U6614 Final Project Proposal — Deliverable 2

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

2022-11-03

1 Research Question and Motivations

Did 2014 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 2014 expansion of the US Medicaid program, 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 panel dataset, which we use to examine 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 (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 avoid 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. After excluding counties in Pennsylvania, 158 Appalachian counties are located in expansion states and 213 Appalachian counties are located in non-expansion states. These numbers imply that the panel is not quite "balanced" between "treatment" and "control" groups as currently constructed, and may require further restructuring (i.e., matching) to ensure balance.

4.2 Variation in Drug Overdose Death Rate

In the histogram below, it is clear that the distribution of county-level drug overdose rates — our outcome variable of interest — differs considerably between counties in expansion states and counties in non-expansion states. Distributions also change year over year, and the distribution "spread" shows that there is considerable variation in overdose death rates within expansion and non-expansion counties.

county_overdose_histogram

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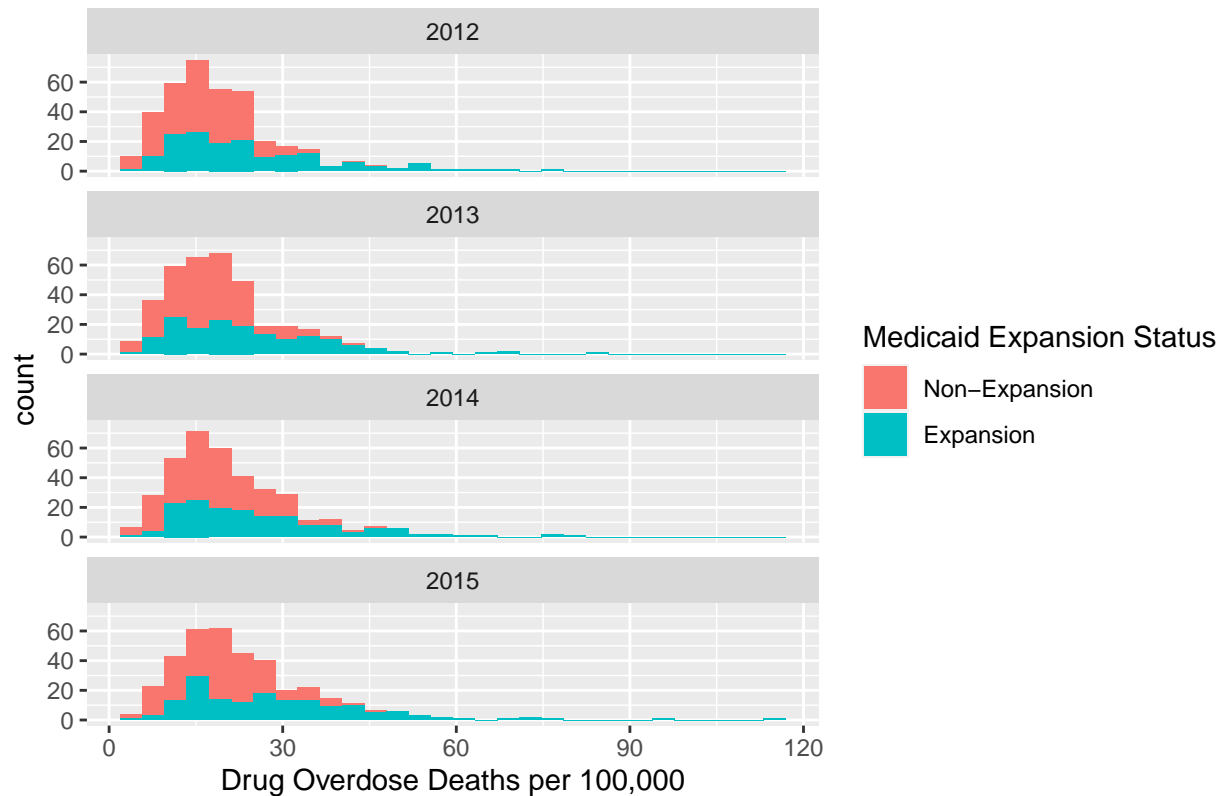
4.3 Variation in Key Control Variables

