Excess all-cause mortality across counties in the United States, March 2020 to December 2021

Eugenio Paglino¹, Dielle J. Lundberg², Ahyoung Cho³, Joe A. Wasserman⁴, Rafeya Raquib², Anneliese N. Luck¹, Katherine Hempstead³, Jacob Bor², Irma T. Elo¹, Samuel H. Preston¹, Andrew C. Stokes²

¹Department of Sociology and Population Studies Center, University of Pennsylvania, Philadelphia, PA
²Department of Global Health, Boston University School of Public Health, Boston, MA
³Department of Political Science, Boston University, Boston, MA
⁴RTI International, Research Triangle Park, NC
⁵Robert Wood Johnson Foundation, Princeton, NJ

ABSTRACT

Official Covid-19 death counts have underestimated the mortality impact of the Covid-19 pandemic in the United States. Excess mortality, which compares observed deaths to deaths expected in the absence of the pandemic, is a useful measure for assessing the total effect of the pandemic on mortality levels. In the present study, we produce county-level estimates of excess mortality for 3,127 counties between March 2020 and December 2021. We fit two hierarchical linear models to county-level death rates from January 2015 to December 2019 and predict expected deaths for each month during the pandemic. We compare observed deaths to these estimates to obtain excess deaths for each county-month. An estimated 936,911 excess deaths occurred during 2020 and 2021, of which 171,168 (18.3%) were not assigned to Covid-19 on death certificates as an underlying cause of death. Urban counties in the Far West, Great Lakes, Mideast, and New England experienced a substantial mortality disadvantage in 2020, whereas rural counties in these regions had higher mortality in 2021. In the Southeast, Southwest, Rocky Mountain, and Plains regions, there was a rural mortality disadvantage in 2020, which was exacerbated in 2021. The proportion of excess deaths assigned to Covid-19 was lower in 2020 (76.3%) than in 2021 (87.0%), suggesting that a larger fraction of excess deaths was assigned to Covid-19 later in the pandemic. However, in rural areas and in the Southeast and Southwest a large share of excess deaths was still not assigned to Covid-19 during 2021.

KEYWORDS: Covid-19, Excess mortality, Geographic inequalities, Rural mortality, Counties

CORRESPONDING AUTHOR: Andrew C. Stokes, 801 Massachusetts Avenue, Crosstown Building 362, Boston, MA, 02118, 617-358-2444, acstokes@bu.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
SIGNIFICANCE

Deaths during the Covid-19 pandemic have been primarily monitored through death certificates containing reference to Covid-19. This approach has missed more than 170,000 deaths related to the pandemic between 2020 and 2021. While the ascertainment of Covid-19 deaths improved during 2021, the full effects of the pandemic still remained obscured in some regions. County-level estimates of excess mortality are useful for studying geographic inequities in the mortality burden associated with the pandemic and identifying specific regions where the full mortality burden was significantly underreported (i.e. Southeast). They can also be used to inform resource allocation decisions at the federal and state levels and encourage uptake of preventive measures in communities with low vaccine uptake.
Introduction

The Covid-19 pandemic has had a substantial impact on mortality in the United States, leading to declines in life expectancy not previously observed since the end of World War II (1, 2). Estimates of excess mortality, which compare observed deaths to those expected in the absence of the pandemic, suggest that the true death toll of the pandemic is much larger than indicated by the official Covid-19 death tallies (3–7). For example, one study estimated that 550,000 excess deaths occurred between March 2020 and February 2021 and that approximately one quarter of these excess deaths were assigned to causes other than Covid-19 (8).

Excess deaths may have been assigned to causes other than Covid-19 for several reasons. Lack of access to testing in the community, combined with the inconsistent use of post-mortem testing for suspected cases, likely resulted in a large share of undiagnosed Covid-19 infections and deaths, especially early in the pandemic and in rural areas (9–12). Additionally, persons with comorbid conditions, such as cardiovascular disease, diabetes, hypertension, and pulmonary disease, may have had their cause of death assigned to the comorbid condition rather than to Covid-19 (13). Unattended Covid-19 deaths occurring in the community may have been especially likely to be assigned to another cause of death (14). Individuals who developed cardiovascular disease or diabetes as a result of the post-acute sequelae of Covid-19 and subsequently died also may not have had their deaths assigned to Covid-19 (15, 16). Finally, excess deaths not assigned to Covid-19 may also reflect deaths indirectly related to the pandemic, including deaths associated with reductions in access to health care, hospital avoidance due to fear of Covid-19 infection, increases in drug overdoses, and economic hardship leading to housing and food insecurity (17–23).
There are multiple benefits to using excess mortality rather than assigned Covid-19 deaths to assess the mortality impact of the Covid-19 pandemic. First, estimates of excess mortality are more comparable spatially than Covid-19-assigned mortality, because states use different definitions to assign Covid-19 deaths and local death investigation systems may have different policies and resources that affect assignment of Covid-19 deaths. Second, as many Covid-19 deaths go uncounted, excess mortality is likely to provide a more accurate measure of the pandemic impact for purposes of resource allocation. Thus, continued tracking of excess mortality across time and space is vital to clarifying the total impact of the pandemic and where its impact is greatest, and to identifying the most appropriate policy responses and interventions.

Prior studies of excess mortality have primarily focused on national and state-level estimates (5, 6), but revealing the true mortality impact of the Covid-19 pandemic at the county-level is especially valuable for several reasons. First, because counties are the administrative unit for death investigation, excess mortality estimates have the potential to help identify counties with substantial Covid-19 under-counts that would benefit from additional training in cause-of-death certification (25). Such estimates may also be valuable for informing local public health workers, community leaders, and residents of the true impact of the pandemic, thus potentially increasing vaccination and uptake of other protective measures (26). These estimates may also be used to target federal and state emergency resources, such as funeral assistance support from the Federal Emergency Management Agency (FEMA). Finally, estimating excess mortality at the county-level also enables analyses of the factors affecting mortality associated with the pandemic, including geographic dimensions like the urban/rural continuum.

