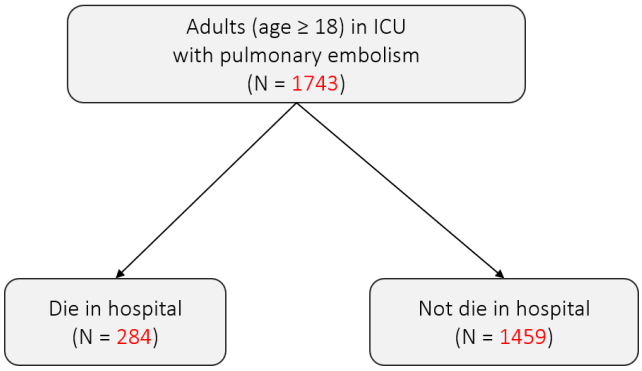
2023 NTHU Data Science for Digital Health Homework 1

(Deadline: Nov. 3 23:59pm)

#1 Predicting Mortality of Pulmonary Embolism Patients in the ICU

(1) 18%



The cohort selected based on the inclusion and exclusion criteria can be acquired by sending the query below:

```
WITH cohort AS (SELECT DISTINCT icu.subject_id, icu.hadm_id, icu.stay_id, icu.hospital_expire_flag, rank()
over (PARTITION BY icu.subject_id ORDER BY stay_id) as rank
FROM `physionet-data.mimiciv_derived..icustay_detail`  as icu -- patients in ICU
  INNER JOIN `physionet-data.mimiciv_hosp.diagnoses_icd` as diagnosis ON diagnosis.subject_id = icu.subject_id
AND diagnosis.hadm_id = icu.hadm_id
  WHERE icu.admission_age >= 18 -- patients age more than 18
    AND (diagnosis.icd_code LIKE '4151%' OR diagnosis.icd_code LIKE 'I26%') -- patients with Pulmonary
Embolism
),
data as (
SELECT cohort.subject_id, cohort.hadm_id, cohort.stay_id, cohort.hospital_expire_flag, patients.gender, CASE
    WHEN admission_age < 20 THEN '<20'
    WHEN admission_age <= 40 AND admission_age > 20 THEN '21~40'
    WHEN admission_age <= 60 AND admission_age > 40 THEN '41~60'
    WHEN admission_age <= 80 AND admission_age > 60 THEN '61~80'
    ELSE '>80' END AS age,
  CASE
    WHEN race LIKE 'WHITE%' THEN 'WHITE'
    WHEN race LIKE 'BLACK%' THEN 'BLACK'
    WHEN race LIKE 'YELLOW%' THEN 'YELLOW'
    WHEN race LIKE 'HISPANIC%' THEN 'HISPANIC'
    ELSE 'OTHER' END AS race,
FROM cohort
  INNER JOIN `physionet-data.mimiciv_derived.icustay_detail` as icu ON icu.subject_id = cohort.subject_id AND
icu.stay_id = cohort.stay_id AND icu.hadm_id = cohort.hadm_id
  INNER JOIN `physionet-data.mimiciv_hosp.patients` as patients ON icu.subject_id = patients.subject_id
  WHERE rank = 1 -- first admission in icu
)
SELECT * FROM data
```

In `cohort`, we first filtered patients with ages more than 18 and selected those with Pulmonary Embolism (PE) with ICD codes. Then, select the first record of each patient.
To obtain in-hospital death data, we can use `icu.hospital_expire_flag` to determine whether patients died in the hospital. When `icu.hospital_expire_flag` equals `1`, the patient dies in the hospital. Otherwise, the patient has not died in the hospital. We grouped the patients with their in-hospital mortality and counted the total number using the dataset obtained from the above.

```
SELECT hospital_expire_flag, COUNT(subject_id) FROM data
GROUP BY hospital_expire_flag
```

(2) 23%

		Outcome		
		Alive	Die	Total
Gender	Male	742	134	876
	Female	717	150	867
Age	0-20	6	0	6
	21-40	173	14	187
	41-60	410	61	471
	61-80	660	141	801
	80+	210	68	278
Race	Black	188	36	224
	White	998	169	1167
	Asian	29	10	39
	Hispanic	45	8	53
	Others	199	61	260

We can also calculate the number of deaths in different genders, age groups, and race groups by grouping.

