

COVID-19

County-level socio-economic disparities in COVID-19 mortality in the USA

Denys Dukhovnov ^{1*} and Magali Barbieri^{1,2}

¹Department of Demography, University of California, Berkeley, CA, USA and ²Mortality, Health, and Epidemiology, French National Institute for Demographic Studies (INED), Aubervilliers, Île de France, France

*Corresponding author. Department of Demography, University of California, Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720, USA. E-mail: denys_dukhovnov@berkeley.edu

Received 30 July 2021; Editorial decision 2 December 2021; Accepted 14 December 2021

Abstract

Background: Preliminary studies have suggested a link between socio-economic characteristics and COVID-19 mortality. Such studies have been carried out on particular geographies within the USA or selective data that do not represent the complete experience for 2020.

Methods: We estimated COVID-19 mortality rates, number of years of life lost to SARS-CoV-2 and reduction in life expectancy during each of the three pandemic waves in 2020 for 3144 US counties grouped into five socio-economic status categories, using daily death data from the Johns Hopkins University of Medicine and weekly mortality age structure from the Centers for Disease Control.

Results: During March–May 2020, COVID-19 mortality was highest in the most socio-economically advantaged quintile of counties and lowest in the two most-disadvantaged quintiles. The pattern reversed during June–August and widened by September–December, such that COVID-19 mortality rates were 2.58 times higher in the bottom than in the top quintile of counties. Differences in the number of years of life lost followed a similar pattern, ultimately resulting in 1.002 (1.000, 1.004) million years in the middle quintile to 1.381 (1.378, 1.384) million years of life lost in the first (most-disadvantaged) quintile during the whole year.

Conclusions: Diverging trajectories of COVID-19 mortality among the poor and affluent counties indicated a progressively higher rate of loss of life among socio-economically disadvantaged communities. Accounting for socio-economic disparities when allocating resources to control the spread of the infection and to reinforce local public health infrastructure would reduce inequities in the mortality burden of the disease.

Key words: Mortality, COVID-19, age-standardized death rates, years of life lost, the USA

Key Messages

- In 2020, the most socio-economically disadvantaged areas in the USA appear to have experienced a 31% heavier mortality burden from the pandemic than the most socio-economically advantaged ones.
- The total number of years of life lost to COVID-19 in 2020 amounted to 5.71 million, equivalent to 16 years and 170 days not lived by an average person dying of the disease, whereas relative to the 2015–2019 average, life expectancy at birth fell by at least 0.85 (0.82, 0.87) years to 78.88 years, close to the 2006 level. Among the most-disadvantaged counties it declined by 1.13 (1.08, 1.19) years in 2020 compared with 0.69 (0.64, 0.75) years in the most socio-economically advantaged counties.
- The gap in life expectancy at birth between the two extreme socio-economic status (SES) quintiles increased from 5.75 to 6.18 years between the 2015–2019 period and the year 2020.
- A clear gradient of years of life lost (YLL) to COVID-19 by SES quintile emerged through 2020: from 0.48, 0.20 and 0.31 million years under the conditions of each of the three waves (0.99 million in total) for the population in the most-affluent quintile to 0.27, 0.36 and 0.76 million years (1.38 million in total) for that in the least-affluent SES quintile, respectively.
- It will be essential to monitor the long-term consequences of the SARS-CoV-2 infection among the least-affluent segments of the US population, for whom the lack of protective factors and limited access to healthcare might be particularly detrimental.

Introduction

Prior research has demonstrated the disproportionate mortality impact of recent natural disasters (heat waves, hurricanes, floods, etc.) on the most socio-economically disadvantaged populations in the USA.^{1–4} Analyses based on preliminary data indicated that this is also likely the case for COVID-19. A strong relationship between socio-economic status (SES) and COVID-19 mortality is expected due to the mechanisms of viral transmission. Low-income populations are over-represented in the essential workforce and within the economic sectors least amenable to remote work, thus increasing potential exposure to the SARS-CoV-2;⁵ in addition, they rely on public transportation more than the general population and they are more likely to live in overcrowded and multigenerational households.⁶ Once infected with the virus, disadvantaged populations experience an increased risk of hospitalization, intensive care and death.^{7,8} This is partly due to the fact that those with lower SES suffer from higher rates of pre-existing conditions that are associated with increased risks of severe symptoms and deaths from COVID. In addition, individuals without paid sick leave or health insurance may delay testing and seeking treatment, further increasing their risks.⁹ Studies conducted on incomplete data or specific geographic locations have shown that COVID-19 cases and deaths are concentrated in areas characterized by low median household income and high levels of income inequality among other factors.^{10–15} More generally, a growing body of evidence has indicated a higher-than-average level of mortality from COVID-19 in some

communities within the USA.^{16–21} Studies in other countries have also shown a strong relationship between COVID-19 mortality and socio-economic deprivation at the area level.^{22,23} Such a relationship is expected due to the particularly complex interaction mechanisms between the host, environment and SARS-Cov-2.

However, due to limitations in the statistical information currently available, no study on socio-economic differentials in COVID-19 mortality at the community level using data for the complete year 2020 and covering the USA as a whole has been published as yet. This paper is the first to provide a country-wide assessment of the disproportionate vulnerability of economically disadvantaged communities to COVID-19 mortality for the entire year.

Methods

We grouped all US counties into five SES quintiles and calculated three mortality indicators for each SES quintile: the age-standardized death rate from COVID-19, the number of years of life lost (YLL) to the disease and the reductions in life expectancy due to COVID-19. The indicators were calculated for each of the three pandemic waves in 2020, namely 1 March to 31 May, 1 June to 31 August and 1 September to 31 December, and for the year 2020 as a whole, by comparison with mortality for the experience over the corresponding months in an average year in 2015–2019, used as the benchmark.

We followed a three-step approach: first, we grouped all US counties into five categories of similar demographic size based on the socio-economic composition of their

populations; second, we calculated all-cause mortality rates for the three different periods in 2015–2019 (March–May, June–August and September–December) and COVID-19-mortality rates for the same periods in 2020 as well as for 2020 as a whole; third, we constructed complete life tables by time period for each SES quintile. We used the results to assess differences across county groupings in the level of COVID-19 mortality and in the number of YLL to the pandemic.

