Appendix to "Understanding Spatial Variation in COVID-19 across the United States"

Klaus Desmet
SMU and NBER

Romain Wacziarg
UCLA and NBER

February 2021

Abstract

This Appendix contains: A) Description and sources of the data used in the analysis. B) Additional tables and figures mentioned in the main text.

A. Data Sources

A1. Dependent Variables

COVID-19 cases and deaths. Daily county-level data on COVID-19 cases and deaths. Source: *New York Times*, `https://github.com/nytimes/covid-19-data`. We adjusted the data in the following ways:

1. The source reports data cumulated for New York City overall (all 5 boroughs/counties together). We apportioned cases and deaths to each of the 5 boroughs/counties by county population shares.

2. The source reports data for all of Kansas City, which is made up of parts of several counties, each independent entries with their own cases and deaths (exclusive of Kansas City). Most of Kansas City is in Jackson County MO, so we added all Kansas City cases and deaths to that county's tally.

3. We did not make any modifications regarding any of the additional geographic specificities as described in the source data: "Counts for Alameda County (CA) include cases and deaths from Berkeley and the Grand Princess cruise ship; counts for Douglas County (NE) include cases brought to the state from the Diamond Princess cruise ship; all cases and deaths for Chicago are reported as part of Cook County (IL); counts for Guam include cases reported from the USS Theodore Roosevelt."

4. The source reports non-monotonic evolutions of cumulative cases and deaths for a very small set of counties, at the very beginning of the pandemic, when there were very few cases and deaths. The reason is unknown. We recoded cases and deaths that subsequently became lower to the level of the later lower number to ensure monotonic cumulative series for all counties.

A2. Independent Variables

Population and age. Age structure of the population by county. Source: U.S. Census Bureau. *2018 American Community Survey 5-Year Estimates*. `https://data.census.gov/cedsci/`.

Population density. Population divided by land area in square miles. Source: U.S. Census Bureau.

Local effective density. Expected density in a one square kilometer around a randomly drawn individual from each county. If all county inhabitants are uniformly distributed across space, this measure is identical to standard population density. If the population is concentrated in a small subset of the county territory, this measure will be larger than standard population density. Own calculations based on 2020 population data from GPW. Source: Center for International Earth Science Information Network, *Gridded Population of the World*, Version 4: Population Count, Revision 11,

Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) (2018). https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11.

Metro county. Classification as large central metro county, large fringe metro country, medium metro county or small metro county. Source: National Center for Health Statistics (NCHS). *Urban-Rural Classification Scheme for Counties 2013.* https://www.cdc.gov/nchs/data_access/urban_rural.htm#Data_Files_and_Documentation

Public transportation. Share of population that goes to work by public transportation. Source: U.S. Census Bureau. *2018 American Community Survey 5-Year Estimates.* https://data.census.gov/cedsci/.

Nursing home residents. Percentage of population who are residents in nursing homes. Source: Centers for Medicare & Medicaid Services. *Nursing Home Compare Datasets: Provider Info.* https://data.medicare.gov/data/nursing-home-compare.

Distance to airport. Data on all international flights to the U.S. in 2019 come from Table T-100 from the Bureau of Transportation Statistics. For each U.S. airport, we take the average number of monthly passengers on direct flights from the top-5 countries in terms of COVID-19 cases on March 15, 2020 (China, Italy, Iran, South Korea and Spain). For each county in the U.S., we then compute the geodesic distance to the closest airport that received at least 250 passengers per month on direct flights from one of these 5 countries. https://www.transtats.bts.gov/

Household income. Log of median household income, 2009-2013. Source: U.S. Census Bureau.

Social capital. Social capital index created using principal component analysis using number of associations and organizations (including non-profits), voter turnout and census response rate in 2014 (variable sk14). Source: Rupasingha, A., S. J. Goetz and D. Freshwater (2006, with updates). https://aese.psu.edu/nercrd/community/social-capital-resources

Race. Black or African American alone, Hispanic or Latino, American Indian and Alaska Native alone, percentage 2014. Source: U.S. Census Bureau.

Education. High school graduate or higher, percentage of persons age 25+, 2009-2013, and bachelor's degree or higher, percentage of persons age 25+, 2009-2013. Source: U.S. Census Bureau.

Housing arrangements. Percent of housing units in multi-unit structures, 2009-2013, and persons per household, 2009-2013. Source: U.S. Census Bureau.

Poverty rate. Fraction of population below the poverty line 2000 Census SF3 Sample Data Table P087. https://opportunityinsights.org/data/.

Gini index within bottom 99%. Gini coefficient minus top 1% income share from Tax Records, Core Sample (Chetty et al., 2014). `https://opportunityinsights.org/data/`.

Smokers and obese. Percentage of the population that smokes and percentage of population that is obese. Source: Bergeron et al. (2016). `https://opportunityinsights.org/data/`.

Risk-adjusted mortality. 30-day risk adjusted mortality for heart attacks, heart failure and pneumonia. Source: Bergeron et al. (2016). `https://opportunityinsights.org/data/`.

Trump vote share in the 2016 general election. Source: Dave Leip's Atlas of U.S. Presidential Elections. `https://uselectionatlas.org/`.

References

[1] Bergeron, A., R. Chetty, D. Cutler, B. Scuderi, M. Stepner, N. Turner (2016), "The Association Between Income and Life Expectancy in the United States, 2001-2014," *Journal of the American Medical Association*, 315, 1750-1766.

[2] Chetty, R., N. Hendren, P. Kline, E. Saez (2014), "Where is the Land of Opportunity: The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 129, 1553-1623.

[3] Rupasingha, A., S. J. Goetz and D. Freshwater (2006, with updates), "The Production of Social Capital in US Counties," *Journal of Socio-Economics*, 35, 83–101.

