Observational Studies 7 (2021); 81-112 Submitted 12/19; Published 4/21

Protocol — Evaluating the Effect of ACA Medicaid Expansion on Mortality During the COVID-19 Pandemic Using County-level Matching

Charlotte Z. Mann **manncz@umich.edu**
Department of Statistics
University of Michigan
West Hall, 1085 S University Ave, Ann Arbor, MI 48109

Ben B. Hansen **bbh@umich.edu**
Department of Statistics
University of Michigan
West Hall, 1085 S University Ave, Ann Arbor, MI 48109

Lauren Gaydosh **lauren.gaydosh@austin.utexas.edu**
Department of Sociology
The University of Texas at Austin
116 Inner Campus Dr Stop G6000, Austin, TX 78712

Timothy Lycurgus **tlycurgu@umich.edu**
Department of Statistics
University of Michigan
West Hall, 1085 S University Ave, Ann Arbor, MI 48109

Abstract

States are able to choose whether to expand Medicaid as part of the Affordable Care Act (ACA); thus it is of interest to understand the impact of this policy choice. In this protocol, we outline a study on the impact of Medicaid expansion as part of the ACA on mortality during the COVID-19 pandemic in the United States. County-level matching using full, optimal matching with a propensity score model is used to estimate causal effects in this observational study. Due to the provisional nature of mortality data in 2020 as reported by the CDC, we outline a modified aligned rank test to account for censored data as well as reporting lags for different states. We aim to make connections between statistical and ethnographic methodologies by particularly examining adjacent counties and similar counties that are in the same region of the US and in vastly different regions of the US. Finally, we aim to add to the growing literature about the effect of ACA Medicaid expansion on mortality by calculating effects, disaggregating by race.

Keywords: Affordable Care Act, aligned rank test, Covid-19, full matching, Medicaid expansion, propensity score, optimal matching

1. Introduction

This article outlines a pre-analysis plan for a study on the impact of Medicaid expansion as part of the Affordable Care Act (ACA) on mortality during the COVID-19 pandemic in the United States. Originally a required part of the ACA, a supreme court ruling in 2012

made adoption of Medicaid expansion to cover individuals at 138% of the federal poverty level optional for states. In states that do not adopt this expansion, eligibility for Medicaid is based on a variety of factors including age, disability status, and family structure, while the expansion allows eligibility entirely based on income (KFF). By mid 2014, twenty-five states in the mainland US adopted the expansion (with the expanded eligibility effective January 1, 2014 in most of these states). There is a growing literature on the effect of ACA Medicaid expansion in the US on general health as well as on mortality (see, for example, Miller et al., 2019; Khatana et al., 2019; Swaminathan et al., 2018; Borgschulte & Vogler, 2020; Black et al., 2019).

While there is strong evidence that Medicaid expansion led to increased insurance coverage, there are inconsistencies and debate in the previous literature about whether ACA Medicaid expansion has detectable effects on mortality (Guth et al., 2020). Borgschulte & Vogler (2020) use propensity score pair matching and find that ACA Medicaid expansion had a significant negative effect on mortality between 2014 and 2017 using an ITOT analysis. Miller et al. (2019); Khatana et al. (2019); Swaminathan et al. (2018) all similarly find that ACA Medicaid expansion was associated with decreased mortality among specific populations. On the other hand, Black et al. (2019) do not find significant associations between ACA Medicaid expansion and mortality and find that their studies of the effect of expansion on mortality at the county level are underpowered. These studies have focused on different populations (in terms of age groups, and comorbidities) and and used distinct methods.

None of these studies of the effect of Medicaid expansion on mortality have considered a breakdown of the effect by race and few other studies examining the effect of Medicaid expansion on healthcare access and general health have added race into the equation (Guth et al., 2020). This leaves a significant hole in the literature and our understanding of the impacts of this major policy on people in the US. It is well documented that Black, Indigenous, and other People of Color (BIPOC) in the US are burdened with worse health outcomes and higher mortality rates than the white US population (Gee & Ford, 2011). There is evidence that BIPOC are disproportionally affected by the COVID-19 pandemic (Cowger et al., 2020). And saliently, as Dr. Jonathan Metzl argues, the "highly unhealthy American politics of race" (Metzl (2019), p. 137) were a major reason that certain states did not expand Medicaid as the rhetoric used to argue against expansion was and is racialized. It is imperative to understand both how this policy has affected different populations in the US as well as the impact on white Americans who "voiced a willingness to die, literally, rather than embrace a law that gave minority or immigrant persons more access to care, even if it helped them as well" (Metzl (2019), p. 124).

In this study, we will estimate the effect of ACA Medicaid expansion on mortality during the COVID-19 pandemic for non-elderly adults aged 20-64 across all states in the mainland US, additionally considering the effects on the US populations by race.[1] This complements the findings of Miller et al. (2019) who evaluated the effects on the newly insured and Borgschulte & Vogler (2020) who evaluated effects for adults in aggregate for the years following Medicaid expansion. We follow the methods of Sommers et al. (2014);

1. The choice of adults aged 20-64 aligns with previous literature on Medicaid expansion, to focus on working age adults.

Borgschulte & Vogler (2020); Black et al. (2019) to match counties in order to evaluate the effect of this state-level policy.

Our primary aim is to estimate the effect of ACA Medicaid expansion on mortality during the COVID-19 pandemic. Some states, including Kansas and Georgia, have been considering expanding Medicaid more recently, but have not yet adopted the expansion. Through ballot initiatives, Oklahoma and Missouri voted to expand Medicaid on June 30, 2020 and August 4, 2020 respectively but it will not go into effect until July 2021 (KFF). Thus, it is also relevant to us to answer the following questions:

I. How many lives could have been saved in states that have not expanded Medicaid if they had expanded Medicaid sometime between 2014 and 2019?

II. How many lives could have been saved in within a year after they may have expanded Medicaid, in states that have not expanded Medicaid?

Our main analysis, assessing the effect of Medicaid expansion anytime between 2014 and 2019 on mortality in 2020, addresses the first question. To address the second question, we estimate the effect of Medicaid expansion by 2014 on mortality in 2014, to understand how expansion could effect mortality within the first year of expansion. We evaluate the effect of ACA Medicaid expansion on mortality within the first year after expansion in 2014 because only two states expanded Medicaid in 2019, so we would have little power to understand the effect on 2020 mortality.

The article is organized as follows: First, we describe the data and variables used, then we describe the matching methodology, followed by description and diagnostics of the resulting county-level matches. Finally, we outline the data and methodology we will use to estimate causal effects of ACA Medicaid expansion on mortality.

2. Baseline Data

We use the report as of August 5, 2020 from the Kaiser Family Foundation (KFF) to identify which states expanded Medicaid in the first half of 2014 and when other states expanded between 2014 and June 2020 (KFF). There is some disagreement in previous studies of which states to include in studies of Medicaid expansion and which to exclude. Rather than excluding states from our analysis that expanded Medicaid after June of 2014 (like Borgschulte & Vogler (2020) and Black et al. (2019)), for matching purposes we group these states with those that have yet to expand Medicaid.[2] That is, we seek matches between a treatment group of counties for which Medicaid was expanded by mid-2014 and a control group consisting of counties for which expansion would occur later or never. This identification of treatment and control groups suits our first analysis, examining 2014 mortality. However, our second outcome of interest is mortality in 2020. For this analysis, we will treat any state that implemented Medicaid expansion on or before June 30, 2020 as part of the treatment group, with differing "doses" of treatment (explained in detail in

2. This includes nine states (ID, IN, LA, ME, MT, NH, PA, UT, VA). Four of these states adopted the Medicaid expansion as part of the ACA in 2019 or 2020 (ID, ME, UT, and VA).

§5.1). States which have implemented Medicaid expansion after June 2020 are considered as control states in both of our analyses.[3]

Following Borgschulte & Vogler (2020), we consider Wisconsin to be part of the treatment group (and expanding as of June 2014), although they did not expand Medicaid as part of the ACA, since Wisconsin covers adults up to 100% of the federal poverty level. We exclude Alaska and Hawaii from our analysis since we are interested in particular in evaluating counties that are close geographically and specifically adjacent. Figure 1 illustrates how states are classified in terms of expansion and also see Appendix A Table 4 for more detailed descriptions of when the states expanded or did not expand Medicaid and how they are treated in our models.

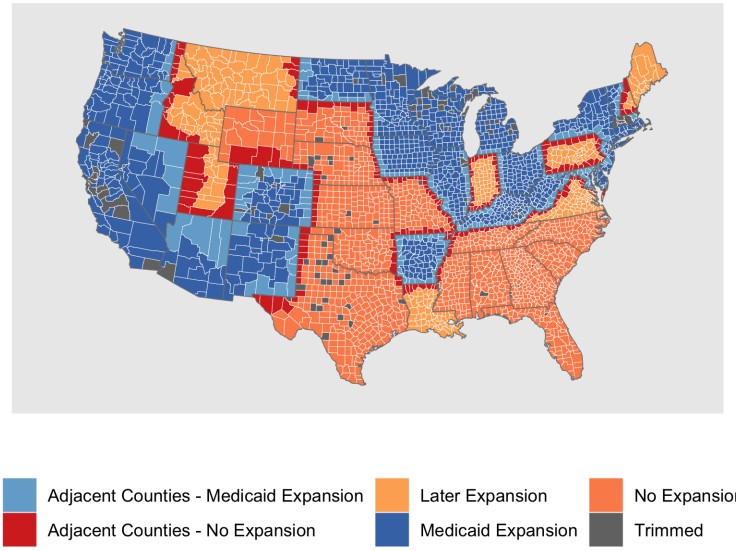

Figure 1: Classification of counties and their inclusion in the outcome analyses. Trimmed counties refers to counties that are excluded due to extreme propensity scores or outlying mortality values, which is described in later sections. "Later expansion" refers to expansion after June 2014.

Mortality counts come from the National Center for Health Statistics (NCHS) restricted use, detailed mortality data, housed through the Center for Disease Control and Prevention (CDC, c).[4] We use all-cause mortality aggregated across 2009-2013 for adults aged 20-64 for our main matching and analysis. Age-adjusted all-cause mortality is calculated for all cross-tabs of gender and race along with Hispanic origin using population values included in the NCHS data (which is from the Census Bureau). We additionally include healthcare amenable deaths, deaths from pneumonia or influenza (flu mortality) and opioid related deaths as covariates in matching. We follow the definitions of healthcare amenable deaths

3. These states include Missouri and Oklahoma, which voted for expansion in 2020 but will not implement eligibility expansion until 2021 as well as Nebraska which voted to expand in 2019 but will not be implemented until October 2020 (KFF).
4. A public use version of this data can be accessed through CDC WONDER, although all cells with fewer than 10 deaths are suppressed.

in Sommers et al. (2014) to identify healthcare amenable deaths as well as CDC definitions of ICD-10 codes to identify deaths from pneumonia or influenza and opioid related deaths (see Appendix A Table 5 for all ICD-10 codes used) as the underlying cause of death. Each of our analyses adjust for these specific causes of death, i.e. they are included as covariates in matching as well as in covariance adjustment in our treatment effect estimation. It is common to examine healthcare amenable deaths in analyses of ACA Medicaid expansion because healthcare amenable deaths are those one might expect to be impacted by increased insurance coverage. Additionally, flu is the closest analogue to COVID-19, so it makes sense to control for historical flu mortality patterns. Opioid-related deaths are included because they increased greatly during the years before and after ACA Medicaid expansion, driving marked rises in deaths from poisoning, which by 2018 was the leading cause of preventable injury-related mortality among adults aged 25-34 and 55-64 (NSC) — the better part of the population meeting the age requirement to gain insurance under Medicaid expansion. We also plan to disaggregate outcome mortality by underlying cause of death for a number of analyses (as described in §5.1).