Grouping counties into SES categories

The grouping of counties into SES categories was implemented following an approach initially developed and validated by Singh.^{24–26} A principal component analysis was applied to 11 variables extracted from the most recent 5-year American Community Survey (2015–2019) from the US Census Bureau²⁷ and calculated at the county level for all US counties. The 11 input variables we used reflect the population composition of each county in terms of the percentage of the population aged ≥ 25 years with < 9 years of education; the percentage with ≥ 4 years of college education; the percentage of households below the federal poverty line; the median household income, including cash benefits such as social-security payments and adjusted to account for differences in local standards of living; the ratio of the average income of the wealthiest quintile of households to the poorest quintile within each county; the unemployment rate; the percentage of the labour force in white-collar occupations; the median housing price; the median gross rent; the percentage of households with no telephone; and the percentage of households with no or incomplete plumbing. After implementing a principal component analysis, we calculated the socio-economic index score as the sum of the product of the correlations between each variable and the first component. Counties were then ranked and classified into quintiles after weighing by their population size. Each SES quintile thus represents $\sim 20\%$ of the total US population.

Recognizing the urban–rural and regional divide in all-cause mortality, we additionally classified the counties within each SES quintile by their respective regions (Northeast, Midwest, South and East) and their metropolitan status (distinguishing between metropolitan and non-metropolitan counties, as defined by the Census Bureau).

Estimating the mortality indicators for each SES quintile and time period

We calculated all three outcome indicators (the age-standardized death rate from COVID, the number of YLL to the disease and reductions in life expectancy for the

population) for all three pandemic waves and for 2020 as a whole and for each SES quintile (and by large US region and metropolitan status). In the absence of complete vital-statistics records by county, month, age and the underlying cause of death for 2020, we implemented an indirect estimation technique to calculate the age-standardized death rates from COVID-19. We combined four sources of data: 2015–2019 average of all-cause deaths by age from the National Center for Health Statistics (NCHS),²⁸ national level weekly COVID-19 mortality data by age from NCHS,²⁹ daily death counts from COVID-19 by county from the Center for Systems Science and Engineering at Johns Hopkins University of Medicine (CSSE)^{30,31} and NCHS exposure counts by age published by the Centers for Disease Control and Prevention (CDC) as bridged-race postcensal estimates.³² The main assumptions on which our method rests are that deaths from COVID-19 were properly reported in the CSSE database; that without COVID-19, all-cause mortality rates (by age and by county) would have been the same in 2020 as in the pre-pandemic 5-year average; and that the age schedule of mortality from COVID-19 has varied over time, but that it was the same for all counties within the same week. We postulated that the granularity of weekly age-specific rates mitigates the overall bias stemming from this approach. All these assumptions are strong but they were necessary for our calculations. We review their implications for our findings in the ‘Discussion’ section. Note that we measured the direct effect of COVID-19 only as we had county-level information on deaths specifically attributed to ICD-10 code U07.1 but not to other causes that might be directly or indirectly related to COVID-19. Detailed explanations, mathematical formalization of all computational steps and all input data sets (comma-separated format) described earlier are provided in the online [Supplementary Material](#) (available as [Supplementary data](#) at *IJE* online). All of the

Table 1 Absolute and relative distribution of COVID-19 deaths by socio-economic status quintile of counties and by period of 2020 in the USA

Period of 2020	SES 1	SES 2	SES 3	SES 4	SES 5	Total
March–May	17 539	12 690	12 773	33 244	28 750	104 996
	5.1%	3.7%	3.7%	9.6%	8.3%	30.3%
June–August	21 000	16 783	15 583	13 535	10 155	77 056
	6.1%	4.8%	4.5%	3.9%	2.9%	22.3%
September–December	50 504	37 797	32 227	26 279	17 732	164 539
	14.6%	10.9%	9.3%	7.6%	5.1%	47.5%
Whole year	89 043	67 270	60 583	73 058	56 637	346 591
	25.7%	19.4%	17.5%	21.1%	16.3%	100%