Table A1 – Summary Statistics

Panel A - Summary Statistics for Various Indicators of Disease Severity (November 30, 2020)

Variable	Obs	Mean	Std. Dev.	Min	Max
Total cases	3,142	4289.398	14474.450	0	401,034
Cases per 100,000	3,142	4639.139	2388.414	0	22,007
Indicator for any case	3,142	0.999	0.031	0	1
Log Cases (IHS)	3,142	7.750	1.562	0	13.595
Log Cases	3,139	7.064	1.545	0	12.902
Total Deaths	3,142	84.785	358.520	0	7,655
Deaths per 100,000	3,142	79.354	71.049	0	764
Indicator for any death	3,142	0.947	0.224	0	1
Log Deaths (IHS)	3,142	3.493	1.726	0	9.636
Log Deaths	2,975	2.981	1.581	0	8.943

Panel B - Summary Statistics for the Baseline Set of 8 Regressors

Variable	Obs	Mean	Std. Dev.	Min	Max
Log Population	3,142	10.275	1.494	4.317	16.129
Log Effective Local Density	3,142	5.488	0.993	0.016	10.028
% people who commute by public transportation	3,141	0.902	3.066	0.000	60.700
Share of people aged 75 or older	3,142	0.079	0.023	0.013	0.241
% nursing home residents in pop.	3,142	0.603	0.448	0.000	5.047
Log km to closest airport w/ flights from top 5 COVID countries	3,142	5.562	1.144	-4.605	8.264
Log household median Income	3,140	10.705	0.242	9.903	11.714
Social Capital Index, 2014	3,139	0.001	1.260	-3.183	21.809

Table A2 - Baseline Specifications of Table 1, with State Fixed Effects
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log Cases, IHS, Nov. 30	**Log Cases, 225 days post-onset**	**Log Deaths, IHS, Nov. 30**	**Log Deaths, 215 days post-onset**
Log population	0.986 (0.009)*** [0.945]	0.921 (0.012)*** [0.887]	0.985 (0.018)*** [0.851]	0.977 (0.025)*** [0.875]
Log effective local density	0.154 (0.012)*** [0.098]	0.154 (0.015)*** [0.105]	0.085 (0.024)*** [0.049]	0.092 (0.034)*** [0.060]
% people who commute by public transportation	0.004 (0.003) [0.008]	0.014 (0.003)*** [0.033]	0.032 (0.005)*** [0.057]	0.029 (0.005)*** [0.086]
Share of people aged 75 & above	-4.502 (0.409)*** [-0.067]	-5.572 (0.537)*** [-0.084]	0.709 (0.831) [0.010]	1.387 (1.126) [0.019]
% nursing home residents in pop.	0.152 (0.018)*** [0.044]	0.160 (0.026)*** [0.044]	0.301 (0.036)*** [0.078]	0.605 (0.072)*** [0.123]
Log km to closest airport w/ flights from top 5 COVID countries	0.006 (0.008) [0.004]	-0.008 (0.009) [-0.006]	-0.011 (0.016) [-0.007]	-0.021 (0.015) [-0.020]
Log household median income	-0.433 (0.037)*** [-0.067]	-0.544 (0.047)*** [-0.095]	-0.590 (0.076)*** [-0.083]	-0.677 (0.095)*** [-0.120]
Social Capital Index, 2014	0.007 (0.008) [0.006]	-0.019 (0.010)* [-0.015]	-0.034 (0.016)** [-0.025]	-0.043 (0.022)** [-0.027]
Constant	1.898 (0.398)***	2.969 (0.493)***	-0.707 (0.810)	-0.659 (1.007)
R^2	0.95	0.92	0.82	0.83
N	3,138	2,756	3,138	1,445

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors in parentheses and standardized betas in brackets.

- Onset day is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).
- All specifications include a set of state fixed effects.

Table A3 - An Investigation of Inequality and Poverty Effects
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log Cases, IHS, November 30	**Log Cases, 225 days since onset**	**Log Deaths, IHS, November 30**	**Log Deaths, 215 days since onset**
Gini Index Within Bottom 99%	-0.506 (0.142)*** [-0.031]	0.326 (0.178)* [0.020]	0.793 (0.238)*** [0.041]	1.273 (0.297)*** [0.076]
Poverty Rate	-0.069 (0.271) [-0.003]	0.921 (0.347)*** [0.043]	2.738 (0.456)*** [0.106]	4.515 (0.598)*** [0.198]
Log household median income	-0.549 (0.072)*** [-0.094]	-0.470 (0.091)*** [-0.083]	-0.178 (0.120) [-0.026]	0.249 (0.149)* [0.044]
R^2	0.87	0.81	0.73	0.75
N	3,027	2,729	3,027	1,440

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors in parentheses and standardized betas in brackets.

- Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).
- All specifications contain an intercept and controls for log population, log effective local density, the percentage of people who commute by public transportation, the share of people aged 75 and above, the percentage of nursing home residents in the population, log kilometers to the closest airport with flights from top 5 COVID countries and the social capital index for 2014.
- There is collinearity between poverty rate and median income (ρ = - 0.75), so care should be exercised in interpreting these coefficients.

Table A4 - An Investigation of Health Effects
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log Cases, IHS, November 30	**Log Cases, 225 days since onset**	**Log Deaths, IHS, November 30**	**Log Deaths, 215 days since onset**
Percentage of the population that smokes	-0.131 (0.186) [-0.006]	-0.375 (0.225)* [-0.017]	-0.296 (0.302) [-0.012]	-0.389 (0.427) [-0.013]
Percentage of the population that is obese	0.502 (0.157)*** [0.029]	0.723 (0.193)*** [0.041]	1.346 (0.255)*** [0.066]	0.568 (0.363) [0.027]
30-day Mortality for Heart Attacks	-0.111 (0.408) [-0.002]	0.296 (0.518) [0.006]	-0.183 (0.662) [-0.003]	1.932 (0.966)** [0.032]
30-day Mortality for Heart Failure	-0.094 (0.673) [-0.001]	-0.815 (0.835) [-0.011]	-3.712 (1.093)*** [-0.043]	-8.655 (1.498)*** [-0.094]
30-day Mortality for Pneumonia	-1.931 (0.601)*** [-0.030]	-0.192 (0.739) [-0.003]	1.996 (0.975)** [0.026]	5.734 (1.294)*** [0.072]
R^2	0.84	0.78	0.71	0.74
N	2,334	2,250	2,334	1,333

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors in parentheses and standardized betas in brackets.

- Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).
- All specifications contain an intercept and controls for the baseline set of 8 variables in Table 1.
- Note the smaller number of observations due to lack of availability of data on obesity and smoking.
- 30-day mortality measures are risk-adjusted so are likely to capture mostly the quality of the health infrastructure / health care system in the county.

Table A5 - An Investigation of Donald Trump Effects (Sample Defined by Calendar Date)

	(1)	(2)	(3)	(4)
	Short Spec., June 29	**Short Spec., November 30**	**Comprehensive Spec., June 29**	**Comprehensive Spec., November 30**
Panel A – Dependent Variable: Log Cases, IHS				
Trump vote share, 2016 general election	-1.527 (0.146)*** [-0.107]	0.401 (0.076)*** [0.041]	0.287 (0.187) [0.021]	1.389 (0.098)*** [0.152]
R^2	0.76	0.86	0.82	0.89
N	3,111	3,111	3,008	3,008
Panel B – Dependent Variable: Log Deaths, IHS				
Trump vote share, 2016 general election	-1.759 (0.162)*** [-0.145]	0.136 (0.122) [0.013]	0.807 (0.222)*** [0.066]	1.971 (0.160)*** [0.185]
R^2	0.59	0.71	0.69	0.79
N	3,111	3,111	3,008	3,008

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors in parentheses and standardized betas in brackets.

- All columns contain an intercept.
- Columns 1 and 2 (the short specification) include only a control for log population.
- Columns 3 and 4 add controls for all 8 regressors in the baseline specification of Table 1, as well as % high school graduate or higher (among persons age 25+), % with bachelor's degree or higher (among persons age 25+), 30-day mortality for heart attacks, 30-day mortality for heart failure, 30-day mortality for pneumonia, Gini index within bottom 99%, poverty rate, top 1% income share, % housing units in multi-unit structures, persons per household and log population density.

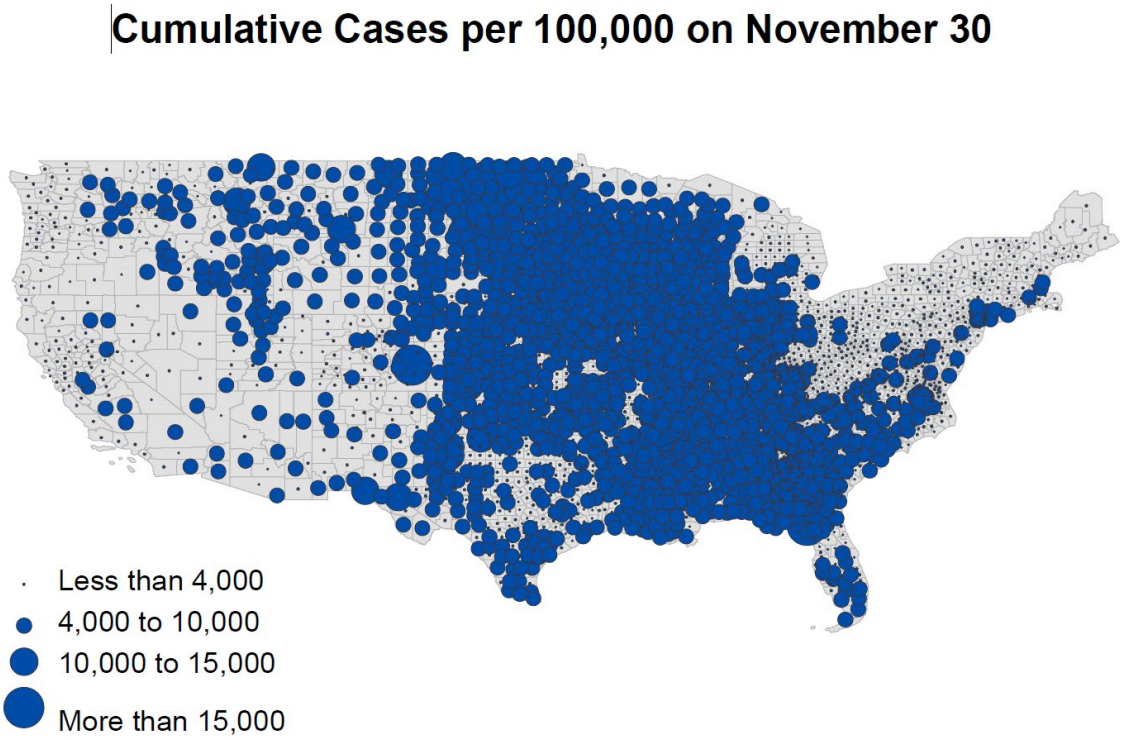
Table A6 - An Investigation of Donald Trump Effects (Sample Defined by Days since Onset)

