Geographic Inequality of COVID-19 in the United States

Johannes Norling*

November 5, 2021

Abstract

The COVID-19 pandemic is unevenly distributed across counties in the United States. Inequality in daily new cases and deaths fell from the start of the pandemic into summer 2020, and has since fluctuated, remaining more equal than the distribution of ethnic and racial groups but less equal than the distribution of poverty. Falling inequality within states accompanies surges in new cases and deaths, indicating that new waves of the pandemic occur at the state level. Although the 1918 influenza pandemic was nearly four times as deadly, deaths during the two pandemics were similarly unevenly distributed.

JEL codes: I14

Keywords: COVID-19, 1918 Influenza, Inequality

* Department of Economics, Mount Holyoke College, 50 College Street, South Hadley, Massachusetts 01075, United States of America; jnorling@mtholyoke.edu. Anjali Mathews provided excellent research assistance.

1. Introduction

Substantial public attention has focused on geographic variation in the COVID-19 pandemic across the United States. News reports have at different times identified hot spots in Montana and the Dakotas (Wines 2021), in the northeast and then southeast (Tauber 2021), in all five boroughs of New York City (Fitzsimmons and Petri 2020, Freytas-Tamura et al. 2020, Saltonstall 2020, Slotnik 2020a, and Slotnik 2020b), and elsewhere. The purpose of this paper is to formally measure geographic variation in experience of COVID-19 and track how it has changed during the pandemic. This paper builds on previous studies that document geographic inequality in mortality within the United States over the past several decades (Vierboom et al. 2019, Couillard et al. 2021, Deryugina and Molitor 2021), and that document other inequalities in experience of COVID-19, such as by race and ethnicity (Raifman and Raifman 2020, Figueroa et al. 2021).

I measure unevenness in daily COVID-19 cases by county using the index of dissimilarity, which calculates the share of people newly diagnosed with COVID-19 who would need to move counties in order for COVID-19 rates to be the same in every county. Larger values of this hypothetical share indicate greater unevenness between counties in experience of COVID-19. At the start of the pandemic in early 2020, cases were concentrated in a handful of counties. By June 2020, inequality across counties had fallen such that less than 40 percent of newly-diagnosed people would have had to move in order to even out COVID-19 rates across all counties. Inequality in daily new cases has since fluctuated, reaching a minimum of 16 percent in December 2020, and has never again exceeded 40 percent. The distribution of new COVID-19 cases across counties has generally been more even than the distribution of ethnic and racial groups between counties. For example, 45 percent of Hispanic residents would need to move to have the same proportion of Hispanic residents in every county. On the other hand, the distribution of new COVID-19 cases has generally been more uneven than the distribution of people in poverty, 16 percent of whom would need to move in order to even out poverty rates across counties.

I separately measure inequality between counties within the same state, and between states. Through October 2021, there have been five surges in average number of new cases nationwide. During each surge, inequality between counties within states fell. As overall cases

declined, inequality within states remained constant or rose. COVID-19 cases surge at the state level, while cases tend to fall in some counties before their neighbors. Inequality between states exhibits no similar pattern, and has risen and fallen both as cases increase and as cases decrease.

I perform similar calculations using deaths due to COVID-19, which tend to follow cases by about two weeks. Deaths are more unevenly distributed, but as with cases, inequality in deaths within states falls as deaths surge, and stays constant or rises as deaths fall. I additionally compare COVID-19 to other leading causes of death. Only heart disease and cancer were deadlier in 2019 than COVID-19 was in 2020, but the distribution of COVID-19 deaths in 2020 was more than 15 percent more uneven than was any other leading cause of death in 2019. In 2020, COVID-19 has been just as deadly but more evenly distributed.

I finally compare COVID-19 to the 1918 influenza pandemic, which lasted until 1920. I use monthly city-level reports of excess deaths in 471 cities during the 1918 influenza pandemic, which I compare to COVID-19 deaths in the 319 present-day counties that contain those 471 cities. 840 out of every 100,000 people in these cities died during the 1918 influenza pandemic, nearly four times the COVID-19 death rate thus far of 216 per 100,000 in the corresponding counties. Yet, the distribution of deaths was similarly uneven during the two pandemics: at the height of both pandemics, between 20 and 30 percent of people would have needed to move in order to even out death rates across cities or counties.