County-level pre-treatment mortality counts are merged with indicators of the health, political leaning, and demographics of each county. We use the percent of votes in a county for Romney, between Romney and Obama, in the 2012 presidential election from the CQ Press Voting and Elections Collection (CQ Press) to measure county-level partisanship ("% voting Republican"). Other health, economic, and demographic variables are collected from the 2019 Area Health Resources Files[5] (AHRF) and the Institute for Health Metrics and Evaluation (IHME, a,b,c,d,e). We additionally include the proportion of households that could be considered to be multi-generational (contain a grandparent living with a grandchild younger than 18 years) from the American Community Survey (ACS) (averaged over 2009-2013), which is of interest when considering COVID-19 mortality. For all pre-treatment covariates, the year of data closest to 2013, or before, is used, as any time before January 2014 is considered pre-treatment. Table 1 provides a summary of the covariates included.

Finally, we use the National Bureau of Economic Research's (NBER) list of county adjacency to indicate which counties are adjacent in our matching procedure.

3. Conceptual Framework

When estimating treatment effects, Rubin and other scholars make the distinction between the design and analysis stages of an experiment (Rubin, 2007, 2008). In this pre-analysis plan, we present the matching we have implemented (the design of the experiment), and lay out a plan for the estimation of causal effects (the analysis stage). To secure the analogy to the design of an experiment, matching and its antecedent statistical calculations were made using only data from the pre-intervention period, i.e. after "blinding" ourselves to outcome variables.

In estimating a causal treatment effect using observational data, researchers are concerned by bias that is introduced through dependence between selection into the treatment group and any covariates that may be related to the outcome. In a randomized experiment, one would expect these variables to be independent. Thus, the goal is to design a study

5. The average PM2.5 was missing for a county in our analysis (Broomfield, CO), so we imputed the value as the mean PM2.5 value from the counties adjacent to Broomfield, CO.

Variable	Source	Years
Population	CDC	2009-2013
Age distribution	CDC	2009-2013
All mortality crosstabs	CDC	2009-2013
Population density	AHRF	2010
% Urban	AHRF	2010
Veteran Population	AHRF	2013
Median household income	AHRF	2013
% persons in poverty	AHRF	2013
Food stamps / SNAP	AHRF	2013
% 18-64 without health insurance	AHRF	2013
Unemployment Rate (16+)	AHRF	2013
PM2.5	AHRF	2013
Total Smoking	IHME	2012
Heavy Drinking prevalence	IHME	2012
Diabetes prevalence	IHME	2012
Hypertension prevalence	IHME	2009
Obesity prevalence	IHME	2011
Sufficient physical activity	IHME	2011
% Voting Republican	CQ Press	2012
% Multigenerational household	ACS	2009-2013

Table 1: Covariates used in modeling and matching and their sources as well as years available. When ranges of years are given, the variable used is the average across those years.

that manipulates the observational data in a way that a researcher can estimate treatment effects, pretending that the data arises from a randomized experiment. We approach this task through propensity score matching. We can asses how "close" to a randomized experiment our analysis, conditional on the matches, is by assessing the balance of observed covariates, conditional on the matches.

Finally, in the design of this study, we aim to make connections between statistical and ethnographic methodologies. We plan to do this by not only evaluating the effect of ACA Medicaid expansion for US states in aggregate, but also to highlight specific pairs or groups of counties

1. That are physically adjacent to one another,

2. That are well-matched and come from states in the same region of the US or neighboring states, and

3. That are well-matched but come from states that are far apart geographically.

We hope to use these vignettes of counties to facilitate extending the work of Metzl, described in his book *Dying of Whiteness*, comparing adjacent Tennessee and Kentucky counties in public health as well as public opinion about the ACA (Metzl, 2019). In addition to the analyses described in this protocol, future work may include *thick description*, or narrative descriptions of the vignettes of counties (Rosenbaum & Silber, 2001). These pairs of counties may also assist collaborators in choosing new locations to conduct interview and focus groups and allow for insight into how the match is performing.

4. Matching Methodology

In this section we describe methods we use to match counties in states that did or did not expand Medicaid as part of the ACA by mid-2014: optimal full matching with propensity scores, propensity score calipers, and penalties on pairings not satisfying any of 1–3 above. This combination of methods is intended to balance objectives of arranging matches for as many counties as could justifiably be matched, maximizing the information content or effective sample size of the matched configuration, securing covariate balance, and presenting multiple matches satisfying each of 1–3 above.

4.1 Evaluating proximity of matches

One of our goals in matching is to match with attention to satisfying 2–3 above, which requires a definition of what we deem to be "well-matched." We use a standard measurement of distance between members of a matched set — the Mahalanobis distance — to quantify how "well-matched" a pair of counties is (Rosenbaum, 2010). Specifically, we first calculate the Mahalanobis distance between all pairs of treatment and control counties in terms of the variables included in the propensity score model (described in §4.2 below). We then calculate the distribution of Mahalanobis distances between adjacent counties, using the lower 20th percentile of this distribution as a cut-off to consider matches to be "close." This method of identifying "well-matched" counties is used as a penalty in the matching procedure described in the following sections.

4.2 Propensity score model

Propensity scores are commonly used in estimation of causal effects and to create matched sets (see Stuart & Green, 2006, for an in-depth explanation of propensity scores). Table 2 shows the covariates used in the propensity score. The intention behind the variables included is to balance the treatment and control groups in terms of covariates we would expect to be associated with a state expanding Medicaid, general mortality, and COVID-19 mortality (such as PM2.5 and % multi-generational households).[6] Before estimating the propensity score, we first exclude 6 counties with outlying baseline mortalities (either healthcare amenable, flu, or opioid) as these covariates directly influence our outcomes of interest. We define outlying counties as those whose baseline value is outside of 1.5 interquartile ranges (IQR) from the median of the distribution and two or more pooled standard deviations from the nearest county with the opposite treatment status. After these exclusions, we fit a logistic regression model of Medicaid expansion status on all of the covariates, weighted by the adult aged 20-64 population size, as our initial propensity score model. This model is then refitted excluding counties that have extreme propensity scores, specifically counties whose (initial) propensity scores differ from those of their comparison group by more than $s_p/4$, where s_p is a weighted, outlier-resistant pooled standard deviation[7] of those scores. Our matching procedures then minimize discrepancies on this

6. Mortality attributed to opioid overdose is included since there is a discussion around whether the effects of Medicaid expansion in terms of health outcomes have largely been offset due to the opioid epidemic (Borgschulte & Vogler, 2020).

7. In this dispersion calculation, the treatment and control groups' standard deviations are replaced with population-weighted median absolute deviations, rescaled for comparability with the ordinary s.d. The

updated propensity score, although the original propensity figures in the determination of an additional matching restriction.

4.3 Full match

We use the `optmatch` package in `R` (Hansen & Klopfer, 2006a) to match counties using a full matching procedure (Hansen, 2004; Rosenbaum, 1991). Thus, we do not restrict our matching of counties only to pairs, but rather allow for groups of control and treatment counties to be matched.

Within calipers of width one-quarter (Rosenbaum & Rubin, 1985) of s_p (as defined above) in the initial propensity score, counties are matched on the updated propensity score (Rubin, 2001), with additional penalties added to the propensity-score discrepancies used in matching if the potential pairing does not satisfy any of 1–3.[8] We additionally add a stability increment to the matched distance in order to incentivize pair matches (Hansen & Klopfer, 2006b).

106 counties are trimmed from our analysis due to the propensity score caliper, as shown in Figure 1. Based on our full matching structure and after trimming the 6 counties due to outlying baseline mortalities and 106 counties from the propensity score caliper, our effective sample size is 828.1. Table 2 shows the weighted average values of each variable for the treatment and control groups as well as the absolute difference[9] before and after matching (weighted by the adult aged 20-64 population) along with two measures of variable dispersion. The first ("Overall") dispersion variable is the pooled standard deviation of county-level measurements of a variable, calculated with weighting for county working age population. The "Adjacent" dispersion variable is the root mean squared of the distances in covariates between adjacent counties, divided by $\sqrt{2}$ for comparability to the pooled standard deviation.[10] Therefore, one can compare the absolute differences to the variability of the covariate across all counties or across only adjacent counties. A permutation test (Hansen & Bowers, 2008) of the combined balance across all covariates indicates overall balance commensurate with what random assignment within matched group would have produced ($p = .55$), so we deem this matching structure satisfactory.

The larger differences of some covariates as compared to the pooled standard deviation observed in Table 2 aligns with our understanding of the characteristics of states that

rescaling multiplies them by $1/\tilde{\Phi}^{-1}(0.5) = 1.4826$, where $\tilde{\Phi}(\cdot)$ is the cumulative distribution function of $|Z|$, $Z \sim \mathcal{N}(0,1)$, ensuring close proximity to the ordinary s.d. in the special case of large samples from $\mathcal{N}(\mu, \sigma^2)$ populations.

8. Specifically, we penalize county pairs that are not adjacent, nor have a Mahalanobis distance greater than the bottom 20th percentile of the Mahalanobis distances between adjacent counties the most and penalize counties that are not adjacent but have a Mahalanobis distance within the bottom 20th percentile of the Mahalanobis distances between adjacent counties with half of that penalty. Adjacent counties have no penalty.
9. Unadjusted and post-matching averages of the treatment group are weighted means over all Medicaid-expanding counties or all expansion counties that were placed into a matched set, respectively, with weights proportional to the size of the county's working-age adult population. Unmatched averages over the control group are calculated similarly, whereas post-matching control group averages are means over the subset of control counties that were matched, each weighted by the product of the ratio of treatment to control counties within its matched set and the size of its working-age adult population.
10. Unlike the pooled standard deviation calculation, however, this calculation is not weighted by the adult aged 20-64 population.

have and have not expanded Medicaid. For instance, percent Hispanic is higher in the treatment group somewhat due to large counties in California and cannot be balanced through matching without excluding large counties in the U.S. such as Los Angeles. Because the distributions of these variables in the expansion and non-expansion counties overlap, we can account imperfect balance for in the outcome analysis through covariance adjustment (See Figure 3 in Appendix A).