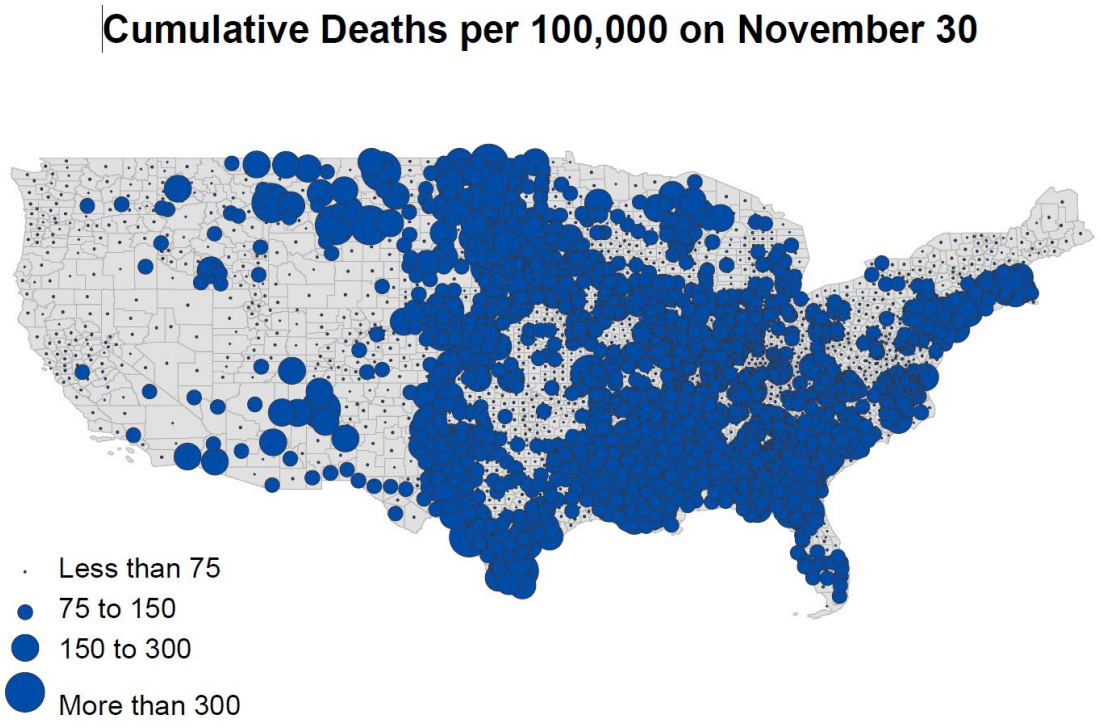
	(1)	(2)	(3)	(4)
	Short Spec., 70 days since onset	**Short Spec., 225 days since onset**	**Comprehensive Spec., 70 days since onset**	**Comprehensive Spec., 225 days since onset**
Panel A – Dependent Variable: Log Cases, IHS				
Trump vote share, 2016 general election	-1.658 (0.152)*** [-0.136]	0.395 (0.093)*** [0.044]	0.541 (0.207)*** [0.044]	2.054 (0.120)*** [0.231]
R^2	0.64	0.77	0.72	0.85
N	3,102	2,744	3,008	2,717
Panel B – Dependent Variable: Log Deaths				
Trump vote share, 2016 general election	-1.496 (0.158)*** [-0.154]	-0.349 (0.158)** [-0.038]	0.352 (0.234) [0.036]	1.846 (0.205)*** [0.200]
R^2	0.46	0.68	0.55	0.81
N	2,655	1,442	2,629	1,436

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors in parentheses and standardized betas in brackets.

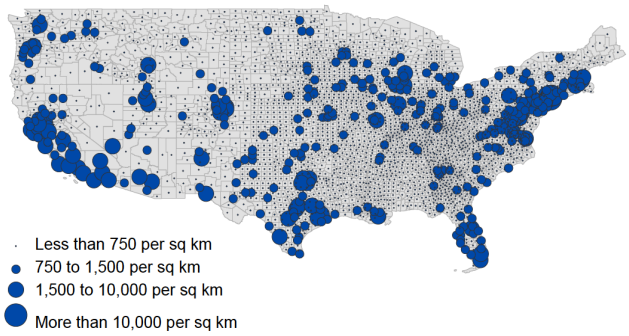
- All columns contain an intercept.
- Columns 1 and 2 (the short specification) include only a control for log population.
- Columns 3 and 4 add controls for all 8 regressors in the baseline specification of Table 1, as well as % high school graduate or higher (among persons age 25+), % with bachelor's degree or higher (among persons age 25+), 30-day mortality for heart attacks, 30-day mortality for heart failure, 30-day mortality for pneumonia, Gini index within bottom 99%, poverty rate, top 1% income share, % housing units in multi-unit structures, persons per household and log population density.
- Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).