2. Data and methodology

2.1. COVID-19 cases and deaths

Johns Hopkins University records the cumulative number of confirmed cases of and deaths due to COVID-19, by county, every day since January 22, 2020 (Dong et al. 2020). I calculate new cases and deaths as the change from the previous day. Due to reporting adjustments, cumulative cases decline in 1.4 percent of county-day observations, and cumulative deaths decline in 0.4 percent of observations. In these instances, I set new cases or deaths to be

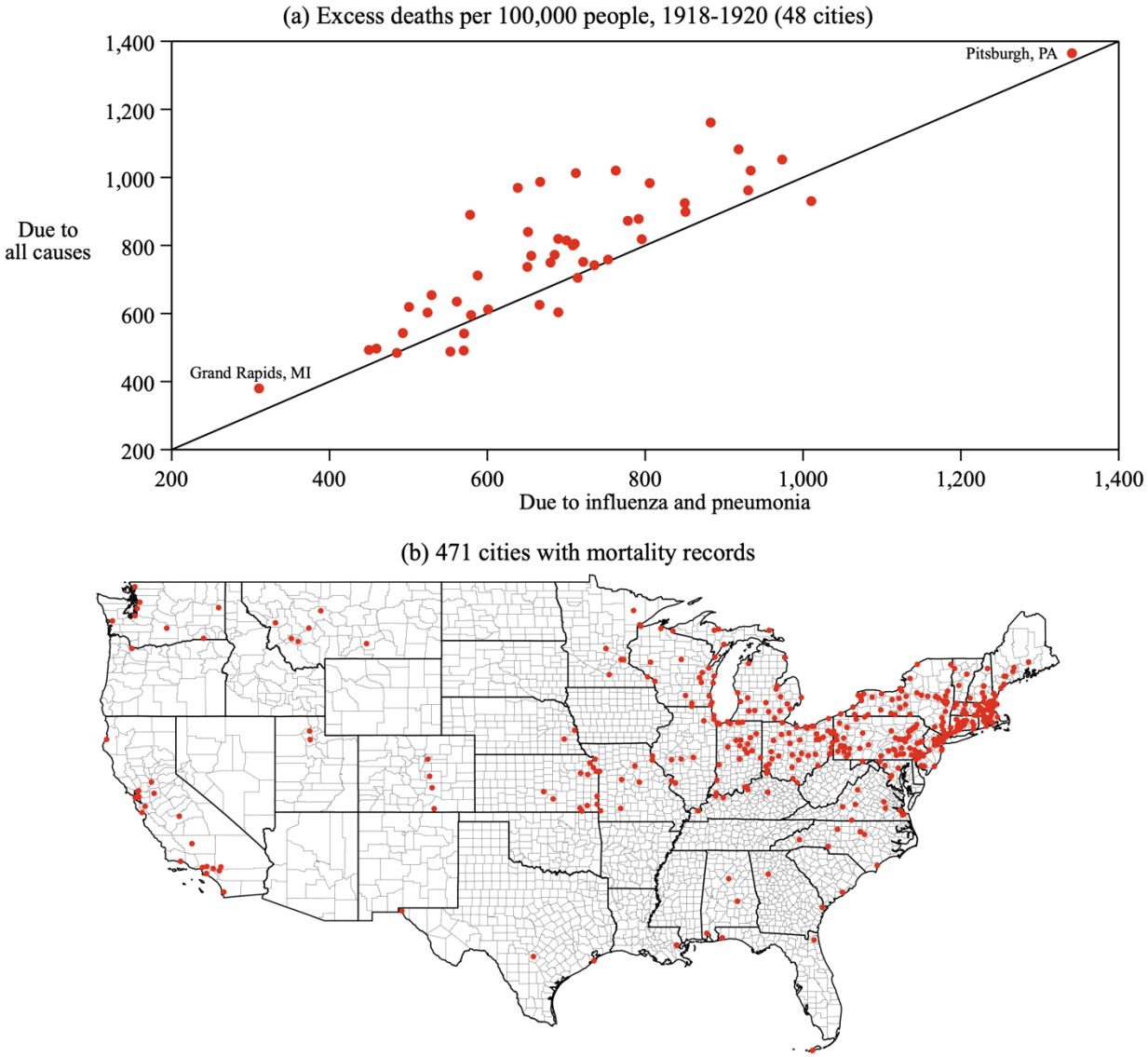
zero.[1] Because of unevenness in reporting by day of the week and around holidays, I use each day's cases and deaths averaged across the prior two weeks. The database also records 2019 county-level population estimates from the US Census Bureau.

