

U6614 Final Project Proposal — Deliverable 2

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

2022-11-02

1 Research Question and Motivations

Did 2014 Medicaid expansion reduce drug overdose deaths in Appalachia?

[@ Nicole, some useful background here could be context for drug overdose deaths in Appalachia. Historically, drug overdoses have been super high in the region, so there should be a ton of good articles out there!]

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.

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. 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 differential as a “treatment” affecting drug overdose incidence in Appalachian communities. Expanded Medicaid eligibility allowed for greater access to low-cost health care in expansion states, therefore potentially enabling people struggling with drug addiction to receive treatment that otherwise would not have been able to receive care, reducing overall deaths from drug overdose.

3 Data Description

Our data examines drug overdose deaths in Appalachian counties over the period 2012-2015. The subset of US counties defined as “Appalachian” is based on the jurisdiction of the Appalachian Regional Council.

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. 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.

[TK, sources currently found in the Appendices]

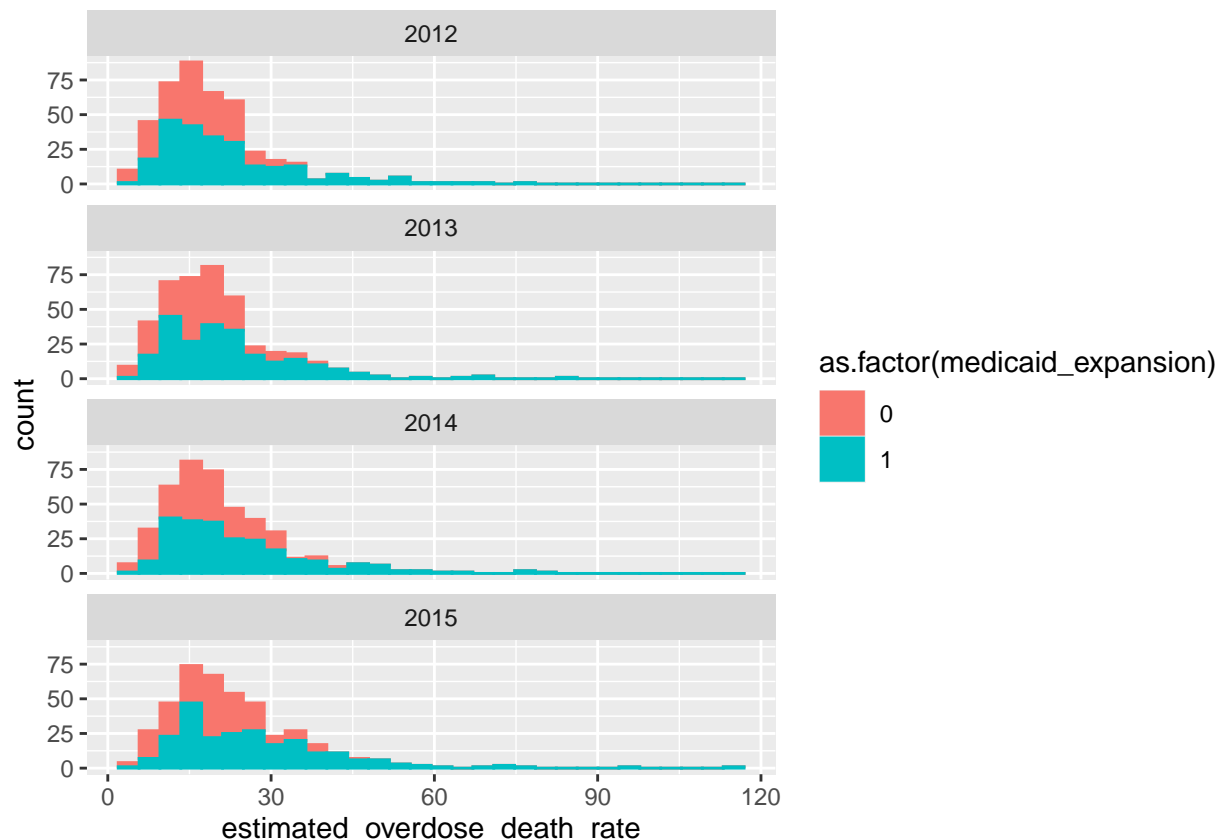
4 Preliminary Exploratory Analysis

Summary statistics for counties in expansion states and for counties in non-expansion states.

```
year_county_panel <- readRDS("yearly_panel.rds")
```

[Note: Can use ggplot facet wrapping and histograms to show distribution between “treatment” and “control” groups for each year]

```
year_county_panel %>%
  ggplot(aes(x = estimated_overdose_death_rate,
             color = as.factor(medicaid_expansion),
             fill = as.factor(medicaid_expansion))) +
  geom_histogram() +
  facet_wrap(~ Year, nrow = 4)
```



5 Empirical Strategy

To evaluate the effect of Medicaid expansion on drug overdose deaths in Appalachian counties, we would like to estimate the following “differences-in-differences” specification:

$$\text{Overdose Death Rate}_{it} = \text{Medicaid Expansion}_{it} + \text{Medicaid Expansion}_{it} * \text{Post Expansion}_t + \text{Male Share}_{it} + \text{Average Age}_{it} + \text{Black Share}_{it} + \text{Median Income}_{it} + \text{Post Expansion}_t + \Gamma_i + u_{it}$$

[Explanation of why we use fixed effects]

Key limitations — 1). both overdose data and ACS data are estimates, which are subject to error, 2). potential imbalance between pseudo “treatment” and “control” groups.

To address the first limitation, we compile two years of data for the pre- and post-“treatment” periods, to smooth out any potential irregularities in single-year estimates.

6 Appendix

6.1 Data Sourcing and Cleaning

6.1.1 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')
```

6.1.2 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/>.

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

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

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

```
# Re-code dates for non-expansion and late expansion states
```

```
appalachian_counties$expansion_date[appalachian_counties$STATE %in% no_expansion] <-  
  NA  
appalachian_counties$expansion_date[appalachian_counties$STATE == "Pennsylvania"] <-  
  make_date(2015, 1, 1)  
appalachian_counties$expansion_date[appalachian_counties$STATE == "Virginia"] <-  
  make_date(2019, 1, 1)
```

6.1.3 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/>.

```
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`)
```

6.1.4 ACS Data

To control for time-variant, county-variant factors, we pull demographic data from the US Census Bureau's American Community Survey, for years 2012-2015. Data is downloaded from the US Census Bureau API using the package tidycensus.

```
census_api_key("b79b301dc87bb0fd551147883a8141dca4e2823e")

# 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 %>%
  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))

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

acs_12_to_15 <- rbind(acs_2012, acs_2013, acs_2014, acs_2015)
acs_12_to_15 <- acs_12_to_15 %>%

```

```
mutate(GEOID = as.numeric(GEOID)) %>%
rename(FIPS = GEOID)
```

6.2 Merge Data into Consolidated Data Set

We merge the Appalachian counties data, overdose data, and ACS data into a single data set.

Create four versions of the Appalachian counties data, one for each year in our study, then append in

```
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_panel <- rbind(appalachian_counties_12,
                                   appalachian_counties_13,
                                   appalachian_counties_14,
                                   appalachian_counties_15)

# 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_12_to_15, by = c("FIPS", "Year"))

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