U6614 Final Project Document

Neil Chin (nkc2124) & Nicole Yahui Wu (yw3551)

2022-11-22

1   Research Question and Motivations

hi sameer

Did Medicaid expansion reduce drug overdose deaths in Appalachia?

Deaths due to drug overdose are a pressing problem facing United States policymakers, with reportedly more than 932,000 people dying due to overdose since 1999 (CDC, 2022). Rate of overdose deaths increased in nearly every US state from 2013-2017, with particularly severe incidence in the Appalachian region (CDC, 2020; CDC, 2021). The increase in deaths is driven primarily by the ongoing US opioid crisis, with the vast majority (>80%) of overdose deaths associated with opioid use (CDC, 2022).

In this context, we hope to assess whether expansion of the social safety net through government policy can be causally linked to reductions in drug overdose death rates. Specifically, we propose to estimate the effects of the expansion of the US Medicaid program using a panel dataset of Appalachian counties, which increased income-based eligibility for healthcare (including drug addiction treatment) in some Appalachian states, on drug overdose deaths. We hope that the findings from this study can assist policymakers in determining whether expanding access to healthcare in low-income areas is a viable policy mechanism for addressing the ongoing drug overdose crisis.

2   Policy Background

The Affordable Care Act (ACA) was passed by the United States Congress and signed into law by President Barack Obama in 2010, drastically changing the policy landscape for health care in the United States. Among the major provisions in the ACA was expanded eligibility for Medicaid (i.e., "Medicaid Expansion"), which allowed states to raise the income-eligibility threshold to 138% of the federal poverty level (KFF, 2022).

Among the 13 states whose boundaries overlap with the broad geographical definition of Appalachia, five states (Kentucky, Maryland, New York, Ohio, and West Virginia) passed legislation mandating the expansion of Medicaid as of January 1st, 2014 (KFF, 2022). Two additional states, Pennsylvania and Virginia, would later expand Medicaid, with the former in 2015 and the latter in 2019. Six states (Alabama, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee) have not expanded Medicaid to-date.

In our study, we identify this policy adoption discrepancy as a "treatment" (i.e., "differential exposure between entities over time") affecting drug overdose incidence in Appalachian communities. Our *hypothesized* causal mechanism is that expanded Medicaid eligibility allowed for greater access to low-cost health care among Appalachian counties in expansion states, therefore enabling people struggling with drug addiction to receive treatment when they otherwise would not have been able to receive care, reducing overall deaths from drug overdose in these areas.

3 Data Description

Our data is a county-year panel dataset, which we use to examine drug overdose deaths in Appalachian counties over the period 2011-2016. In total, this represents 2,538 county-year observations (i.e., 423 counties across six years).

The subset of US counties defined as "Appalachian" is based on the jurisdiction of the Appalachian Regional Council (ARC, 2021). Accordingly, our units of observation for this study are counties, with the representative population being people living in the Appalachian region. We focus on counties as units of observation because the Appalachian region is defined at the county-level and to avoid potential selection bias and small sample size issues that would confound state-level analysis. We intentionally focus on Appalachian counties to both increase sample size and to construct pseudo "control" and "treatment", assuming that Appalachian counties are largely similar in terms on non-measurable, unobserved characteristics across state lines.

Identification of state-level Medicaid expansion is based on tracking done by the Kaiser Family Foundation (KFF, 2022). Due to the fact that Pennsylvania implemented Medicaid expansion a year after other expansion states, we exclude Pennsylvania counties from our analysis to ensure that there is a distinct time-based differential in treatment (i.e., pre- and post-2014 periods). Similarly, we consider Virginia to be a non-expansion state for the purposes of this study, because it only enacted Medicaid expansion in 2019, well after the end of our study period.

Data on drug overdose death rates (i.e., deaths per 100,000 residents) comes from estimates modeled by the National Center for Health Statistics (NCHS), which are available at the county-level for the period 2003-2020. Unfortunately, county-level statistics on overdose deaths based on final counts of cause of death reporting only became available in 2020, outside of the time frame of this study.