2.2. 1918 influenza pandemic deaths

Unlike the SARS-CoV-2 coronavirus that causes COVID-19, the H1N1 virus that caused the 1918 influenza pandemic was not directly observable at the time (Jester et al. 2019). Additionally, records of people who became sick but recovered were not widely maintained. Only deaths were consistently recorded, typically as either influenza or pneumonia. Counts of influenza and pneumonia deaths by week were recorded in just 43 cities (Markel et al. 2007), and by month in just 50 cities (Collins et al. 1930). Instead, I use monthly records of deaths due to all causes in 471 cities (US Census Bureau 1916-1925). For each city and month between 1918 and 1920, I estimate pandemic deaths as the number of deaths minus the median number of deaths in that same city and month in 1914, 1915, 1916, 1917, 1921, and 1922 (the only neighboring years for which similar mortality records were maintained). As depicted in panel (a) of Figure 1, this measure of all-cause excess deaths during the pandemic is correlated with excess deaths due to influenza and pneumonia over the same period in the 48 cities with deaths recorded by cause. Using all-cause excess deaths offers a larger number of cities with which to measure geographic inequality. As depicted in panel (b) of Figure 1, these 471 cities are located in 319 present-day counties in 37 states. The 471 cities contained 37 percent of the population of the United States in 1918, and the 319 counties contained 39 percent of the population in 2019.

[1] I make several additional revisions: The Johns Hopkins University database does not consistently track cases and deaths by county in Utah, so for Utah I use similar records from the New York Times (2021) that track daily, county-level record of cases and deaths in Utah. Cases and deaths are reported for all of Kansas City, Missouri together, which lies in four counties: Cass, Clay, Jackson, and Platte. I allocate Kansas City's cases and deaths to each county in proportion to its population. Similarly, Alaska's Bristol Bay borough and Lake and Peninsula borough are reported together, as are Massachusetts' Dukes county and Nantucket county. I again allocate cases and deaths to each county according to population.

Figure 1: Deaths during the 1918 influenza pandemic



2.3. Index of dissimilarity

The index of dissimilarity measures the unevenness of the distribution of two groups across a collection of units. (Massey and Denton [1988] describe other measure of segregation, such as centralization and clustering.) The index is most commonly used to measure residential segregation by neighborhood within a city, such as by race (Cutler et al. 1999) or place of birth (Cutler et al. 2008). I use the index to measure unevenness of COVID-19 cases and deaths by county within the United States. I use a standard formulation (Eriksson and Ward 2019):

$$D = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{a_i}{A} - \frac{b_i}{B}\right| \quad (1)$$

D is the index of dissimilarity. The population consists of A people newly diagnosed with COVID-19 and B other people. Each of the N=3,142 counties, indexed by i, contains a_i people newly diagnosed with COVID-19 and b_i other people. The index equals zero if the proportion of people newly diagnosed with COVID-19 is evenly distributed across counties according to their populations. The index approaches a maximum value of one if all cases of COVID-19 are concentrated in a single county. The index has a convenient interpretation: the minimum share of people newly diagnosed with COVID-19 who would have to move counties in order for COVID-19 rates to the same in every county.

I similarly calculate the index of dissimilarity between states using the equation 1, where i indexes each of the N=50 states. I also calculate the index between counties within a state. The average of each of these 50 within-state values, weighted by the number of newly diagnosed people in each state, yields the minimum share of people newly diagnosed with COVID-19 nationwide who would need to move to a different county within the same state in order for the COVID-19 to be the same in every county within states (but possible varying across states).[2]