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

```
# @ Nicole, then use this chunk to display the table of descriptive statistics

# e.g., for each x variable (median income, median age, male share, black share, etc.) calculate and di
# basic descriptive statistics (median, median, standard error, min, max)
```

[@ Nicole, maybe try to summarize each table here in 1-2 sentences, noting if any of the variables seem to lack variation (i.e., small min-max spread, small standard errors). Also worth noting skew (i.e., mean-median spread)]

4.3.2 Post-Expansion Descriptive Statistics in Expansion States

```
# @ Nicole, then use this chunk to display the table of descriptive statistics

# e.g., for each x variable (median income, median age, male share, black share, etc.) calculate and di
# basic descriptive statistics (median, median, standard error, min, max)
```

[@ Nicole, maybe try to summarize each table here in 1-2 sentences, noting if any of the variables seem to lack variation (i.e., small min-max spread, small standard errors). Also worth noting skew (i.e., mean-median spread)]

4.3.3 Pre-Expansion Descriptive Statistics in Non-Expansion States

```
# @ Nicole, then use this chunk to display the table of descriptive statistics

# e.g., for each x variable (median income, median age, male share, black share, etc.) calculate and di
# basic descriptive statistics (median, median, standard error, min, max)
```

[@ Nicole, maybe try to summarize each table here in 1-2 sentences, noting if any of the variables seem to lack variation (i.e., small min-max spread, small standard errors). Also worth noting skew (i.e., mean-median spread)]

4.3.4 Post-Expansion Descriptive Statistics in Non-Expansion States

```
# @ Nicole, then use this chunk to display the table of descriptive statistics

# e.g., for each x variable (median income, median age, male share, black share, etc.) calculate and di
# basic descriptive statistics (median, median, standard error, min, max)
```

[@ Nicole, maybe try to summarize each table here in 1-2 sentences, noting if any of the variables seem to lack variation (i.e., small min-max spread, small standard errors). Also worth noting skew (i.e., mean-median spread)]

5 Empirical Strategy

The primary variation that we seek to exploit through our analysis is the differential state-level adoption of Medicaid expansion in the Appalachian region, with policy variation thus filtering down to the county-level. 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:

$Overdose\ Death\ Rate_{it} = Medicaid\ Expansion_i + Medicaid\ Expansion_i * Post\ Expansion_t + Median\ Income_{it} + Median\ Age_{it} + Male\ Share_{it} + Black\ Share_{it} + White\ Share_{it} + Hispanic\ Share_{it} + Asian\ Share_{it} + Post\ Expansion_t + \Gamma_i + u_{it}$

where $Overdose\ Death\ Rate_{it}$ is deaths attributed to drug overdose per 100,000 county residents for county i at time t, $Medicaid\ Expansion_i$ is an indicator (0-1) for if the county is located in an expansion state, $Median\ Income_{it}$ is the median income of the county, $Median\ Age_{it}$ is the median age of county residents, $Male\ Share_{it}$ is the male share (0-100) of county residents, $Black\ Share_{it}$, $White\ Share_{it}$, $Hispanic\ Share_{it}$, and $Asian\ Share_{it}$ measure county-level racial composition (0-100).

Additionally, we include the term $Post\ Expansion_t$, which indicates (0-1) observations measured prior to 2014 Medicaid Expansion, to control for unobserved county-invariant factors that might have changed between the pre- and post-expansion time periods (i.e., "time fixed effects"). Such factors would include events such as periodic economic shocks that affect the entire Appalachian region in certain years, which potentially affect the rate of overdose deaths. We further include an array of county-specific dummy variables Γ_i that control for unobserved time-invariant factors that are specific to individuals counties (i.e., "county fixed effects"). 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.

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 datasets 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   Appendix

6.1   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.

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/.

```r
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/.

```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 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 (ACS), 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",
```

```r
                             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 %>%
```

```r
  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. 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.

```r
# 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_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"))

# 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")
```