[Query algorithms for structured data in MIMIC II are performed using [Oracle SQL].

\* The following SQL code is composed of separate small tables queried from the main relational database(MIMIC-II) by the “select-from-where” statements. Each small tables select a sub-cohort by imposing conditions in the “where” statement, i.e. narrowing down to the target study population according to the study questions raised by clinicians. Each small tables can interact with each other or the main tables in the database by being the source of the “from” statement.

\* The following query is a truncated version showing only key steps is cohort selection.

-- Get the first ICU stay of each adult by imposing conditions on time stamps of hospital admission or ICU admission, some structured demographic data were also selected

with first\_admissions as

(select icud.subject\_id,

icud.hadm\_id as first\_hadm\_id,

icud.icustay\_id as first\_icustay\_id,

icud.hospital\_admit\_dt as first\_hadm\_admit\_dt,

icud.hospital\_disch\_dt as first\_hadm\_disch\_dt,

icud.icustay\_intime as first\_icustay\_intime,

icud.icustay\_outtime as first\_icustay\_outtime,

extract(day from icud.icustay\_intime-icud.hospital\_admit\_dt) as days\_btw\_hosp\_icu\_admit,

extract(day from icud.hospital\_disch\_dt-icud.icustay\_outtime) as days\_btw\_hosp\_icu\_disch,

case

when lower(d.admission\_source\_descr) like '%emergency%' then 'Y'

else 'N'

end as ED\_admission,

case

when icud.icustay\_admit\_age>150 then 91.4

else round(icud.icustay\_admit\_age,1)

end as age\_first\_icustay,

icud.gender,

d.ethnicity\_descr as race,

round(icud.icustay\_los/60/24,2) as first\_icustay\_los,

icud.icustay\_first\_service,

icud.sapsi\_first as first\_icustay\_admit\_saps,

case

when msa.hadm\_id is not null then 'Y'

else 'N'

end as first\_hadm\_sepsis

from mimic2v26.icustay\_detail icud

left join mimic2v26.demographic\_detail d on icud.hadm\_id=d.hadm\_id

left join mimic2devel.martin\_sepsis\_admissions msa on icud.hadm\_id=msa.hadm\_id

where icud.subject\_icustay\_seq=1

and icud.icustay\_age\_group='adult'

)

--select \* from first\_admissions;

-- several covariates were selected in the following step. These covariates can then be used in the regression model for adjustment if clinicians feel like these covariates are possible confounders for the study outcome. Besides using clinicians’ subjective judgment, there are other objective methods to determine if a covariate is a possible confounder, which is beyond the scope of this chapter.

, raw\_icu\_admit\_labs as

(select distinct r.subject\_id,

case

when l.itemid=50090 then 'serum\_cr'

when l.itemid=50159 then 'serum\_sodium'

when l.itemid=50655 then 'urine\_protein'

when l.itemid=50264 then 'urine\_cr'

when l.itemid=50277 then 'urine\_sodium'

when l.itemid=50276 then 'urine\_protein\_cr\_ratio'

when l.itemid=50177 then 'bun'

when l.itemid=50149 then 'potassium'

when l.itemid=50083 then 'chloride'

when l.itemid=50172 then 'bicarb'

when l.itemid=50383 then 'hematocrit'

when l.itemid=50468 then 'wbc'

when l.itemid=50140 then 'magnesium'

when l.itemid=50148 then 'phosphate'

when l.itemid=50079 then 'calcium'

when l.itemid=50010 then 'lactate'

when l.itemid=50018 then 'ph'

when l.itemid=50428 then 'platelets'

when l.itemid=50060 then 'albumin'

when l.itemid=50112 then 'glucose'

when l.itemid=50399 then 'inr'

when l.itemid=50439 then 'pt'

when l.itemid=50440 then 'ptt'

when l.itemid=50115 then 'haptoglobin'

when l.itemid=50134 then 'ldh'

when l.itemid=50370 then 'd-dimer'

end as lab\_type,

first\_value(l.value) over (partition by l.hadm\_id, l.itemid order by l.charttime) as lab\_value,

first\_value(l.charttime) over (partition by l.hadm\_id order by l.charttime) as icu\_admit\_lab\_time

from first\_admissions r

join mimic2v26.labevents l

on r.first\_hadm\_id=l.hadm\_id

and l.itemid in (50090,50159,50655,50264,50277,50276,50177,50149,50083,50172,50383,50468,50140,50148,50079,50010,50018,50428,50060,50112,50399,50439,50440,50115,50134,50370)

and l.charttime between r.first\_icustay\_intime - interval '12' hour and r.first\_icustay\_intime + interval '12' hour

)

--select \* from raw\_icu\_admit\_labs order by 1,2;

-- Get peak creatinine values from first ICU stays, This is an example that certain variables, although stored as structured data in the database, might require multiple conditions being imposed in multiple steps in order to select the desired value that’s relevant to the study question. For example, each patient would have multiple creatinine values being measured during each ICU course. The study question might call for the first available value or the highest numeric value for research purpose.

, peak\_creat\_first\_icustay as

(select distinct r.subject\_id,

count(\*) as num\_cr\_first\_icustay,

max(l.valuenum) as cr\_peak\_first\_icustay

from first\_admissions r

join mimic2v26.labevents l on r.first\_icustay\_id=l.icustay\_id and l.itemid=50090

group by r.subject\_id

)

--select \* from peak\_creat\_first\_icustay;

-- Sometimes, the variable required by the study question does not exist as either structured or unstructured data, but rather as a calculated result of available data. One example would be “acute kidney injury(AKI) stage by hourly urine output.” In this case, it would require additional algorithm to calculate hourly urine output and determine if a patient sustained AKI and was in what stage of the AKI. Using database language like SQL to compose these algorithms could be extremely complicated and buggy. We would recommend using other interpreted language like Matlab to accomplish such task, and using SQL to export filtered raw data for further processing.

-- Total urine output during the first ICU stay

, uo\_first\_icustay as

(select r.subject\_id,

sum(ie.volume) as urine\_first\_icustay

from first\_admissions r

join mimic2v26.ioevents ie on r.first\_icustay\_id=ie.icustay\_id

where ie.itemid in ( 651, 715, 55, 56, 57, 61, 65, 69, 85, 94, 96, 288, 405, 428, 473, 2042, 2068, 2111, 2119, 2130, 1922, 2810, 2859, 3053, 3462, 3519, 3175, 2366, 2463, 2507, 2510, 2592, 2676, 3966, 3987, 4132, 4253, 5927 )

group by r.subject\_id

)

--select \* from uo\_first\_icustay;

, uo\_first\_icustay\_24h as

(select r.subject\_id,

sum(ie.volume) as urine\_first\_icustay\_24h

from first\_admissions r

join mimic2v26.ioevents ie on r.first\_icustay\_id=ie.icustay\_id and ie.charttime between r.first\_icustay\_intime and r.first\_icustay\_intime + interval '24' hour

where ie.itemid in ( 651, 715, 55, 56, 57, 61, 65, 69, 85, 94, 96, 288, 405, 428, 473, 2042, 2068, 2111, 2119, 2130, 1922, 2810, 2859, 3053, 3462, 3519, 3175, 2366, 2463, 2507, 2510, 2592, 2676, 3966, 3987, 4132, 4253, 5927 )

group by r.subject\_id

)

--select \* from uo\_first\_icustay\_24h;