3. Findings

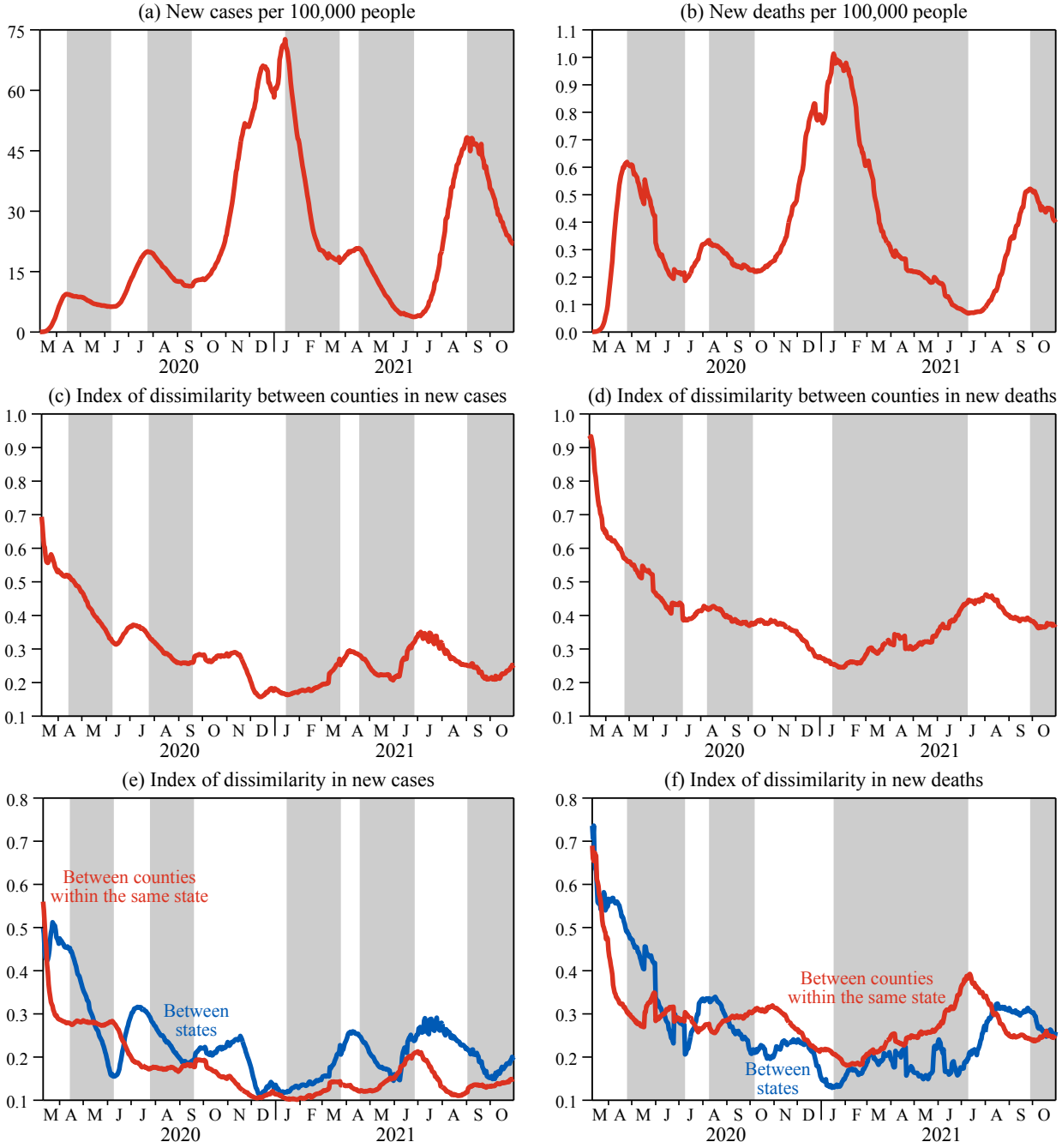
3.1. Inequality in cases of COVID-19

Panel (a) of Figure 2 shows the number of newly diagnosed cases of COVID-19 per 100,000 people in the United States from March 11, 2020 (when the World Health Organization declared COVID-19 a pandemic) through October 2021. Periods of rising cases are identified with white a background, and periods of falling cases are identified with a gray background.

[2] The index of dissimilarity is not decomposable: between-state and within-state values of the index do not necessarily sum to the overall between-county index. However, the changes over time in between-state and within-state values of the index, which I present in the next section, are similar using the Theil index, an alternative measure of unevenness that is decomposable. I use the index of dissimilarity because its value has a meaningful interpretation.

(Peaks and troughs in cases are defined as days in which new cases are higher or lower than any other day within an eight-week window.) I repeat the same white and gray backgrounds to identify periods of rising and falling cases on later graphs. There have been five distinct surges in cases, the longest of which lasted four months, from September 2020 into January 2021, and culminated in the highest daily rate of 73 new cases per 100,000 people.

