NBER WORKING PAPER SERIES

EXCESS DEATHS IN THE UNITED STATES DURING THE FIRST YEAR OF COVID-19

Christopher J. Ruhm

Working Paper 29503
http://www.nber.org/papers/w29503

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2021

Research reported in this publication was supported by the National Institute on Aging and the Office of the Director, National Institutes of Health, under Award Number P01AG005842. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. I thank Neil Mathur and Daniel Podratsky for helpful research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Excess Deaths in the United States During the First Year of COVID-19
Christopher J. Ruhm
NBER Working Paper No. 29503
November 2021
JEL No. I10,I13,I18

ABSTRACT

Accurately determining the number of excess deaths caused by the COVID-19 pandemic is hard. The most important challenge is accurately estimating the counterfactual count of baseline deaths that would have occurred in its absence. This analysis used new methods to: estimate this baseline metric; calculate excess deaths during the first year of the U.S. COVID-19 pandemic; and examine plausibility of the excess death estimates obtained in this and prior analyses. Total, group-specific and cause-specific excess deaths in the U.S. from March 2020 through February 2021 were calculated using publicly available data covering all deaths from March 2009 through December 2019 and provisional data from January 2020 through February 2021. The estimates indicate that there were 646,514 (95% CI: 597,514 to 695,520) excess deaths in the U.S. during this period, with 83.4% (95% CI: 77.5% - 90.2%) of these attributed directly to COVID-19. There were substantial differences across population groups and causes in the ratio of actual-to-baseline deaths, and in the contribution of COVID-19 to excess mortality. Prior research has frequently underestimated baseline deaths and so has overstated excess mortality and the percentage of it attributed to non-COVID-19 causes.

Christopher J. Ruhm
Frank Batten School of
Leadership and Public Policy
University of Virginia
235 McCormick Rd.
P.O. Box 400893
Charlottesville, VA 22904-4893
and NBER
ruhm@virginia.edu

Introduction

Determining the death toll of the coronavirus (COVID-19) pandemic is hard. In its early stages, testing was sporadic, and some mortality was probably attributed to other causes (1). Misclassification likely continued as the epidemic progressed, particularly since provisional data were often incomplete and so likely to undercount fatalities. Some deaths probably also occurred indirectly, such as those due to reductions emergency department or hospital visits occurring because the health care system was stressed beyond capacity or individuals avoided medical care to reduce exposure to the virus (2–4). Conversely, mortality from other sources may have declined. For example, traffic fatalities typically fall during economic slowdowns (5), and some studies suggest this occurred during the pandemic (6). Similarly, social distancing possibly reduced deaths from transmissible diseases such as influenza (7).

The standard method for addressing these combined effects is to calculate excess deaths, here estimated as the difference between actual deaths and the baseline count expected if the COVID-19 pandemic had not occurred. Excess deaths have been computed across a wide variety of settings including measuring the mortality effects of obesity, HIV, hurricanes and heatwaves (8–11), as well as for COVID-19 (12–20).

A key challenge is determining the counterfactual baseline number to which observed mortality counts are compared. In the context of COVID-19, multiple methods have been employed and no single procedure is necessarily universally preferable. For instance, approaches used to compute excess deaths over short periods (e.g. weeks) or for limited geographic areas (like states) may be less well suited to calculating excess national deaths over a full year.

Prior investigations have typically modeled deaths over (arbitrarily determined) prior periods and used the specifications with the best fit among the models tested to estimate baseline deaths. However, the plausibility of the resulting counterfactual mortality rates or counts has not been examined. The current study first developed a simple criterion for examining the reliability of such excess mortality estimates. The method involved examining how the difference between estimated (counterfactual) baseline deaths and actual mortality over the same period one-year earlier compares to historical changes in year-over-year fatality counts. The results indicate that earlier studies frequently overestimated the number of excess U.S. deaths during the COVID-19 pandemic, and also the share of these deaths attributed to causes other than COVID-19. Next, new procedures were used to provide estimates of excess mortality from March 2020 through February 2021, the first year of the significant COVID-19 mortality in the United States. (A few COVID-19 deaths occurred during the first two months of 2020 and the virus may have entered the country as early as December 2019 (21)).

