**

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)

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

FROM NAACCR\_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 variables

JOIN concept AS c2

ON c2.vocabulary\_id = 'NAACCR' AND (c2.concept\_code = s.naaccr\_item\_number OR c2.concept\_code = c1.concept\_code || '@' || s.naaccr\_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.naaccr\_item\_number || '@' || s.naaccr\_item\_value OR c3.concept\_code = c1.concept\_code || '@' || s.naaccr\_item\_number || '@' || s.naaccr\_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 NAACCR\_DATA\_SOURCE as sd

on s.person\_id = sd.person\_id and sd.naaccr\_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.naaccr\_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 NAACCR\_DATA\_SOURCE AS s

where s.naaccr\_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

* 1. **Ingest diagnostic modifiers with values as number**

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)

select row\_number() over() as measurement\_id, -- use your sequence generator instead

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

FROM NAACCR\_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.naaccr\_item\_number OR c2.concept\_code = c1.concept\_code || '@' || s.naaccr\_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.naaccr\_item\_number || '@' || s.naaccr\_item\_value OR c3.concept\_code = c1.concept\_code || '@' || s.naaccr\_item\_number || '@' || s.naaccr\_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 NAACCR\_DATA\_SOURCE as sd

on s.person\_id = sd.person\_id and sd.naaccr\_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.naaccr\_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 NAACCR\_DATA\_SOURCE AS s

where s.naaccr\_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

1. **Post-ETL of disease into EPISODE, EPISODE\_EVENT, and MEASUREMENT**
   1. **Ingest first disease occurrence in 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\_occurence.condition\_occurrence\_id

and m.measurement\_concept\_id = 32528 -- ‘First disease occurrence’ concept

* 1. **Ingest connection between EPISODE record and CONDITION\_OCCURRENCE record**

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

* 1. **Ingest a “clean” set of EPISODE modifiers into MEASUREMENT**

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)

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 ,

999999 as episode\_event\_field\_concept\_id -- ‘condition\_occurrence.condition\_occurrence\_id’ concept to be created

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