We again assessed remaining outliers by looking at overlap for the full distribution of counties across each variable on which we match, flagging outlying counties for each variable based on the same criteria described above. We do not remove any counties beyond the 6 excluded as a result of outlying baseline mortalities. Instead, most discrepancies are easily explained after closer inspection. For instance, two counties, Crowley, Colorado and Pulaski, Georgia are outliers with respect to percent Male. Crowley, a small county of 5000 that is 75% male, has a men's correctional facility housing roughly 1500 inmates. Pulaski, the outlier in the other direction with a population that is only 41% male, has a women's correctional facility. Rather than excluding these counties to achieve better balance on % Male, we allow covariance adjustment to correct for outliers.

In addition to achieving reasonable balance, this matching structure addresses our attention to satisfying 1–3 of §3. The matching structure includes over 48 matches that are considered close in terms of the Mahalanobis distance and, as Figure 2 shows, there are additionally 18 1:1 matches of adjacent counties.

4.4 Re-evaluating matching structure for later expanded states

The matching structure described above uses treatment defined by states who adopted ACA Medicaid expansion by June 2014. This places in the control group nine states which expanded Medicaid between mid-2014 and 2020 (see Figure 1 and Appendix A). This is not a problem for our analysis of the effect of Medicaid expansion on 2014 mortality. By 2020, however, nine more states expanded Medicaid, increasing the size of the treatment group while reducing that of the control group. We therefore separately evaluated the extent which our match mitigates differences in potentially confounding baseline (2013 and prior) variables between 2020 treatment and control groups.[11] We were satisfied with the balance using re-defined test and control groups, so chose to use the same matching structure for both analyses. See the Appendix A Table 6 for detailed summaries of balance.

11. We additionally considered adding an additional propensity score caliper using a propensity score model with this new treatment assignment vector as the outcome, but did not find that this assisted with balance, so chose to only use the original propensity score model for clarity and simplicity.

	No Adjustment			After Matching			Dispersion	
	Control	Treat.	Dif.	Control	Treat.	Dif.	Overall	Adjacent
% White	80.0	80.8	0.8	84.2	80.8	-3.4	14.1	8.3
% Black	15.9	10.5	-5.4	10.3	10.8	0.5	12.4	6.4
% Hispanic	12.9	15.9	2.9	8.5	15.3	6.8	15.3	5.2
% Male	48.4	48.5	0.2	48.3	48.5	0.2	1.6	2.4
% 20-34	28.2	28.1	-0.1	26.1	28.1	2.0	5.4	5.1
% 35-44	17.9	18.0	0.0	17.4	17.9	0.5	2.2	1.7
% 45-54	19.3	19.6	0.3	19.9	19.6	-0.3	1.8	1.5
% 55-64	16.3	16.4	0.1	17.1	16.4	-0.7	1.9	1.9
All Mortality	373.6	314.5	-59.1	354.8	316.5	-38.3	101.5	90.0
20-34 Mortality	111.5	91.0	-20.5	107.3	91.5	-15.8	35.3	52.1
35-44 Mortality	194.3	159.9	-34.5	180.5	161.0	-19.5	61.9	80.7
45-54 Mortality	452.5	377.4	-75.1	410.0	379.7	-30.2	122.7	119.5
55-64 Mortality	923.2	791.7	-131.5	842.9	795.4	-47.5	204.4	181.5
White Male Mortality	398.8	345.3	-53.5	373.7	346.1	-27.6	95.4	86.1
White Female Mortality	232.0	198.3	-33.7	218.2	198.7	-19.5	59.8	57.8
Black Male Mortality	517.4	467.0	-50.4	472.6	469.6	-2.9	224.8	550.8
Black Female Mortality	322.2	296.0	-26.2	316.8	297.3	-19.5	229.5	740.4
Other Race Male Mortality	182.2	194.0	11.9	216.7	194.0	-22.8	186.8	562.7
Other Race Female Mortality	111.7	115.6	3.9	108.8	115.2	6.3	114.5	294.2
Healcare Amenable (non-flu) Mortality	195.6	165.0	-30.6	176.9	166.2	-10.7	49.3	49.1
Opioid Mortality	21.8	20.7	-1.1	24.7	20.8	-3.9	10.7	20.1
Flu Mortality	4.7	3.9	-0.8	4.4	3.9	-0.4	3.2	10.3
Population Density	926.6	3267.6	2341.0	1184.3	3430.6	2246.4	6867.4	1085.2
% Urban	76.4	85.2	8.8	77.3	85.1	7.7	23.9	25.6
% Veteran	7.7	6.2	-1.5	7.6	6.2	-1.3	2.5	1.7
Median Income	50579.9	58156.3	7576.4	57560.7	57676.6	115.9	13998.3	7251.3
% Poverty	16.7	15.3	-1.4	14.1	15.3	1.1	5.4	4.3
% SNAP	15.7	14.1	-1.6	13.8	14.2	0.3	6.5	5.0
% No Health Insurance	23.1	18.3	-4.9	17.1	18.4	1.2	6.7	3.3
Unemployment Rate	6.9	7.9	1.0	7.2	7.8	0.7	2.0	1.5
PM2.5	0.1	0.1	-0.0	0.1	0.1	-0.0	1.7	0.2
Smoking	21.3	19.0	-2.2	21.4	19.2	-2.2	4.3	2.4
Heavy Drinking	7.6	8.7	1.0	8.0	8.6	0.7	1.8	1.3
Diabetes	15.0	13.8	-1.2	13.7	13.8	0.2	2.1	1.3
Male Hypertension	37.5	35.8	-1.7	36.3	35.9	-0.5	3.2	1.9
Female Hypertension	40.0	38.3	-1.8	38.6	38.4	-0.2	3.4	2.0
Male Obesity	35.4	32.5	-2.9	34.5	32.6	-1.9	4.3	2.2
Female Obesity	37.8	34.6	-3.2	36.0	34.7	-1.3	5.1	3.3
Male Physical Activity	54.2	57.9	3.7	56.3	57.8	1.5	5.2	3.7
Female Physical Activity	49.8	55.0	5.2	52.3	54.7	2.3	6.0	3.4
% Republican	54.3	42.1	-12.2	48.6	42.3	-6.2	14.8	9.1
% Multigenerational Households	0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.1
Population	5.2	5.5	0.4	5.2	5.5	0.3	0.6	0.4

Table 2: Differences between control and treatment (ACA Medicaid expansion by June 2014) counties before and after matching along with the pooled, weighted standard deviation of the covariates between treatment and control counties ("Overall") and the root mean squared of the distances in covariates between adjacent counties, divided by $\sqrt{2}$ for comparability ("Adjacent"). The test of overall balance results in a p-value of .55 which is insignificant, indicating good balance.

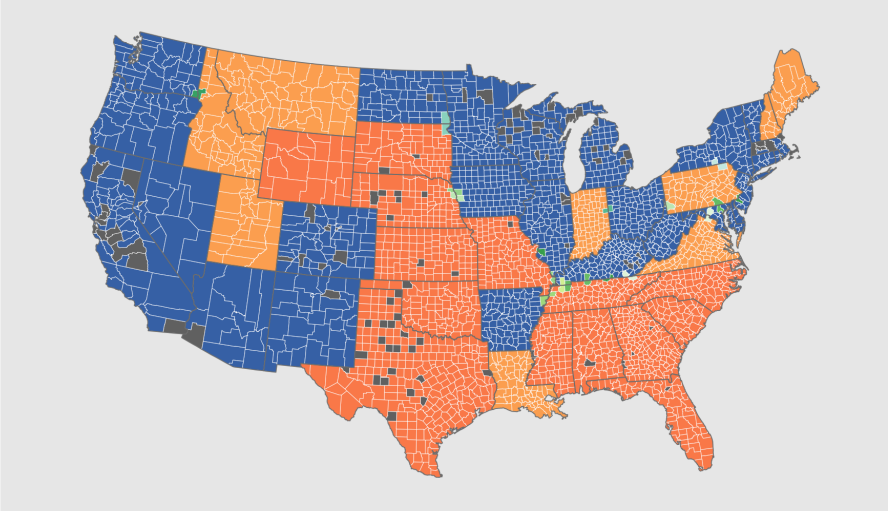

Figure 2: One-to-one matches with two counties that are adjacent to a county of the other treatment condition (filled in green colors).

5. Treatment Effect Estimation

As discussed above, with the same matched sets, we plan to do two separate estimates of two treatment effects: (I) The effect of Medicaid expansion by June 2020 on mortality during the COVID-19 pandemic (hereafter the "2020 analysis"); and (II) The effect of Medicaid expansion by June 2014 on mortality in 2014 (hereafter the "2014 analysis"). In this section, we describe our pre-analysis plan for treatment effect estimation of these two treatment effects.

5.1 Mortality data

5.1.1 2020 Mortality data

For the 2020 analysis, we will use cumulative 2020 mortality from January through December 2020. As of March 19, 2021 the NCHS is reporting cumulative COVID-19 attributed mortality at the county level for counties that have reported 10 or more deaths to the CDC (CDC, b). This publicly available data from the CDC is updated weekly and the cumulative counts begin the first week of January, 2020. The data additionally contains provisional overall cumulative death counts for these same counties, and we will use these provisional all-cause death counts as our primary outcome, with the deaths attributed to COVID-19 as a secondary outcome.[12] Through March 3, 2021, the 2020 all cause mortality

12. We chose to use overall deaths rather than deaths attributed to COVID-19 because different counties and states can use varying requirements for reporting a death as relating to COVID-19 (ranging from

output was censored, where mortality is suppressed for counties that have fewer than 10 deaths officially attributed to COVID-19. We note that starting on March 10, 2021, the CDC began reporting the provisional all-cause mortality for all counties with at least one death attributed to any cause since January 1, 2020 (CDC, b). However, since we will be considering deaths between January and December of 2020, we will still be reliant on the data that was censored. As of January 6, 2021 (the dataset which includes the entirety of deaths in 2020) this dataset included 1,810 (uncensored) counties (out of around 3,100 counties in the mainland US). These uncensored counties account for around 94% of the total adult population in the data and as expected, under-represents primarily rural counties. The average value of the "% Urban" variable across all counties is 41.5% while the average value of "% Urban" for the currently uncensored counties is 55%. Because the data was only reported for around two thirds of counties through the end of 2020, we will need to consider how to analyze considerably censored data in the outcome analysis.

Additionally, the CDC notes that different states report deaths within different time periods, which complicates any between-state comparisons. We plan to address this complication in the outcome analysis using data from the FluView Surveillance System maintained by the CDC, which provides weekly state-level pneumonia, influenza, and all-cause mortality counts and is updated weekly (NCHS).[13] Because the data is updated weekly, and counts from previous weeks are updated as reporting continues over time, we can use the data to estimate the number of weeks that mortality counts tend to lag in a state. We plan to use the all-cause mortality counts in the FluView data as our best approximation for the lags in the county level COVID-19 and all-cause mortality counts, so we are assuming that these lags are similar. Our proposed approach is described in Appendix B. We will also determine states that have outlying lag times if our calculated average lag is greater than 10 weeks or is highly irregular, and counties from these states will not be compared to other counties in our test statistic (described in detail in Section 5.2.5). We plan to do a final pull the provisional COVID-19 mortality data on March 17th, 2021. This date will allow us to have 10 weeks of data past the end of 2020, so that we can implement either of the approaches to adjust for the lag in data reporting as described below to analyze mortality over the entire year of 2020.