---

1 Recent guidance from the Council of State and Territorial Epidemiologists (CTSE) provides states with a standardized definition for Covid-19-associated deaths. Prior to the release of this guidance on December 22, 2021, there was no standard definition for reporting of Covid-19-associated deaths among Covid-19 cases and state procedures have varied widely (24).
One prior study generated predictions of excess mortality at the county level but was limited to data from 2020 and pooled small counties together to increase precision (4). Expanding these estimates to 2021 is critical because the geography of the pandemic has likely changed markedly since 2020 as a result of policy shifts, the availability of vaccines, and the emergence of new variants. In the present study, we developed two hierarchical models to estimate excess mortality for 3,127 harmonized counties for the period from March 2020 to December 2021. We also evaluated the extent to which excess deaths are accounted for in official Covid-19 death tallies as an indicator of potential under-reporting of Covid-19 mortality. In addition to generating county-level estimates of excess mortality, we produced several aggregations of the county estimates to examine differences in excess mortality by U.S. Bureau of Economic Analysis (BEA) region and across metropolitan (metro) vs. non-metropolitan (nonmetro) areas.

**Results**

Across the United States, 459,764 excess deaths occurred between March and December 2020, and 477,147 excess deaths occurred in 2021. This equals 936,911 excess deaths during calendar years 2020 and 2021, of which 765,743 (81.7%) were assigned to Covid-19 as an underlying cause of death and 171,168 (18.3%) were not assigned to Covid-19. In 2020, excess death rates were highest in nonmetro areas (207 deaths per 100,000 residents) followed by large central metro areas (173 deaths per 100,000 residents), whereas in 2021, excess death rates were highest in nonmetro areas (227 deaths per 100,000 residents) followed by small or medium metros (163 deaths per 100,000 residents). Regionally, the excess death rate in 2020 was highest in the Mideast (206 deaths per 100,000 residents) followed by the Southwest (188 deaths per 100,000 residents).

---

2 To reconcile difference in FIPS code across all of our data sources we tracked county FIPS codes back to their 1990 values reversing the changes operated by the Census Bureau over time as detailed here: [https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2020.html](https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2020.html)
residents). In 2021, however, the excess death rate was highest in the Southeast (205 deaths per 100,000 residents) and the Southwest (198 deaths per 100,000 residents). The areas with the highest excess death rates in 2020 were large central metros in the Mideast and nonmetro areas in the Southwest and Southeast. In 2021, the areas with the highest excess death rates were nonmetro areas in the Southwest, nonmetro areas in the Southeast, and small or medium metros in the Southwest and Southeast (Table 1). Supplemental Table S1 provides estimates of Covid-19 and excess mortality rates for each state in 2020 and 2021. Excess death rates were highest in Mississippi (301 deaths per 100,000 residents) followed by Arizona (246 deaths per 100,000 residents) in 2020 and in West Virginia (298 deaths per 100,000 residents) followed by Mississippi (271 deaths per 100,000 residents) in 2021.

Figure 1 shows excess death rates across all counties in the United States. Between 2020 and 2021, excess deaths shifted from the Northeast and Midwest to the South and to the West and from metro to nonmetro areas. In 2020, excess mortality was higher in metro counties in the Far West, Great Lakes, Mideast, and New England (Table 1). In 2021, however, excess mortality was higher in nonmetro areas than metro areas in these regions. In the Southeast, Southwest, Rocky Mountain, and Plains regions, excess mortality was higher in nonmetro areas in both 2020 and 2021. Nationally, excess mortality was higher in nonmetro areas than metro areas, and the disparity between nonmetro and metro areas was greater in 2021.

Supplemental Figure S1 shows actual and expected death rates for the U.S. by month during 2020 and 2021. Three peaks in mortality are apparent: (1) early 2020, (2) end of 2020 / start of 2021, and (3) end of 2021. Figure 2 breaks down trends in excess death rates within each BEA region by month throughout the period. Excess death rates peaked in the Mideast in early 2020, primarily in large metro areas. Excess death rates also increased markedly in New England.
and the Great Lakes during this time. Around the end of 2020, a second peak resulted in high excess death rates in the Far West, Great Lakes, Southwest, and Southeast regions. A third peak was observed in most of the regions near the end of 2021. Supplemental Figure S2 shows actual and expected death rates for the largest county in each BEA region - metro combination by month during 2020 and 2021. Some counties only experienced one peak in mortality (e.g. Kings County, New York), whereas others experienced three distinct peaks (e.g. Navajo County, Arizona).

In absolute terms, the Southeast was the region with the most excess deaths in both 2020 and 2021 followed by the Mideast and Great Lakes in 2020 and by the Southwest in 2021 (Figure 3). While New York was the state with the most excess deaths in 2020, Texas, California, and Florida had the most excess deaths in 2021. Excess mortality was also less concentrated in large metro areas and large fringe areas in 2021 than it was in 2020. Supplemental Figure S3 provides a comparison of states and regions in terms of excess death rates, rather than absolute counts.

Figure 4 plots the proportion of excess deaths assigned to Covid-19 across counties in the United States. In both 2020 and 2021, counties across the country reported substantial numbers of excess deaths not assigned to Covid-19 as an underlying cause of death. In total, 76.3% of excess deaths were assigned to Covid-19 in 2020, whereas in 2021, 87.0% of excess deaths were assigned to Covid-19. This equals 109,069 excess deaths that were not assigned to Covid-19 in 2020 and 62,099 excess deaths that were not assigned to Covid-19 in 2021, for a total of 171,168 deaths. Despite the increase in assignment of Covid-19 deaths from 2020 to 2021, many regions still had areas with a low proportion of excess deaths assigned to Covid-19 during 2021, such as
nonmetro areas in the Far West (69.5%) and Southeast (71.4%). The Southeast and Southwest were the regions with the lowest overall assignment of excess deaths to Covid-19 in 2021.