Figure 2: Geographic inequality in COVID-19



Panel (c) of Figure 2 shows the index of dissimilarity in new cases each day. As the virus first appeared in the United States, its distribution was uneven, so most people with COVID-19 would have needed to move in order for COVID-19 rates to be equal across counties. The distribution of cases across counties became more even as the virus spread. Since June 2020, inequality has fluctuated, reaching a nadir in December 2020, when just 16 percent of newly-diagnosed people would have need to move to even out COVID-19 rates across counties. To understand the magnitude these measurements, I compare them to other characteristics that vary by county (US Census Bureau 2020a and 2020b). Since May 2020, new COVID-19 cases have been more evenly distributed than ethnic or racial groups. For example, the index of dissimilarity between counties is 0.45 for Hispanic people and 0.43 for Black people. On the other hand, the index of dissimilarity for people in poverty is 0.16, an evenness matched by new COVID-19 cases only in December 2020.

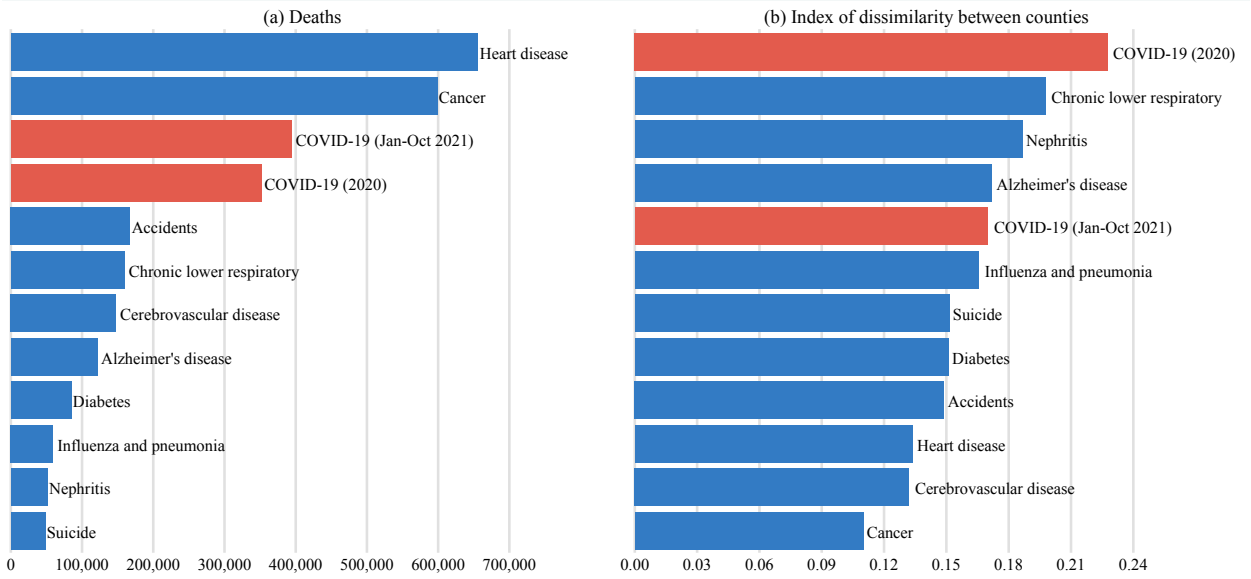
I finally separately calculate inequality in new cases between counties within the same state, and inequality between states, presented in panel (e) of Figure 2. During every surge in cases, inequality between counties within states fell. During every period of falling cases, within-state inequality stayed constant or rose. This evidence highlights the usefulness of measuring the geographic inequality in COVID-19 cases to understand the evolution of the pandemic. Cases tend to surge at the state level, while cases tend to decline at the county level. The index of dissimilarity between states shows no similar pattern, sometimes rising and sometimes falling during surges in cases and during declines in cases.

3.2. Inequality in deaths due to COVID-19

Panel (b) of Figure 2 shows the daily number of new deaths due to COVID-19 per 100,000 people in the United States. Again, the white and gray backgrounds indicate periods when deaths surged and fell. Daily deaths peaked in January 2021 at 1.01 per 100,000 people. Panel (d) shows the index of dissimilarity in deaths between counties. Deaths due to COVID-19 are more unevenly distributed than are cases. Panel (f) compares inequality between counties within the same state to inequality between states. As with cases, inequality within states falls as deaths surge and stays constant or rises as deaths fall, and inequality between states is not consistently associated with surges and falls in new deaths.