Figure A1 – Maps of the Variables Used in the Analysis

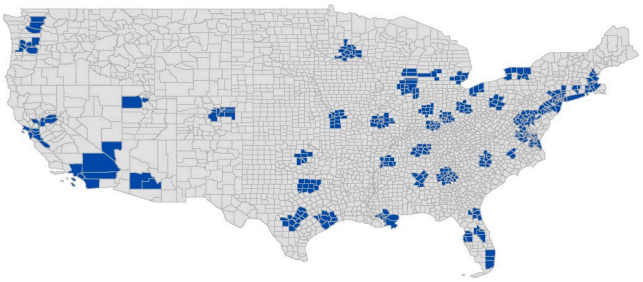




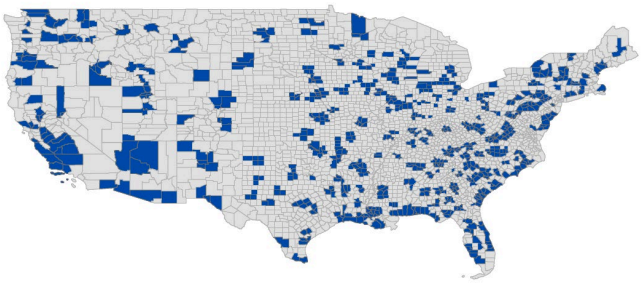
Effective Local Population Density



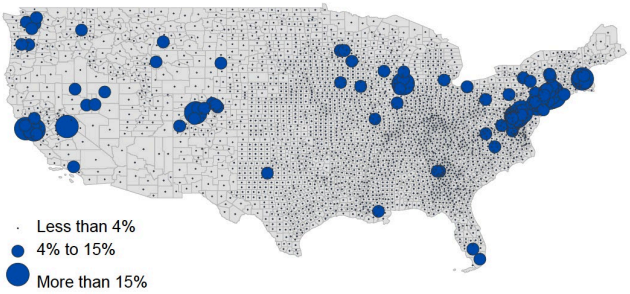
Large Metro



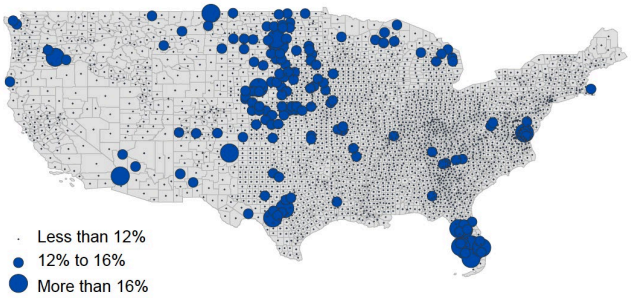
Medium or Small Metro



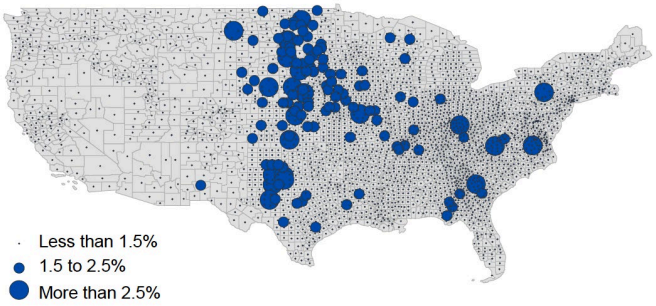
Commute by Public Transportation



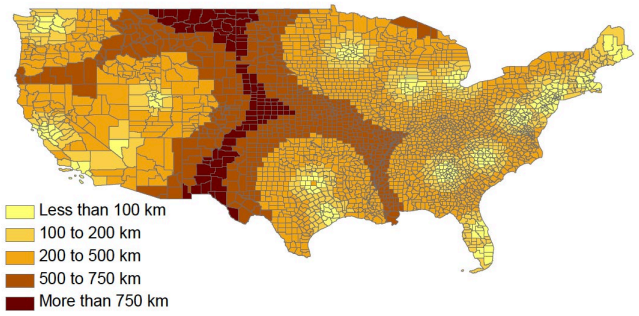
Population Aged 75+



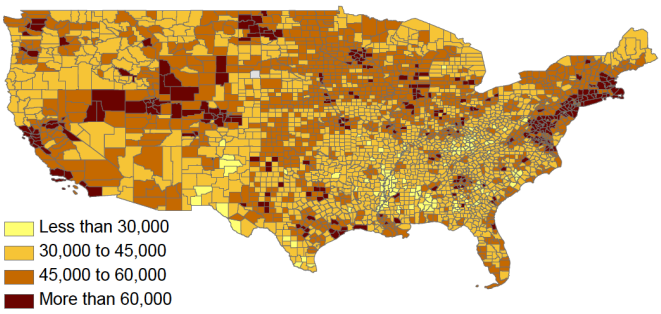
Nursing Homes Residents



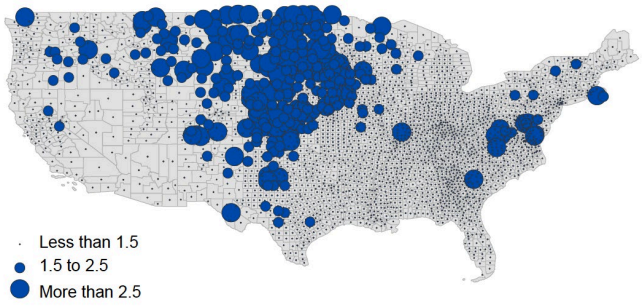
Distance to Airport with Flights to High-COVID Countries