In 2020, assignment of excess deaths to Covid-19 was similar in large central metro areas (76.3%) compared to nonmetro areas (74.9%), whereas in 2021, assignment was lower in nonmetro areas (80.0%) than large central metro areas (89.1%). Despite this overall trend, several regions (Mideast and New England) had lower assignment in nonmetro areas in 2020 than in metro areas. This suggests that assignment of excess deaths to Covid-19 in nonmetro areas was low in many regions across the study period. In 2021, assignment of excess deaths to Covid-19 was particularly low in nonmetro areas in the Southeast. Supplemental Figure S4 identifies the states with the lowest assignment of excess deaths to Covid-19. Many of these states were in the Southeast and included South Carolina, Louisiana, Mississippi, and North Carolina. In contrast, several states in New England including Massachusetts, New Jersey, Connecticut, and Rhode Island reported more Covid-19 deaths than excess deaths. Overall, few states reported more deaths assigned to Covid-19 than our estimates of excess deaths (5 states in 2020 and 14 states in 2021).

Supplemental Figure S5 plots excess death rates for each county against their Covid-19 death rates. Counties above the 45-degree reference line represent areas where the excess death rate was higher than the Covid-19 death rate, indicating there were excess deaths not assigned to Covid-19 in these counties. In 2020, the majority of counties in all regions except New England were above the 45-degree line, demonstrating that there were excess deaths not assigned to Covid-19 in these areas. Excess deaths not assigned to Covid-19 occurred in counties across the metro-nonmetro continuum. In 2021, the majority of counties in the Southeast, Southwest, Far West, and Rocky Mountains regions remained above the 45-degree line, demonstrating that they
still had excess deaths not assigned to Covid-19. In particular, the counties with the highest proportion of excess deaths not assigned to Covid-19 tended to be nonmetro counties. Most counties in the Mideast, Plains, and Great Lakes, however, no longer had excess deaths not assigned to Covid-19 in 2021.

Discussion

In the present study, we produced estimates of excess deaths associated with the Covid-19 pandemic from March 2020 to December 2021 across 3,127 harmonized counties in the United States. Our study found that nearly 940,000 excess deaths occurred in the U.S. between 2020 and 2021, of which more than 170,000 were not assigned to Covid-19 on death certificates. This indicates that excess deaths were 22% higher than deaths assigned to Covid-19 during this period.

Prior studies of excess mortality have largely produced estimates for the year 2020 (3–6, 27), leaving patterns of excess mortality during 2021 under-studied. The Center for Disease Control and Prevention (CDC) however has reported a provisional estimate of approximately 944,000 excess deaths in the U.S. from March 2020 to December 2021, which is very close to our estimate (28). Woolf et al. identified 522,368 excess deaths from March 1, 2020 to January 2, 2021, which is higher than our estimate of approximately 470,000 deaths for the year 2020 (6). Islam et al. reported 458,000 excess deaths during 2020, which is close to our estimate (29). A prior estimate by Stokes et al. found 438,386 excess deaths in 2020, which is lower than our estimate due to differences in methods and time horizons for predicting expected deaths (4).

There are multiple potential advantages to using county-level data to generate estimates of excess mortality at the state and national levels. Predicting expected mortality at the state or
national levels assumes that all areas within the state or nation have the same background trends of mortality. However, in actuality, different regions and metro and nonmetro counties of the U.S. have experienced varied long-term mortality trends (8, 30, 31). Our approach may produce more reliable estimates of expected mortality in the absence of the pandemic, because the projections build on historical trends at the county-level.

Another key advantage of producing excess mortality estimates at the county-level is the opportunity to examine differences across the urban-rural continuum, which is not possible with state-level data. By exploring how excess mortality varies by region and time across the urban-rural continuum, we gained several novel insights about patterns of mortality and the assignment of Covid-19 as an underlying cause on death certificates during the pandemic.

One major finding of this study is that there were similar numbers of excess deaths in 2020 and 2021, which is noteworthy as vaccinations were available for much of 2021. Despite the strong efficacy of vaccines, gaps in uptake likely contributed to high excess mortality in 2021, which may persist into the future if these vaccination gaps are not closed. This finding may also reflect federal and state governments’ decision to prioritize individual-level interventions over population-based strategies designed to protect the communities at greatest risk for Covid-19 death, such as financial support for family and medical leave, improved ventilation of schools and workplaces, and vaccine delivery programs organized in coordination with community partners (32).

We found substantial variation in excess mortality by US region and across the urban-rural continuum, which differed between 2020 and 2021. In the Far West, Great Lakes, Mideast, and New England, there was a substantial urban mortality disadvantage in 2020, which was reversed in 2021 to yield a rural mortality disadvantage. In the Southeast, Southwest, Rocky Mountain,
and Plains regions, there was a rural mortality disadvantage in 2020, which was exacerbated in 2021. This suggests that the pandemic has impacted rural areas heavily, especially in 2021, suggesting a need for increased preventative measures in these areas where vaccination remains low (33).

Another finding of this study is that excess death rates exceeded Covid-19 death rates across most counties in all BEA regions during 2020 except New England. This indicates that there were excess deaths not assigned to Covid-19 reported in these regions during 2020. Between 2020 and 2021, Covid-19 and excess mortality then converged in some regions of the country (e.g. Mideast and Great Lakes). In other regions (e.g. Southeast and Southwest), excess death rates continued to exceed Covid-19 death rates in most counties, indicating that excess deaths not assigned to Covid-19 persisted. This finding may have several explanations. In the Mideast, the gap between Covid-19 and excess mortality in 2020 likely reflected the fact that the pandemic affected this region early in the pandemic when access to testing was extremely limited and the clinical manifestations of Covid-19 were unclear (27). The elimination of this gap by 2021 suggests that as the pandemic progressed, the Mideast was able to more effectively capture Covid-19 deaths. It is also possible that some indirect effects of the pandemic, such as disruptions in health care access, also decreased during this time. In other regions such as the Southeast and Southwest, the gaps between Covid-19 and excess mortality persisted into 2021, which may relate to the continued lack of Covid-19 testing in many Southeastern and Southwestern counties even as the pandemic progressed (34–36).