Data

Data on the universe of U.S. deaths from 2009-2019 came from the *Multiple Cause of Death* files, which provided information from death certificates on demographic characteristics and cause of death, categorized using four-digit *International Classification of Diseases, Tenth Revision* (ICD-10) codes (22). Since final data on deaths in 2020 and 2021 were not available at the time of this study, information for January 2020 through February 2021 was obtained from the Centers for Disease Control and Prevention, as part of their *Provisional Death Counts for Coronavirus Disease (COVID-19)* series (23). For the main analysis, separate data sets supplied monthly death counts from various causes (24); weekly counts by sex and age (25); and weekly counts by race and Hispanic origin (26), all updated on November 3, 2021. Additional provisional death data sets (27,28) were used

when examining the excess death estimates from prior studies, as described the Appendix. Weekly death reports do not exactly correspond to those from calendar months or years, requiring adjustments that are also detailed there.

Total mortality counts were computed as were those by sex, race/ethnicity, age, and selected causes. For brevity, whites, blacks, and other races below refer to non-Hispanics. The causes of death examined include: COVID-19; heart disease; malignant neoplasms (cancer); cerebrovascular disease (stroke); Alzheimer's disease; diabetes; chronic lower respiratory disease; influenza or pneumonia; kidney disease; all external causes; unintentional injuries (accidents); motor vehicle accidents; drug overdoses; intentional self-harm (suicide); and assault (homicide). Table 1 indicates the ICD-10 codes corresponding to each of these causes, which were largely determined by availability in the provisional mortality data.

Methods

Year-Over-Year Changes in Deaths

Death counts were calculated for each month from March 2009 through February 2021 and then aggregated over 12-month periods starting in March of each year. These March through February periods are hereafter usually denoted by the year with the majority of months (e.g. 2020 refers to March 2020 through February 2021) for expositional convenience. Year-over-year percentage changes in the counts were computed, as were average annual percent changes over the 2009-2019 period.

Prior Estimates of Excess Deaths Attributed to COVID-19

Estimates of excess deaths and the share attributed to COVID-19 were obtained from four prior studies (17–20) that covered periods from at least the start of the U.S. pandemic through the end of calendar year 2020. Two of these investigations (17,18) built upon

methods developed in prior analyses covering shorter periods (12–15), so that problems identified here also apply to those earlier investigations. Baseline deaths were computed as the difference between actual and excess fatalities, using information provided in the original investigations where possible but with total deaths calculated using the data sources described above, when not supplied by the authors.

Lagged mortality counts, defined as those occurring over the same time-period but 12-months earlier than the original study, were constructed, as were percentage changes in baseline deaths versus actual lagged fatalities. Ninety-five percent confidence intervals (95% CI) or bounds on the estimates were determined, to the extent possible given the information supplied in the original analyses. The implied year-over-year percentage changes were then compared with historical growth in annual death counts to assess their plausibility.

Further adjustments were sometimes required to provide comparable estimates. For instance, one of the studies (20) reported as COVID-19 only fatalities where this was the underlying cause of death, whereas the others included cases where COVID-19 was either an underlying or contributory cause. Additional information on the methods, data used, and adjustment procedures is provided in the Appendix.

New Estimates of Excess Deaths From March 2020 Through February 2021

New estimates on excess deaths during the first year of the COVID-19 pandemic were obtained using the following procedure. First, year-over-year changes in monthly total death counts were calculated. These revealed substantial excess volatility during January, relative to other months: the standard deviation in the year-over-year monthly change from 2009-2019 was almost five times as high in January as the average of the other eleven months and more than twice as large as for the next highest month (December). This volatility was

particularly large for influenza/pneumonia deaths, which are known to exhibit considerable annual variation (29), but also present for some other causes (Appendix Table A1). To address the challenges posed by this largely idiosyncratic variation, the estimates of baseline mortality below were based on adjusted annual mortality counts obtained by calculating the number of deaths during December and February through November of each (March through February) year and then multiplying this by the average ratio of deaths over the full-year (including January) versus 11-month periods (excluding January) from 2009-2019. This preserved the overall number of deaths over the 11-years but removes the annual January volatility.

Second, these adjusted mortality counts were regressed on pre-COVID-19 linear or quadratic time trends. For total mortality, and all causes and groups, the last year included in the regression models was 2019, while the first year was allowed to vary between 2009 and 2016. The decision of whether to use a linear or quadratic trend, as well as the analysis time period was determined, by estimating both models for all potential starting years and then choosing the specification with the smallest root mean squared error ($RMSE$), computed using a degrees of freedom correction that penalized shorter panels and quadratic (relative to linear) specifications; 2016 was the latest potential starting year to ensure that the estimation model contained a minimum of five years. The use of Poisson, rather than regression, models was considered but preliminary analysis provided no indication that doing so improved model fit.

