# Characterize a Given OMOP CDM Database

## Initial Set-Up

### Activating Project Environment

```
knitr::opts_chunk$set(echo = TRUE, message = F, results = 'hold')
options(scipen=999)
devtools::load_all()
```

```
## i Loading MentalHealthEquity
```

```
library(MentalHealthEquity)
```

### Packages Used in Analysis

`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 @table.@concept_name, count(*)
  FROM person
      LEFT JOIN (select LOCATION_ID, STATE FROM location) AS location ON
        person.location_id = location.location_id
      LEFT JOIN (select * FROM concept WHERE DOMAIN_ID = 'Race'
              AND STANDARD_CONCEPT = 'S') AS race ON
        person.race_concept_id = race.concept_id
      LEFT JOIN (select * FROM concept WHERE DOMAIN_ID = 'Ethnicity'
              AND STANDARD_CONCEPT = 'S') AS ethnicity ON
        person.ethnicity_concept_id = ethnicity.concept_id
      LEFT JOIN (select * FROM concept WHERE DOMAIN_ID = 'Gender'
              AND STANDARD_CONCEPT = 'S') AS gender ON
        person.gender_concept_id = gender.concept_id
  GROUP BY @table.@concept_name
  ORDER BY count(*) DESC
"
```

# 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)
dates
```

```
##   FIRST_VISIT LAST_VISIT
## 1  2008-01-01 2010-12-31
```

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")
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
```

**Age Breakdown**

```
##   MIN_AGE MAX_AGE
## 1      25     101
```

The age of the patient at the time of visit ranged from years of age to years of age.

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

**Racial Breakdown**

```
##                     CONCEPT_NAME   COUNT
## 1                          White 1926708
## 2 Black or African American  247723
## 3                         <NA>  152425
```

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

| CONCEPT_NAME | COUNT |
| --- | --- |
| White | 1926708 |
| Black or African American | 247723 |
| NA | 152425 |

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

**Gender Breakdown**

```
##   CONCEPT_NAME   COUNT
## 1       FEMALE 1292861
## 2         MALE 1033995
```

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

| CONCEPT_NAME | COUNT |
| --- | --- |
| FEMALE | 1292861 |
| MALE | 1033995 |

```
conn <- connect(connectionDetails)
caresite_sql <- "
  SELECT concept.concept_name, 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
```

**Care Site Breakdown**

```
##      CONCEPT_NAME     COUNT
## 1 Inpatient Visit 111637582
```

In the dataset, the care site setting Inpatient Visit had 111637582 visits