We are considering two possible approaches to adjust for the determined reporting lag: (A) in our outcome analysis, using a method that only compares mortality between states that have similar lag times in mortality reporting or (B) using our understanding of the lag patterns to estimate mortality counts that are comparable across states. Approach (A) is described in detail in §5.2.5 as well as how we will chose between these two approaches. If approach (B) is chosen, and only then, we will utilize the FluView data to estimate mortality

including deaths with strong physical suspicion of COVID-19 to only including deaths with the ICD-10 code) and we also expect that there is variation between hospitals and mortuaries about the human tendency to over or under count deaths as relating to COVID-19. Therefore, we find that evaluating all mortality in 2020 is likely a more reliable measure. In addition to not involving the same level of possible measurement error, early research has found excess deaths from certain conditions that are not COVID-19 as a "symptom" of the pandemic (Woolf et al., 2020).

13. The CDC actually recommends referring to the FluView Surveillance System as a a potential way to evaluate the reporting lag.

counts that are approximately comparable between states rather than just to calculate a reporting lag.[14] Otherwise, we will use the provisional 2020 mortality counts as is.

The same provisional all-cause mortality data at the county-level is available by race, for counties with more than 100 COVID-19 deaths, and suppressing any cells that are under 10 deaths (CDC, a). We therefore plan to use the same outcome analysis method described in the following sections using this data, for subgroup analyses by race as secondary analyses. We are additionally interested in evaluating the treatment effect in US counties and counties with "supermajority white" population.[15] We specify the test statistic used in these subgroup analyses in §5.2.7. Finalized 2020 county-level mortality data from the CDC will likely not be released until December 2021, so we will rely on provisional counts and plan to revisit the analysis with finalized data when it is available.

5.1.2 2014 Mortality data

For the 2014 analysis, we plan to use the restricted use, detailed mortality files from the NCHS (the same county-level morality data as used in matching), which will not be censored. This data also has information on all of the multiple causes of deaths that are listed on an individuals death certificate, in addition to the primary cause of death. We will have access to mortality counts by race and ethnicity in the 2014 mortality data, so will estimate treatment effects for subgroups within these categories. As in the 2020 analyses, all-cause mortality is a primary endpoint. In the 2014 analysis, we also consider two additional primary endpoints within the broad category of healthcare amenable deaths: healthcare amenable deaths attributed to flu; and all other healthcare amenable deaths. We plan to do this because healthcare amenable deaths are those one might expect to be impacted by increased insurance coverage and influenza and pneumonia mortality is the closest categorization in the ICD-10 codes in 2014 to COVID-19. As with the 2009-2013 data, we use Sommers et al's (2014) definition of healthcare amenable deaths, as well as the common definition of flu mortality as deaths with a pneumonia or influenza ICD-10 code listed as primary cause of death, to identify healthcare amenable deaths and those attributed to flu. Finally, we will consider deaths attributed to opioids as a secondary outcome because they increased during the years before and after ACA Medicaid expansion.[16]

14. Specifically, to do this, we plan to utilize the fact that the FluView data reports mortality data for each week of the year, that is updated as the data is updated weekly. Therefore, in a dataset that is of cumulative deaths, of the total cumulative deaths that are added every week, we expect that a proportion of them are actually attributed to the preceding weeks mortality. Thus, we plan to estimate an adjustment factor that we interpret as the average percent of the additional mortality reported in each week (t) that should be attributed to previous weeks ($t-1$, $t-2$, $t-3$, ...). We would then estimate cumulative mortality for week $T-x$ (where T is the week we pull the data and x is determined as the longest lag time in the states, excluding outlier lag times), by using these percentages to subtract mortality that would be attributed to weeks $T-x+1$,...s in the cumulative mortality counts in the data pulled at each week s ($T-x+1 \leq s \leq T$).
15. Specifically, we consider a county to be "supermajority white" if the proportion of the population that is white (aggregated across 2009-2013) is in the top 97.5th percentile of all counties (weighted by county population size), which is greater than 98.43% white.
16. We considered removing opioid related deaths from the other causes of death because the opioid epidemic has overlapped with the period post-Medicaid expansion (Borgschulte & Vogler, 2020). However, excluding opioid related mortality ultimately only excluded .3% of deaths attributed to influenza or pneumonia and .4% of other healthcare amenable deaths in 2013; so we elected to avoid this complication.

5.2 Estimation

To address the issue of estimation using censored data as well as to include covariate adjustment, we rely on two separate methods described by Rosenbaum — randomization inference for covariance adjustment and for partially ordered outcomes (Rosenbaum, 2002a,b). The same method will be used in both the 2014 and 2020 analyses, with differences noted as appropriate in the sections following.

5.2.1 Notation and mode of inference

To establish notation and an overview of these methods, we first discuss the binary treatment case, applicable to analyses with 2014 mortality as the outcome. For analyses with 2020 mortality as the outcome, where there is variation by state in the duration of time since Medicaid expansion, we adapt these methods to accommodate nonbinary treatments occurring in integer doses. These adaptations will be described in §5.2.6.

Assume we have K matched sets with $n_k \geq 2$ counties in each matched set, so that $\sum_{k=1}^{K} n_k = N$, the total number of counties. In the potential outcomes tradition of causal analysis, we assume that for the N counties, each county, indexed by the matched set $k \in \{1, \ldots, K\}$ it is a member of, and the index within that matched set, $i = 1, \ldots, n_k$, has two potential outcomes, y_{ki}^t and y_{ki}^c, which would be observed if the county was in the treatment group or the control group, respectively. We only observe one of these outcomes, depending on the actual treatment assignment. For analyses with 2014 mortality as outcome, let $\mathbf{Z} = (Z_{11}, ...Z_{Kn_K})^T$ be the vector of (binary) treatment assignments where Z_{ki} is an indicator of whether county i in matched set k was "assigned to treatment," i.e. fell within a state that had expanded Medicaid by mid-June 2014. Then, the observed outcome for county i in matched set k, a mortality count, is defined as $Y_{ki} = Z_{ki}y_{ki}^t + (1 - Z_{ki})y_{ki}^c$. For analyses with 2020 mortality as outcome, treatment assignment is an integer random variable $D_{ki} \in \{0, 1, \ldots, 7\}$ recording the number of years since the state's Medicaid expansion as of July 1, 2020 and $Y_{ki} = \sum_{d=0}^{7} \mathcal{I}\left[D_{ki} = d\right] y_{ki}^{(d)}$, where $y_{ki}^{(d)}$ is the potential outcome under dose d).

Tests and confidence intervals will follow the Fisher randomization inference paradigm, with hypotheses other than the strict null formulated in §5.2.2. We assume a constant, multiplicative treatment effect for all analysis. Therefore, for the analyses of 2014 mortality with a binary treatment effect, we assume that there is a constant δ such that for all $k = 1, ..., K$ and $i = 1, ..., n_k$, $y_{ki}^t = \delta y_{ki}^c$. For the analyses using the 2020 mortality outcome, with treatment assignment as the random integer variable D_{ki}, we assume a constant treatment effect across doses (there is a constant δ such that for all $k = 1, ..., K$ and $i = 1, ..., n_k$, $y_{ki}^d = \delta y_{ki}^0$, when $d > 0$). Therefore, for example, the strict null for the primary 2014 analysis hypothesis is:

$$H_0 : y_{ki}^t = y_{ki}^c, k = 1, ..., K, i = 1, ..., n_k$$

Where y_{ki}^t is the potential 2014 all cause mortality outcome if county i in matched set k is in a state that expanded Medicaid by June 2014, with the matched structure defined in §4.3.

All inferences will condition on treatment assignment margins within matched sets. For analyses with binary treatment assignment Z, this means conditioning on $\{\sum_{i=1}^{n_k} Z_{ki} : k =$

$1, \ldots, K\}$; for analyses with integer treatment assignments D, it means conditioning on

$$\{\sum_{i=1}^{n_k} \mathcal{I}\left[D_{ki} \geq d\right] : d = 1, \ldots, 7; k = 1, \ldots, K\}.$$

For primary analyses assuming hidden biases to be absent, this means we will make statistical inferences against permutation distributions, with the random permutations of treatment assignment labels occurring within matched sets, and independently across matched sets.

5.2.2 Adjusted Outcomes

Following Rosenbaum (2002a), under the assumption that there is a constant, multiplicative treatment effect, i.e. that there is a constant δ such that for all $k = 1, ..., K$ and $i = 1, ..., n_k$, $y^t_{ki} = \delta y^c_{ki}$, a test statistic for testing the null hypothesis $H_0 : \delta = \delta_0$ can be calculated by adjusting the observed outcome. Specifically, first, adjusted outcomes $\tilde{Y}_{ki} = Y_{ki}\delta_0^{-Z_{ki}}$ are calculated (Rosenbaum, 2002a).

With the constant treatment effect assumption as described in the previous section §5.2.1, in the analyses with the 2020 mortality outcome, the adjusted outcomes for testing the null hypothesis $H_0 : \delta = \delta_0$ are calculated as $\tilde{Y}_{ki} = Y_{ki}\delta_0^{-I(D_{ki}>0)}$ (where I is the indicator function).

The outcomes are adjusted in the sense of dividing the null hypothesized treatment effect from them. The residuals from a regression of the adjusted outcomes on covariates are then calculated. We will adapt this approach to our specific data setting, as described in the following section.

5.2.3 Covariance Adjustment Modeling

With our covariance adjustment, we are in a bit of a bind to control for mortality in the years prior to the outcome mortality, while also not biasing the estimation by adjustment that uses mortality counts after the treatment in 2014. In order to address this issue, we will consider two modeling options, which only differ by which years of data are used as the outcome in the model fit, as described below.

First, instead of fitting a model on 2014 or 2020 mortality, we will instead fit a model on 2013 mortality. The 2013 morality does not need to be adjusted by the hypothesized treatment effect because it is pre-treatment. To facilitate subgroup analyses (§5.2.7), we will fit a model where the observation is at the county, race, and age group level. We will fit a LASSO-regularized quasi-Poisson GLM, using cross-validation to select the penalty parameter.[17] This model will include all covariates included in Table 1 with all mortality variable averaged over the 5 years prior to the year of the dependent variable (2008-2012 in the 2013 model) in addition to the group (county/race/age group) specific all-cause and health amenable mortality (with no opioid-related multiple causes of death) for each of the 2 years prior (2011 and 2012 for the 2013 model) and the interaction between age group and all of these variables. Then, once a subset of covariates is determined by the LASSO, we will fit a negative binomial model on the 2013 mortality with the subset of covariates and interactions.