Excess deaths exceeded deaths assigned to Covid-19 in rural areas across many regions, especially in 2021. Many counties in the Southeast and Southwest also had low levels of assignment of excess deaths to Covid-19 throughout the pandemic. Rural counties may have
under-counted Covid-19 deaths throughout the pandemic as a result of the exceptionally high death rates many rural areas faced during the Winter surge of 2020-2021 and the subsequent Delta surge in Summer/Fall 2021 (37–41). Other contributing factors may have included under-resourced health care systems that were unable to care for patients with Covid-19 and/or other non-Covid-19 conditions, under-resourced death investigation systems in which medical examiners or coroners did not pursue post-mortem Covid-19 testing, and partisan beliefs regarding the Covid-19 pandemic that may have influenced cause of death assignment and the likelihood of testing (19, 25).

Although our study does not distinguish between uncounted Covid-19 deaths and deaths indirectly related to the pandemic, emerging literature suggests that a large share of the excess deaths not assigned to Covid-19 likely represent uncounted Covid-19 deaths (10, 11, 42). For example, one recent study by Lee et al. found that approximately 90% of excess mortality between March 2020 and April 2021 could be attributed to the direct effects of SARS-CoV-2 infection (42). This possibility is also supported by investigative reporting during the pandemic which has documented widespread irregularities in cause of death assignment resulting from over-burdened and under-resourced death investigation systems (43). Discrepancies between Covid-19 death rates and excess death rates are problematic because they have the potential to mislead scientists and policymakers about which areas were most heavily affected during the pandemic. Failure to accurately capture Covid-19 deaths also points to an urgent need to modernize the death investigation system in the United States, including expanding budgets for medical examiner officers and eliminating the archaic coroner system (25).

In New England, we observed a different pattern around assignment of Covid-19 deaths than in the other BEA regions. In this region, a large share of counties had higher Covid-19 than
excess death rates, a pattern that became more pronounced between 2020 and 2021. Several explanations may exist for this pattern including that other causes of death (i.e. influenza) declined in these areas or that the economically privileged status of many of these counties shielded their residents from the negative indirect effect of the pandemic by allowing them to work-from-home and avoid household crowding. Finally, it is possible that deaths were over-assigned to Covid-19 in these areas due to different cause of death assignment protocols for Covid-19. For example, until March 2022, Covid-19 deaths in the state of Massachusetts included any death that occurred within 60 days of a Covid-19 diagnosis, which differed from other states and guidelines from the Council of State and Territorial Epidemiologists that recommended states use a 30 day window (44).

The study had several limitations. First, the study relied on publicly available data, which were subject to suppression of death counts fewer than 10 in a given county-month. We addressed this limitation by pooling information across different geographical levels through the use of hierarchical models and by taking advantage of the additional information provided by yearly death counts, however, our estimates remain uncertain in areas with small populations and few deaths. Second, some counties and states have experienced prolonged reporting delays of Covid-19 deaths, which could affect our estimates of the proportion of excess deaths assigned to Covid-19, particularly in more recent months. Third, we were unable to distinguish between excess deaths that represented uncounted Covid-19 deaths and excess deaths indirectly related to the pandemic. Future research should examine this distinction to clarify the extent to which excess deaths not assigned to Covid-19 reported in this study represent under-reporting of Covid-19 deaths. Fourth, our study examined all-cause mortality and did not explore differences in trends using cause-specific death rates. Assessing geographic and temporal differences in
excess death rates by cause-of-death would allow for a deeper understanding of the mechanisms
driving trends in excess mortality overall and is an important direction for future work. Fifth, we
used underlying cause of death data to identify deaths assigned to Covid-19 and thus did not
identify deaths where Covid-19 was listed as a contributing cause. Finally, due to data
limitations, our model does not account for differences in age structure between counties. Since
the pandemic has affected older populations more significantly, some differences in mortality
observed between counties may simply reflect differences in their age distribution.

In conclusion, the present study generated novel estimates of Covid-19 and excess mortality
for 3,127 harmonized county units over the period from March 2020 to December 2021. In
contrast to official Covid-19 death tallies, which are subject to differential underreporting and
fail to capture indirect pandemic effects, the present estimates more fully account for the true toll
of the pandemic across local areas and provide a more comparable measure of the Covid-19
mortality burden. As such, these estimates may be useful for additional work to investigate the
determinants of excess mortality throughout the pandemic and may also be useful for
communicating Covid-19 risks with local communities where the direct tallies have hidden the
full extent of the pandemic’s consequences.

Materials & Methods

Yearly and monthly death counts at the county level were extracted from CDC WONDER online
tool. See Methods Supplement for further details about data extraction procedures. We
extracted death counts by all causes of death and from Covid-19. Causes of death were selected
from the Multiple Cause of Death database using the provisional counts for 2020 and 2021 and
the final counts for 2015-2019. A death was assigned to Covid-19 when Covid-19 was listed as
the underlying cause of death using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) code U07.1. For all years except 2020, yearly deaths include all deaths between January to December. For 2020, yearly deaths include only deaths that occurred between March and December to better reflect the pandemic period. All death rates computed for 2020 were adjusted to account for the shorter exposure period.

To convert the number of deaths into rates, we used publicly available yearly county-level population estimates from the Census Bureau. To compute monthly rates, we assumed linear growth between each two time points. For the August-December period in 2021, for which no population estimates are available, we assumed the population remained constant at its value in July.