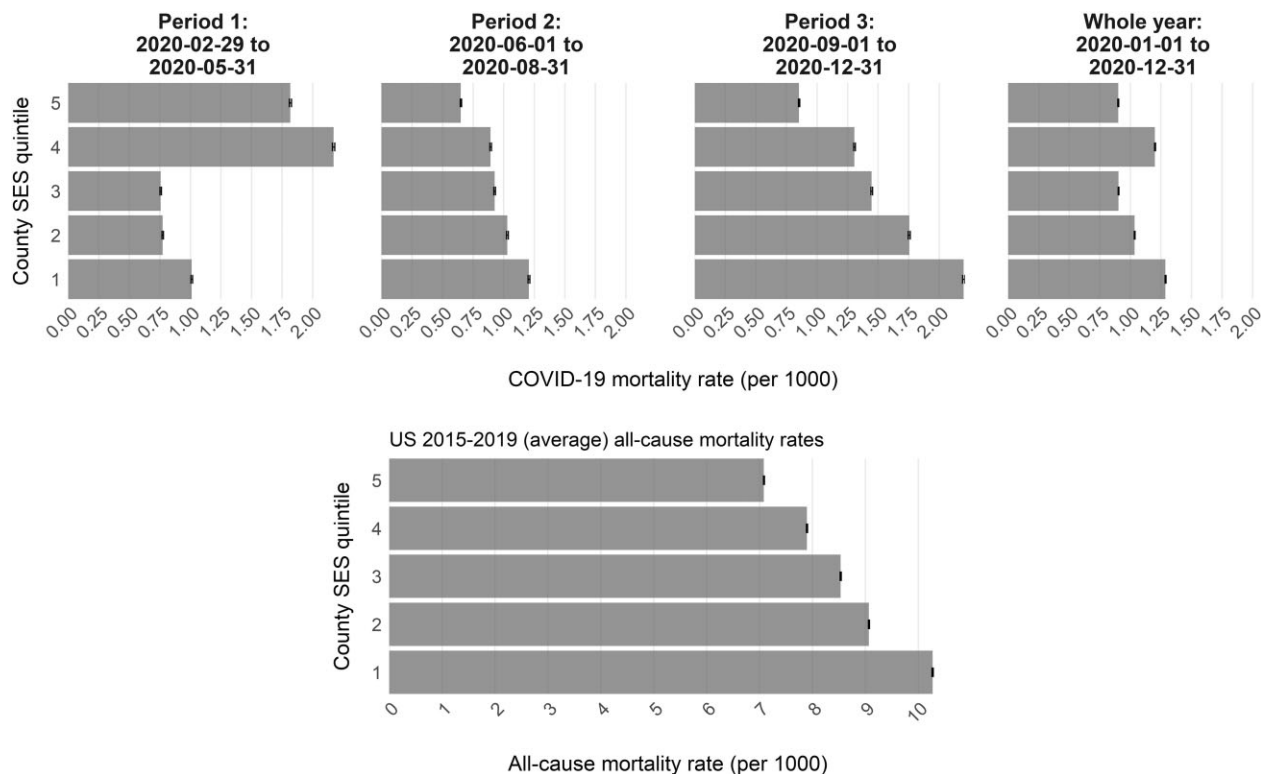


Figure 1 Age-standardized COVID-19 mortality rates (per 1000) by socio-economic status quintile (1 = lowest, 5 = highest) of counties and time period in 2020, and pre-pandemic all-cause mortality rates (per 1000) in the USA, with 95% confidence intervals

analyses were conducted in R, version 4.0.2 (R Project for Statistical Computing).

Results

The Johns Hopkins database included a total of 346 591 US deaths directly attributed to COVID-19 in 2020 as of 22 September 2021 excluding Puerto Rico and the US overseas territories. Of these, as shown in Table 1, 104 996 (or 30.3%) occurred in the months of March–May, 77 056 (22.2%) in June–August and 164 591 (47.5%) in September–December. These numbers correspond to the following death rates from COVID-19: 1.28, 0.95 and 1.53 per 1000, for each successive pandemic wave in 2020. The death rate from COVID-19 for 2020 as a whole was 1.07 per 1000, with COVID-19 mortality representing 12.4% of all-cause mortality.

Figure 1 shows the age-standardized death rates from COVID-19 for each SES quintile of counties by time period for 2020 and from all causes in 2015–2019. During the first pandemic wave (March–May), there was a J-shaped relationship between the SES and COVID-19 mortality of the counties: the highest death rate (2.17 per 1000) was in the second most-affluent (fourth) SES quintile of counties, followed by the most socio-economically advantaged (fifth quintile, 1.82), the least advantaged (first quintile, 1.01) and then the second and third (0.77 and 0.76 per 1000,

respectively). At this stage, COVID-19 mortality represented 20–21% of total mortality in the top two quintiles, 8% in the two middle quintiles and 9% in the lowest quintile during the first wave.

During the following 3 months (June–August), the pattern reversed: COVID-19 mortality was lowest in the top quintile (0.65 per 1000) and highest in the bottom quintile (1.21 per 1000). During the last 4 months of 2020, the summer gradient of COVID-19 mortality by SES quintile became much more pronounced, with the highest rates in the most-disadvantaged quintile (2.20 per 1000, the highest for any quintile over all three waves of the pandemic), gradually declining to 0.85 per 1000 in the most-advantaged. The share of COVID-19 in total mortality during the third wave ranged from 11% in the top quintile to 18% in the bottom quintile. For the whole year, a similar, albeit more muted pattern was apparent, with rates ranging from 1.29 per 1000 in the least-privileged quintile to 0.9 per 1000 in the most-privileged. The pattern in the last 4 months of the year roughly mirrors the SES differentials in all-cause mortality before the pandemic, with a death rate from COVID-19 in the bottom quintile 31% higher than in the top, compared with the 32% higher death rate from all causes between the bottom and top quintiles in the 2015–2019 average.

Between the spring and the autumn of 2020, the number of counties with at least one death attributed to COVID-19