Third, baseline deaths were estimated from the preferred regression specification just described, with the trend and, if applicable, trend squared variables set to their 2020-year values. Ninety-five percent confidence intervals were computed using the standard error of the regression predicted value.

Finally, excess deaths were computed as the difference between actual 2020 mortality and the estimated baseline number of fatalities.

Additional details are provided in the Appendix.

Plausibility of the Excess Death Estimates

Plausibility of the excess death estimates, both from prior studies and the original analysis presented here, was evaluated by examining how the implied change in estimated baseline deaths (for total mortality or for the specified group or cause) during the analysis timespan, versus actual fatalities during the same period one-year earlier, compared to actual changes in year-over-year deaths from 2009-2019. These are often referred to below as "baseline changes: and "historical changes". Specifically, the average year-over-year change, during 2009-2019, was calculated, as was the range of these historical changes. Next the implied difference between baseline 2020 mortality counts and actual 2019 deaths, the baseline changes, were calculated with the former obtained from the regression procedure discussed above, as were the 95% CI's. Close accordance between the estimated baseline and actual historical changes suggests that the estimation models provide plausible results.

When evaluating prior studies estimates of excess deaths, lagged mortality was calculated starting and ending on the same day of the week as the analysis period. For example, one study (17) evaluated excess COVID-19 deaths from March 1, 2020 to January 2, 2021. Lagged actual deaths here were calculated for March 3, 2019 to January 4, 2020, so that both periods contained the same number of days, started on the first Sunday in March and ended on the first Saturday in January. This procedure ensured that the counts were not affected by day-of-the week or seasonal differences in death rates. For the original analysis, covering March 1, 2020 through February 28, 2021, the lag period was March 1, 2019 through February 29, 2020, with a correction for the extra leap-year day in the February 2020.

The Appendix further details the methods.

Results

Year-over-Year Changes in Total Deaths

Figure 1 shows the year-over-year percentage change in total deaths over 12-month periods from 2009-2019, as well as corresponding results for mortality adjusted to eliminate excess January volatility. Annual counts increased from 2.44 million in 2009 to 2.87 million in 2019. Growth averaged 1.65% per year but with considerable variability. The standard deviation of the 12-month change (not shown) was 1.70%, with minimum and maximum changes of -1.21% and 3.51%. Positive year-over-year growth occurred in eight of ten years, with reductions exceeding 0.6% happening just once, from 2017 to 2018.

Average growth in adjusted fatalities counts was virtually the same as the unadjusted change (1.63%) but the standard deviation was reduced by almost 40% (to 1.05%). Positive year-over-year growth was observed in 9 of the 10 years, with annual increases of 1.8% to 3.0% in 7 out of 10 years.

These patterns imply that positive year-over-year mortality growth is the norm and that baseline death estimates indicating declines relative actual deaths 12-months earlier are probably unreliable.

Prior Estimates of Excess Deaths During COVID-19 Pandemic

Table 2 summarizes the results of four previous estimates of excess mortality during the COVID-19 pandemic. Estimated excess deaths totaled between 453,521 and 912,345, with studies covering longer and later periods providing larger numbers, as expected. The primary or point estimates indicate that the share of excess mortality attributed to COVID-19, ranged from 63% to 75% in three of the investigations (17–19) and 83% in the fourth (20). The Woolf et al. study (17) has extremely narrow confidence intervals (less than a

0.4% difference in the upper and lower bounds of the 95% CI for excess deaths), as do the widely cited related prior investigations (12–14) that used similar methods. These contrast with the substantial variability of actual year-over-year mortality counts documented above.

The last column of the table shows the implied change in estimated baseline versus 12-month lagged actual deaths. For three of the studies (17–19) , the primary estimates suggest negative baseline changes, in sharp contrast with the positive historical growth observed for 8 of the 10 years from 2010-2019. Conversely, baseline deaths in one investigation (20) were 0.86% (95% CI, -0.48% to 2.20%) higher than actual deaths one-year earlier and the upper-bound for another (18) also showed positive year-over-year growth. These analyses also attributed the highest share of excess deaths directly to COVID-19, as expected since the denominator over which these shares were being calculated was lower. The estimated negative baseline change in in deaths from another study (19) far exceeded the largest historical decrease observed over any of the prior 10 years.