Panel (a) of Figure 3 compares total deaths due to COVID-19 in 2020 and thus far in 2021 to the ten leading causes of death in 2019 (Centers for Disease Control and Prevention 2021b). Annually, COVID-19 has been deadlier than all but two other causes of death, heart disease and cancer. Panel (b) compares inequality between counties in each cause of death.[3] Deaths due to COVID-19 in 2020 had an index of dissimilarity of 0.227 and were more unevenly distributed than any other cause of death in 2019. Deaths due to COVID-19 became more evenly distributed in 2021, equivalent to inequality in other leading causes of death.

Figure 3: COVID-19 deaths compared to 10 leading causes of death in 2019



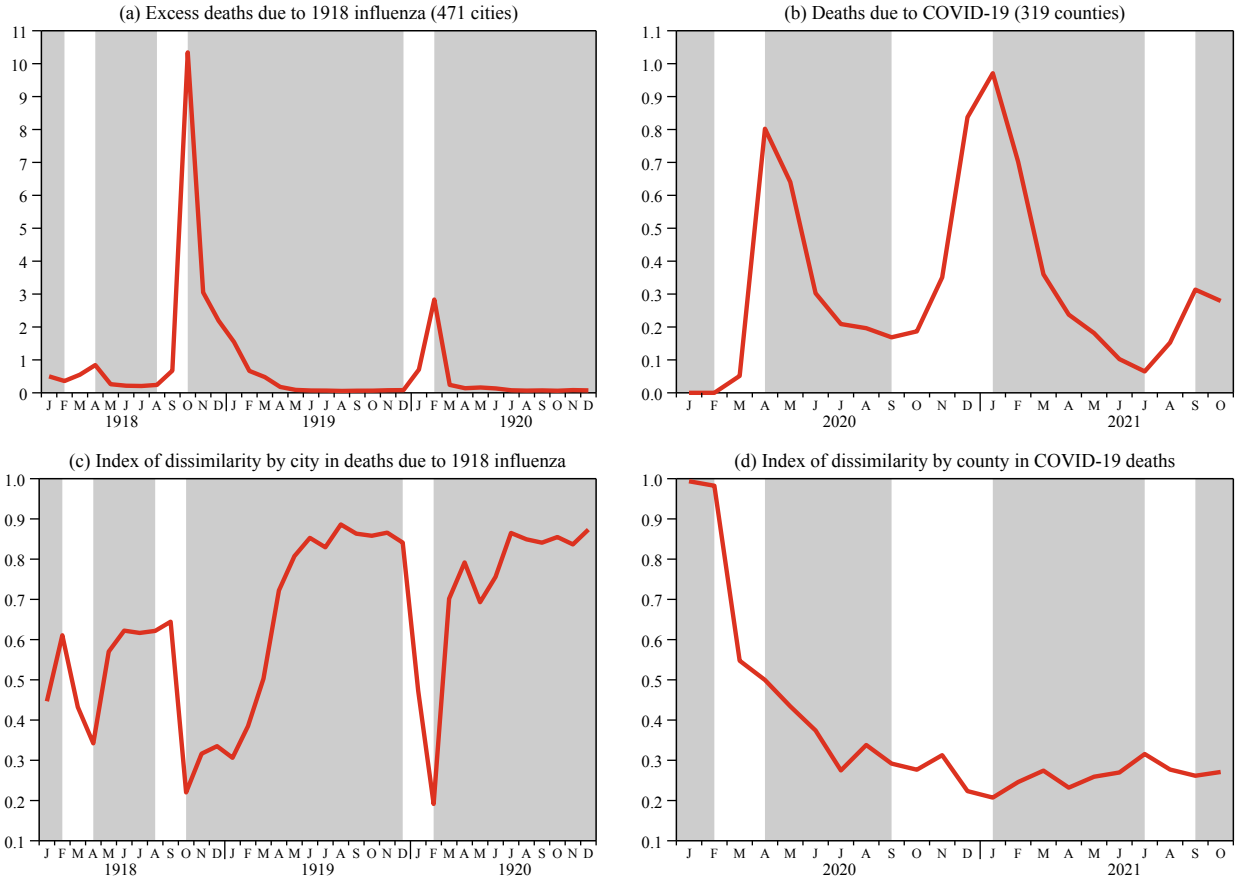
3.3. Inequality in deaths due to the 1918 influenza pandemic