County-level demographic covariates are taken from the US Census Bureau American Community Survey (ACS). Using these covariates, we hope to control for omitted variable bias stemming from county-level factors such as median income, median age, and sex and race compositions. In particular, we expect that median income would relate negatively to overdose deaths, as higher income communities would be better able to provide drug addiction treatment, all else equal. We include controls for median age as drug overdose death incidence tends to vary by age group (KFF, 2022). Drug overdose deaths are also more common among individuals identified as male than female (CDC, 2022). Furthermore, access to drug treatment has been shown to differ according to race (NIDA, 2019), potentially leading to differentials in drug overdose deaths depending on racial composition of counties.

Given that both the NCHS data on overdose death rates and ACS control variables are estimated at the county-level, we expand our study period to the two years prior to Medicaid expansion and the two years after (i.e., 2012-2015), in order to smooth over any potential estimation errors. We also hope that this larger time frame will capture any lags in treatment effects, given that reductions in drug overdose deaths due to expanded access to health care may not be reflected in the data until more than a year after Medicaid expansion.
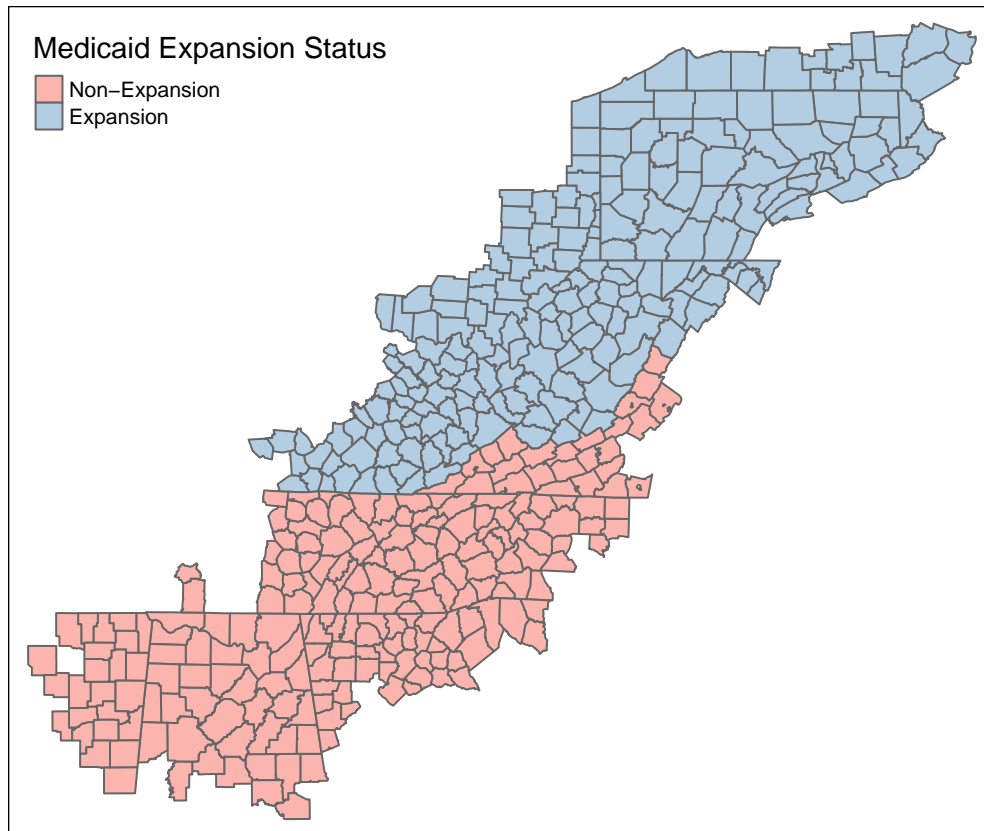
4 Preliminary Exploratory Analysis

4.1 Variation in Policy "Treatment"

In total, 210 Appalachian counties are located in expansion states and 213 Appalachian counties are located in non-expansion states.