New Estimates of Excess Deaths

Table 3 and Figure 2 summarize the estimates of overall excess deaths, and those for demographic subgroups, obtained using the methods described above; Table A3 provides additional detail. There were 646,514 (95% CI: 597,507 to 695,520) excess deaths from March 2020 through February 2021, a 22% (95% CI: 20% to 25%) increase in over the counterfactual baseline. COVID-19 was as an underlying or contributing cause in 538,924 fatalities, or 83.4% (95% CI: 77.5% to 90.2%) of excess mortality.

Gender differences in the ratio of observed versus baseline deaths were fairly modest, although males may have been somewhat more affected than females. Race disparities were larger. Whites had the most excess deaths, but these represented a smaller relative increase

than for other groups: the ratio of observed-to-baseline deaths was 1.17 (95% CI: 1.14 to 1.19) for non-Hispanic whites compared to 1.31 to 1.51 (with fairly similar sized 95% CI's) for blacks, Hispanics and other nonwhites. Among whites, 89.7% (95% CI: 78.4% to 104.7%) of the excess deaths were attributed to COVID-19, with considerably lower percentages for blacks (70.7%, 95% CI: 67.6% to 74.0%) and other nonwhites (60.6%, 95% CI: 59.5% to 61.7%).

Actual-to-baseline death ratios were fairly similar for the 25-44, 45-64 and ≥65 year olds, ranging from 1.21 to 1.27, with 95% CI's between 1.20 and 1.23, but a considerably lower 1.12 (95% CI: 1.07 to 1.16) for persons younger than 25. Almost all excess deaths of seniors were attributed to COVID-19, compared with monotonically decreasing shares at younger ages.

Table 4, Figure 3, and Table A3 provide corresponding information for (non-COVID-19) causes of death. Heart disease had the most excess fatalities (33,133; 95% CI: 26,312 to 39,953) but this represented a smaller actual-to-baseline death ratio (1.05, 95% CI: 1.04 to 1.06) than for diabetes or Alzheimer's disease. There were no excess cancer deaths and lower respiratory disease mortality was 5% (95% CI: 2% to 8%) below the baseline number, with a similar but less precisely reduction estimated for influenza/pneumonia fatalities.

Excess mortality from external causes was frequently large in percentage terms. The ratio of actual-to-baseline deaths was 1.16 (95% CI: 1.11 to 1.21) for all accidents, 1.29 (95% CI: 1.20 to 1.39) for fatal drug overdoses, 1.27 (95% CI: 1.16 to 1.40) for homicides, and 1.10 (95% CI: 1.08 to 1.12) for vehicle deaths. Conversely, suicides were 4,347 (95% CI: 3,477 to 5,237) *below* the baseline count, a relative decline of 9% (95% CI: 7% to 10%).

Success of Estimation Procedures in Predicting Baseline Deaths

Evidence on the plausibility of the excess death estimates was obtained by examining whether the baseline changes were consistent with historical year-over-year death counts from 2009-2019. The results are summarized in Figure 4, with additional details in Appendix Table A4. The range of historical changes was substantial in percentage terms for all groups and causes, and extremely large in some cases. For example, the historical changes ranged from 1.8% to 6.7% for Hispanics, -2.5% to 8.7% for <25 year olds, -11.7% to 7.7% for influenza/pneumonia, -3.0% to 23.3% for drug fatalities, and -7.3% to 12.5% for homicides.

The estimated baseline changes were usually fairly similar to observed historical changes: the 95% CI's always overlapped the range of observed historical year-over-year growth, and usually included the average of these. These results provide support for the methods used here to estimate the excess deaths. However, there were two cases (lower respiratory disease and vehicle fatalities) where the baseline changes in death counts fell by more than 1.0%, while the average year-over-year historical increase was greater than 1.0%, and one (persons <25) and where both were negative but the former was considerably more so than the latter (-3.9% versus -0.9%). In these cases, the estimates presented above may have understated excess deaths.

Discussion

Excess mortality provides important information for understanding consequences of the COVID-19 pandemic (30–32) but it is difficult to estimate and the , with the results are particularly dependent on the procedures used to calculate the counterfactual baseline number of deaths that would have occurred without the pandemic. This study evaluated excess death estimates obtained from prior research and provided new calculations on excess mortality from March 1, 2020 through February 28, 2021.

