Medicaid Expansion and Racial/Ethnic and Sex Disparities in Cardiovascular Diseases over 6 Years: A Generalized Synthetic Control Approach

Roch A Nianogo, Fan Zhao, Stephen Li, Akihiro Nishi, Sanjay Basu

Table of contents

Appendix Sections

Appendix Tables

Appendix Figures

eSection 1. Missingness in the CVD outcome data and dealing with missing in the analysis

The CVD outcome data were fully observed for the men, women, White subpopulations as well as for the overall population. However, for the Black subpopulation, the CVD outcome data were partially missing (i.e. between 4 and 19 yearly CVD outcome data were missing, see **eFigure 1**) for four states (Alaska, New Mexico, Rhode Island, Utah) and completely missing (i.e. all 20 yearly outcome data were missing) for nine states (Hawaii, Idaho, Maine, Montana, New Hampshire, North Dakota, South Dakota, Vermont, Wyoming).

For the Hispanic subpopulation, the outcome data were minimally missing (i.e., with fewer than three yearly outcome data were missing, see **eFigure 2**) for six (6) states (Kansas, Maryland, Missouri, North Carolina, Oregon, Utah), partially missing (i.e. between 4 and 19 yearly CVD outcome data were missing, see **eFigure 2**) for twelve (12) states (Alabama, Arkansas, Idaho, Iowa, Kentucky, Louisiana, Minnesota, Nebraska, Rhode Island, South Carolina, Tennessee, Wisconsin) and completely missing (i.e. all 20 yearly outcome data were missing) for eleven (11) states (Alaska, Delaware, Maine, Mississippi, Montana, New Hampshire, North Dakota, South Dakota, Vermont, West Virginia, Wyoming).

For both the Black and the Hispanic subpopulation, we excluded the states with partially or completely missing data (see **eFigure 1** and **eFigure 2**). For the Hispanic subpopulation, we further imputed CVD outcome data for the six states with minimal missing (Kansas, Maryland, Missouri, North Carolina, Oregon, Utah). We imputed the outcome data with the closest outcome data (previous/next year) given that excluding these six states will have prevented the generalized synthetic control method implemented via the *gsynth* package from working.

The missingness was in fact imposed at the CDC WONDER database level. [35] For confidentially reasons, statistics representing fewer than ten persons were suppressed and when the death count was less than 20, the corresponding rates were marked as “unreliable”. [35] The unreliable rates were then converted to missing in our analysis.

To ensure the robustness of our findings as it relates to the potential influence of the missingness, we conducted a sensitivity analysis by running an analysis of the White subgroup restricted to the non-missing states used in the Black subgroup as well as in the Hispanic subgroup and using these estimates for estimating the triple difference. This yielded overall similar conclusions (similar direction but with less precision, **eTable 3**) suggesting missingness had a minor influence of our findings

eSection 2. Steps in implementing the generalized synthetic control

There are several steps used for estimating effects using the GSCM, described in detail in Xu and Garber et al[34, 36] and implemented via the gsynth package.[37]

In step 1, an interactive-fixed-effects (IFE) model is estimated using the control data only (i.e., non-expansion states) to obtain three sets of parameters: $\hat{\beta}$, the vector of parameters corresponding to observed covariates X_{it}, $\hat{F}$, the vector of latent factors (i.e., a vector of time-varying, state-fixed parameters), $\widehat{\Lambda}$, the vector of factor loadings (i.e., a vector of time-fixed state-varying parameters). The number of latent factors is determined using cross-validation. The vector of latent factors is interacted with the vector of factor loadings to form the IFE.

In step 2, the factor loadings ($\hat{\lambda}_i$), were estimated for each expansion state i by minimizing the mean squared error of the predicted outcome in the pre-treatment periods.

In step 3, the treated counterfactuals are predicted (i.e. the projected number of CVD deaths per 100,000 persons in expansion state i at time t had they not adopted the Medicaid expansion) based on the parameters obtained in step 1 and 2.

In step 4, the treatment effect of the expansion state i at time t is given by the difference between the observed outcome in the expansion state, $Y_{it}(1)$ and its estimated counterfactual, $\hat{Y}_{it}(0)$ (i.e., synthetic control) as $\hat{\delta}_{it} = Y_{it}(1) - \hat{Y}_{it}(0)$.