We grouped counties into 4 metropolitan-nonmetropolitan categories (large central metro, large fringe metro, medium or small metro, nonmetro) based on the 2013 NCHS Rural-Urban Classification Scheme for Counties (45) and into 8 Bureau of Economic Analysis (BEA) regions (Far West, Great Lakes, Mideast, New England, Plains, Rocky Mountains, Southeast, Southwest) (30). Next, we stratified each region by each metropolitan- nonmetropolitan category, leading to 32 geographic units. We also grouped small counties into county-sets according to the United States Census Bureau’s County Sets classification. County-sets were used, in addition to counties, within our model to improve the precision of estimates for small counties in the study. See Methods Supplement for further details about the geographic classifications used in this study.

---

3Covid-19 is listed as the underlying cause of death in approximately 92% of cases in which it is mentioned somewhere on the death certificate (28).
4Population data were obtained from the following sources for 2010-2020 and 2020-2021: 
https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/totals/co-est2020.csv
https://www2.census.gov/programs-surveys/popest/datasets/2020-2021/counties/totals/co-est2021-alldata.csv
To model the monthly county-level number of deaths for 100,000 residents (DR), we estimated two different hierarchical linear models. The first model predicts monthly DR for a county directly while the second one predicts yearly DR for a county and then distributes the corresponding number of deaths over the year by month according to weights computed over the All-Causes of Death data at the national level. Both models were estimated using the lme4 package for the R language (46) and fitted using monthly mortality data for the period January 2015 - December 2019. The need for these two different models arises as a result of the suppression procedure applied in the CDC WONDER tool to all data points (county-months in our case) with fewer than 10 deaths. As a result of suppression, small counties that rarely exceed 10 deaths in a given month have very few data points, and these data points are not representative of normal mortality conditions (that is why they were not suppressed). In other words, data points are not missing at random and non-missing data points for small counties reflect higher-than-normal mortality. The yearly-level model, by making use of the additional information on the yearly number of deaths, leads to more accurate predictions in counties with a high proportion of missing data points. Combining these two models led to a better overall performance compared to using either one in isolation.

The monthly-level model expresses the monthly number of deaths for every 100,000 residents (DR) as a function of time (in years), month (with dummy variables), and an intercept allowed to vary across counties, county-sets, and states. To make the model more flexible, the time slope is also allowed to vary across county sets. Formally:\(^5\):

\[^5\text{Here we use the notation employed in Raudenbush and Bryk (2002) (47).}\]
Where:

\[
DR_{c,cs,s,m,y} = \pi_{0c,cs,s} + \pi_{1cs} Year + \sum_{m=1}^{12} \theta_m \text{Month}_m + \varepsilon_{c,cs,s,m,y}
\]

Level 2: \(\pi_{0c,cs,s} \sim \mathcal{N}(\gamma_{0cs,s}, \sigma_{0C}^2)\)

Level 3: \(
\begin{bmatrix}
\gamma_{0cs,s} \\
\gamma_{1cs}
\end{bmatrix}
\sim \mathcal{N}
\left(
\begin{bmatrix}
\gamma_{0s} \\
\gamma_{1CS}
\end{bmatrix},
\begin{bmatrix}
\sigma_{0CS}^2 & \sigma_{01CS} \\
\sigma_{10CS} & \sigma_{1CS}^2
\end{bmatrix}
\right)
\)

Level 4: \(\gamma_{0s} \sim \mathcal{N}(\gamma_0, \sigma_{0S}^2)\)

\[
\varepsilon_{c,cs,s,m,y} \sim \mathcal{N}(0, \sigma_{\varepsilon}^2)
\]

In the equations above, the subscripts \(c, cs, s, m, \) and \(y\) indicate county, county-set, state, month, and year respectively. The capital letters C, CS, and S indicate that a term is specific to the county, county-set, and state level equations respectively. Using county-sets as an intermediate level between counties and states helped us overcome estimation difficulties with counties with few data points due to suppression. Even when the county level intercept cannot be estimated with precision, the estimate will be pulled toward the mean of counties in the same state and county-set, which we were able to estimate more precisely. The yearly-level model follows a structure similar to the monthly model but, due to the smaller sample size, it is less complex. Yearly death rates are modeled as a function of time (in years) and an intercept, both allowed to vary across counties.
Level 1: \[ DR_{c,y} = \pi_0c + \pi_1c Year_y + \varepsilon_{c,y} \]
Level 2:
\[
\begin{bmatrix}
\pi_0c \\
\pi_1c
\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}
\gamma_0 \\
\gamma_1
\end{bmatrix}, \begin{bmatrix}
\sigma_0^2 & \sigma_{01} \\
\sigma_{10} & \sigma_1^2
\end{bmatrix}\right)
\]

Where:

\[ \varepsilon_{c,y} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \]

In both models, the intercepts and the slopes are allowed to be correlated. To obtain the number of deaths, we multiplied the estimated death rate by the corresponding population. To obtain the monthly number of deaths from the annual number of deaths for the yearly model, we first computed the average proportion of deaths occurring in each month over the 2015-2019 period and then distributed annual deaths accordingly.

To decide whether to use the yearly model or the monthly model, we computed the average percentage difference between the predicted yearly deaths for the period 2015-2019 and the actual yearly deaths. We then used the estimate from the yearly model for all counties in which the difference was larger than 10%. Applying this decision rule, we used the yearly model for 790 counties and the monthly model for the remaining 2322 counties. Further details about the model section are provided in the **Methods Supplement**.

