**Background Variables for OUD Matching**

*Objective*

To create a benchmark for OUD prevalence via propensity score matching, we will match regions on their baseline OUD prevalence information – extracted from NSDUH restricted-use files at the RDC – as well as demographic information. We would like to extract the demographic information from the American Community Survey.

* survey frequency: five-year
* geographic level: Census tract
* survey date: 2018
* states: all states, including DC but excluding:
  + Puerto Rico
  + Alaska
  + Hawaii
  + [other exceptions?]

It is prohibitively time-consuming to download files for all the variables and states of interest manually, so we will query the ACS API to download each of the following table codes or table code groups from each state listed above. The ACS guide to table contents and codes is [here](https://api.census.gov/data/2010/acs/acs5/variables.html).

For each query we will need the state, county, and tract FIPS codes for later merges.

The ACS tables contain population counts in each category. We need to summarize counts to a level of interest specified below. We will not calculate percentages at this stage because we will need to aggregate the data to the survey sampling region level by combining information across census tracts.

*ACS Table Codes for Matching Variables*

* + B01001: sex by age (get whole group)
* total number of males
* total number of females
* total number in age categories
  + TBD
  + B02001: race (get whole group)
* total number in each racial category
  + marital status for population 15 years and over
* codes to extract:
  + B12001\_001E: total population 15 years and over
  + B12001\_002E: total population, male, 15 years and over
  + B12001\_003E: male, never married
  + B12001\_004E: male, now married
  + B12001\_009E: male, widowed
  + B12001\_010E: male, divorced
  + B12001\_011E: total population, female, 15 years and over
  + B12001\_012E: female, never married
  + B12001\_013E: female, now married
  + B12001\_018E: female, widowed
  + B12001\_019E: female, divorced
* summarize to:
  + total population 15 years and over
  + total population 15 years and over, never married
  + total population 15 years and over, now married
  + total population 15 years and over, widowed
  + total population 15 years and over, divorced
  + English-speaking ability for population 5 years and over: B16001 (get whole group)
* broken out by language spoken at home
* within language spoken at home, categories are:
  + speak English "very well"
  + speak English less than "very well"
* summarize to:
  + total population 5 years and over
  + total population speaking English “very well”
  + ratio of income to poverty level in past 12 months (get whole group C17002):
* no need to summarize further than what’s reported
* levels reported:
  + C17002\_001E: total
  + C17002\_002E: under 0.5
  + C17002\_003E: 0.5 to 0.99
  + C17002\_004E: 1 to 1.24
  + C17002\_005E: 1.25 to 1.49
  + C17002\_006E: 1.5 to 1.84
  + C17002\_007E: 1.85 to 1.99
  + C17002\_008E: 2.00 and over
  + individual income in the past 12 months
* B06010\_001E: total
* B06010\_002E: no income
* B06010\_003E: with income
* B06010\_004E: loss, or income $1 to $9,999
* B06010\_005E: income $10,000 to $14,999
* B06010\_006E: income $15,000 to $24,999
* B06010\_007E: income $25,000 to $34,999
* B06010\_008E: income $35,000 to $49,999
* B06010\_009E: income $50,000 to $64,999
* B06010\_010E: income $65,000 to $74,999
* B06010\_011E: income $75,000 or more
  + public assistance of food stamps/SNAP in past 12 months
* B19058\_001E: total
* B19058\_002E: total with cash, public assistance, or food stamps/SNAP
* B19058\_003E: total with no cash, public assistance, or food stamps/SNAP
  + veteran status for civilian population 18 years and over
* B21001\_002E: total civilian population 18 years and over
* B21001\_002E: total veterans
* B21001\_003E: total non-veterans
  + employment status for the population 16 years and over: B23001 (get whole group)
* broken out by sex, age, and employment status
* summarize across sex and age to:
  + total population 16 years and over
  + total in labor force, Armed Forces
  + total in labor force, civilian, employed
  + total in labor force, civilian, unemployed
  + total not in labor force
  + educational attainment
* B23006\_001E: total population 25-64 years
* B23006\_002E: less than high school graduate, population 25-64 years
* B23006\_009E: high school graduate, population 25-64 years
* B23006\_023E: bachelor's degree or higher, population 25-64 years

*Population Density*

We will want to match on the population density in each census tract, but this information is not available in the ACS tables. However, it is available from the Census planning database, saved [here](file:///\\mathematica.Net\NDrive\Project\50828_OCP\Restricted\DC1\OUD%20benchmarking\Data\) (data set and codebook). The relevant variables are:

* GIDTR: FIPS tract code, comprising State code (first two digits), County code (next three digits), and Tract code (last six digits)
* State: FIPS state code (two digits)
* State\_name: State name
* County: FIPS county code (three digits)
* County\_name: county name
* Tract: six-digit Census tract code
* LAND\_AREA: land area in square miles
* Tot\_Population\_CEN\_2010: total population in the tract in 2010 Census
* Tot\_Population\_ACS\_12\_16: total population in the tract in 2012-2016 ACS

We will:

1. Calculate population density in each census tract
   1. Divide Tot\_Population\_ACS\_12\_16 by LAND\_AREA
   2. Maybe divide by 10,000 to put values on reasonable scale?
2. Merge to ACS data by GIDTR