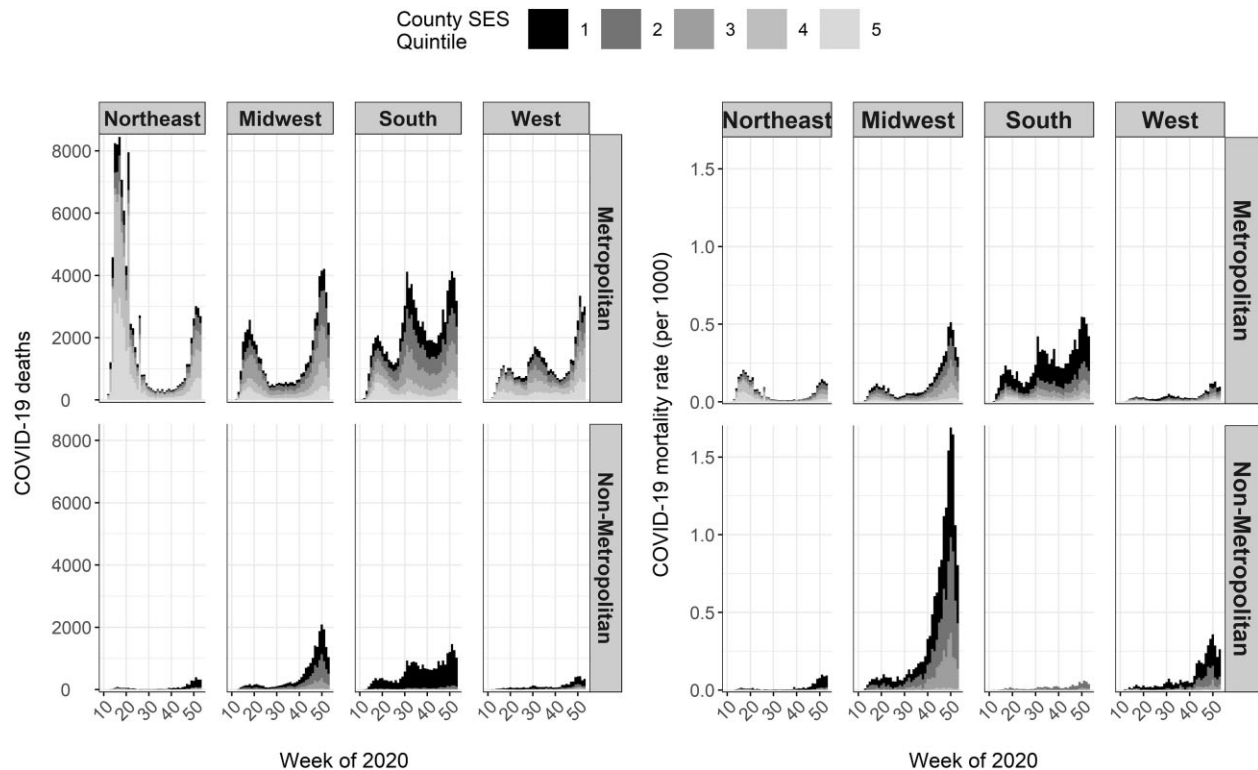


Figure 2 Weekly COVID-19 deaths (left panel) and age-standardized mortality rate (per 1000, right panel) in 2020 in the USA, by socio-economic status quintile of counties (1 = lowest, 5 = highest), US Census geographic region and county metropolitan status

increased progressively in all quintiles, but much more so in the least-affluent than in the most-affluent quintile (Figure 2). By the end of May, the proportion of counties with at least one COVID-19 death had reached 51.7, 48.6, 56.5, 75.7 and 90.7% in each quintile from the least to most affluent. By the end of summer, the proportions had reached 76.3, 63.1, 67.3, 83.3 and 91.3%, and by the end of year, 97.0, 95.0, 94.2, 96.4 and 94.7%.

As shown in Figure 2, during the spring of 2020, deaths were highly concentrated in the metropolitan Northeast, followed by other metropolitan areas in the Midwest and South. Over the summer, the number of COVID-19 deaths declined in the Northeast and Midwest but increased in both urban and rural areas of the South and, to a lower extent, in urban areas of the West. During the fall, COVID-19 deaths increased everywhere but the age-standardized death rates from the virus exhibited a particularly pronounced increase in the rural Midwest. In this region, 80% of all deaths concentrated in the most-disadvantaged counties (those in the first and second SES quintiles). In the urban South, where the pandemic continuously increased in intensity throughout 2020, counties in the two most-deprived quintiles persistently contributed over two-thirds of the age-standardized regional death rate. We present a spatio-temporal representation of this pattern in Supplementary Figure S1 (available as Supplementary data at *IJE* online).

Overall, we estimated that the pandemic resulted in the loss of 5.71 million years of life over the course of 2020, of which 29.6% were lost in March–May, 24.7% in June–August and 45.8% in September–December. Figure 3 represents a decomposition of the number of YLL due to COVID-19 in 2020 by age group, SES quintile of counties and time of year. The results corroborate our earlier findings in that the early wave of the pandemic disproportionately affected the more affluent counties (fourth and fifth quintiles) with 61% of all YLL, a proportion declining to only 33% and 29% in the summer and autumn, respectively. Notably, the population aged 65–74 years contributed the largest number of YLL (1.43 million), representing a quarter of all potential years of life, with the majority lost by those dying in disadvantaged counties, particularly towards the second half of 2020. The gradual decline in YLL among those aged ≥ 75 years was expected due to the naturally decreasing number of years of remaining life and the smaller proportion of survivors.

Lastly, in Table 2, we summarize the differences in life expectancy as well as the number of YLL to COVID-19 for each of the five SES quintiles in the life tables corresponding to each time period. According to our calculations, the total reduction of life expectancy in the US population due to COVID-19 in 2020 was 0.85 years. It ranged from 0.57 to 1.13 years across all quintiles and reflected an increasingly clear gradient of YLL by SES quintile from the first to

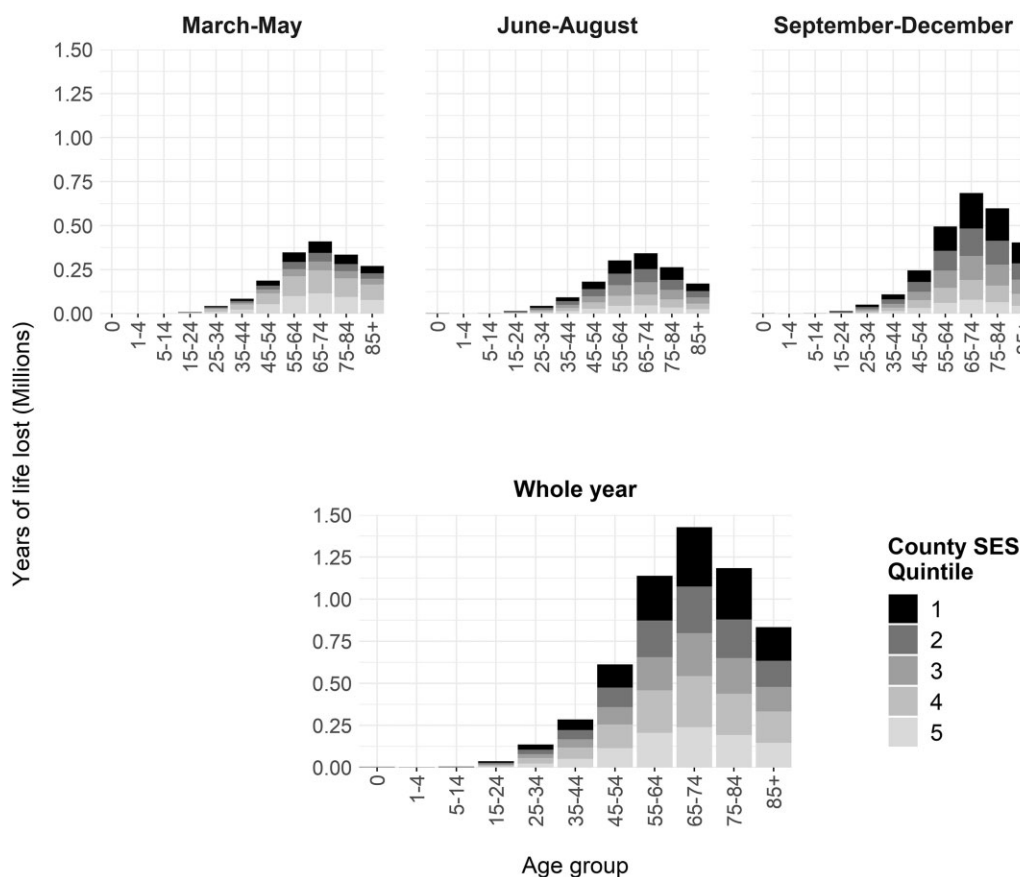


Figure 3 Estimated years of life lost (YLL, in millions) due to COVID-19 in 2020 in the USA, by age group, socio-economic status quintile of counties (1 = lowest, 5 = highest) and time period of 2020

the third study period. The life-expectancy reductions would have reached 1.95 years in the bottom quintile, gradually declining to 0.61 years in the most socio-economically advantaged quintile, if the mortality conditions of the third epidemic wave (September–December) had been experienced continuously throughout 2020.