We obtained confidence intervals for the death rates by sampling from the distribution of the models’ fixed effects 1000 times and using these samples to compute 1000 different predicted rates. We then computed the 2.5 and 97.5 percentiles of the resulting distribution. The intervals thus obtained do not reflect all of the models’ uncertainty but only the portion due to fixed effects. However, they are consistent with the historical variability in the death rates and can be used to get a sense of the estimates’ variability.
This study used de-identified publicly available data and was exempted from review by the Boston University Medical Center Institutional Review Board. Programming code was developed using R, version 4.1.0 (R Project for Statistical Computing) and Python, version 3.7.13 (Python Software Foundation).
Acknowledgements

The authors would like to thank Robert N. Anderson and Farida B. Ahmad from the National Center for Health Statistics for technical assistance with the provisional mortality files used in the present study. We also thank Elif Coskun, Mikas Hansen, Sylvia Lutze, Pia MacDonald, and Steele Myrick for additional technical and administrative support. Additionally, we are grateful to Nahid R. Bhadelia and Erica Augustine from the Center for Emerging Infectious Diseases Policy and Research at Boston University for their guidance and support with translation and dissemination of study findings; Gus Wezerek from the New York Times for input on the analytic strategy and data visualization; and members of the Documenting Covid-19 Project team at Muckrock for their insightful investigative reporting on uncounted Covid-19 deaths. The authors gratefully acknowledge financial support from the Robert Wood Johnson Foundation (77521), the National Institute on Aging (R01-AG060115-04 and R01-AG060115-04S1) and the W.K. Kellogg Foundation. The views expressed in this paper are those of the authors and not necessarily the views of the study sponsors.

Conflicts of Interest

The authors report that they have no conflicts of interests to disclose.

Data Availability

Data used in the present manuscript are publicly available from the Centers for Disease Control and Prevention and Census Bureau. Additional details about the data and programming code for replicating the analyses of the present study can be accessed at the linked GitHub repository:

https://github.com/Vital-Stats-Integrity-Proj/monthly_county_level_excess_mortality
References


31. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Committee on National Statistics, Committee on Population, Committee on Rising Midlife Mortality Rates and Socioeconomic Disparities, High and Rising Mortality Rates Among Working-Age Adults (National Academies Press, 2021).

32. E. Wrigley-Field, K. M. Berry, A. C. Stokes, J. P. Leider, “Pandemic of the unvaccinated”? At midlife, white people are less vaccinated but still at less risk of Covid-19 mortality in Minnesota. medRxiv (2022) https://doi.org/10.1101/2022.03.02.22271808.


Table 1. Excess mortality, Covid-19 mortality, and the ratio of Covid-19 to excess mortality by metropolitan-nonmetropolitan status and BEA region

<table>
<thead>
<tr>
<th>BEA Region</th>
<th>Metro Status</th>
<th>Number of Deaths</th>
<th>Rates per 100,000 PY</th>
<th>Number of Deaths</th>
<th>Rates per 100,000 PY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Excess</td>
<td>COVID</td>
<td>Ratio</td>
<td>Excess</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
<td>459,764</td>
<td>350,695</td>
<td>0.763</td>
<td>167</td>
</tr>
<tr>
<td>Total</td>
<td>Large Central</td>
<td>166,578</td>
<td>111,101</td>
<td>0.708</td>
<td>173</td>
</tr>
<tr>
<td>Total</td>
<td>Large Fringe</td>
<td>98,568</td>
<td>72,432</td>
<td>0.736</td>
<td>140</td>
</tr>
<tr>
<td>Total</td>
<td>Medium/Small</td>
<td>135,180</td>
<td>97,669</td>
<td>0.723</td>
<td>164</td>
</tr>
<tr>
<td>Total</td>
<td>Non Metro</td>
<td>79,458</td>
<td>59,473</td>
<td>0.749</td>
<td>207</td>
</tr>
</tbody>
</table>

Notes: Death rates were calculated by aggregating deaths and population over counties within each BEA region and metropolitan-non-metropolitan area. Estimates for 2020 correspond to the period March - December 2020. Estimates for 2021 refer to the period January to December, 2021.
Figure 1. County excess mortality rates per 100,000, 2020-2021

Notes: Panels A and B show the geographic distribution of excess death rates in 2020 (A) and 2021 (B) as estimated by comparing the expected number of deaths from our model to the actual number of deaths. Panels C and D report excess and COVID deaths rates for the counties with the highest excess deaths rates 2020 and 2021 respectively. Counties with less than 30,000 residents and less than 60 COVID deaths across the two years were excluded from the rankings in the barplots.
Figure 2. Monthly Excess Deaths Rates by BEA Region and Metro Status, 2019-2021

Notes: This graph shows aggregated trends in excess mortality at the monthly level between 2019-2021 stratified by BEA region and metro status.
Figure 3. Excess Deaths by BEA Region and Metro Status, and by State, 2020-2021

Notes: The top panel presents absolute excess deaths by BEA region disaggregated by metro status (left) and state (right) for 2020; the bottom panel shows the same for 2021.
Figure 4. Percentage of Excess Deaths not Assigned to Covid-19, 2020-2021

Notes: Panels A (2020) and B (2021) show the geographic distribution of proportion of excess deaths not assigned to COVID in 2020 and 2021. Panels C (2020) and D (2021) report excess and COVID deaths rates for the counties with the lowest COVID to excess ratios in 2020 and 2021. Counties with less than 30,000 residents and less than 60 COVID deaths across the two years were excluded from the rankings in the barplots.
Supplementary Information


Table S2. Performance of different models on training and test data.

Figure S1. Actual and expected mortality trends by month at the national level, 2015-2021

Figure S2. Time-series plots for the largest counties in each BEA region and metro-status combination

Figure S3. Excess Death Rates by BEA Region and Metro Status, and by State, 2020-2021

Figure S4. State-level rankings of COVID-19 to excess mortality ratios for 2020 and 2021

Figure S5. A Comparison of Excess and Covid-19 Deaths across BEA Regions and Metro Status.