An important finding is that previous investigations have probably often understated the counterfactual number of baseline deaths and, consequently, overestimated both excess mortality and the share of it attributed to causes other than COVID-19. Key evidence for this is that the baseline death estimates used in these studies frequently implied lower mortality than that observed over the same period one year earlier but that such year-over-year decreases, historically, have rarely occurred. Instead deaths almost always increased over time, but with considerable year-over-year variability. This further implies that confidence intervals on excess deaths should inherently be substantial and that estimates yielding small confidence intervals are dubious.

New estimates of excess deaths from March 2020 through February 2021, the first year of COVID-19 in the U.S., were provided with goodness-of-fit measures used to determine the preferred estimation specification, a choice that was allowed to vary across total mortality and groups or causes of death. In contrast to some earlier investigations, the baseline changes in deaths obtained using this method were consistent with historical patterns of year-over-year mortality growth and the results frequently suggested lower actual-to-baseline death ratios and higher shares of excess deaths attributed to COVID-19 than in prior research.

The analysis indicates that, in its first year, the COVID-19 pandemic was responsible for 646,514 (95% CI, 597,507 to 695,520) excess deaths in the U.S., a 22% (95% CI: 20% to 25%) rise over baseline. The percentage increases were relatively small for non-Hispanic whites and persons younger than 25, consistent with prior research (15,17,33). Although seniors did not have particularly large percentage growth in excess mortality, the absolute number of excess deaths (452, 051, 95% CI: 430,292 to 473,051) was by far the greatest for them.

Eighty three percent of excess deaths were directly attributed to COVID-19, but with a wide 95% confidence interval (77.5% to 90.2%) reflecting the substantial variability in historical mortality counts. There are two likely reasons why the estimated COVID-19 share of excess deaths was higher here than in most prior research (12–15,17–19). First, detection of COVID-19 involvement may have been poor early in the epidemic, due to disruptions in the medical system and broader economy (34), an issue of particular importance for studies focused on periods early in the pandemic. Second, the understatement of baseline deaths in most previous analyses, discussed extensively above, mechanically induced a lower estimated COVID-19 share of excess mortality.

Excess deaths from non-COVID-19 sources nevertheless remain important. Additional heart and cerebrovascular disease fatalities are consistent with health system capacity constraints and with the reluctance of some individuals to seek medical care due to fears of infection (4,35,36). Interestingly, even though cancer screenings declined early in the epidemic (37), and there were predictions of higher deaths from this cause (38), the analysis provides no evidence that the pandemic raised cancer mortality. This absence may reflect substantial delays between the timing of screenings or treatment and deaths, or other factors such as the rapid rebound in health system capacity for temporarily deprioritized procedures (39). Excess diabetes fatalities could indicate its role as a comorbidity, raising the risk of death following COVID-19 infection, adverse effects of the virus on glycemic control, or increased difficulty in disease management during a period of social isolation and relatively limited medical support (40). Elevated Alzheimer's disease mortality may have resulted because affected individuals frequently lived in long-term care facilities, where COVID-19 deaths were common, due to reductions in access to medical care, disrupted routines, or social isolation, and possibly particularly adverse consequences of coronavirus infection (41–44).

Some results for external causes of death were surprising. Drug fatalities and suicides have been described as deaths of despair (45), although this characterization has been questioned (46), and numerous commentators expected such deaths to rise during the pandemic (47–50). While high excess mortality from drug overdoses and homicides is consistent with despair; the decline in suicides, also observed elsewhere (16), is not. Certain types of risk-taking may have increased, as reflected by the growth in vehicle fatalities even while miles driven declined (51). Other evidence suggests that heavy drinking increased and physical activity declined during the pandemic (52,53), although these might not be fully reflected by the causes of death examined here.

Many study limitations reflected the necessity of using provisional mortality data which were incomplete, reported with delays and provided information for only a restricted set of causes (54). Some fatalities attributed to COVID-19 may have been misclassified and its involvement in other deaths could have been missed. The weekly death reports, used for parts of this analysis, did not exactly correspond with calendar months or years, requiring additional adjustments. As emphasized, there is inherent uncertainty in the calculation of baseline deaths, and the share of excess mortality directly attributable to the virus may have changed over the course of the epidemic. The methods developed here almost certainly improved upon earlier national estimates of excess deaths occurring over a lengthy timespan, but they may be less useful for corresponding calculations at the state-level or over shorter durations. Finally, this research did not evaluate adverse consequences other than mortality, such as increased morbidity, declining economic circumstances, and some costs of social isolation.