We provided the full output of our analysis as comma-separated files prefixed as ‘results’ in the online [Supplementary Material](#) (available as [Supplementary data](#) at *IJE* online). These files include the breakdown of Johns Hopkins-reported COVID-19 deaths by county SES grouping and time period, age-standardized mortality rates, complete life tables (including results of a sensitivity analysis for two methods of calculating life-table probability of death) as well as YLL estimates for all county SES groupings and time periods covered in this analysis.

Discussion

We found evidence of increasing variations in COVID-19 mortality across SES groupings of counties throughout the course of the pandemic. This is not to be confused with a

possible relationship between *individual* socio-economic characteristics and COVID-19 mortality. The results we presented are only relevant to the aggregate level at which the study was conducted. These appear to demonstrate a clear socio-economic gradient of COVID-19 mortality across communities in the USA. We recognize the need for more granular, yet encompassing spatial analysis, as proximity and adjacency matter for the spillover and containment effects of an infectious spread. However, our focus in this paper has been on establishing an association between areal socio-economic characteristics and COVID-19 mortality, and measuring its differential impact specifically in terms of deaths and not cases.

The first 3 months of the pandemic marked an unusual shift in mortality patterns in which many of the socio-economically well-off areas were affected the most. Initially the virus had not fully spread across the USA, affecting primarily densely populated, intense-economic-activity, high-transit and cosmopolitan areas. The burden of COVID-19 mortality was thus borne disproportionately by a small number of wealthy urban areas, especially those located in the Connecticut–New Jersey–New York

Table 2 Life expectancy at birth, $e(0)$, in 2015–2019 and standard errors, and difference (in years) in $e(0)$ before and after adding deaths from COVID-19 with 95% confidence interval; simulated years of life lost (YLL in millions over all ages), by time period, for the country as a whole and by socio-economic status quintile of counties in the USA

Life expectancy at birth and years of life lost to COVID-19 by time period	Socio-economic status quintile of counties					Diff. Q1–Q5	National average
	Q1	Q2	Q3	Q4	Q5		
March–May							
$e(0)$ 2015–2019	75.71	77.76	78.53	79.86	81.47	–5.76	78.57
Standard error	0.0454	0.0485	0.0486	0.0519	0.0527	0.0696	0.0227
$e(0)$ 2020 with COVID-19	74.87	77.32	78.17	77.69	79.61	–4.74	77.50
Standard error	0.0445	0.0471	0.0468	0.0501	0.0504	0.0672	0.0218
Diff. [$e(0)$ wo.-w. COVID-19]	–0.84	–0.43	–0.36	–2.18	–1.86	1.02	–1.07
Diff. in $e(0)$: CI upper bound	–0.95	–0.55	–0.47	–2.29	–1.97	1.02	–1.12
Diff. in $e(0)$: CI lower bound	–0.72	–0.33	–0.25	–2.07	–1.76	1.04	–1.02
YLL mean (millions)	0.26	0.20	0.20	0.55	0.48	–0.21	1.69
YLL CI upper bound	0.27	0.20	0.20	0.55	0.48	–0.21	1.69
YLL CI lower bound	0.26	0.20	0.20	0.54	0.48	–0.21	1.69
June–August							
$e(0)$ 2015–2019	76.43	78.46	79.23	80.63	82.32	–5.89	79.32
Standard error	0.0449	0.0479	0.0479	0.0510	0.0516	0.0684	0.0228
$e(0)$ 2020 with COVID-19	75.19	77.53	78.50	79.80	81.88	–6.69	78.47
Standard error	0.0438	0.0462	0.0460	0.0489	0.0488	0.0656	0.0217
Diff. [$e(0)$ wo.-w. COVID-19]	–1.24	–0.93	–0.73	–0.83	–0.43	–0.81	–0.85
Diff. in $e(0)$: CI upper bound	–1.36	–1.04	–0.85	–0.95	–0.55	–0.81	–0.90
Diff. in $e(0)$: CI lower bound	–1.12	–0.81	–0.61	–0.71	–0.31	–0.81	–0.80
YLL mean (millions)	0.36	0.31	0.28	0.26	0.20	0.16	1.42
YLL CI upper bound	0.36	0.31	0.29	0.26	0.20	0.16	1.42
YLL CI lower bound	0.36	0.31	0.28	0.26	0.20	0.16	1.42
September–December							
$e(0)$ 2015–2019	75.97	77.93	78.79	80.04	81.66	–5.69	78.78
Standard error	0.0392	0.0420	0.0421	0.0450	0.0457	0.0602	0.0197
$e(0)$ 2020 with COVID-19	74.01	76.41	77.58	78.85	81.04	–7.03	77.43
Standard error	0.0384	0.0408	0.0406	0.0433	0.0434	0.0579	0.0188
Diff. [$e(0)$ wo.-w. COVID-19]	–1.95	–1.52	–1.20	–1.19	–0.61	–1.34	–1.36
Diff. in $e(0)$: CI upper bound	–2.05	–1.62	–1.30	–1.29	–0.71	–1.34	–1.40
Diff. in $e(0)$: CI lower bound	–1.85	–1.42	–1.11	–1.10	–0.51	–1.34	–1.31
YLL mean (millions)	0.76	0.59	0.52	0.44	0.31	0.45	2.61
YLL CI upper bound	0.76	0.59	0.52	0.44	0.31	0.45	2.61
YLL CI lower bound	0.75	0.59	0.51	0.44	0.31	0.45	2.60
January–December							
$e(0)$ 2015–2019	75.88	77.90	78.72	80.00	81.63	–5.75	78.73
Standard error	0.0232	0.0250	0.0251	0.0270	0.0275	0.0360	0.0114
$e(0)$ 2020 with COVID-19	74.75	77.14	78.15	78.92	80.93	–6.18	77.89
Standard error	0.0225	0.0240	0.0240	0.0256	0.0260	0.0344	0.0109
Diff. [$e(0)$ wo.-w. COVID-19]	–1.13	–0.77	–0.57	–1.08	–0.69	–0.44	–0.85
Diff. in $e(0)$: CI upper bound	–1.19	–0.82	–0.62	–1.14	–0.75	–0.44	–0.87
Diff. in $e(0)$: CI lower bound	–1.08	–0.71	–0.51	–1.02	–0.64	–0.44	–0.82
YLL mean (millions)	1.38	1.10	1.00	1.25	0.99	0.39	5.70
YLL CI upper bound	1.38	1.10	1.00	1.25	0.99	0.39	5.72
YLL CI lower bound	1.38	1.09	1.00	1.25	0.99	0.39	5.70