Methods Supplement. Details on data extraction, geographic classifications, model selection and estimation
Table S1. Excess mortality, Covid-19 mortality, and the ratio of Covid-19 to excess mortality by US state

<table>
<thead>
<tr>
<th>State</th>
<th>2020</th>
<th>2021</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Deaths</td>
<td>Rates per 100,000 PY</td>
</tr>
<tr>
<td></td>
<td>Excess</td>
<td>COVID</td>
</tr>
<tr>
<td>Alabama</td>
<td>9,002</td>
<td>6,543</td>
</tr>
<tr>
<td>Alaska</td>
<td>476</td>
<td>275</td>
</tr>
<tr>
<td>Arizona</td>
<td>14,755</td>
<td>8,447</td>
</tr>
<tr>
<td>Arkansas</td>
<td>4,906</td>
<td>3,529</td>
</tr>
<tr>
<td>California</td>
<td>42,806</td>
<td>31,339</td>
</tr>
<tr>
<td>Colorado</td>
<td>6,432</td>
<td>4,546</td>
</tr>
<tr>
<td>Connecticut</td>
<td>5,826</td>
<td>5,782</td>
</tr>
<tr>
<td>Delaware</td>
<td>1,439</td>
<td>1,008</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>1,121</td>
<td>831</td>
</tr>
<tr>
<td>Florida</td>
<td>27,159</td>
<td>19,230</td>
</tr>
<tr>
<td>Georgia</td>
<td>14,921</td>
<td>9,431</td>
</tr>
<tr>
<td>Hawaii</td>
<td>199</td>
<td>142</td>
</tr>
<tr>
<td>Idaho</td>
<td>1,927</td>
<td>1,202</td>
</tr>
<tr>
<td>Illinois</td>
<td>20,979</td>
<td>15,717</td>
</tr>
<tr>
<td>Indiana</td>
<td>10,496</td>
<td>8,520</td>
</tr>
<tr>
<td>Iowa</td>
<td>4,492</td>
<td>4,315</td>
</tr>
<tr>
<td>Kansas</td>
<td>3,964</td>
<td>3,287</td>
</tr>
<tr>
<td>Kentucky</td>
<td>5,925</td>
<td>4,112</td>
</tr>
<tr>
<td>Louisiana</td>
<td>10,384</td>
<td>6,520</td>
</tr>
<tr>
<td>Maine</td>
<td>454</td>
<td>419</td>
</tr>
<tr>
<td>Maryland</td>
<td>8,275</td>
<td>5,995</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>7,222</td>
<td>9,329</td>
</tr>
<tr>
<td>Michigan</td>
<td>16,246</td>
<td>11,398</td>
</tr>
<tr>
<td>Minnesota</td>
<td>5,891</td>
<td>5,211</td>
</tr>
<tr>
<td>Mississippi</td>
<td>7,431</td>
<td>4,460</td>
</tr>
<tr>
<td>Missouri</td>
<td>10,296</td>
<td>7,129</td>
</tr>
<tr>
<td>Montana</td>
<td>1,315</td>
<td>1,125</td>
</tr>
<tr>
<td>Nebraska</td>
<td>2,146</td>
<td>2,079</td>
</tr>
<tr>
<td>Nevada</td>
<td>3,997</td>
<td>3,243</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>747</td>
<td>746</td>
</tr>
<tr>
<td>New Jersey</td>
<td>17,156</td>
<td>16,497</td>
</tr>
<tr>
<td>New Mexico</td>
<td>4,192</td>
<td>2,651</td>
</tr>
<tr>
<td>North Carolina</td>
<td>14,721</td>
<td>35,723</td>
</tr>
<tr>
<td>North Dakota</td>
<td>1,218</td>
<td>1,203</td>
</tr>
<tr>
<td>Ohio</td>
<td>17,793</td>
<td>13,605</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>6,309</td>
<td>4,882</td>
</tr>
<tr>
<td>Oregon</td>
<td>2,315</td>
<td>1,444</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>17,221</td>
<td>16,907</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>1,265</td>
<td>1,085</td>
</tr>
<tr>
<td>South Carolina</td>
<td>9,397</td>
<td>5,277</td>
</tr>
<tr>
<td>South Dakota</td>
<td>1,505</td>
<td>1,481</td>
</tr>
<tr>
<td>Tennessee</td>
<td>10,929</td>
<td>6,831</td>
</tr>
<tr>
<td>Texas</td>
<td>41,209</td>
<td>30,871</td>
</tr>
<tr>
<td>Utah</td>
<td>1,879</td>
<td>1,367</td>
</tr>
<tr>
<td>Vermont</td>
<td>220</td>
<td>135</td>
</tr>
<tr>
<td>Virginia</td>
<td>8,805</td>
<td>5,799</td>
</tr>
<tr>
<td>Washington</td>
<td>3,204</td>
<td>3,296</td>
</tr>
<tr>
<td>West Virginia</td>
<td>2,567</td>
<td>1,478</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>7,549</td>
<td>5,429</td>
</tr>
<tr>
<td>Wyoming</td>
<td>862</td>
<td>461</td>
</tr>
</tbody>
</table>

Notes: Death rates were calculated by aggregating deaths and population over counties within each state.

Figure S1. Actual and expected mortality trends by month at the national level, 2015-2021

Notes: Monthly trends in actual and expected deaths at the national level are obtained by aggregating the estimates from the county level.
Figure S2. Time-series plots for the largest counties in each BEA region and metro-status combination

Notes: The plot shows the largest counties within each BEA region and metro status category. Time series of actual and expected all-cause deaths per 100,000 are plotted over 2015-2021.
Figure S3. Excess Death Rates by BEA Region and Metro Status, and by State, 2020-2021

Notes: The top panel shows excess death rates by state organized by BEA region for 2020 (left) and 2021 (right). The bottom panel shows excess death rates by metro status organized by BEA region for 2020 (left) and 2021 (right).
Figure S4. State-level rankings of COVID-19 to excess mortality ratios for 2020 and 2021.