Panels (a) and (b) of Figure 4 show the daily average number of deaths per 100,000 people in each month of the 1918 influenza pandemic and the COVID-19 pandemic. Again, the 1918 values are estimates of excess deaths from 471 cities, and COVID-19 deaths are from only

[3] These calculations sum all deaths by cause in a year. For COVID-19, this aggregation evens out some of the day-to-day inequality in panel (d) of Figure 2. Additionally, publicly-available records suppress counties with small death counts in 2019, so the group of counties used to calculate each index of dissimilarity varies slightly.

the 319 counties that contain these 471 cities. The 1918 pandemic had three distinct surges in deaths, in early 1918, late 1918, and early 1920, again highlighted with white backgrounds. 10.3 out of every 100,000 people died from influenza in October 1918, ten times peak COVID-19 mortality in January 2020.

Figure 4: Inequality in the 1918 influenza and COVID-19 pandemics



Panels (c) and (d) of Figure 4 show the index of dissimilarity in deaths by month during each pandemic. As deaths surged in both pandemics, deaths were similarly distributed: between 20 and 30 percent of deaths would have needed to occur elsewhere in order for the death rate to be the same nationwide. As deaths declined, inequality rose, particularly in 1918. Measurement error may explain the high levels of inequality in the last half of 1919 and in much of 1920: because deaths are only estimated in excess of a baseline median in nearby years, any noise in this estimate may dwarf actual pandemic deaths when pandemic deaths were rare.

The 1918 influenza pandemic was deadlier than the COVID-19 pandemic has been thus far. There were 840 excess deaths per 100,000 people in the 471 cities between 1918 and 1920, nearly four times the 216 deaths due to COVID-19 per 100,000 people in the 391 corresponding counties through October 2021. Yet the distribution of deaths was similarly uneven during both pandemics. The index of dissimilarity between cities in cumulative excess deaths between 1918 and 1920 was 0.142, similar to the index of dissimilarity of 0.151 between the corresponding counties in COVID-19 deaths.

4. Conclusion

The COVID-19 pandemic is unevenly distributed across counties in the United States. Inequality in daily new cases and deaths fell from the start of the pandemic into summer 2020, and has since fluctuated, remaining more equal than the distribution of ethnic and racial groups but less equal than the distribution of poverty. Falling inequality within states accompanies surges in new cases and deaths, indicating that new waves of the pandemic occur at the state level. Although the 1918 influenza pandemic was nearly four times as deadly, deaths during the two pandemics were similarly unevenly distributed.

As new virus variants emerge, vaccines become more widely available, and public health policies (such as masking requirements) change, the distribution of COVID-19 across counties may also change. It is also plausible that COVID-19 has changed the distribution of other causes of death since the start of 2020. For example, the 2020-2021 flu season was unusually mild (Centers for Disease Control and Prevention 2021a). Finally, the index of dissimilarity could similarly be used to measure inequality in COVID-19 between countries, although such calculations may be sensitive to differential reporting between countries.

Works cited

Centers for Disease Control and Prevention. 2021a. 2020-2021 Flu Season Summary. Available at <https://www.cdc.gov/flu/season/faq-flu-season-2020-2021.htm>, accessed November 2, 2021.

Centers for Disease Control and Prevention. 2021b. CDC Wonder. Available at <http://wonder.cdc.gov>, accessed October 19, 2021.

Collins, Selwyn D., W. H. Frost, Mary Gover, and Edgar Sydenstricker. 1930. "Mortality from Influenza and Pneumonia in 50 Large Cities of the United States, 1910-1929." *Public Health Reports (1896-1970)*, 45 (39): 2277-2328.

Couillard, Benjamin K., Christopher L. Foote, Kavish Gandhi, Ellen Meara, and Jonathan Skinner. 2021. "Rising Geographic Disparities in US Mortality." *Journal of Economic Perspectives*, 35 (4): 123-146.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107 (3): 455-506.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation." *Review of Economics and Statistics*, 90 (3): 478-497.