We map Appalachian counties by Medicaid expansion status below:

```
expansion_map
```

Note that Virginia did enact Medicaid expansion in 2019, after the end of our study period, and thus Appalachian counties in Virginia are considered to be "non-expansion" in our analysis. Clearly, at the state level there is a degree of "North-South" bias in terms of which states elected to expand Medicaid. However, our empirical strategy (as disucssed in the next section) aims to minimize any associated omitted variable bias with county-level fixed effects.

4.2   Variation in Drug Overdose Death Rate

In the figure below, we plot the weighted-average of county-level overdose death rate over time (i.e., "Years since Medicaid Expansion"), separating counties in expansion states from counties in non-expansion states. It appears that expansion counties and non-expansion counties experience near-parallel trends prior to Medicaid expansion. Following expansion, however, we observe an *steepening trend* of overdose death rates in *expansion counties* relative to *non-expansion counties*, indicating that Medicaid expansion had a *positive* effect on overdose death rates, contradicting the direction of our hypothesized causal effect.

figure1

County–Level Drug Overdose Death Rates by Medicaid Expansion Status
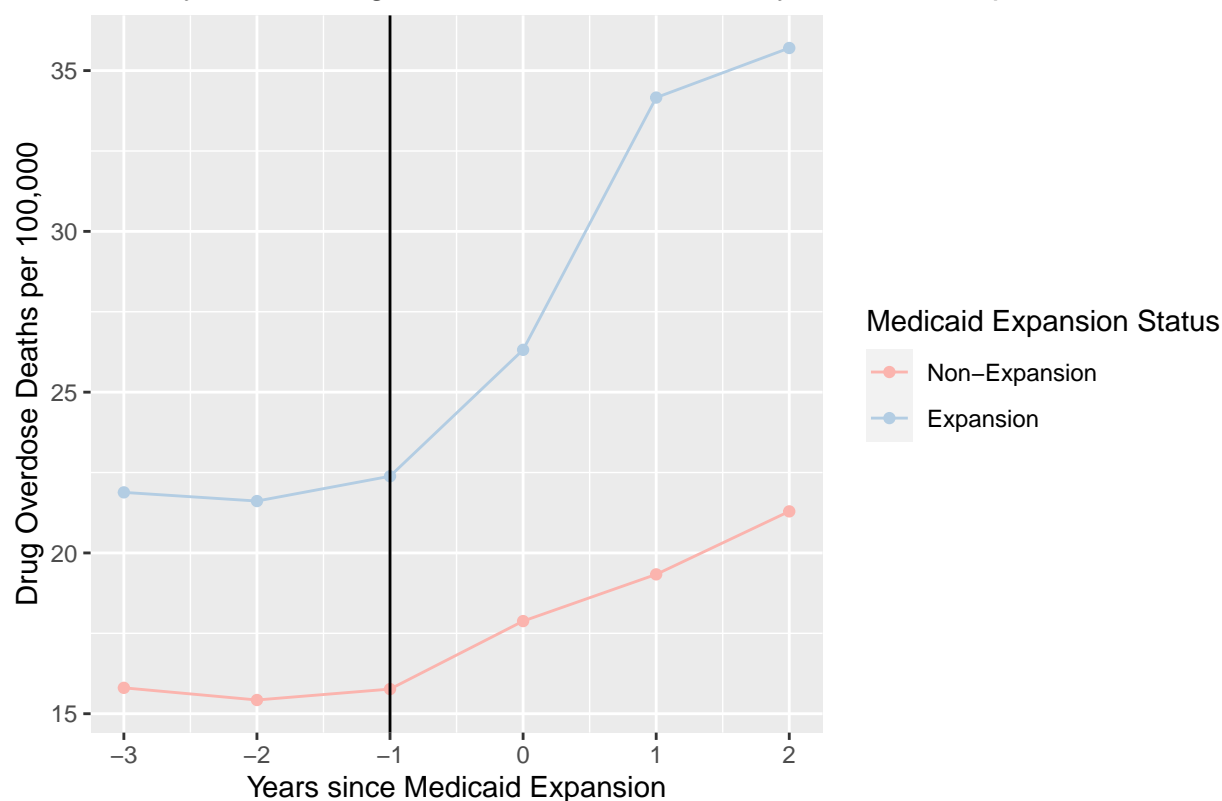
4.3 Variation in Key Continuous Variables