Notes: These graphs present state-level rankings of Covid-19 to excess mortality ratios for 2020 and 2021. The dotted vertical line represents a Covid-19 to excess ratio of 1.0, indicating that Covid-19 deaths fully account for estimated excess deaths. States with very few Covid-19 deaths over 2020-2021 were excluded (Alaska, Hawaii, Vermont) from the graph.
Figure S5. A Comparison of Excess and Covid-19 Deaths across BEA Regions and Metro Status.

Notes: Panels A and B show the relationship between COVID and excess deaths in 2020 and 2021 respectively. Counties with less than 30,000 residents were excluded. The regression lines reflect the coefficients of a set of regressions of COVID deaths on excess deaths within each BEA region using population as weights.
Methods Supplement

Data Extraction from CDC WONDER

The CDC WONDER online database query system found at https://wonder.cdc.gov/ was used to extract all mortality data used in this project. To obtain death counts for all-causes mortality, we used the Multiple Cause of Death (Final) database from 1999-2020. We obtained two sets of extracts, one for data at the county-year level and one for data at the county-month level.

For county-year extracts, the data request was submitted for the time period of interest using the request form with the following settings changed from the default:

- Tab 1. Organize table layout: Group results by County and by Year
- Tab 4. Select time period of death: specific period
- Tab 6. Select underlying cause of death: *All* (All Causes of Death)
- Tab 8. Other options: checking Export Results and Show totals. The request generates a text file.

For county and month for the years 1999 through 2019, the data was extracted using six month time periods due to limitations of the CDC Wonder servers. The data request was submitted for each time period of interest using the request form with the following settings changed from the default:

- Tab 1. Organize table layout: Group results by County and by Month
- Tab 4. Select time period of death: specific period (6 months at a time)
- Tab 6. Select underlying cause of death: *All* (All Causes of Death)
- Tab 8. Other options: checking Export Results and Show totals. The request generates a text file.
To extract data for the time periods of March 2020 to December 2021, we used the Multiple Cause of Death (Provisional) database from 2018 – Last Month database. The data requests were submitted for each time period of interest using the request form with the following settings changed from the default:

- Tab 1. Organize table layout: Group results by County and by Year
- Tab 4. Select time period of death: March 2020 to December 2021
- Tab 6. Select underlying cause of death: *All* (All Causes of Death)
- Tab 8. Other options: checking Export Results and Show totals. The request generates a text file.

To extract counts of COVID deaths for 2020 and 2021 at the county-year level we used the Multiple Cause of Death (Provisional) database from 2018 – Last Month database. The data requests were submitted for each time period of interest using the request form with the following settings changed from the default:

- Tab 1. Organize table layout: Group results by County and by Year
- Tab 4. Select time period of death: March 2020 to December 2021
- Tab 8. Other options: checking Export Results and Show totals. The request generates a text file.

**Geographic Classifications**

**USDA/ERS/NCHS Metropolitan-Nonmetropolitan Categories:**

- **Large central metros**: counties in metropolitan statistical areas with a population of more than 1 million.
- **Large fringe metros**: counties that surrounded the large central metros
- **Small or medium metros**: counties in metropolitan statistical areas with a population between 50,000 and 999,999.
- **Nonmetropolitan areas**: all other counties.

**BEA Regions:**

- **New England**: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
- **Mideast**: Delaware, District of Columbia, Maryland, New Jersey, New York, and Pennsylvania
- **Great Lakes**: Illinois, Indiana, Michigan, Ohio, and Wisconsin
- **Plains**: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota
- **Southeast**: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia
- **Southwest**: Arizona, New Mexico, Oklahoma, and Texas
- **Rocky Mountain**: Colorado, Idaho, Montana, Utah, and Wyoming
- **Far West**: Alaska, California, Hawaii, Nevada, Oregon, and Washington

**County-Sets:**

- Counties with populations greater than or equal to 50,000 were not grouped into county-sets
- Counties with populations less than 50,000 were grouped with contiguous counties to form county-sets with populations greater than 50,000
- County-sets did not cross state borders

**Model Selection and Estimation**

Before settling on the models used in the paper (Random Time), we tested a simple model (Base), as well as a more complex one (Random Time Plus). Model Base shares the fixed effect of the Random Time model described in the methods section but has a simpler random effects structure including only random intercepts for counties. Model Random Time Plus borrows the random effects structure of the Random Time model but allows the month dummy variables (used to capture seasonality) to vary across census regions. We hypothesized that, because seasonal patterns in mortality are not uniform across the United States, such a model would have
achieved a better fit. To evaluate the model performance, we trained them on data for the period 2014-2018 and then tested them on data for 2019. Table S2 reports the Root Mean Squared Error for all models. While the Random Time Plus model outperformed the two simpler alternatives in the training data, the Random Time model achieved the best performance in the test data. This suggests that the Random Time Plus Model was overfitting the training data to a larger extent compared to the simpler models. Looking at the AIC and BIC confirms our insights. According to the BIC, we should select the Random Time model. According to the AIC the Random Time Plus model should instead be preferred. Overall, because the Random Time model outperformed the Base Model in all data sets, has the lowest BIC, and is the best performing model on the test data, we chose to use it as the final model in the paper.

**Table S2. Model Comparison**

<table>
<thead>
<tr>
<th>Performance</th>
<th>Base</th>
<th>Random Time</th>
<th>Random Time Plus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Data (2019)</td>
<td>17.981</td>
<td>17.955</td>
<td>17.961</td>
</tr>
<tr>
<td>Overall</td>
<td>17.231</td>
<td>17.187</td>
<td>17.175</td>
</tr>
<tr>
<td>BIC</td>
<td>1002787</td>
<td>1002431</td>
<td>1002579</td>
</tr>
<tr>
<td>AIC</td>
<td>1002642</td>
<td>1002248</td>
<td>1002048</td>
</tr>
</tbody>
</table>