De Freytas-Tamura, Kimiko, Winnie Hu, and Londsey Rogers Cook. 2020. "'It's the Death Towers': How the Bronx Became New York's Virus Hot Spot." *New York Times*, May 26, 2020. Available at <https://www.nytimes.com/2020/05/26/nyregion/bronx-coronavirus-outbreak.html>, accessed October 28, 2021.

Deryugina, Tatyana, and David Molitor. 2021. "The Causal Effects of Place on Health and Longevity." *Journal of Economic Perspectives*, 35 (4): 147-170.

Dong, Ensheng, Hongru Du, and Lauren Gardner. 2020. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time." *Lancet*, 20 (5): P533-P534. Available at <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series>, accessed November 1, 2021.

Eriksson, Katherine, and Zachary Ward. 2019. "The Residential Segregation of Immigrants in the United States from 1850 to 1940." *Journal of Economic History*, 79 (4): 989-1026.

Figueroa, Jose F., Riski K. Wadhera, Winta T. Mehtsun, Kristen Riley, Jessica Phelan, and Ashish K. Jha. 2021. "Association of Race, Ethnicity, and Community-Level Factors with COVID-19 Cases and Deaths across U.S. Counties." *Healthcare*, 9: 100495.

Fitzsimmons, Emma G., and Alexandra E. Petri. 2020. "How N.Y.C.'s Conservative Bastion Became a Virus Hot Spot." *New York Times*, November 11, 2020. Available at <https://www.nytimes.com/2020/11/11/nyregion/staten-island-second-wave.html>, accessed October 28, 2021.

Markel, Howard, Harvey B. Lipman, J. Alexander Navarro, Alexandra Sloan, Joseph R. Michalsen, Alexandra Minna Stern, and Martin S. Cetron. 2007. "Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic." *JAMA*, 298 (6): 644-654.

Massey, Douglas S., and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces*, 67 (2): 281-315.

New York Times. 2021. Coronavirus (Covid-19) Data in the United States. Available at <https://github.com/nytimes/covid-19-data>, accessed November 1, 2021.

Raifman, Matthew A., and Julia R. Raifman. 2020. "Disparities in the Population at Risk of Severe Illness from COVID-19 by Race/Ethnicity and Income." *American Journal of Preventive Medicine*, 59 (1): 137-139.

Saltonstall, Gus. 2020. "ZIP Code in Upper Manhattan Hot Spot Sees COVID Rate Pass 8%." *Patch*, December 28, 2020. Available at <https://patch.com/new-york/washington-heights-inwood/was-heights-zip-code-has-highest-covid-rate-borough-over-8>, accessed October 28, 2021.

Slotnik, Daniel E. 2020a. "A New Battle Against the Virus Is Set for 9 Zip Codes." *New York Times*, October 5, 2020. Available at <https://www.nytimes.com/2020/10/05/nyregion/coronavirus-zip-codes-nyc.html>, accessed October 28, 2021.

Slotnik, Daniel E. 2020b. "Coronavirus in N.Y.C.: Latest Updates." *New York Times*, October 6, 2020. Available at <https://www.nytimes.com/2020/10/06/nyregion/coronavirus-nyc.html>, accessed October 28, 2021.

Tauber, Eugene. 2021. "Pennsylvania Coronavirus Update: Southeast Region Becoming New Hot Spot as Northeast Moderates." *Morning Call*, April 20, 2021. Available at <https://www.mcall.com/coronavirus/mc-nws-coronavirus-apr-20-update-20210420-5davi4kpvzh77lua5rzuj4pcae-story.html>, accessed October 28, 2021.

US Census Bureau. 1916-1925. *Mortality Statistics*. Washington, D.C.: Government Printing Office.

US Census Bureau. 2020a. County Population by Characteristics: 2010-2019. Available at <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html>, accessed December 9, 2020.

US Census Bureau. 2020b. Small Area Income and Poverty Estimates (SAIPE). Available at <https://www.census.gov/data-tools/demo/saipe/#/?map_geoSelector=aa_c>, accessed December 9, 2020.

Vierboom, Yana C., Samuel H. Preston, and Arun S. Hendi. 2019. "Rising Geographic Inequality in Mortality in the United States." *SSM – Population Health*, 9: 100478.

Wines, Michael. 2021. "Montana and the Dakotas Were Hot Spots. Until They Weren't." *New York Times*, April 22, 2021. Available at <https://www.nytimes.com/2021/02/20/world/covid-coronavirus-dakotas.html>, accessed October 28, 2021.