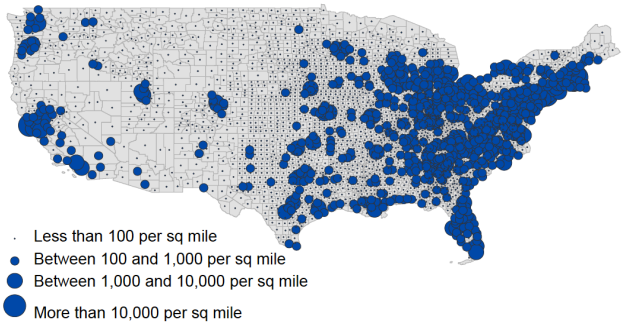
Median Household Income



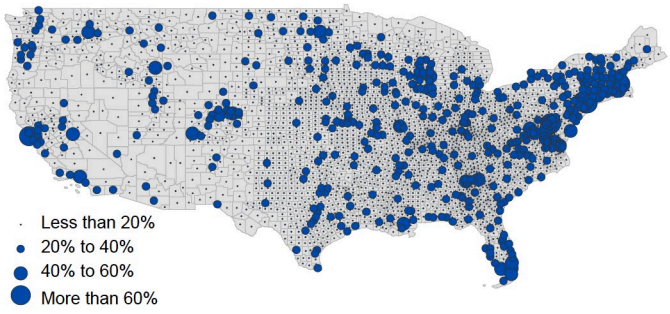
Social Capital



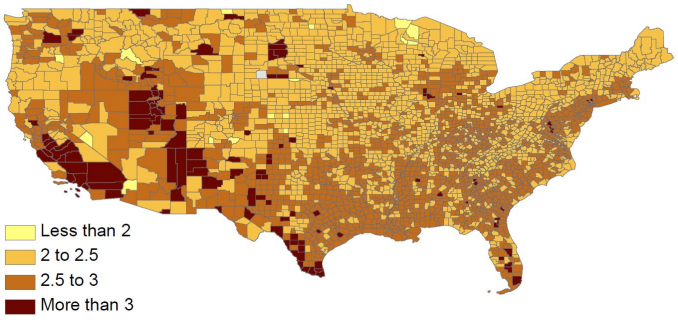
Population Density



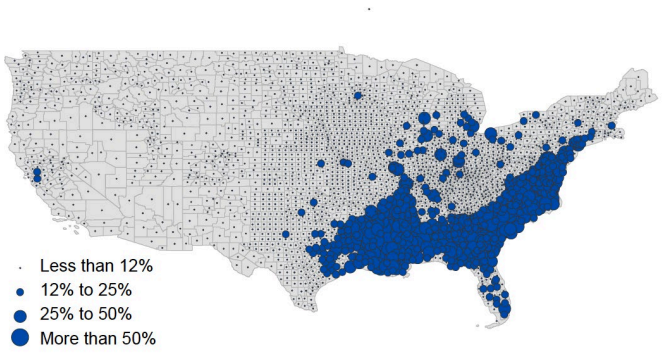
Multi-Unit Housing



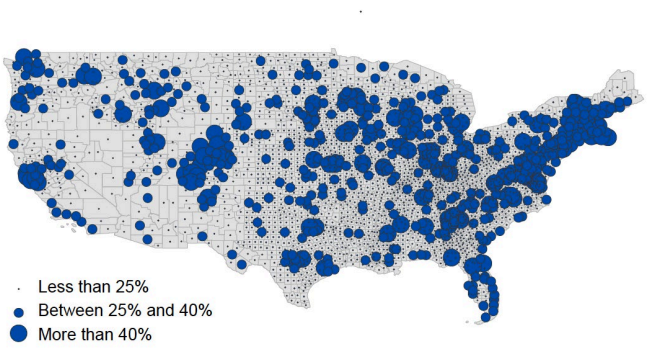
Household Size



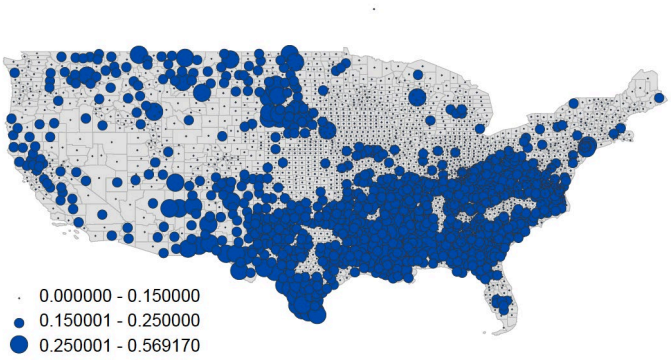
Share of African American Population



Share Bachelor's Degree or More



Poverty Rate



Trump 2016 Vote Share

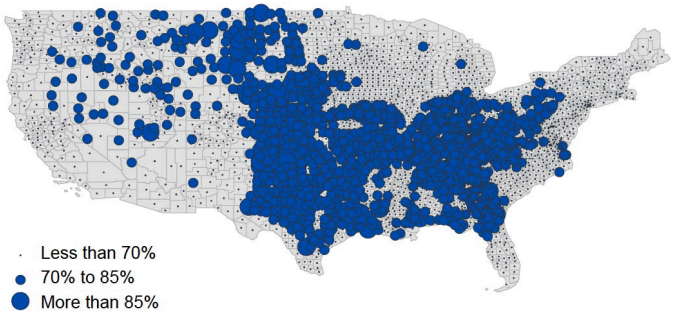


Figure A2A - Effects on Log Cases, by Days Since Onset

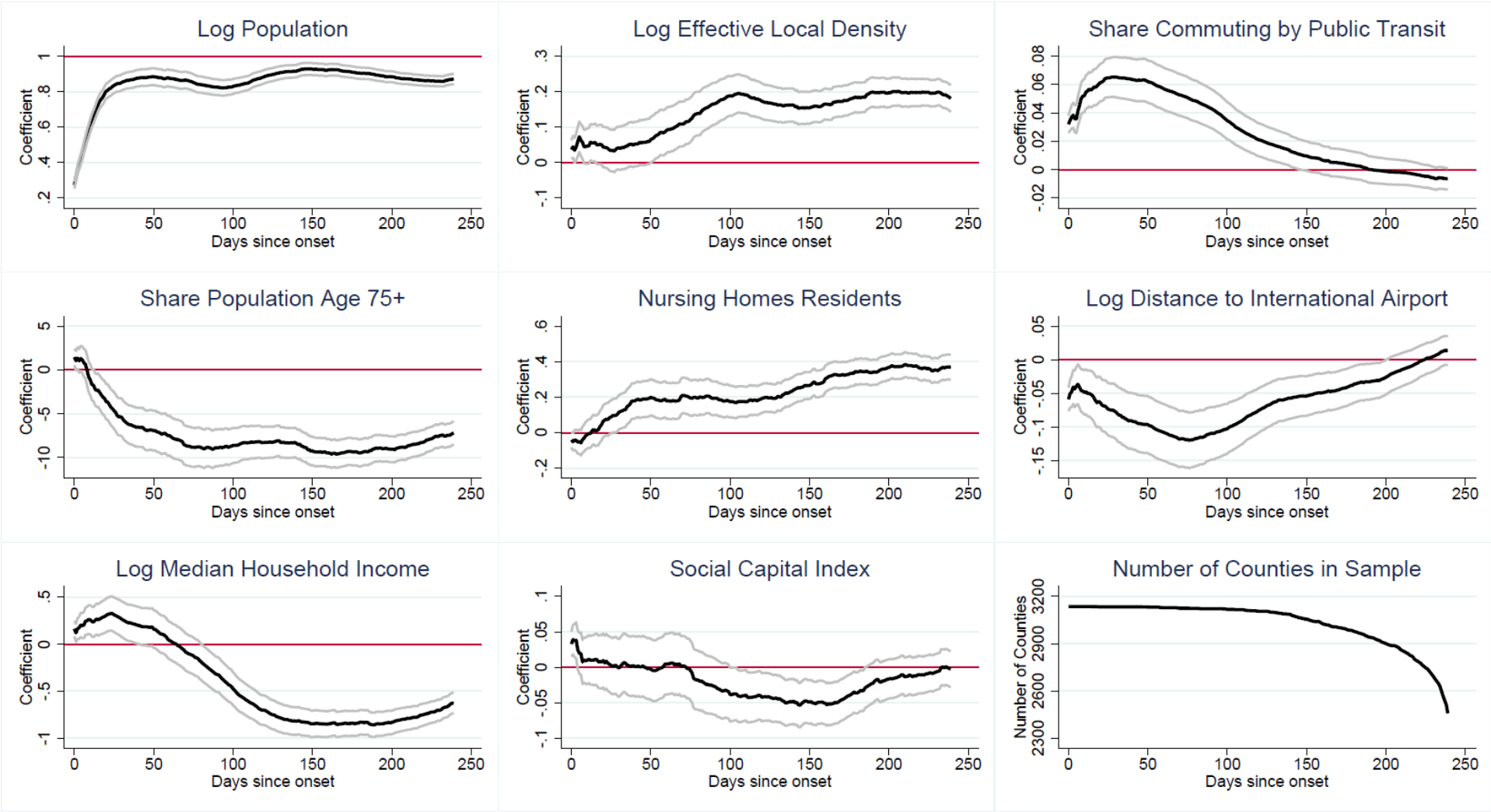


Figure A2B - Effects on Log Deaths, by Days Since Onset

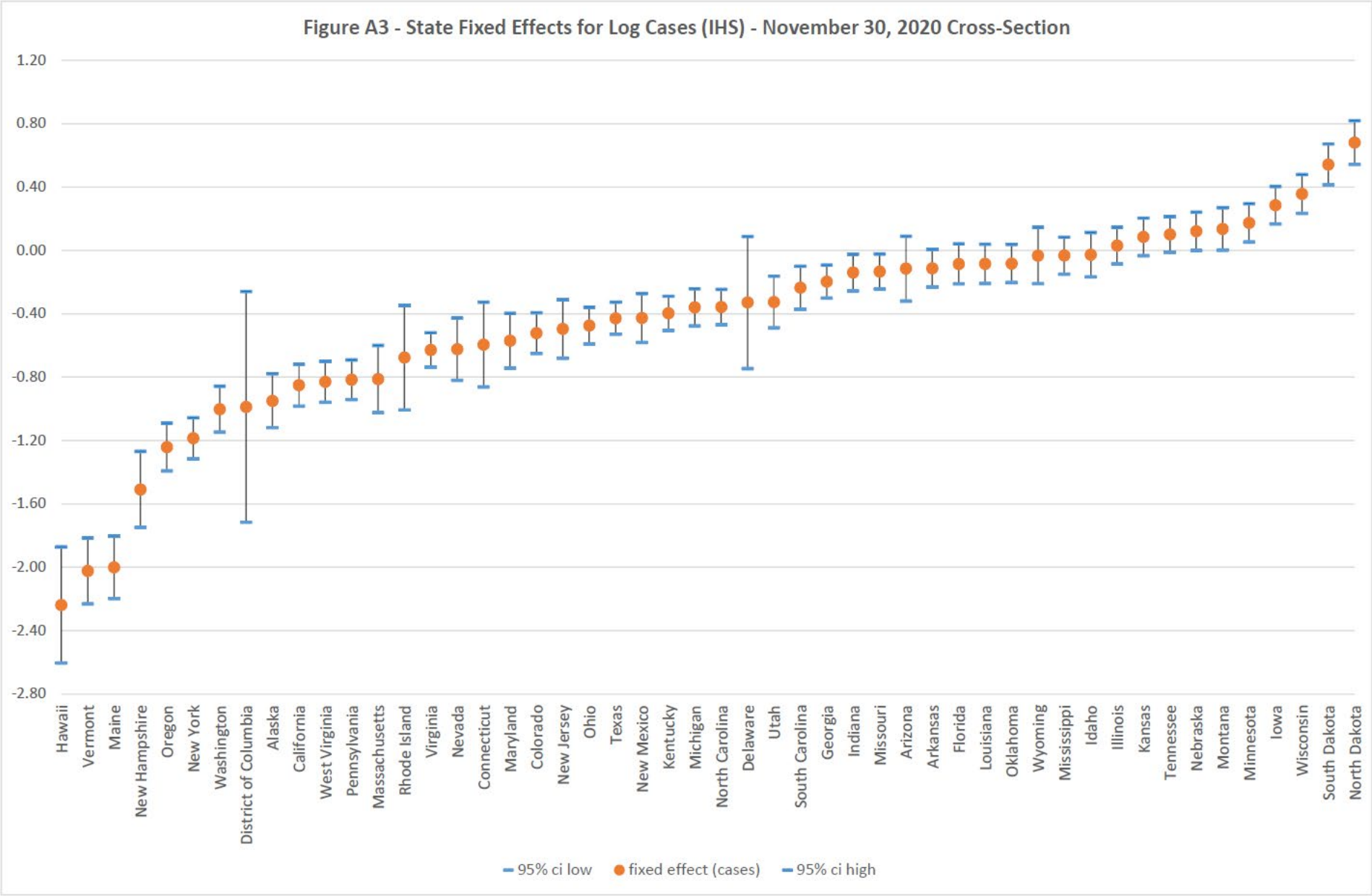