We also explore descriptive statistics key variables across counties in expansion states and counties in non-expansion states, both before and after expansion occurs.

4.3.1 Pre-Expansion Descriptive Statistics in Expansion States

table1

	mean	sd	min	max	range
estimated__overdose__death__rate	22.29	13.28	4.34	102.47	98.13
med__income	39505.08	8143.29	19344.00	65304.00	45960.00
med__age	41.52	3.15	26.30	51.40	25.10
male__share	49.90	2.13	44.82	69.71	24.89
black__share	2.47	2.84	0.00	23.21	23.21
hisp__share	1.52	1.61	0.00	13.91	13.91
white__share	94.16	4.54	69.00	100.00	31.00
asian__share	0.52	0.88	0.00	9.54	9.54

4.3.2 Post-Expansion Descriptive Statistics in Expansion States

4

table2

	mean	sd	min	max	range
estimated_overdose_death_rate	28.67	15.58	4.76	113.94	109.18
med_income	40128.62	8338.29	18972.00	69753.00	50781.00
med_age	42.15	3.28	27.00	52.40	25.40
male_share	49.95	2.31	44.75	78.49	33.74
black_share	2.44	2.96	0.00	31.64	31.64
hisp_share	1.58	1.55	0.00	14.64	14.64
white_share	93.93	4.64	59.46	99.57	40.12
asian_share	0.56	0.95	0.00	10.20	10.20

4.3.3 Pre-Expansion Descriptive Statistics in Non-Expansion States

table3

	mean	sd	min	max	range
estimated_overdose_death_rate	16.88	7.54	2.79	47.58	44.79
med_income	39148.58	8058.30	21798.00	88262.00	66464.00
med_age	40.88	3.99	25.10	58.20	33.10
male_share	49.15	1.30	45.51	56.51	11.01
black_share	10.55	14.46	0.00	83.13	83.13
hisp_share	3.88	4.02	0.00	32.22	32.22
white_share	83.21	14.70	15.21	99.87	84.66
asian_share	0.67	1.06	0.00	10.77	10.77

4.3.4 Post-Expansion Descriptive Statistics in Non-Expansion States

table4

	mean	sd	min	max	range
estimated_overdose_death_rate	19.86	8.68	2.93	51.29	48.36
med_income	40169.64	7917.76	26231.00	91842.00	65611.00
med_age	41.82	4.24	25.00	59.80	34.80
male_share	49.19	1.37	44.57	56.91	12.33
black_share	10.63	14.50	0.00	82.11	82.11
hisp_share	4.18	4.16	0.00	33.50	33.50
white_share	82.53	14.79	15.42	99.82	84.40
asian_share	0.73	1.16	0.00	11.14	11.14

We observe that mean drug overdose death rates are higher in expansion counties than non-expansion counties, both before and after Medicaid expansion. However, both groups of counties experience a relative increase in drug overdose death rates over the period 2011-2016.

It is further noteworthy that the mean of black share of counties in non-expansion states is much higher than that of counties in expansion states regardless of time period. The means of black share of counties in non-expansion states for both pre-expansion and post-expansion are both over 10%, but the means of black share of counties in expansion states for both pre and post-expansion are 2.19% and 2.21%, respectively. Moreover, the range of median income for counties in non-expansion states is much higher than that of counties in expansion states regardless of years. Specifically, the range of median income of counties in non-expansion states for pre- and post-expansion are 65380 and 62585, respectively, and the range of median income of counties in expansion states for pre- and post-expansion are 45680 and 47531. This means that the gap between the rich counties and the poor counties is larger in non-expansion states than in expansion states.