References

1. Gill JR, DeJoseph ME. The Importance of Proper Death Certification During the COVID-19 Pandemic. *JAMA* [electronic article]. 2020;324(1):27. (https://jamanetwork.com/journals/jama/fullarticle/2767262). (Accessed May 27, 2021)

2. Baum A, Schwartz MD. Admissions to Veterans Affairs Hospitals for Emergency Conditions During the COVID-19 Pandemic. *JAMA*. 2020;324(1):96–99.

3. Jeffery MM, D'Onofrio G, Paek H, et al. Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. *JAMA Intern Med* [electronic article]. 2020;180(10):1328. (https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2768777). (Accessed May 27, 2021)

4. Solomon MD, McNulty EJ, Rana JS, et al. The Covid-19 Pandemic and the Incidence of Acute Myocardial Infarction. *N Engl J Med* [electronic article]. 2020;383(7):691–693. (http://www.nejm.org/doi/10.1056/NEJMc2015630). (Accessed May 27, 2021)

5. Ruhm CJ. Are Recessions Good for Your Health? *The Quarterly Journal of Economics* [electronic article]. 2000;115(2):617–650. (https://academic.oup.com/qje/article-lookup/doi/10.1162/003355300554872). (Accessed May 27, 2021)

6. Waetjen, David, Shilling, Fraser. Rapid Reporting of Vehicle Crash Data in California to Understand Impacts from COVID-19 Pandemic on Traffic and Incidents. 2021;(https://escholarship.org/uc/item/078193ht). (Accessed May 27, 2021)

7. Jones N. How COVID-19 is changing the cold and flu season. *Nature* [electronic article]. 2020;588(7838):388–390. (https://www.nature.com/articles/d41586-020-03519-3). (Accessed May 27, 2021)

8. Flegal KM, Graubard BI, Williamson DF, et al. Excess deaths associated with underweight, overweight, and obesity. *JAMA*. 2005;293(15):1861–1867.

9. Buehler JW, Devine OJ, Berkelman RL, et al. Impact of the human immunodeficiency virus epidemic on mortality trends in young men, United States. *Am J Public Health*. 1990;80(9):1080–1086.

10. Santos-Lozada AR, Howard JT. Use of Death Counts From Vital Statistics to Calculate Excess Deaths in Puerto Rico Following Hurricane Maria. *JAMA* [electronic article]. 2018;320(14):1491. (http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2018.10929). (Accessed May 27, 2021)

11. Nitschke M, Tucker GR, Hansen AL, et al. Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: a case-series analysis. *Environ Health* [electronic article]. 2011;10(1):42. (http://ehjournal.biomedcentral.com/articles/10.1186/1476-069X-10-42). (Accessed May 27, 2021)

12. Woolf SH, Chapman DA, Sabo RT, et al. Excess Deaths From COVID-19 and Other Causes, March-April 2020. *JAMA*. 2020;324(5):510–513.

13. Weinberger DM, Chen J, Cohen T, et al. Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020. *JAMA Intern Med*. 2020;180(10):1336–1344.

14. Woolf SH, Chapman DA, Sabo RT, et al. Excess Deaths From COVID-19 and Other Causes, March-July 2020. *JAMA*. 2020;324(15):1562–1564.

15. Rossen LM, Branum AM, Ahmad FB, et al. Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity - United States, January 26-October 3, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(42):1522–1527.

16. Faust JS, Du C, Mayes KD, et al. Mortality From Drug Overdoses, Homicides, Unintentional Injuries, Motor Vehicle Crashes, and Suicides During the Pandemic, March-August 2020. *JAMA*. 2021;

17. Woolf SH, Chapman DA, Sabo RT, et al. Excess Deaths From COVID-19 and Other Causes in the US, March 1, 2020, to January 2, 2021. *JAMA*. 2021;

18. Rossen LM, Branum AM, Ahmad FB, et al. Notes from the Field: Update on Excess Deaths Associated with the COVID-19 Pandemic - United States, January 26, 2020-February 27, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(15):570–571.