region.³³ In the following months, the pattern reversed, mirroring the typical gradient of mortality increasing with socio-economic disadvantage. Northeast coastal regions effectively curbed the rampant spread of the epidemic following the devastation of the spring months, possibly thanks to the timely and selective shutdowns and social-distancing policies implemented before and during the summer. By contrast, the dynamic in the Southern states and in the West endured and further increased in magnitude during the last 3 months of 2020. It was accompanied by a dramatic increase in the number of deaths, particularly in the most socio-economically disadvantaged rural areas.

Our overall estimate of the number of YLL is 5.71 million, translating to ~16 years and 170 days not lived by an average person dying of COVID-19. The corresponding reduction in life expectancy at birth due to COVID-19 in the overall US population for 2020 as a whole (0.85 years) is far below the 1.5 years estimate calculated by NCHS from preliminary data.³⁴ This difference can be primarily attributed to the fact that NCHS used 2019 mortality as their benchmark, whereas we used the 2015–2019 average mortality to avoid year-to-year fluctuations in mortality due to influenza and other seasonal diseases. Yet, we know that after plateauing between 2010 and 2014, life expectancy at birth declined in the USA up to 2017, experiencing a rebound in 2018 and 2019. This means that our benchmark is closer to the mortality conditions of 2020 under COVID-19 than that of NCHS, since mortality for the average of 2015–2019 was higher than for 2019. Our earlier estimates (not presented here) comparing the decline from 2019 to 2020 amounted to an overall reduction in life expectancy at birth of 1.38 years—a number comparable to the NCHS estimates and similar multinational studies.³⁵ Regardless, the COVID-19 pandemic in 2020 effectively moved the national average to 77.88 years, 15 years back in time. It also increased the pre-existing US international disadvantage in life expectancy. Indeed, recent research estimated that the USA experienced a larger loss in life expectancy in 2020 than 28 other high-income countries with good-quality data, including Chile and many Eastern European countries.³⁵ The finding confirms the results of an earlier study, which showed that the decrease in the mean length of life in the USA in 2020 was more than eight times larger than the average for 16 peer countries.³⁶ In the comparison countries, the average loss was 0.22 years, which represents about one-third of the loss in the least-affected SES (0.57 years) and one-fifth in the most-affected quintiles (1.14 years), relative to our estimates using the 2015–2019 average. This gap would have been 0.4 years greater for each quintile had we used only 2019 data as our benchmark.

Limitations

The fact that our estimate of the average loss in life expectancy at birth in 2020 is lower than that of NCHS but higher than those in other studies reflects not only differences in the benchmark used to estimate the degree of the mortality excess directly attributable to COVID-19, but also variations in the methodological approaches and sources of data.^{16,37,38} At this time (October 2021), different sources indicate a disparate number of COVID-19 deaths. In addition, it is unclear in many cases whether deaths are classified by county of residence or by county of occurrence, including for the CSSE data set that we used. We assumed the latter in our analysis, but if the former were in fact the case, it would affect our results negatively only if a high enough number of people died in a county classified in a different SES quintile than the county in which they resided. However, since it is more likely that more affluent counties are better equipped with tertiary care units (on the front line for the treatment of severe COVID-19 cases), using deaths classified by place of occurrence instead of place of residence would lead to an overcounting of COVID-19 deaths in those counties and undercounting in less-affluent counties. Further, CSSE reported close to 40 000 fewer COVID-19 deaths in 2020 than the CDC preliminary data for the corresponding geographic aggregate. This could be due to differences in source or definition, in particular regarding the inclusion of deaths of ‘probable’ or ‘presumed’ COVID-19 by NCHS. Misclassification (COVID-19 deaths mistakenly attributed to other causes) and under-reporting are likely to vary from place to place as suggested by NCHS.³⁴ Prior research suggested that 30–40% of the excess deaths might have been correctly or incorrectly attributed to causes other than COVID-19.^{39,40} However, preliminary research also suggested that such effects were disproportionately borne by the most vulnerable communities.^{15,41} Taking into account the overall mortality excess of 2020 (directly and indirectly attributable to the pandemic) could thus have only magnified our results.

Furthermore, our analysis likely underestimated the socio-economic gradient in COVID-19 mortality due to the assumption of a uniform age schedule of COVID-19 mortality (age pattern, similar across all counties, though varying from one week to the next). There is some indication of the greater susceptibility of underprivileged populations to COVID-19 at younger ages.¹⁶ Since deaths at younger ages result in more YLL, this would also further strengthen our findings on the differential impact of the pandemic on the expectation of life at birth by SES category. On the other hand, many of the people who died of COVID-19 had higher mortality risks than the general

population even before the pandemic because of a higher prevalence of co-morbidities (obesity, diabetes and hypertension in particular), which are expected to be especially prevalent within the most-deprived segments of the population. Had we been able to take these increased risks into account rather than assuming that, within each SES grouping, everyone not dying of COVID-19 had the same chance of dying in 2020, differences in the number of YLL across SES quintiles would have been lower than we found.