It is also of interest that variation between counties across each variable of interest appears to be fairly substantial, with the possible exception of male share, based on measured standard deviations. Further exploratory analysis will need to consider whether this is sufficient "within" county variation, in addition to "between" county variation, to support a panel data analysis.

5   Empirical Strategy

The primary variation that we seek to exploit through our analysis is the differential in state-level adoption of Medicaid expansion across the Appalachian region, with policy variation at the state-level thus filtering down to the county-level. To evaluate the effect of Medicaid expansion on drug overdose deaths in Appalachian counties, we estimate the following "differences-in-differences" specification:

$$Overdose\ Death\ Rate_{it} = \beta Treat_{it} + X_{it}\gamma + \nu_i + \tau_t + \varepsilon_{it}$$

where $Overdose\ Death\ Rate_{it}$ is deaths attributed to drug overdose per 100,000 county residents for county i at time t, $Treat_{it}$ is a binary variable that indicates "treatment" status (i.e., Medicaid expansion) for a county-year, X_{it} is a vector of time varying controls (e.g., median income, median age, male population share, racial composition) for potential determinants of overdose death rates outside of our policy variation of interest.

Additionally, we include an array of county fixed effects, ν_i, that control for unobserved time-invariant factors that are specific to individuals counties. An example of one such factor would be if, throughout the entire 2012-2015 period, a specific county had its own drug treatment program that reduced drug overdose deaths compared to other counties, all else equal. We further include τ_t, a year fixed effect, to control for unobserved county-invariant factors that might have changed between each year included in our panel. Such factors would include events such as periodic economic shocks that affect the entire Appalachian region in certain years, which potentially could be deterministic of the rate of overdose deaths. Finally, ε_{it} is the idiosyncratic error term.

To support this analysis, we compile publicly-available data from the Appalachian Regional Council (ARC), Kaiser Family Foundation (KFF), National Center for Health Statistics (NCHS), and the US Census Bureau. From the sources, data is largely already available at the county-level, thus we are not required to perform any aggregation or dis-aggregation steps to make the data suitable for use. Furthermore, variables of interest are entirely quantitative, thus cleaning needs are minimal. All source data include county-specific FIPS codes as merging indices. The only intermediate data transformation we perform is the calculation of county-level racial composition shares, based on Census Bureau population data.

We note two key limitations to the internal validity of our study, the first being potential imbalance between pseudo "treatment" (i.e., counties in expansion states) and "control" (i.e., counties in non-expansion states) groups and the second being possible omitted variable bias due to differential county-level availability of drug treatment facilities. On the first issue, we are considering employing a matching strategy to pair "treatment" counties in expansion states with "control" counties in non-expansion states in a way that mimics random assignment. On the second issue, we will attempt to collect additional data that allows us to control for the number of drug treatment facilities per 100,000 residents at the county-level. If this data is not directly available, we will also look for potential proxy measures.

6   Difference-in-Difference Estimation Results

The estimation results of our specification are displayed below:

	Model 1
treat	6.826***
	(0.457)
med_income	−0.0004***
	(0.000 03)
med_age	−0.134
	(0.091)
male_pop	0.000 02
	(0.000 01)
black_pop	−0.000 06***
	(0.000 01)
hisp_pop	−0.000 06
	(0.000 05)
asian_pop	0.0008***
	(0.0001)
Num.Obs.	2538
R2	0.103
R2 Adj.	−0.029
RMSE	1457.53

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

```
estimation_results
```

Note that estimates are weighted by county population.

7   Discussion

TL;DR Medicaid expansion had some pretty awful unintended consequences. Most likely explanation: Medicaid expansion increased access to prescription opioids, resulting in higher rates of drug overdose deaths.

8   Appendix

8.1   Data Sourcing and Cleaning

We compile publicly-available data from the Appalachian Regional Council (ARC), Kaiser Family Foundation (KFF), National Center for Health Statistics (NCHS), and the US Census Bureau into a single county-year panel data set.