19. Institute for Health Metrics and Evaluation. Estimation of total mortality due to COVID-19. 2021;(http://www.healthdata.org/special-analysis/estimation-excess-

mortality-due-covid-19-and-scalars-reported-covid-19-deaths). (Accessed June 2, 2021)

20. Sanmarchi F, Golinelli D, Lenzi J, et al. Exploring the Gap Between Excess Mortality and COVID-19 Deaths in 67 Countries. *JAMA Netw Open*. 2021;4(7):e2117359.

21. Althoff KN, Schlueter DJ, Anton-Culver H, et al. Antibodies to SARS-CoV-2 in *All of Us* Research Program Participants, January 2-March 18, 2020. *Clinical Infectious Diseases* [electronic article]. 2021;ciab519. (https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciab519/6294073). (Accessed June 15, 2021)

22. Centers for Disease Control and Prevention. National Center for Health Statistics Mortality Data on CDC WONDER. *CDC Wonder*. 2021;(https://wonder.cdc.gov/mcd.html). (Accessed March 15, 2021)

23. National Center for Health Statistics. COVID-19 Death Data and Resources - National Vital Statistics System. 2021;(https://www.cdc.gov/nchs/covid19/covid-19-mortality-data-files.htm). (Accessed March 17, 2021)

24. Centers for Disease Control and Prevention. Monthly Provisional Counts of Deaths by Select Causes, 2020-2021 (updated 11/3/2021). (https://data.cdc.gov/NCHS/Monthly-Provisional-Counts-of-Deaths-by-Select-Cau/9dzk-mvmi). (Accessed November 5, 2021)

25. Centers for Disease Control and Prevention. Provisional COVID-19 Deaths by Week, Sex, and Age (updated 11/3/2021). (https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Week-Sex-and-Age/vsak-wrfu). (Accessed November 5, 2021)

26. Centers for Disease Control and Prevention. Weekly Counts of Deaths by Jurisdiction, and Race and Hispanic Origin (updated 11/3/2021). (https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-Jurisdiction-and-Race-a/qfhf-uhaa). (Accessed November 5, 2021)

27. Centers for Disease Control and Prevention. Weekly Counts of Deaths by State and Select Causes, 2014-2019 (updated 7/22/21). (https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-State-and-Select-Causes/3yf8-kanr). (Accessed September 15, 2021)

28. Centers for Disease Control and Prevention. Weekly Provisional Counts of Deaths by State and Select Causes, 2020-2021 (updated 11/3/2021).

(https://data.cdc.gov/NCHS/Weekly-Provisional-Counts-of-Deaths-by-State-and-S/muzy-jte6). (Accessed November 5, 2021)

29. Petrova VN, Russell CA. The evolution of seasonal influenza viruses. *Nat Rev Microbiol* [electronic article]. 2018;16(1):47–60. (http://www.nature.com/articles/nrmicro.2017.118). (Accessed November 10, 2021)

30. Zylke JW, Bauchner H. Mortality and Morbidity: The Measure of a Pandemic. *JAMA*. 2020;324(5):458–459.

31. Bauchner H, Fontanarosa PB. Excess Deaths and the Great Pandemic of 2020. *JAMA*. 2020;324(15):1504–1505.

32. Garber AM. Learning From Excess Pandemic Deaths. *JAMA*. 2021;

33. Kathail P, Chin A. Further Evidence that COVID-19 Disproportionately Impacts African American, Hispanic, and Low-Income Populations - Datavant. (https://datavant.com/2020/06/18/further-evidence-that-covid-19-disproportionately-impacts-african-american-hispanic-and-low-income-populations/). (Accessed June 1, 2021)

34. Stuart B. How The COVID-19 Pandemic Has Affected Provision Of Elective Services: The Challenges Ahead | Health Affairs. (https://www.healthaffairs.org/do/10.1377/hblog20201006.263687/full/). (Accessed June 1, 2021)

35. Moroni F, Gramegna M, Ajello S, et al. Collateral Damage: Medical Care Avoidance Behavior Among Patients With Myocardial Infarction During the COVID-19 Pandemic. *JACC: Case Reports* [electronic article]. 2020;2(10):1620–1624. (https://linkinghub.elsevier.com/retrieve/pii/S2666084920303697). (Accessed June 2, 2021)

36. Uchino K, Kolikonda MK, Brown D, et al. Decline in Stroke Presentations During COVID-19 Surge. *Stroke*. 2020;51(8):2544–2547.