In step 5, uncertainty estimates (standard errors and confidence intervals) are obtained via bootstrapping.

eSection 3. Estimation of the triple difference-in-difference, its standard error and confidence interval

To obtain the difference in mean difference between the groups (DMD), as well as the standard error and the confidence interval of the DMD, we first estimated the effect of the Medicaid expansion on CVD mortality in each subgroup (e.g. Black subgroup, White subgroup) by running separate GSCM models because the missingness pattern and sample size for each subgroup was different. Once the effect point estimates and standard error for each subgroup were obtained from the GSCM model, we obtained the following:

- the difference in mean difference (DMD) between the groups as:

$$\mathrm{DMD}_{\text{Black vs White}} = \mathrm{MD}_{\text{Black}} - \mathrm{MD}_{\text{White}}$$

- the standard error of the difference in mean difference ($\mathrm{SE}_{\mathrm{DMD}}$) as:

$$\mathrm{SE}_{\mathrm{DMD}} = \sqrt{\mathrm{SE}^2_{\text{White}} + \mathrm{SE}^2_{\text{Black}}}$$

- the confidence interval of the difference in mean difference ($\mathrm{CL}_{\mathrm{DMD}}$) as:

$$\mathrm{CL}_{\mathrm{DMD}} = \mathrm{DMD}_{\text{Black vs White}} \pm 1.96 \cdot \mathrm{SE}_{\mathrm{DMD}}$$

We did likewise for other subgroups.

eTable 1. STROBE Statement checklist

	Item No	Recommendation
Title and abstract	1	Page 1
Introduction		
Background/rationale	2	Page 3-6
Objectives	3	Page 6
Methods		
Study design	4	Page 10
Setting	5	Page 10
Participants	6	Page 10
Variables	7	Page 8-9
Data sources/ measurement	8	Page 10
Bias	9	Page 12
Study size	10	Page 10
Quantitative variables	11	Page 8-9
Statistical methods	12	Page 10
Results		
Participants	13	Page 13
Descriptive data	14	Page 13
Outcome data	15	Page 13
Main results	16	Page 13-14
Other analyses	17	
Discussion		
Key results	18	Page 15
Limitations	19	Page 17
Interpretation	20	Page 15-16
Generalizability	21	Page 19
Other information		
Funding	22	Anonymized for review

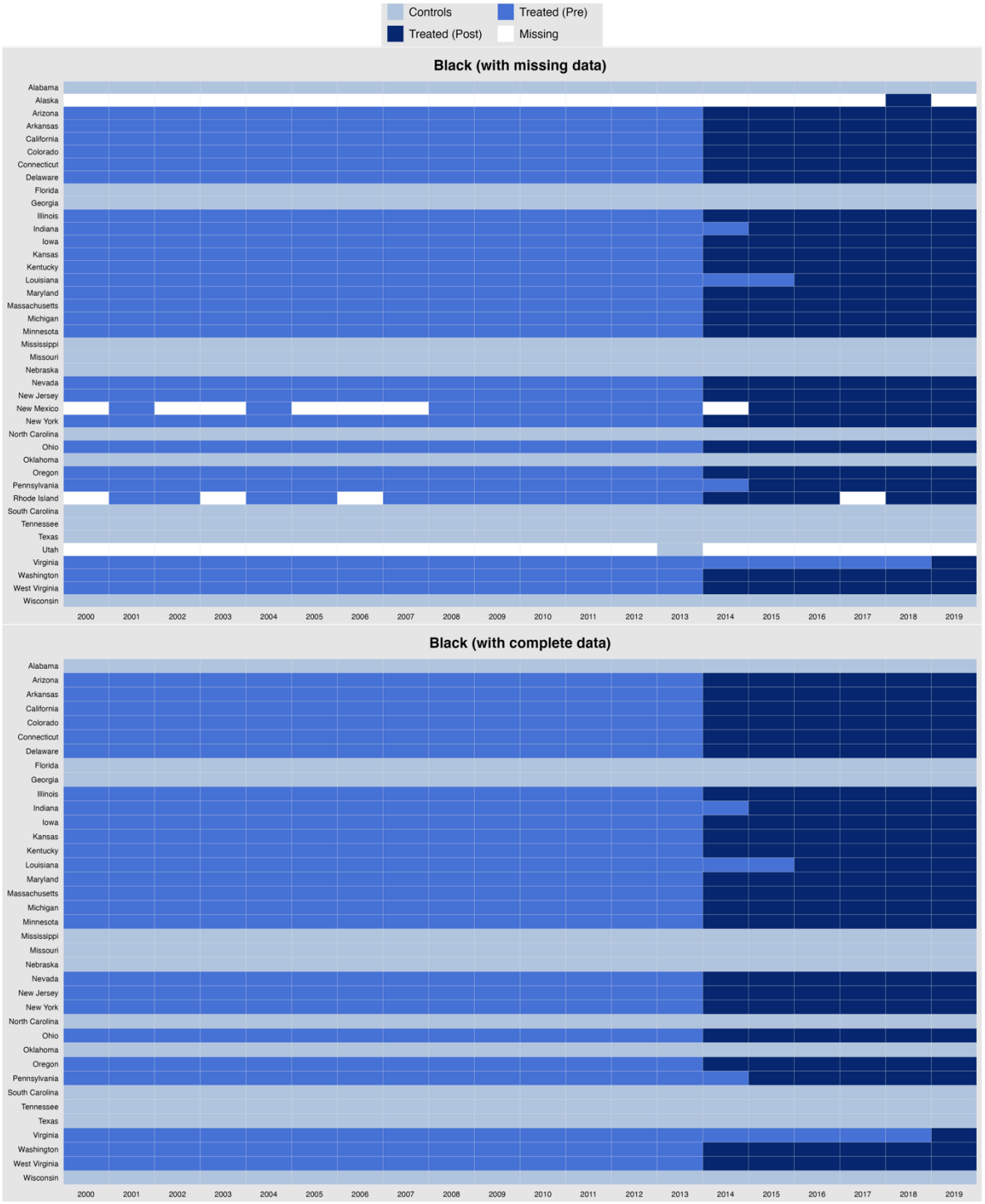
eTable 2. Overall annual effect of the Medicaid expansion on CVD deaths per 100,000 persons for the overall population and for the Black, Hispanic, White, Men and women subpopulations.