The use of the 2015–2019 American Community Survey data to construct our SES quintiles ignores any change incurred by the economic consequences of the pandemic on the distribution of the counties within each group in 2020 in terms of the increased poverty, inequality and selective migration associated with shelter-in-place policies and travel restrictions. This, however, is unlikely to have affected our findings if these changes affected counties proportionately to their degree of deprivation (i.e. the more deprived a county, the more severe the negative economic impact of the pandemic). Indeed, this would result in an increase in the range of the county-level socio-economic scores but not the county ranking per se, and thus would not fundamentally modify their classification into the five socio-economic groupings and the subsequent mortality estimates for each grouping.

A full spatial-dependence component is missing from our analysis. As infectious diseases spread, proximity and adjacency undoubtedly play a role in dissemination of the virus. We conducted some exploratory analysis of spatial dependency (see [Supplementary Material](#), available as [Supplementary data](#) at *IJE* online). However, it is important to note the limitations that we faced and the fact that several considerations would need to be addressed in our specific case to make such an analysis meaningful. First, counties are fairly large, heterogenous and variable-sized administrative units, where borders do not take into an account the proximity, access, availability and utilization of various resources, nor specific individual behaviours. Indeed, coastal regions in the USA, despite being 2500 miles apart, are often much more similar socio-economically and better connected through modern transportation means than nearby suburban and rural counties. The amount of data and nuance required to carry out such an analysis merits a separate paper. With our limited data, we showed (see [Supplementary Material](#), available as [Supplementary data](#) at *IJE* online) that overall, there is a weak spatial correlation between the number of deaths and the rate of mortality from COVID-19 in one county in one time period and its neighbours in a previous period. Although a number of isolated clusters were present, there is not enough information to establish the true extent, much less any causal links, of spatial diffusion of the COVID-19 disease.

Overall, the data limitations and study biases lean in opposite directions as some understated and others overstated the size of the differences in the mortality risks from COVID-19 by SES category of counties. More precise estimates will be possible when complete and detailed data become available but we remain confident that our current findings of the emergence of a strong mortality gradient by SES category of counties will hold. Furthermore, current indications of a similar gradient in vaccination rates suggest that the observed inequalities in COVID-19 mortality for 2020 will have increased in 2021. To better control the pandemic and mitigate its impact on the general US population, current policies to strengthen testing, increase vaccination efforts or invest in public health infrastructure must proactively target the most socio-economically disadvantaged areas, which bear the heaviest mortality burden from the pandemic.

Ethics approval

All the secondary data used in the present study have been retrieved from publicly accessible sources, with the exception of the anonymized NCHS restricted-use mortality files. Although restricted, access to and subsequent analysis of the data from the latter source do not involve participation of or interaction with human subjects. These data were at no point publicly disclosed in the raw form. Given these conditions, our study does not fall under the definition of human-subjects research, as per advice of the local ethics review entity [Berkeley Committee for the Protection of Human Subjects (OPHS)]. As a consequence, this study does not require ethics approval.

Author contributions

Both authors collaborated on all aspects of the paper, although M.B. developed the initial concept and research design for the study whereas D.D. contributed the most to developing specific methods as well as extracting and processing the input data. Both worked together on analysing and interpreting the results. D.D. wrote the preliminary draft of the article and M.B. substantially contributed to revisions. Both authors carefully reviewed and approved the final version of the manuscript.

Data availability

The input data underlying this article are available in the article and as comma-separated files in its online [Supplementary Material](#), available as [Supplementary data](#) at *IJE* online. The input data as well as the computer code to reproduce the results have been also made available on GitHub: https://github.com/denysdukhovnov/COVID_disparities_USA.git.

Supplementary data

[Supplementary data](#) are available at *IJE* online.

Funding

This work was supported by a grant from the National Institutes of Health [grant number P30AG012839] via the University of California, Berkeley Center for the Economics and Demography of Aging and by direct awards to Magali Barbieri by the National Institute on Aging [grant number R03AG058110] and by the Society of Actuaries. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Acknowledgements

We thank the anonymous reviewers of this Journal as well as Ronald Lee and Dana Gleit for their comments and suggestions on earlier versions of this manuscript. We are also grateful to our colleagues at the University of California, Berkeley, Celeste Winant and Carl Boe, for their help in developing and implementing the computer code for the analysis and to Timothy Riffe at Universidad del País Vasco, Spain, for a stimulating discussion on estimation methods.

Conflict of interest

None declared.