8.1.1   Appalachian Counties Data

The Appalachian Regional Commission defines 423 counties in 13 states (West Virginia, Alabama, Georgia, Kentucky, Maryland, Mississippi, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, and Virginia) as demarcating the Appalachian region. We adopt this geographical definition in our research approach.

More information available at: https://www.arc.gov/appalachian-counties-served-by-arc/.

```
appalachian_counties <- read_excel('data/appalachian_counties_ARC_2021.xlsx')
```

8.1.2 Medicaid Expansion

Of the 13 states, five states (Kentucky, Maryland, New York, Ohio, and West Virginia) passed legislation mandating the expansion of Medicaid as of January 1st, 2014. Two additional states, Pennsylvania and Virginia, would later expand Medicaid, with the former in 2015 and the latter in 2019. Six states (Alabama, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee) have not expanded Medicaid to-date.

We add a dummy variable into our data to reflect this policy difference. Information on state-level Medicaid expansion available at: https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/.

```r
# Create dummy variable for medicaid expansion and a variable for date of expansion

expansion <- c("Kentucky", "Maryland", "New York", "Ohio", "West Virginia",
               "Pennsylvania")
no_expansion <- c("Alabama", "Georgia", "Mississippi", "North Carolina",
               "South Carolina", "Tennessee", "Virginia")

appalachian_counties <- appalachian_counties %>%
  mutate(medicaid_expansion = as.numeric(STATE %in% expansion),
         expansion_date = make_date(2014, 1, 1))

# Re-code dates for Pennsylvania (late expansion)

appalachian_counties$expansion_date[appalachian_counties$STATE == "Pennsylvania"] <-
  make_date(2015,1,1)
```

8.1.3 Overdose Deaths

Our data on drug overdose deaths comes from estimates modeled by the National Center for Health Statistics (NCHS), which are available at the county-level for the period 2003-2020. Source link: https://www.cdc.gov/nchs/data-visualization/drug-poisoning-mortality/.

```r
overdose_data <- read_csv("data/NCHS_-_Drug_Poisoning_Mortality_by_County__United_States.csv")

# Keep columns for FIPS codes, year, and modeled overdose deaths

overdose_data <- overdose_data %>%
  select(FIPS, Year, `Model-based Death Rate`) %>%
  rename(estimated_overdose_death_rate = `Model-based Death Rate`)
```