17. We plan to use the `glmnet` function from the `glmnet` package in `R`.

We will then re-fit this model four times, twice with mortality outcomes from 2014 and again twice more with the outcome variable taken from 2020. Each of these fits use the same pre-2014 covariates and excludes intervention group counties. That is, the model is fit to those counties that had not expanded Medicaid as of June 30, 2014 or June 30, 2020, respectively. The covariates will be those previously selected with the help of the LASSO, but with baseline mortality variables from Table 1 averaged over 2009-2013 and the two years of mortality data at the group level being for years 2012 and 2013.[18] The model will be fit constraining the coefficients on each variable to remain the same as the model fit on 2013 mortality, but allowing the intercept to vary. Then, the model will be fit again without this constraint. We will finally conduct a likelihood ratio test at the .1 level, comparing the versions of the model with and without predetermined slopes, to test the null hypothesis that the model fit on 2013 mortality is as good a fit as the model fit on 2014 or 2020 mortality, respectively.

Should this null hypothesis be sustained, we will use covariate slopes as estimated with 2013 mortality outcomes to predict mortality in 2014 or 2020. (The decision is made separately for analyses of 2014 and 2020 outcome data, in either case decided by the corresponding likelihood ratio test.) In this case, the same mortality count predictions $\{\hat{Y}_{ki} : k, i\}$ contribute to the calculation of residuals (cf. (1) in §5.2.4 below), regardless of the value of δ_0 for which $H_0 : \delta = \delta_0$ is under test (cf. §5.2.2). Should the likelihood ratio test reject the slopes fitted with 2013 mortality outcomes, we will be forced to instead re-fit a separate version of the outcome model for each value of δ_0 and corresponding hypothesis $H_0 : \delta = \delta_0$. These fits involve outcomes of form $\{\tilde{Y}_{ki} : k, i\}$, as described in §5.2.2. Let $\hat{Y}_{ki}$ denote the fitted model prediction for a county, aggregated as appropriate to the county or subgroup county population level, depending on the analysis, and $\hat{\mathbf{Y}}$ denote the vector of fitted values.

5.2.4 Alignment within matched sets

In each test of a hypothesized value for δ_0, residuals of the covariance adjustment procedure will be aligned [19] in the sense of subtracting the mean residual from within each matched set. The residuals we plan to calculate are on the scale of the linear predictors (i.e. the log of the outcome). For purposes of this alignment (only), counties for which the outcome is censored (so that a residual is unavailable) are treated as having 0 residuals. These aligned residuals e_{ki} will be used to calculate the test statistic. To make concrete in our notation:

$$e_{ki} = \log(\max(\min(\hat{\mathbf{Y}}), \tilde{Y}_{ki})) - \log(\hat{Y}_{ki}) - \frac{1}{n_k}\sum_{j=1}^{n_k}\left[\log(\max(\min(\hat{\mathbf{Y}}), \tilde{Y}_{ki})) - \log(\hat{Y}_{kj})\right] \quad (1)$$

Note that as described above, if the outcome mortality for county ki is censored, we define $\log(\max(\min(\hat{\mathbf{Y}}), \tilde{Y}_{ki})) - \log(\hat{Y}_{ki}) \equiv 0$.

18. We use the same years of data for both 2014 and 2020 because any mortality post 2014 could be affected by treatment.

19. In the special case of pair matching without censoring, tests using as a test statistic the treatment-group sum of ranks of *aligned* residuals, calculating ranks without reference to matched sets, are equivalent to signed-rank tests. In this sense alignment within matched sets enables generalization of signed-rank and similar nonparametric tests from pair to full matching.

5.2.5 Partially ordered outcomes

With censored data, it is useful to consider the censored outcomes as a partially ordered set. Therefore, to accommodate the issue of censored data in the 2020 analysis, we will define a partial ordering to apply to our definition of a test statistic, noting that for the uncensored 2014 data, while we will apply the same methods, this will be a total order rather than a partial order.

Let the outcome be $r = (c, e)$, with c an indicator of whether the county in question had fewer than 10 COVID-19 deaths recorded by the CDC and e the aligned residual of a mortality count (in the 2014 analysis, $c = 0$ for all counties). In order to account for the lag in reporting mortality across states, as mentioned in §5.1, we are considering either adjusting the test statistic, through our definition of the partial ordering, or through interpolation. We therefore define two possible partial orderings here (A) accounts for lag in the partial ordering and (B) relies on interpolation of the data. Therefore, in the case of the 2020 outcome data, the residual value differs between (A) and (B) (with $\tilde{e}_{ki}$ in (B) below representing the residual estimated using the interpolated 2020 mortality data), but in the 2014 analysis, the residuals are the same value in the two partial orderings ($\tilde{e}_{ki} = e_{ki}$) since there will be no interpolation of the 2014 outcome data regardless of the choice of partial ordering.

(A) Define as follows the partial ordering ($\lesssim$) on outcomes $\{r_{ki}$: counties i within matched sets $k\}$ where s_{ki} is defined as the lag in weeks for the state that county i within matched set k resides:

1. $r_{ki} \lesssim r_{\ell j}$ if $c_{ki} = 1$ and $c_{\ell j} = 0$
2. $r_{ki} \lesssim r_{\ell j}$ if $c_{ki} = c_{\ell j} = 0$ and both of the following hold:
   (a) $e_{ki} \leq e_{\ell j}$
   (b) $s_{ki} \leq s_{\ell j}$

(B) Define as follows the partial ordering ($\lesssim$) on outcomes $\{r_{ki}$: counties i within matched sets $k\}$:

1. $r_{ki} \lesssim r_{\ell j}$ if $c_{ki} = 1$ and $c_{\ell j} = 0$
2. $r_{ki} \lesssim r_{\ell j}$ if $c_{ki} = c_{\ell j} = 0$ and $\tilde{e}_{ki} \leq \tilde{e}_{\ell j}$.

In other words, a censored outcome is always considered to be lower in the ordering than a non-censored outcome. In (A) we are adding the additional stipulation that a county will only be considered to have higher mortality than another county if the population adjusted residual is greater and the reporting lag is no less than the other county.

We will chose between these two partial orderings in our outcome analysis by implementing both in our 2014 analysis and comparing the relative widths of the confidence intervals that result from each method. If the width of the confidence intervals using partial ordering (A) are no more than 10% wider than the intervals using partial ordering (B), then we will calculate our test statistics using partial ordering (A). This would avoid doing any data imputation, as we will only impute the 2020 mortality data if partial ordering (B) is used.

We are additionally introducing weighting by a specialized adjustment of county population size (m_{ki}) of a county to the aligned rank test in order to account for the fact that we are using aggregated, rather than individual data. These adjusted population sizes take the age distribution of a county into account. Specifically we will multiply the (normalized) 2018 nationwide age specific mortality rates with the age group population in that county and sum these values across age groups.

Let I_{ki} be an indicator that takes a value of 1 if county i within matched set k is in a state that is considered to have outlier lag in mortality reporting and define

$$u_{ki\ell j} = \begin{cases} m_{ki}m_{\ell j} & \text{if } r_{\ell j} \lesssim r_{ki} \text{ and } I_{\ell j} = I_{ki} = 0 \\ -m_{ki}m_{\ell j} & \text{if } r_{ki} \lesssim r_{\ell j} \text{ and } I_{\ell j} = I_{ki} = 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$q_{ki} = \sum_{\ell=1}^{K}\sum_{j=1}^{n_\ell} u_{ki\ell j}.$$

This adapts Rosenbaum (2002b, §2.8.4) to entertain comparison across as well as within matched sets. By the argument of Mantel (1967) reviewed there, for binary random variables $(Z_{ki} : k; 1 \le i \le n_k)$,

$$\sum_{k,l=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_\ell} Z_{ki}(1 - Z_{\ell j})u_{ki\ell j} = \sum_{k=1}^{K}\sum_{i=1}^{n_k} Z_{ki}q_{ki}.$$

At this point, we have defined a test statistic for a weighted aligned rank test with a binary treatment variable.

5.2.6 Incorporating "doses"

For purposes of the 2020 analysis, counties are considered to have had a "dose" of the Medicaid expansion treatment equal to approximately the number of years since expansion, with expansion between July 1, 2019 and June 30, 2020 considered a dose of 1 year, between July 1, 2018 and June 30, 2019 considered dose of 2 years, etc. Thus doses d_{ki} range from 0 to 7 (see Appendix A Table 4 for the dose assignments for each state).

Building upon the previous section, to incorporate treatment "dose", our tests will use the statistic:

$$\begin{aligned} W = \sum_{k,\ell=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_\ell}(D_{ki} - D_{\ell j})_+ u_{ki\ell j} &= \sum_{d=0}^{7}\sum_{k,\ell=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_\ell}\mathcal{I}\left[D_{ki} > d\right]\mathcal{I}\left[D_{\ell j} \le d\right]u_{ki\ell j} \\ &= \sum_{d=0}^{7}\sum_{k,\ell=1}^{K}\sum_{i=1}^{n_k}\mathcal{I}\left[D_{ki} > d\right]\sum_{j=1}^{n_\ell} u_{ki\ell j} \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} D_{ki}q_{ki}, \end{aligned}$$

where "$(x)_+$" denotes x's positive part $\max(x, 0)$.[20]

20. The same Mantel (1967) argument justifies progression from the first to the second line.

We will reject the null hypothesis when the realized value of the test statistic (W) falls in either the upper or lower tail of the distribution arising by, independently, for each i, permuting realized values of $(D_{i1}, \ldots, D_{in_k})$. Note that this adapts the methods of Rosenbaum (1997) and Rosenbaum (2002b, sec. 2.8) to full matching and treatments with doses.

For the 2014 analysis, we define $D_{ki} = Z_{ki} \in (0, 1)$, and therefore W is equivalent to the statistic outlined in §5.2.5, without incorporating doses. In this way, we use the same test statistic in both analyses.

5.2.7 Subgroup analyses

In addition to analyses for 2020 and 2014 which use mortality disaggregated by race as the outcome, we are interested in evaluating the effect of Medicaid expansion on primarily white communities. We therefore take an approach to define counties as "supermajority white" as described in §5.1 and will compare outcomes in these counties to the counties that they were matched with. In other words, we plan to compare the outcomes in these subgroups to similar counties, as defined by our matching structure.

Define the indicator s_{ki} to be equal to 1 if county ki is 'supermajority white" and 0 otherwise. Then, define $\mathcal{K} = \{k : \sum_{i=1}^{n_k} s_{ki} > 0\}$ (in other words, $\mathcal{K}$ is the set of matched sets for which there is at least one county that is "supermajority white"). Then, we redefine q_{ki} as

$$q_{ki} = \mathcal{I}\left[s_{ki} = 1\right] \sum_{\ell \in \mathcal{K}} \sum_{j=1}^{n_\ell} u_{ki\ell j}.$$

And make a slight adjustment to the test statistic W, so it is only defined over $\mathcal{K}$:

$$\sum_{k \in \mathcal{K}} \sum_{i=1}^{n_k} D_{ki} q_{ki}$$

Using this adjusted test statistic, we will still reject the test using the same method as described in the previous section.