	Annual adjusted Mean Difference (95% CI)						
Time since Medicaid expansion adoption	0	1	2	3	4	5	6
Overall	0.40 (-3.38, 4.17)	-5.09 (-9.34, -0.85)	-7.61 (-13.43, -1.79)	-2.81 (-9.33, 3.71)	-4.18 (-10.76, 2.41)	-3.98 (-11.23, 3.28)	-1.55 (-9.88, 6.77)
Black	-1.48 (-16.15, 13.19)	-6.19 (-23.10, 10.72)	-14.67 (-32.74, 3.40)	-4.66 (-23.60, 14.29)	-5.44 (-26.24, 15.37)	-1.61 (-25.21, 22.00)	1.45 (-24.93, 27.82)
Hispanic	-2.56 (-18.05, 12.93)	-6.31 (-27.21, 14.60)	-7.93 (-31.09, 15.23)	-4.26 (-31.20, 22.68)	0.32 (-28.57, 29.21)	-6.78 (-37.26, 23.70)	-0.18 (-34.13, 33.78)
White	0.02 (-3.63, 3.68)	-5.42 (-10.99, 0.15)	-4.48 (-10.09, 1.13)	-2.16 (-8.10, 3.77)	-1.88 (-7.63, 3.87)	-3.06 (-9.81, 3.69)	-1.67 (-9.49, 6.15)
Men	1.50 (-5.36, 8.36)	-4.72 (-12.59, 3.15)	-10.35 (-20.28, -0.42)	-2.40 (-12.92, 8.13)	-8.16 (-19.81, 3.49)	-5.74 (-18.13, 6.65)	-4.16 (-17.59, 9.28)
Women	-1.58 (-6.37, 3.20)	-5.50 (-9.92, -1.09)	-4.36 (-9.87, 1.15)	-2.91 (-8.15, 2.34)	-2.06 (-7.69, 3.58)	-4.57 (-10.57, 1.44)	-0.08 (-5.49, 5.34)

eTable 3. Sensitivity analysis evaluating the impact of using heterogenous samples vs homogenous samples on both adjusted mean differences and difference in mean differences

	Number of states included	Adjusted Mean Difference, (95%CI)	Difference in Mean Difference, (95%CI)	
White	50	-3.18 (-8.30, 1.94)[a]	Reference 1	
	37 (same as Black sample)	-2.15 (-9.48, 5.17)	Reference 2	
	27 (same as Hispanic sample)	-11.44 (-18.60, -4.29)	Reference 3	
Black	37	-5.36 (-22.63, 11.91)[a]	-2.18 (-20.19, 15.83)[a,b]	-3.20 (-21.96,15.56)[c]
Hispanic	27	-4.28 (-30.08, 21.52)[a]	-1.10 (-27.40, 25.20)[a,b]	7.17 (-19.60, 33.94)[d]

[a]*Results included in the main manuscript.*
[b]*using Reference 1*
[c]*using Reference 2*
[d]*using Reference 3*
Note that using reference 1 implies heterogenous samples (i.e., different samples for Whites [n=50], Blacks[n=37] and Hispanics [27]) and using reference 2 and 3 implies homogenous samples (i.e., similar samples for Whites [n=37] vs Blacks[n=37] and Whites [n=27] vs Hispanics [27])



eFigure 1. Analytical data structure of the Medicaid expansion states and control states highlighting the missingness for the Black subpopulations for the outcome data.

eFigure 2. Analytical data structure of the Medicaid expansion states and control states highlighting the missingness for the Hispanic subpopulations for the outcome data.

eFigure 3. Annual difference in mean difference between the effect of the Medicaid expansion on CVD deaths per 100,000 persons