- Gender

```
SELECT gender, hospital_expire_flag, COUNT(subject_id) FROM data
GROUP BY gender, hospital_expire_flag
```

- Age

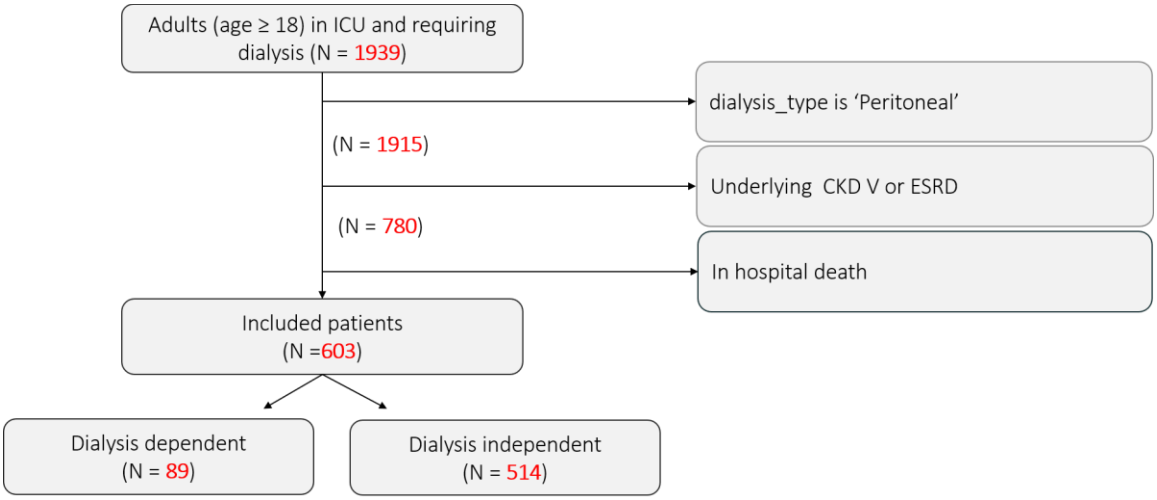
```
SELECT age, hospital_expire_flag, COUNT(subject_id) FROM data
GROUP BY age, hospital_expire_flag
```

- Race

```
SELECT race, hospital_expire_flag, COUNT(subject_id) FROM data
GROUP BY race, hospital_expire_flag
```