5.2.8 Treatment effect estimation

We will calculate a confidence set for the multiplicative treatment effect δ by inverting the test as described in §5.2.6. We will then calculate a Hodges-Lehmann estimate. This estimation allows for an evaluation of treatment effect and accounts for covariate adjustment, censoring and considering treatment in doses. Table 3 reports whether we designate each outcome as a primary or secondary outcome, for each outcome by cause of death and population. We will use a max-t correction (Hothorn et al., 2008) to control for multiplicity in the treatment effect estimation for every outcome that we designate as a secondary outcome. Thus, for each cause of mortality, we will implement a max-t correction to account for multiplicity when conducting multiple tests on that outcome for different populations (e.g. all adults, Black adults, and white adults).

	2014 Outcome Status				2020 Outcome Status	
	Cause of Death				Cause of Death	
Subgroup	**All Cause**	**Healthcare Amenable**	**Flu**	**Opioid**	**All Cause**	**COVID-19**
All	Primary	Primary	Primary	Secondary	Primary	Primary
American Indian / Alaskan Native	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary
Asian	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary
Black	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary
Hispanic	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary
White	Secondary	Secondary	Secondary	Secondary	Secondary	Secondary
Supermajority white	Primary	Primary	Primary	Secondary	Primary	Primary

Table 3: Designation of primary or secondary outcome for the mortality outcomes for the 2014 and 2020 analyses. "Supermajority white" refers to the subgroup analysis described in § 5.2.7. Boxes illustrate the manner in which multiplicity correction will be implemented, controlling for multiplicity for secondary outcomes within each cause of death.

6. Discussion

Randomized experiments are not feasible to evaluate the impact of nationwide policy changes and therefore we must use statistical techniques to estimate causal treatment effects from observational data as if the data actually arose from a randomized experiment. In studies of the effect of ACA Medicaid expansion, we find county-level matching to be a persuasive method for estimating causal effects and reducing bias. There are limitations to propensity score matching, namely that reduction of bias depends on which covariates are observed and assumes that any covariates that are dependent on the treatment assignment have been observed. Now with only 12 states remaining who have not expanded Medicaid, it is critical to understand how expansion could have effected outcomes during the Covid-19 pandemic and whether expansion could save lives in the future. Particularly, we aim to understand how racialized political rhetoric has impacted health outcomes for different racial groups in the US in addition to primarily white communities, which is not yet addressed in the studies of Medicaid expansion.

Acknowledgments

This project is a variation of one suggested to us by Jonathan Metzl. We thank him for helpful comments and suggestions. We also thank Liz Neyens, Jordy Berne, and Mark Fredrickson for their consult and suggestions. This project was supported by US National Science Foundation grant DMS-1646108.

Appendix A. Supplemental Tables and Figures

State	Date Implemented	Treatment Group 2014	2020
Alabama		C	C [0]
Alaska	9/1/2015	C	T [5]
Arizona	1/1/2014	T	T [7]
Arkansas	1/1/2014	T	T [7]
California	1/1/2014	T	T [7]
Colorado	1/1/2014	T	T [7]
Connecticut	1/1/2014	T	T [7]
Delaware	1/1/2014	T	T [7]
District of Columbia	1/1/2014	T	T [7]
Florida		C	C [0]
Georgia		C	C [0]
Hawaii	1/1/2014	T	T [7]
Idaho	1/1/2020	C	T [1]
Illinois	1/1/2014	T	T [7]
Indiana	2/1/2015	C	T [6]
Iowa	1/1/2014	T	T [7]
Kansas		C	C [0]
Kentucky	1/1/2014	T	T [7]
Louisiana	7/1/2016	C	T [4]
Maine	1/10/2019	C	T [2]
Maryland	1/1/2014	T	T [7]
Massachusetts	1/1/2014	T	T [7]
Michigan	4/1/2014	T	T [7]
Minnesota	1/1/2014	T	T [7]
Mississippi		C	C [0]
Missouri		C	C [0]
Montana	1/1/2016	C	T [5]
Nebraska		C	C [0]
Nevada	1/1/2014	T	T [7]
New Hampshire	8/15/2014	C	T [6]
New Jersey	1/1/2014	T	T [7]
New Mexico	1/1/2014	T	T [7]
New York	1/1/2014	T	T [7]
North Carolina		C	C [0]
North Dakota	1/1/2014	T	T [7]
Ohio	1/1/2014	T	T [7]
Oklahoma		C	C [0]
Oregon	1/1/2014	T	T [7]
Pennsylvania	1/1/2015	C	T [6]
Rhode Island	1/1/2014	T	T [7]
South Carolina		C	C [0]
South Dakota		C	C [0]
Tennessee		C	C [0]
Texas		C	C [0]
Utah	1/1/2020	C	T [1]
Vermont	1/1/2014	T	T [7]
Virginia	1/1/2019	C	T [2]
Washington	1/1/2014	T	T [7]
West Virginia	1/1/2014	T	T [7]
Wisconsin	*	T	T [7]
Wyoming		C	C [0]

Table 4: Date of ACA medicaid expansion implementation by state and classification as treatment or control group for the 2014 and 2020 analysis. Letters in brackets indicate the 2020 treatment "dose" value, d. The 2014 classifications are used in the matching procedure.

Condition(s)	ICD-10 Codes
Influenza and pneumonia	J09-J18
All opioid poisoning	X40-X44, X60-64, X85, Y10-Y14
Healthcare amenable	
Tuberculosis	A16-19, B90
Other infections	A35-A37, A40-41, A80, B05
Malignant neoplasm of colon and rectum	C18-C21
Malignant neoplasm of skin	C44
Malignant neoplasm of breast	C50
Malignant neoplasm of cervix or uterus	C53-C55
Malignant neoplasm of testis	C62
Hodgkin's disease	C81
Leukemia	C91-C95
Disorders of thyroid gland	E00-E07
Diabetes Mellitus	E10-E14
Epilepsy	G40-G41
Chronic rheumatic heart diseases	I05-I09
Hypertensive diseases	I10-I13, I15
Ischemic heart diseases	I20-I25
Cerebrovascular diseases	I60-I69
All respiratory diseases	J00-J98
Gastric and duodenal ulcers	K25-K27
Diseases of appendix	K35-K38
Hernia	K40-K46
Diseases of gallbladder and biliary tract	K80-K83
Glomerular diseases	N00-N07
Renal failure	N17-N19
Pregnancy, childbirth and the puerperium	O00-O99
Misadventures to patients during surgical and medical care	Y60-Y69, Y83-Y84

Table 5: ICD-10 codes used in defining influenza and pneumonia, opioid, and healthcare amenable mortality.

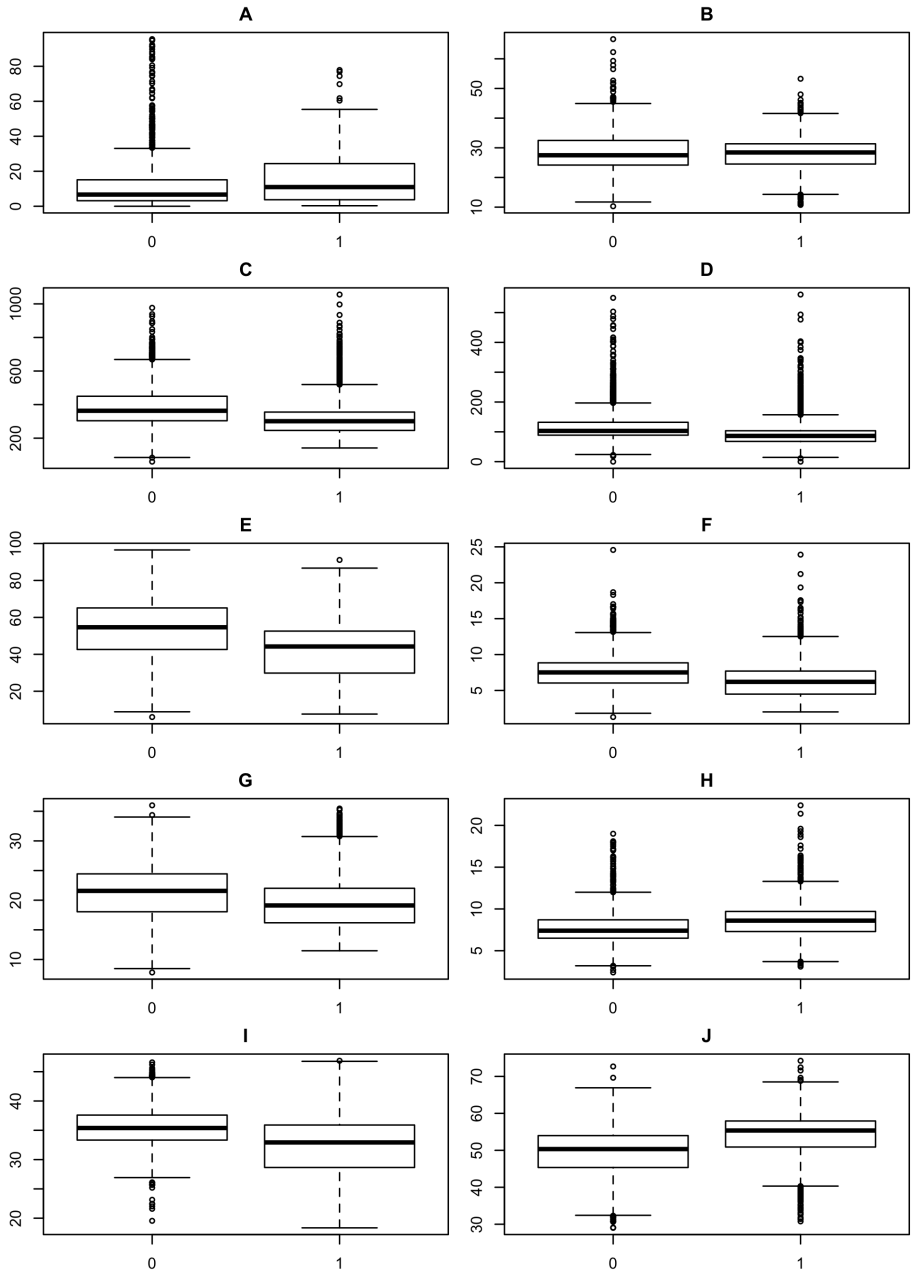

Figure 3: Distributions of variables with larger standardized mean differences, omitting unmatched counties: (A) % Hispanic, (B) % 20-34, (C) Working Age Adult Mortality, (D) 20-34 Mortality, (E) % Republican, (F)% Veteran, (G) Smoking, (H) Heavy Drinking (I) Male Obesity, (J) Female Physical Activity. On the horizontal axis, 1 indicates the treatment group (Medicaid expansion by June 2014). Unmatched counties are those separated from their comparison group by: more than 2 s.d.'s of flu and pnemonia, health-care amenable exclusive of flu or opioid-related mortality; or more than 1/4 s.d. of the propensity score.