8.1.4 ACS Data

To control for time-variant, county-variant factors, we pull demographic data from the US Census Bureau's American Community Survey (ACS), for years 2011-2016. Data is downloaded from the US Census Bureau API using the package tidycensus.

```r
census_api_key("b79b301dc87bb0fd551147883a8141dca4e2823e")

# Pull 2011 ACS county-level data

acs_2011 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
```

```r
                          male_pop = "B01001_002",
                          med_income = "B19013_001",
                          med_age = "B01002_001"),
                 year = 2011)

acs_2011 <- acs_2011 %>%
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2011,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
         male_share = 100 * (male_pop/population))


# Pull 2012 ACS county-level data

acs_2012 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
                                  male_pop = "B01001_002",
                                  med_income = "B19013_001",
                                  med_age = "B01002_001"),
                    year = 2012)

acs_2012 <- acs_2012 %>%
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2012,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
         male_share = 100 * (male_pop/population))

# Pull 2013 ACS county-level data

acs_2013 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
                                  male_pop = "B01001_002",
                                  med_income = "B19013_001",
                                  med_age = "B01002_001"),
                    year = 2013)

acs_2013 <- acs_2013 %>%
```

```r
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2013,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
         male_share = 100 * (male_pop/population))

# Pull 2014 ACS county-level data

acs_2014 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
                                  male_pop = "B01001_002",
                                  med_income = "B19013_001",
                                  med_age = "B01002_001"),
                    year = 2014)

acs_2014 <- acs_2014 %>%
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2014,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
         male_share = 100 * (male_pop/population))

# Pull 2015 ACS county-level data

acs_2015 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
                                  male_pop = "B01001_002",
                                  med_income = "B19013_001",
                                  med_age = "B01002_001"),
                    year = 2015)

acs_2015 <- acs_2015 %>%
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2015,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
```

```
        male_share = 100 * (male_pop/population))

# Pull 2016 ACS county-level data

acs_2016 <- get_acs(geography = "county",
                    variables = c(population = "B01001_001",
                                  white_pop = "B01001H_001",
                                  hisp_pop = "B01001I_001",
                                  asian_pop = "B02001_005",
                                  black_pop = "B02001_003",
                                  male_pop = "B01001_002",
                                  med_income = "B19013_001",
                                  med_age = "B01002_001"),
                    year = 2016)

acs_2016 <- acs_2016 %>%
  select(-moe, -NAME) %>%
  spread(key = variable, value = estimate) %>%
  mutate(Year = 2016,
         white_share = 100 * (white_pop/population),
         hisp_share = 100 * (hisp_pop/population),
         asian_share = 100 * (asian_pop/population),
         black_share = 100 * (black_pop/population),
         male_share = 100 * (male_pop/population))

# Append into single ACS data frame, rename GEOID as FIPS

acs_11_to_16 <- rbind(acs_2011, acs_2012, acs_2013, acs_2014, acs_2015, acs_2016)
acs_11_to_16 <- acs_11_to_16 %>%
  mutate(GEOID = as.numeric(GEOID)) %>%
  rename(FIPS = GEOID)
```

8.2   Merge Data into Consolidated Data Set

We merge the Appalachian counties data, overdose data, and ACS data into a single data set. We save the resulting panel dataframe and county-level dataframe as .rds files, so that they can be loaded at the beginning of this file and elsewhere.

```
# Create four versions of the Appalachian counties data, one for each year in our study,
# merge with other data, then append into a panel

appalachian_counties_11 <- appalachian_counties %>%
  mutate(Year = 2011)
appalachian_counties_12 <- appalachian_counties %>%
  mutate(Year = 2012)
appalachian_counties_13 <- appalachian_counties %>%
  mutate(Year = 2013)
appalachian_counties_14 <- appalachian_counties %>%
  mutate(Year = 2014)
appalachian_counties_15 <- appalachian_counties %>%
  mutate(Year = 2015)
appalachian_counties_16 <- appalachian_counties %>%
  mutate(Year = 2016)
```

```r
appalachian_counties_panel <- rbind(appalachian_counties_11,
                                    appalachian_counties_12,
                                    appalachian_counties_13,
                                    appalachian_counties_14,
                                    appalachian_counties_15,
                                    appalachian_counties_16)

# Create variable for time since expansion and treatment variable

appalachian_counties_panel <- appalachian_counties_panel %>%
  mutate(t_expansion = Year - year(expansion_date),
         treat = ifelse(t_expansion >= 0, medicaid_expansion, 0))

# Merge panel with overdose data

appalachian_counties_overdose_panel <- appalachian_counties_panel %>%
  mutate(FIPS = as.numeric(FIPS)) %>%
  left_join(overdose_data, by = c("FIPS", "Year"))

# Merge panel with ACS data (save as an RDS file)

yearly_counties_panel_overdose_acs <- appalachian_counties_overdose_panel %>%
  left_join(acs_11_to_16, by = c("FIPS", "Year"))

# Turn expansion dummy variable into a factor

yearly_counties_panel_overdose_acs$medicaid_expansion <-
  factor(yearly_counties_panel_overdose_acs$medicaid_expansion,
                              levels = c(0,1),
                              labels = c("Non-Expansion", "Expansion"))

# Save panel and county datasets as .rds files

saveRDS(yearly_counties_panel_overdose_acs, file = "yearly_panel.rds")
saveRDS(appalachian_counties, file = "appalachian_counties.rds")
```