Characterize a Given OMOP CDM Database

Initial Set-Up

Activating Project Environment

```
knitr::opts_chunk$set(echo = TRUE, message = F, results = 'hold', cache = T)
options(scipen=999)
devtools::load_all()
```

```
## i Loading MentalHealthEquity
```

```
library(MentalHealthEquity)
```

Packages Used in Analysis

```
library(DatabaseConnector)
```

`DatabaseConnector`

Package Description:

This package will be used to create connections to the database and execute queries to obtain characterizations about the cohorts in the database.

Introduction

Background

The intent of this report is to investigate the characterization statement:

> **Characterization Statement 1:** Characterize the individuals being seen for mental health care services (related to depression, bipolar disorder, and suicidality) at least one time – including hospitalization events.

This characterization statement is founded on the central research topic for this study:

> **Research Topic:** Based on [CLAIMS], we see X% of all persons with at least one claim indicating [DEPRESSION/BIPOLAR DISORDER/SUICIDALITY] are not seen again.

By which the phrases "…all persons…" refer to those seen by patience care provider, etc. and "…are not seen again." implies lack of adherence to care.

Data Analysis Preparation

Creating Initial Connection

Defining connection details for connecting to a given database:

```
connectionDetails <- createConnectionDetails(
  dbms = "postgresql",
  server = "data.hdap.gatech.edu/synpuf_v5",
  user = keyring::key_list("synPUF")[1,2],
  password = keyring::key_get("synPUF", keyring::key_list("synPUF")[1,2]),
  port = 5434,
  pathToDriver = "../utils"
)
```

Creating Initial OMOP Tables

In this case, the schema follows the OMOP CDM v5 schema:

For this analysis, we will work with the following tables from the schema:

Per the [Book of OHDSI], here are the break downs for these tables:

- PERSON -
- LOCATION -
- OBSERVATION_PERIOD -
- CONDITION_OCCURRENCE -

General Characterization of Data

```
sql <- "
  SELECT {@x} ? {@table.@concept_name, count(DISTINCT PERSON_ID)} : {*}
  FROM person
      LEFT JOIN (SELECT LOCATION_ID, STATE FROM location) AS location ON
        person.location_id = location.location_id
      LEFT JOIN (SELECT CONCEPT_ID,
                    CONCEPT_NAME as RACE_CONCEPT_NAME
                  FROM concept WHERE DOMAIN_ID = 'Race'
                  AND STANDARD_CONCEPT = 'S') AS race ON
        person.race_concept_id = race.concept_id
      LEFT JOIN ( SELECT CONCEPT_ID,
                    CONCEPT_NAME as ETHNICITY_CONCEPT_NAME
                  FROM concept WHERE DOMAIN_ID = 'Ethnicity'
                    AND STANDARD_CONCEPT = 'S') AS ethnicity ON
        person.ethnicity_concept_id = ethnicity.concept_id
      LEFT JOIN (SELECT CONCEPT_ID,
                    CONCEPT_NAME as GENDER_CONCEPT_NAME
                  FROM concept WHERE DOMAIN_ID = 'Gender'
                  AND STANDARD_CONCEPT = 'S') AS gender ON
        person.gender_concept_id = gender.concept_id
  {@x}? {GROUP BY @table.@concept_name
        ORDER BY COUNT DESC} :
        {LIMIT 10}
"
```

Cohort Populations

General Cohort

Overall Trends

```
conn <- connect(connectionDetails)
date_sql <- "
  SELECT  MIN(visit.visit_end_date) as first_visit,
          MAX(visit.visit_end_date) as last_visit
  FROM visit_occurrence visit
"
dates <- renderTranslateQuerySql(conn, date_sql)
disconnect(conn)
```

The dates in the dataset range from 2008-01-01 and 2010-12-31

Patient Spread This calculates the number of patients in the dataset:

```
conn <- connect(connectionDetails)
count <- querySql(conn, sql = "SELECT COUNT(DISTINCT PERSON_ID) FROM PERSON")
disconnect(conn)
```

In this dataset, there are 2326856 patients. To further examine this data, we can break them down across the following axes:

- State
- Age
- Race
- Gender
- Care setting

```
conn <- connect(connectionDetails)
state <- renderTranslateQuerySql(conn, sql, table = "location",
                                 concept_name = "state", x = T)
disconnect(conn)
```