#2 Renal Recovery Prediction Following Initiation of Renal Replacement Therapy

(1) 36%



The cohort can be obtained by:

```
WITH c AS (
  SELECT DISTINCT icu.subject_id, icu.hadm_id, icu.stay_id, icu.hospital_expire_flag
  FROM `physionet-data.mimiciv_derived..icustay_detail` as icu -- patients in ICU
  INNER JOIN `physionet-data.mimiciv_hosp.poe` as poe ON poe.subject_id = icu.subject_id AND poe.hadm_id = icu.hadm_id
  WHERE poe.order_type = 'Hemodialysis'
    AND icu.admission_age >= 18 -- patients age more than 18
),
cohort AS (
  SELECT DISTINCT subject_id, hadm_id, stay_id, rank() over (PARTITION BY c.subject_id ORDER BY c.stay_id) as rank FROM
c
  WHERE NOT EXISTS (
      SELECT 1
      FROM `physionet-data.mimiciv_derived.rrt` AS rrt
      WHERE rrt.stay_id = c.stay_id
            AND (rrt.dialysis_type = 'Peritoneal')
      )
  AND NOT EXISTS (
      SELECT 1
      FROM `physionet-data.mimiciv_hosp.diagnoses_icd` as diagnosis
      WHERE diagnosis.subject_id = c.subject_id AND diagnosis.hadm_id = c.hadm_id
            AND (diagnosis.icd_code LIKE '40301%' OR diagnosis.icd_code LIKE '40311%' OR diagnosis.icd_code LIKE
'40391%' OR diagnosis.icd_code LIKE '40402%' OR diagnosis.icd_code LIKE '40403%' OR diagnosis.icd_code  LIKE '40412%'
OR diagnosis.icd_code  LIKE '40413%' OR diagnosis.icd_code LIKE '40492%' OR diagnosis.icd_code LIKE '40493%' OR
diagnosis.icd_code LIKE '5856%' OR diagnosis.icd_code LIKE 'I120%' OR diagnosis.icd_code LIKE 'I1311%' OR
diagnosis.icd_code LIKE 'I132%' OR diagnosis.icd_code LIKE 'N186%')
      )
    AND hospital_expire_flag = 0
),
last_dialysis AS (
  SELECT subject_id, hadm_id, MAX(ordertime) as last_order
  FROM `physionet-data.mimiciv_hosp.poe` as poe
  GROUP BY subject_id, hadm_id
),
data AS (
  SELECT DISTINCT cohort.subject_id, cohort.hadm_id, cohort.stay_id, patients.gender,
  CASE
    WHEN admission_age < 20 THEN '<20'
    WHEN admission_age <= 40 AND admission_age > 20 THEN '21~40'
    WHEN admission_age <= 60 AND admission_age > 40 THEN '41~60'
    WHEN admission_age <= 80 AND admission_age > 60 THEN '61~80'
    ELSE '>80' END AS age,
  CASE
    WHEN race LIKE 'WHITE%' THEN 'WHITE'
    WHEN race LIKE 'BLACK%' THEN 'BLACK'
    WHEN race LIKE 'YELLOW%' THEN 'YELLOW'
    WHEN race LIKE 'HISPANIC%' THEN 'HISPANIC'
    ELSE 'OTHER' END AS race,
  CASE WHEN date_diff(last_dialysis.last_order, icu.icu_outtime, hour) >= 72 THEN 0 ELSE 1 END AS dependency
FROM cohort
 INNER JOIN `physionet-data.mimiciv_derived.icustay_detail` as icu ON icu.subject_id = cohort.subject_id AND
icu.stay_id = cohort.stay_id AND icu.hadm_id = cohort.hadm_id
 INNER JOIN `physionet-data.mimiciv_hosp.patients` as patients ON icu.subject_id = patients.subject_id
 INNER JOIN last_dialysis ON last_dialysis.subject_id = icu.subject_id AND last_dialysis.hadm_id = icu.hadm_id
 WHERE rank = 1 -- first admission in icu
)
SELECT * FROM data
```

First, we selected patients under 18 who needed initiation of inpatient dialysis in the ICU. Excluded those who had already initiated peritoneal dialysis, had chronic kidney disease (CKD) stage V or end-stage renal disease (ESRD), and died before discharge. Labeled the dialysis dependency by calculating the time from the last dialysis order to ICU discharge. If the time is longer than 72 hours, define it to be independent. Otherwise, it is dialysis dependent. For the number of dialysis-dependent and independent patients, use the following code.

```
SELECT dependency, COUNT(subject_id) FROM data
GROUP BY dependency
```

(2) 23%

		Outcome		
		Not recover	Recover	Total
Gender	Male	325	61	386
	Female	189	28	217
Age	0-20	0	1	1
	21-40	58	5	63
	41-60	181	27	208
	61-80	233	42	275
	80+	42	14	56
Race	Black	58	9	67
	White	332	60	392
	Asian	11	1	12
	Hispanic	18	4	22
	Others	95	15	110

Similarly, calculate the number of deaths in different genders, age groups, and race groups by grouping.

- Gender

```
SELECT gender, dependency, COUNT(subject_id) FROM data
GROUP BY gender, dependency
```

- Age

```
SELECT age, dependency, COUNT(subject_id) FROM data
GROUP BY age, dependency
```

- Race

```
SELECT race, dependency, COUNT(subject_id) FROM data
GROUP BY race, dependency
```