	No Adjustment			After Matching			Dispersion	
	Control	Treat.	Dif.	Control	Treat.	Dif.	Overall	Adjacent
% White	78.6	81.2	2.6	80.6	81.3	0.8	14.1	8.3
% Black	17.2	10.9	-6.3	14.8	11.2	-3.6	12.3	6.4
% Hispanic	15.9	13.9	-2.0	15.6	13.4	-2.2	15.4	5.2
% Male	48.3	48.5	0.2	48.4	48.5	0.0	1.6	2.4
% 20-34	28.4	28.0	-0.4	28.0	28.0	0.0	5.3	5.1
% 35-44	18.2	17.9	-0.3	18.0	17.8	-0.2	2.2	1.7
% 45-54	19.2	19.6	0.4	19.2	19.6	0.4	1.7	1.5
% 55-64	16.1	16.5	0.4	16.3	16.5	0.2	1.9	1.9
All Mortality	381.3	322.0	-59.2	394.0	324.1	-69.9	102.0	90.0
20-34 Mortality	112.8	94.2	-18.6	118.1	94.8	-23.2	35.7	52.1
35-44 Mortality	198.0	164.6	-33.4	207.1	165.8	-41.2	62.4	80.7
45-54 Mortality	465.6	385.5	-80.1	477.8	387.8	-90.0	122.7	119.5
55-64 Mortality	948.6	804.6	-144.0	962.9	808.3	-154.5	203.9	181.5
White Male Mortality	408.3	350.9	-57.4	418.7	351.9	-66.8	95.3	86.1
White Female Mortality	237.7	202.0	-35.7	247.1	202.5	-44.6	59.8	57.8
Black Male Mortality	531.0	470.2	-60.9	513.6	472.5	-41.1	224.4	550.8
Black Female Mortality	327.6	298.5	-29.2	351.8	299.6	-52.1	229.4	740.4
Other Race Male Mortality	182.9	191.5	8.6	216.6	191.3	-25.3	186.8	562.7
Other Race Female Mortality	111.1	115.2	4.0	125.5	114.8	-10.8	114.5	294.2
Healcare Amenable (non-flu) Mortality	201.5	168.0	-33.5	205.7	169.1	-36.6	49.2	49.1
Opioid Mortality	20.9	21.3	0.4	22.6	21.5	-1.1	10.7	20.1
Flu Mortality	4.9	4.0	-0.9	5.6	4.0	-1.6	3.2	10.3
Population Density	785.5	2892.3	2106.8	676.9	3002.1	2325.2	6886.8	1085.2
% Urban	77.0	83.3	6.3	72.5	83.1	10.6	24.1	25.6
% Veteran	7.5	6.5	-1.0	7.7	6.6	-1.2	2.5	1.7
Median Income	48682.9	57608.9	8926.0	48247.0	57196.2	8949.2	13882.6	7251.3
% Poverty	17.6	15.1	-2.5	17.9	15.1	-2.8	5.4	4.3
% SNAP	16.5	14.0	-2.6	16.9	14.0	-2.9	6.4	5.0
% No Health Insurance	25.3	18.2	-7.2	24.3	18.2	-6.0	6.2	3.3
Unemployment Rate	7.1	7.7	0.6	7.3	7.6	0.3	2.0	1.5
PM2.5	0.1	0.1	0.0	0.1	0.1	0.0	1.7	0.2
Smoking	21.2	19.5	-1.8	21.9	19.6	-2.3	4.4	2.4
Heavy Drinking	7.5	8.5	1.0	7.6	8.5	0.9	1.8	1.3
Diabetes	15.5	13.8	-1.7	15.5	13.8	-1.7	2.0	1.3
Male Hypertension	37.7	36.0	-1.6	37.7	36.1	-1.6	3.2	1.9
Female Hypertension	40.3	38.5	-1.8	40.4	38.6	-1.8	3.4	2.0
Male Obesity	35.5	33.0	-2.6	36.0	33.0	-3.0	4.3	2.2
Female Obesity	38.1	35.1	-3.0	38.5	35.1	-3.3	5.2	3.3
Male Physical Activity	53.6	57.5	3.9	53.4	57.4	4.0	5.2	3.7
Female Physical Activity	49.1	54.4	5.3	49.2	54.1	4.9	6.1	3.4
% Republican	55.1	44.1	-11.0	55.5	44.4	-11.2	15.1	9.1
% Multigenerational Households	0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.1
Population	5.2	5.4	0.2	5.1	5.4	0.3	0.6	0.4

Table 6: Differences between control and treatment (ACA Medicaid expansion by June2020) counties before and after matching along with the pooled, weighted standard deviation of the covariates between treatment and control counties ("Overall") and the root mean squared of the distances in covariates between adjacent counties, divided by $\sqrt{2}$ for comparability ("Adjacent").

Appendix B. State Mortality Reporting Lag

In this appendix we will describe and give examples for our proposed method to calculate the average lag seen in all-cause mortality reporting by state using the FluView data. At the time that the protocol is being written, we have data through the week ending September 12, 2020. For simplicity in the discussion here, we will refer to weeks as they are numbered in the data, with week 1 indicating the first week in January 2020 and week 37 indicating the week ending September 12. We will refer to the week when the dataset was pulled as the "reporting week" and the week in the data that mortality is assigned to as simply "week" in the figures and tables.

The weekly mortality data from each state is updated weekly. Therefore, the mortality counts attributed to a specific week of the year, can change week to week as the dataset is updated. We are interested in understanding how quickly the mortality counts attributed to a certain week in a state seem to level-off to a constant number, as the data from each state is updated weekly. The CDC suggested using the FluView calculated values of the proportion of deaths that have been reported, out of the number of deaths that are expected for a certain state and week, according to modeling from historical data. We decide, instead to do a direct calculation ourselves using the reported all-cause mortality counts because this calculated percent from the FluView system is capped at 100% of expected, which is reached even as the mortality counts increase, since mortality is at a historically high point due to the pandemic.

Figures 4 and 5 visualize our conceptualization of the data lag for a subset of example states. These figures show the proportion of the maximum mortality count that we observe in the data for a specific week, over all reporting weeks we have available. Since the maximum mortality count reported for a specific week tends to be for the latest reporting week (week 37), we only show these values for weeks 31-36.[21]

Figure 4 illustrates what the reporting lag looks like for most states, which have a more or less consistent pattern in the lag, with mortality counts filled in at similar rates for previous weeks as the reporting week increases. We pulled these states specifically to also show how the lag differs between states, with Maine reporting a large portion of the deaths attributed to a week during the same reporting week, while in Ohio, mortality counts attributed to a week are not filled in until later reporting weeks. It is also important to note, as in the case with Indiana, that the pattern is not necessarily totally consistent over time.

Figure 5 illustrates what the reporting lag looks like for a handful of states for which the rate at which deaths counts are filled in over time is highly irregular. However, for these states, the mortality counts for earlier weeks in the year do seem to level out. Therefore, our calculation needs to consider these unstable patterns. Additionally, for this reason, we do not present a final calculation of lag times for each state for this protocol, but rather will wait until we can gather more weeks of data and include this calculation in our final manuscript.

We chose to calculate the reporting lag by evaluating at which week the mortality counts appear to have stabilized (in other words, that the mortality count doesn't change much between reporting weeks), for each reporting week. To do this, we calculate the

21. Reporting week 32 is missing because we did not get a data pull for that reporting week)

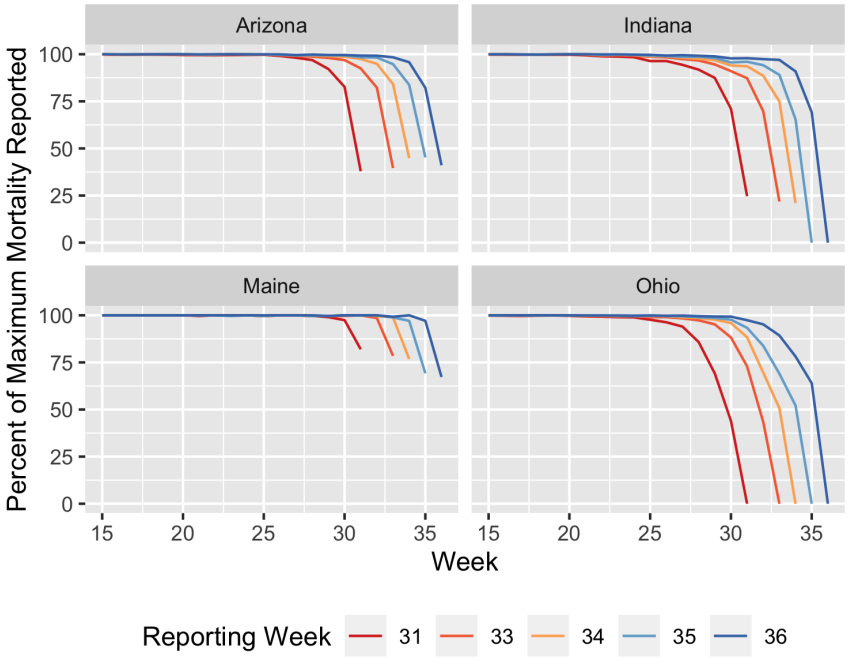

Figure 4: Percent of maximum mortality reported for a given week in the data, by the reporting week (i.e. week of the data pull). The color of the line represents the reporting week, while the x-axis is the week of the year for which the mortality count is attributed. For example, a value of 80% indicates that only a subset of the deaths for that weeks that are reported at some point by the CDC are reported in that reporting week.

percent change in mortality count for a given week between the previous reporting week and the current reporting week. We consider the counts to have stabilized in a given week if the change is less than 1%. We then determine the maximum week ($week_{max}$) that is considered stabilized and the minimum week ($week_{min}$) that is considered not stabilized. In a perfect world, the pattern of filling in mortality counts would be constant (and thus $week_{max} < week_{min}$), so we could just use the maximum stabilized week to determine the lag. However, since this is not always the case, if $week_{max} < week_{min}$, then $week_{min} - 1$ will be considered the first stabilized week. Table 7 shows this calculation for a state that has a consistent reporting lag (Arizona), while Table 8 shows this calculation for a state that has an inconsistent reporting lag (Connecticut).

As shown in Table 7 and Table 8 as well, we plan to calculate the lag in weeks by subtracting the first stabilized week from the reporting week. Finally, since we understand that the lag is not consistent across all reporting weeks, we will take the average lag across all reporting weeks.