References

- Fothergill A, Peek LA. Poverty and disasters in the United States: a review of recent sociological findings. *Nat Hazards* 2004;32: 89–110.
- Bolin B, Kurtz LC. Race, class, ethnicity, and disaster vulnerability. In: Rodríguez H, Donner W, Trainor J (eds). *Handbook of Disaster Research*. 2018, pp. 181–203.
- Dash N. Race and ethnicity. In: Thomas DSK, Phillips BD, Lovekamp WE, Fothergill A (eds). *Social Vulnerability to Disasters*, 2nd edn. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2013, pp. 113–28.
- Laska S, Morrow BH. Social vulnerabilities and Hurricane Katrina: an unnatural disaster in New Orleans. *Mar Technol Soc J* 2006;40:16–26.
- Rogers TN, Rogers CR, VanSant-Webb E, Gu LY, Yan B, Qeadan F. Racial disparities in COVID-19 mortality among essential workers in the United States. *World Med Health Policy* 2020;12:311–27.
- Rollston R, Galea S. COVID-19 and the social determinants of health. *Am J Health Promot* 2020;34:687–89.
- Alcendor DJ. Racial disparities-associated COVID-19 mortality among minority populations in the US. *J Clin Med* 2020;9:2442.
- Ravi K. Ethnic disparities in COVID-19 mortality: are comorbidities to blame? In: *The COVID-19 Reader*. New York (US): Routledge, 2020, pp. 105–06.
- Khalatbari-Soltani S, Cumming RC, Delpierre C, Kelly-Irving M. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J Epidemiol Community Health* 2020;74:620–23.
- Krieger N, Waterman PD, Chen JT. COVID-19 and overall mortality inequities in the surge in death rates by zip code characteristics: Massachusetts, January 1 to May 19, 2020. *Am J Public Health* 2020;110:1850–52.
- Chen JT, Krieger N. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county versus zip code analyses. *J Public Health Manag Pract* 2021;27:S43–56.
- Clouston SA, Natale G, Link BG. Socio-economic inequalities in the spread of coronavirus-19 in the United States: a examination of the emergence of social inequalities. *Soc Sci Med* 2021;268: 113554.
- Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ* 2020;728:138884.
- Seligman B, Ferranna M, Bloom DE. Social determinants of mortality from COVID-19: a simulation study using NHANES. *PLoS Med* 2021;18:e1003490.
- Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: a county-level analysis. *PLoS Med* 2021;18:e1003571.
- Andrasfay T, Goldman N. Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. *Proc Natl Acad Sci USA* 2021; 118:e2014746118.
- Cheng KJ, Sun Y, Monnat SM. COVID-19 death rates are higher in rural counties with larger shares of Blacks and Hispanics. *J Rural Health* 2020;36:602–08.
- Cordes J, Castro MC. Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spat Spatiotemporal Epidemiol* 2020;34:100355.
- Fielding-Miller RK, Sundaram ME, Brouwer K. Social determinants of COVID-19 mortality at the county level. *PLoS One* 2020;15:e0240151.
- Kim SJ, Bostwick W. Covid19 social vulnerability and racial inequality in COVID-19 deaths in Chicago. *Health Educ Behav* 2020;47:509–13.
- Pro G, Hubach R, Wheeler D *et al*. Differences in US COVID-19 case rates and case fatality rates across the urban-rural continuum. *Rural Remote Health* 2020;20:6074.
- Caul S. *Deaths Involving COVID-19 by Local Area and Socioeconomic Deprivation: Deaths Occurring Between 1 March and 31 July 2020*. [Internet] London (UK): UK Office for National Statistics, 2020. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deaths-involving-covid-19-by-local-area-and-deprivation/deaths-occurring-between-1-march-and-31-july-2020> (22 September 2021, date last accessed).
- de Souza CD, do Carmo RF, Machado MF. The burden of COVID-19 in Brazil is greater in areas with high social deprivation. *J Travel Med* 2020;27:taaa145.
- Singh GK, Siahpush M. Widening socioeconomic inequalities in US life expectancy, 1980–2000. *Int J Epidemiol* 2006;35: 969–79.
- Singh GK, Miller BA, Hankey BF, Feuer EJ, Pickle LW. Changing area socioeconomic patterns in US cancer mortality, 1950–1998: Part I—all cancers among men. *J Natl Cancer Inst* 2002;94:904–15.
- Singh GK, Miller BA, Hankey BF. Changing area socioeconomic patterns in US cancer mortality, 1950–1998: Part II—lung and colorectal cancers. *J Natl Cancer Inst* 2002;94:916–25.

27. U.S. Census Bureau. *Data from: 2015–2019 American Community Survey 5-year Public Use Microdata Samples*. 2020. <https://www.census.gov/programs-surveys/acs/data.html> (5 January 2021, date last accessed).
28. National Center for Health Statistics. *Data from: Restricted-use Detailed Mortality File 2015–2019*. 2020. *Compiled from Data Provided by the 57 Vital Statistics Jurisdictions Through the Vital Statistics Cooperative Program* (7 January 2021, date last accessed).
29. National Center for Health Statistics. Center for Disease Control. *Data from: Provisional Death Counts for Coronavirus Disease 2019 (COVID-19). Weekly Updates by Select Demographic and Geographic Characteristics*. <https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-W/vsak-wrfu> (22 September 2021, date last accessed).
30. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis* 2020;**20**:533–34.
31. Johns Hopkins Center for Systems Science and Engineering. *Johns Hopkins University 2020. Data from: Covid-19 Data Repository*. <https://github.com/CSSEGISandData/COVID-19> (22 September 2021, date last accessed).
32. National Center for Health Statistics. *Data from: Vintage 2020 Postcensal Estimates of the Resident Population of the United States (April 1, 2010, July 1, 2010–July 1, 2020), by Year, County, Single-year of Age (0, 1, 2, ..., 85 years and over), Bridged Race, Hispanic Origin, and Sex. Prepared Under a Collaborative Arrangement with the U.S. Census Bureau*. http://www.cdc.gov/nchs/nvss/bridged_race.htm (22 September 2021, date last accessed).
33. Woolf SH, Chapman DA, Sabo RT, Zimmerman EB. Excess deaths from COVID-19 and other causes in the US, March 1, 2020, to January 2, 2021. *JAMA* 2021;**325**:1786–89.
34. Arias E, Tejada-Vera B, Ahmad F, Kochanek KD. *Provisional Life Expectancy Estimates for 2020. Vital Statistics Rapid Release; No 15*. Hyattsville, MD: National Center for Health Statistics, 2021.
35. Aburto JM, Schöley J, Kashnitsky I *et al*. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int J Epidemiol* 2021;**51**:63–74.
36. Woolf SH, Masters RK, Aron LY. Effect of the covid-19 pandemic in 2020 on life expectancy across populations in the USA and other high income countries: simulations of provisional mortality data. *BMJ* 2021;**373**:n1343.
37. Goldstein JR, Lee RD. Demographic perspectives on the mortality of COVID-19 and other pandemics. *Proc Natl Acad Sci USA* 2020;**117**:22035–41.
38. Heuveline P, Tzen M. Beyond deaths per capita: comparative COVID-19 mortality indicators. *BMJ Open* 2021;**11**:e042934.
39. Faust JS, Krumholz HM, Du C *et al*. All-cause excess mortality and COVID-19–related mortality among US adults aged 25–44 years. *JAMA* 2021;**325**:785–87.
40. Kiang MV, Irizarry RA, Buckee CO, Balsari S. Every body counts: measuring mortality from the COVID-19 pandemic. *Ann Intern Med* 2020;**173**:1004–07.
41. Chen YH, Glymour MM, Catalano R *et al*. Excess mortality in California during the coronavirus disease 2019 pandemic, March to August 2020. *JAMA Intern Med* 2021;**181**:705–07.