Figure A3 - State Fixed Effects for Log Cases (IHS) - November 30, 2020 Cross-Section
1.20
0.80
0.40
0.00
-0.40
-0.80
-1.20
-1.60
-2.00
-2.40
-2.80
Hawaii
Vermont
Maine
New Hampshire
Oregon
New York
Washington
District of Columbia
Alaska
California
West Virginia
Pennsylvania
Massachusetts
Rhode Island
Virginia
Nevada
Connecticut
Maryland
Colorado
New Jersey
Ohio
Texas
New Mexico
Kentucky
Michigan
North Carolina
Delaware
Utah
South Carolina
Georgia
Indiana
Missouri
Arizona
Arkansas
Florida
Louisiana
Oklahoma
Wyoming
Mississippi
Idaho
Illinois
Kansas
Tennessee
Nebraska
Montana
Minnesota
Iowa
Wisconsin
South Dakota
North Dakota
95% ci low
fixed effect (cases)
95% ci high

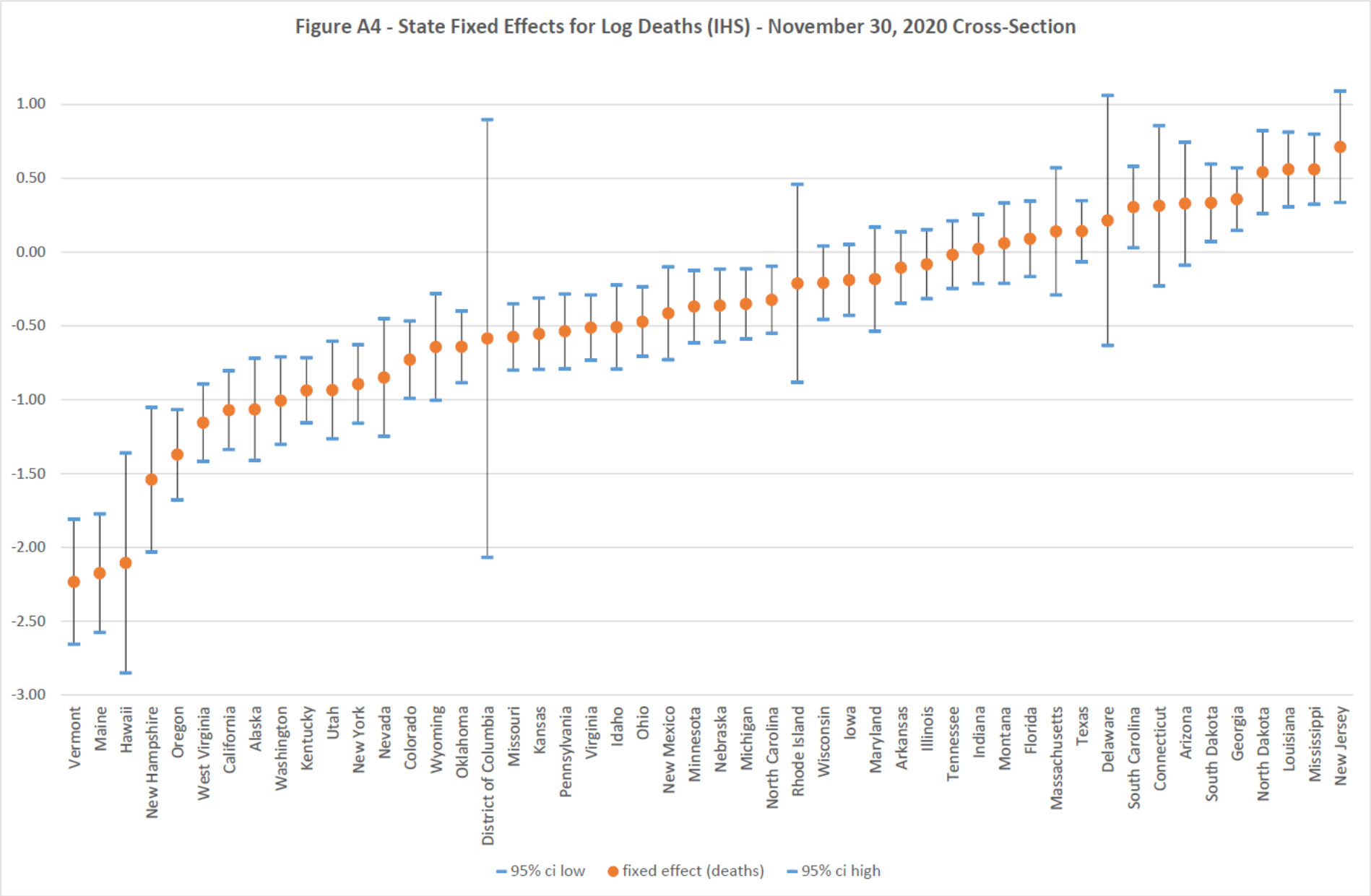


Figure A4 - State Fixed Effects for Log Deaths (IHS) - November 30, 2020 Cross-Section
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
-3.00
Vermont
Maine
Hawaii
New Hampshire
Oregon
West Virginia
California
Alaska
Washington
Kentucky
Utah
New York
Nevada
Colorado
Wyoming
Oklahoma
District of Columbia
Missouri
Kansas
Pennsylvania
Virginia
Idaho
Ohio
New Mexico
Minnesota
Nebraska
Michigan
North Carolina
Rhode Island
Wisconsin
Iowa
Maryland
Arkansas
Illinois
Tennessee
Indiana
Montana
Florida
Massachusetts
Texas
Delaware
South Carolina
Connecticut
Arizona
South Dakota
Georgia
North Dakota
Louisiana
Mississippi
New Jersey
95% ci low
fixed effect (deaths)
95% ci high