(a)

Week	Deaths	% Change	% Maximum
25	1525	0.00	99.9
26	1583	0.00	99.9
27	1753	0.00	99.6
28	1912	0.00	99.8
29	1931	0.00	99.6
30	1800	0.00	99.6
31	1730	0.01	99.3
32	**1525**	**0.01**	**99.2**
33	1456	0.04	98.4
34	1357	0.14	95.8
35	1118	0.81	82.1
36	413		41.1

(b)

Reporting Week	Week Stable	Weeks Lag	Mean Lag
33	26	7	
34	29	5	
35	31	4	
36	32	4	
37	32	5	5.00

Table 7: Example calculation of reporting lag for Arizona. Table (a) illustrates how to determine the week when the mortality counts are considered to have stabilized, for reporting week 36. % Change indicates the percent change in mortality count between reporting week 35 and reporting week 36 for the given week. % Maximum represents the percent of the maximum mortality observed over all reporting weeks. Here $week_{max} = 32 < week_{min} = 33$ so the week stable is 32. Table (b) illustrates how the weeks lag can differ for different reporting weeks and how the mean lag would be calculated.

(a)

Week	Deaths	% Change	% Maximum
21	806	0.00	99.6
22	748	0.00	99.7
23	642	0.03	98.6
24	**569**	**0.01**	**93.4**
25	529	0.01	93.6
26	**575**	**0.01**	**97.1**
27	521	0.06	99.0
28	570	0.11	97.1
29	532	0.18	7.4
30	526	0.28	96.5
31	440	0.48	93.8
32	216	Inf	87.8
33	144	Inf	85.2
34	0		
35	0		

(b)

Reporting Week	Week Stable	Weeks Lag	Mean Lag
33	20	13	
34	25	9	
35	24	11	
36	22	14	
37	22	15	12.40

Table 8: Example calculation of reporting lag for Connecticut. Table (a) illustrates how to determine the week when the mortality counts are considered to have stabilized, for reporting week 35. % Change indicates the percent change in mortality count between reporting week 34 and reporting week 35 for the given week. % Maximum represents the percent of the maximum mortality observed over all reporting weeks. In this example, $week_{max} = 26 > week_{min} = 25$ so the week stable is 25-1 = 24. Table (b) illustrates how the mean lag would be calculated.

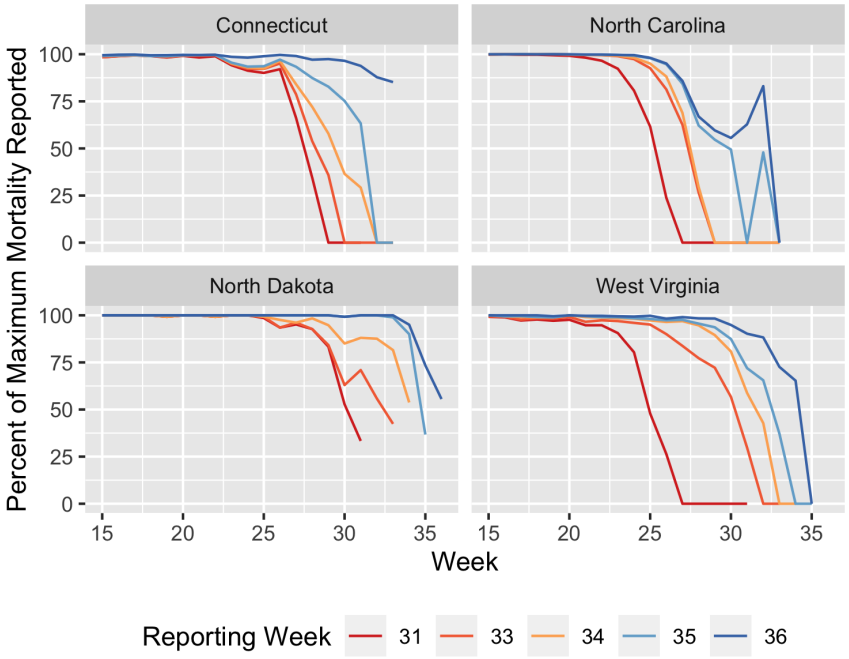

Figure 5: Percent of maximum mortality reported for a given week in the data, by the reporting week (i.e. week of the data pull).

References

ACS (2013). American community survey 5-year estimates survey tables (s1002) (2008-2013). https://data.census.gov/cedsci/table.

AHRF (2019). Area health resources files. https://data.hrsa.gov/data/download. Date Accessed: 7/7/2020.

Black, B., Hollingsworth, A., Nunes, L., & Simon, K. (2019). The effect of health insurance on mortality: Power analysis and what we can learn from the affordable care act coverage expansions. Working Paper 25568, National Bureau of Economic Research. Available from: http://www.nber.org/papers/w25568.

Borgschulte, M., & Vogler, J. (2020). Did the aca medicaid expansion save lives? *Journal of Health Economics*, *72*, 102333. Available from: http://www.sciencedirect.com/science/article/pii/S0167629619306228.

CDC (2020c). Restricted-use vital statistics data. Available from: https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm.

CDC (2021a). Provisional COVID-19 Death Counts by County and Race | Data | Centers for Disease Control and Prevention. Available from: https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-County-and-Ra/k8wy-p9cg.

CDC (2021b). Provisional COVID-19 Death Counts in the United States by County | Data | Centers for Disease Control and Prevention.
Available from: https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-in-the-United-St/kn79-hsxy.

Cowger, T. L., Davis, B. A., Etkins, O. S., Makofane, K., Lawrence, J. A., Bassett, M. T., & Krieger, N. (2020). Comparison of Weighted and Unweighted Population Data to Assess Inequities in Coronavirus Disease 2019 Deaths by Race/Ethnicity Reported by the US Centers for Disease Control and Prevention. *JAMA Network Open*, *3*(7), e2016933–e2016933.
Available from: https://doi.org/10.1001/jamanetworkopen.2020.16933.

CQ Press (2020). Voting and elections collection. https://library.cqpress.com/elections/. Date Accessed: 6/11/2020.

Gee, G. C., & Ford, C. L. (2011). Strustural racism and health inequalities: Old issues, new directions. *Du Bois Review: Social Science Research on Race*, *8*(1), 115–132.

Guth, M., Garfield, R., & Rudowitz, R. (2020). The effects of medicaid expansion under the aca: Updated findings from a literature review.
Available from: https://www.kff.org/report-section/the-effects-of-medicaid-expansion-under-the-aca-updated-findings-from-a-literature-review-report/.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association*, *99*(467), 609–618.
Available from: https://doi.org/10.1198/016214504000000647.

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, *23*(2), 219–236.

Hansen, B. B., & Klopfer, S. O. (2006a). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*(3), 609–627.

Hansen, B. B., & Klopfer, S. O. (2006b). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*(3), 609–627.
Available from: http://www.stat.lsa.umich.edu/\%7Ebbh/hansenKlopfer2006.pdf.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *50*(3), 346–363.

IHME (2013c). United states hypertension estimates by county 2001-2009. http://ghdx.healthdata.org/record/ihme-data/united-states-hypertension-estimates-county-2001-2009.

IHME (2013d). United states physical activity and obesity prevalence by county 2001-2011. http://ghdx.healthdata.org/record/ihme-data/united-states-physical-activity-and-obesity-prevalence-county-2001-2011.

IHME (2014e). United states smoking prevalence by county 1996-2012. http://ghdx.healthdata.org/record/ihme-data/united-states-smoking-prevalence-county-1996-2012.

IHME (2015b). United states alcohol use prevalence by county 2002-2012. http://ghdx.healthdata.org/record/ihme-data/united-states-alcohol-use-prevalence-county-2002-2012.

IHME (2016a). Diagnosed and undiagnosed diabetes prevalence by county in the u.s., 1999-2012. http://ghdx.healthdata.org/record/ihme-data/united-states-diabetes-prevalence-county-1999-2012.

KFF (2020). Status of State Action on the Medicaid Expansion Decision. Date Accessed: 8/5/2020.
Available from: https://www.kff.org/health\-reform/state\-indicator/state\-activity\-around\-expanding\-medicaid\-under\-the\-affordable\-care\-act/.

Khatana, S. A. M., Bhatla, A., Nathan, A. S., Giri, J., Shen, C., Kazi, D. S., Yeh, R. W., & Groeneveld, P. W. (2019). Association of Medicaid Expansion With Cardiovascular Mortality. *JAMA Cardiology*, *4*(7), 671–679.
Available from: https://doi.org/10.1001/jamacardio.2019.1651.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics*, (pp. 65–78).

Metzl, J. (2019). *Dying of Whiteness*. Basic Books.

Miller, S., Altekruse, S., Johnson, N., & Wherry, L. R. (2019). Medicaid and mortality: New evidence from linked survey and administrative data. Working Paper 26081, National Bureau of Economic Research.
Available from: http://www.nber.org/papers/w26081.

NBER (2010). County adjacency (2010). https://data.nber.org/data/county-adjacency.html.

NCHS (2020). Pneumonia and influenza mortality surveillance from the national center for health statistics mortality surveillance system. https://gis.cdc.gov/grasp/fluview/mortality.html.

NSC (2020). Top 10 preventable injuries. https://injuryfacts.nsc.org/all-injuries/deaths-by-demographics/top-10-preventable-injuries/data-details/.

Rosenbaum, P. R. (1991). A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*(3), 597–610.
Available from: http://www.jstor.org/stable/2345589.

Rosenbaum, P. R. (1997). Signed rank statistics for coherent predictions. *Biometrics*, *53*(2), 556–566.
Available from: http://www.jstor.org/stable/2533957.

Rosenbaum, P. R. (2002a). Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, *17*(3), 286–327.

Rosenbaum, P. R. (2002b). *Observational Studies*. Springer-Verlag, second ed.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer Verlag.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33–38.

Rosenbaum, P. R., & Silber, J. H. (2001). Matching and thick description in an observational study of mortality after surgery. *2*(2), 217–232.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169–188.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, *26*(1), 20–36.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, *2*(3), 808–40.

Sommers, B. D., Long, S. K., & Baicker, K. (2014). Changes in mortality after massachusetts health care reform. *Annals of Internal Medicine*, *160*(9), 585–593. PMID: 24798521.
Available from: `https://www.acpjournals.org/doi/abs/10.7326/M13-2275`.

Stuart, E. A., & Green, K. M. (2006). Using full matching to estimate causal effects in non-experimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. Tech. rep., Johns Hopkins University.

Swaminathan, S., Sommers, B. D., Thorsness, R., Mehrotra, R., Lee, Y., & Trivedi, A. N. (2018). Association of Medicaid Expansion With 1-Year Mortality Among Patients With End-Stage Renal Disease. *JAMA*, *320*(21), 2242–2250.
Available from: `https://doi.org/10.1001/jama.2018.16504`.

Woolf, S. H., Chapman, D. A., Sabo, R. T., Weinberger, D. M., & Hill, L. (2020). Excess Deaths From COVID-19 and Other Causes, March-April 2020. *JAMA*.
Available from: `https://doi.org/10.1001/jama.2020.11787`.