State Breakdown The top states are: CA, FL, TX, NY, PA, IL, OH, MI, NC, NJ

STATE	COUNT
CA	201651
FL	155040
TX	133809
NY	132059
PA	102191
IL	86557
OH	86325
MI	79556
NC	78912
NJ	63475

```
conn <- connect(connectionDetails)
age_sql <- "
```

```
  SELECT  MIN(YEAR(visit.visit_end_date) - person.year_of_birth) as min_age,
          MAX(YEAR(visit.visit_end_date) - person.year_of_birth) as max_age
  FROM person
    JOIN visit_occurrence visit ON
      person.person_id = visit.person_id
"
age <- renderTranslateQuerySql(conn, age_sql)
disconnect(conn)
```

Age Breakdown The age of the patient at the time of visit ranged from 25 years of age to 101 years of age.

```
conn <- connect(connectionDetails)
race <- renderTranslateQuerySql(conn, sql, table = "race",
                                concept_name= "RACE_CONCEPT_NAME", x = T)
disconnect(conn)
```

Racial Breakdown In the dataset, there are 1926708 White persons, 247723 Black or African American persons, 152425 NA persons

RACE_CONCEPT_NAME	COUNT
White	1926708
Black or African American	247723
NA	152425

```
conn <- connect(connectionDetails)
gender <- renderTranslateQuerySql(conn, sql, table = "gender",
                                  concept_name = "GENDER_CONCEPT_NAME", x = T)
disconnect(conn)
```

Gender Breakdown In the dataset, there are 1292861 FEMALE persons, 1033995 MALE persons

GENDER_CONCEPT_NAME	COUNT
FEMALE	1292861
MALE	1033995

```
conn <- connect(connectionDetails)
caresite_sql <- "
  SELECT concept.concept_name, count(*) VISIT_COUNT,
    count(DISTINCT visit.person_id) PERSON_COUNT
  FROM visit_occurrence visit
    JOIN (SELECT CONCEPT_NAME , CONCEPT_ID
          FROM concept
          WHERE DOMAIN_ID = 'Visit'
              AND STANDARD_CONCEPT = 'S') concept ON
      visit.visit_concept_id = concept.concept_id
  GROUP BY concept.concept_name
  ORDER BY count(*) DESC
"
```

```
care_site <- renderTranslateQuerySql(conn, caresite_sql)
disconnect(conn)
```

Care Site Breakdown In the dataset, there are 1984582 unique patients with an indicated care site. The care site setting Inpatient Visit had 111637582 visits

CONCEPT_NAME	VISIT_COUNT	PERSON_COUNT
Inpatient Visit	111637582	1984582

Inpatient Cohort

Overall Trends

Patient Spread This calculates the number of patients in the dataset:

```
conn <- connect(connectionDetails)
inpatient_sql <- "
  SELECT @query
  FROM (SELECT *
        FROM visit_occurrence visit
        JOIN
          ( SELECT CONCEPT_NAME , CONCEPT_ID
            FROM concept
            WHERE DOMAIN_ID = 'Visit'
            AND STANDARD_CONCEPT = 'S') concept
        ON visit.VISIT_CONCEPT_ID = concept.CONCEPT_ID
        WHERE LOWER(concept.CONCEPT_NAME) LIKE '%@a%'
        ) VISIT_INFO
  JOIN
    ( SELECT *
      FROM person
          LEFT JOIN (SELECT LOCATION_ID, STATE FROM location) AS location ON
            person.location_id = location.location_id
          LEFT JOIN (SELECT CONCEPT_ID,
                        CONCEPT_NAME as RACE_CONCEPT_NAME
                      FROM concept WHERE DOMAIN_ID = 'Race'
                      AND STANDARD_CONCEPT = 'S') AS race ON
            person.race_concept_id = race.concept_id
          LEFT JOIN ( SELECT CONCEPT_ID,
                        CONCEPT_NAME as ETHNICITY_CONCEPT_NAME
                      FROM concept WHERE DOMAIN_ID = 'Ethnicity'
                        AND STANDARD_CONCEPT = 'S') AS ethnicity ON
            person.ethnicity_concept_id = ethnicity.concept_id
          LEFT JOIN (SELECT CONCEPT_ID,
                        CONCEPT_NAME as GENDER_CONCEPT_NAME
                      FROM concept WHERE DOMAIN_ID = 'Gender'
                      AND STANDARD_CONCEPT = 'S') AS gender ON
            person.gender_concept_id = gender.concept_id) DEMO_INFO
  USING(PERSON_ID)
  {@x} ? {GROUP BY DEMO_INFO.@concept_name
          ORDER BY COUNT DESC}
"
inpatient <- renderTranslateQuerySql(conn,
                                     inpatient_sql,
                                     x = F, a = "inpatient",
                                     query = "COUNT(*) VISIT_COUNT,
                                     COUNT(DISTINCT PERSON_ID) PERSON_COUNT
                                     ")
disconnect(conn)
```

In the inpatient cohort, there are a total of 1984582 distinct patients, and there are a total of 111637582 entries. Compared to the original unique person count in the overall dataset, which is 2326856, 342274 are not accounted for in the inpatient count.

To further examine this data, we can break them down across the following axes:

- State
- Age
- Race
- Gender
- Care setting

```
conn <- connect(connectionDetails)
inpatient_state <- renderTranslateQuerySql(conn, inpatient_sql, x = T, a = "inpatient",
                        query = "DEMO_INFO.state, COUNT(DISTINCT PERSON_ID)",
                        concept_name = "state")
disconnect(conn)
```

State Breakdown In the inpatient cohort, the top 5 states by person count are

STATE	COUNT
CA	178254
FL	140385
TX	123530
NY	119443
PA	90893
IL	79577
OH	77830
MI	70896
NC	65822
NJ	56971

```
conn <- connect(connectionDetails)
inpatient_age <- renderTranslateQuerySql(conn, inpatient_sql,
                                   x = F, a = "inpatient",
                                   query = "
                MIN(YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth) as min_age,
                MAX(YEAR(VISIT_INFO.visit_end_date) - DEMO_INFO.year_of_birth)
                  as max_age,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth < 18 THEN 1 END) AS less_than_18,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth BETWEEN 18 AND 24 THEN 1 END) AS age_18_to_24,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth BETWEEN 25 AND 34 THEN 1 END) AS age_25_to_35,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth BETWEEN 35 AND 44 THEN 1 ELSE 0 END) AS age_35_44,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth BETWEEN 45 AND 54 THEN 1 ELSE 0 END) AS age_45_54,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth BETWEEN 55 AND 64 THEN 1 ELSE 0 END) AS age_55_64,
                SUM(CASE WHEN YEAR(VISIT_INFO.visit_end_date) -
                  DEMO_INFO.year_of_birth >= 65 THEN 1 ELSE 0 END) AS age_65_and_more")
disconnect(conn)
```

Age Breakdown In the inpatient dataset, it takes in the visit encounters, and the ages for the individuals were found according to the time of the inpatient visit. The youngest individual was of age 25, and the oldest individual was of age 101

```
conn <- connect(connectionDetails)
inpatient_race <- renderTranslateQuerySql(conn, inpatient_sql,
                                          x = T,
                                          a = "inpatient",
                                          concept_name = "RACE_CONCEPT_NAME",
                                          query = "
                                          DEMO_INFO.RACE_CONCEPT_NAME,
                                          COUNT(DISTINCT PERSON_ID)")
disconnect(conn)
```

Racial Breakdown In the inpatient cohort, there are 1664596 White persons, 198723 Black or African American persons, 121263 NA persons

RACE_CONCEPT_NAME	COUNT
White	1664596
Black or African American	198723
NA	121263

```
conn <- connect(connectionDetails)
inpatient_gender <- renderTranslateQuerySql(conn, inpatient_sql,
                        x = T, a = "inpatient",
                        concept_name = "GENDER_CONCEPT_NAME",
                        query = "
                        DEMO_INFO.GENDER_CONCEPT_NAME,
                        COUNT(DISTINCT PERSON_ID)
                        ")
disconnect(conn)
```

Gender Breakdown In the inpatient cohort, there are 1131807 FEMALE persons, 852775 MALE persons

GENDER_CONCEPT_NAME	COUNT
FEMALE	1131807
MALE	852775