37. Basu P, Alhomoud S, Taghavi K, et al. Cancer Screening in the Coronavirus Pandemic Era: Adjusting to a New Situation. *JCO Global Oncology* [electronic article]. 2021;(7):416–424. (https://ascopubs.org/doi/10.1200/GO.21.00033). (Accessed June 2, 2021)

38. Sharpless NE. COVID-19 and cancer. *Science*. 2020;368(6497):1290.

39. Saraswathula A, Gourin CG, Stewart CM. National Trends in US Otolaryngology Surgical Volume During the Early COVID-19 Pandemic. *JAMA Otolaryngol Head Neck Surg* [electronic article]. 2021;147(4):397. (https://jamanetwork.com/journals/jamaotolaryngology/fullarticle/2776451). (Accessed June 2, 2021)

40. Muniangi-Muhitu H, Akalestou E, Salem V, et al. Covid-19 and Diabetes: A Complex Bidirectional Relationship. *Front Endocrinol (Lausanne)*. 2020;11:582936.

41. Brown EE, Kumar S, Rajji TK, et al. Anticipating and Mitigating the Impact of the COVID-19 Pandemic on Alzheimer's Disease and Related Dementias. *Am J Geriatr Psychiatry*. 2020;28(7):712–721.

42. Paulin E. Is Isolation Killing America's Nursing Home Residents? *AARP*. (https://www.aarp.org/caregiving/health/info-2020/covid-isolation-killing-nursing-home-residents.html). (Accessed June 2, 2021)

43. Wan W. Pandemic Isolation has Killed Thousands of Alzheimer's Patients While Families Watch from Afar. *The Washington Post*. 2020;(https://www.washingtonpost.com/health/2020/09/16/coronavirus-dementia-alzheimers-deaths/). (Accessed June 2, 2021)

44. Cronin CJ, Evans WN. Nursing Home Quality, COVID-19 Deaths, and Excess Mortality. National Bureau of Economic Research; 2020 (Accessed June 3, 2021).(https://www.nber.org/papers/w28012). (Accessed June 3, 2021)

45. Case A, Deaton A. Deaths of despair and the future of capitalism. Princeton: Princeton University Press; 2020.

46. Ruhm CJ. Drivers of the fatal drug epidemic. *J Health Econ*. 2019;64:25–42.

47. Aronson L. Opinion | For Older People, Despair, as Well as Covid-19, Is Costing Lives. *The New York Times*. 2020;(https://www.nytimes.com/2020/06/08/opinion/coronavirus-elderly-suicide.html). (Accessed June 2, 2021)

48. Gunnell D, Appleby L, Arensman E, et al. Suicide risk and prevention during the COVID-19 pandemic. *The Lancet Psychiatry* [electronic article]. 2020;7(6):468–471. (https://linkinghub.elsevier.com/retrieve/pii/S2215036620301711). (Accessed May 13, 2021)

49. Petterson S, Westfall JM, Miller BF. Projected Deaths of Despair from COVID-19. Well Being Trust; 2020 (Accessed June 2, 2021).(https://wellbeingtrust.org/wp-content/uploads/2020/05/WBT_Deaths-of-Despair_COVID-19-FINAL-FINAL.pdf). (Accessed June 2, 2021)

50. Mulligan CB. Deaths of Despair and the Incidence of Excess Mortality in 2020. National Bureau of Economic Research; 2020 (Accessed June 3, 2021).(https://www.nber.org/papers/w28303). (Accessed June 3, 2021)

51. National Safety Council. Motor Vehicle Deaths in 2020 Estimated to be Highest in 13 Years, Despite Dramatic Drops in Miles Driven - National Safety Council. (https://injuryfacts.nsc.org/motor-vehicle/overview/preliminary-monthly-estimates/data-details/). (Accessed June 2, 2021)

52. Pollard MS, Tucker JS, Green HD. Changes in Adult Alcohol Use and Consequences During the COVID-19 Pandemic in the US. *JAMA Netw Open.* 2020;3(9):e2022942.

53. Stockwell S, Trott M, Tully M, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. *BMJ Open Sport Exerc Med* [electronic article]. 2021;7(1):e000960. (https://bmjopensem.bmj.com/lookup/doi/10.1136/bmjsem-2020-000960). (Accessed June 3, 2021)

54. Ahmad FB, Cisewski JA, Miniño A, et al. Provisional Mortality Data — United States, 2020. *MMWR Morb. Mortal. Wkly. Rep.* [electronic article]. 2021;70(14). (http://www.cdc.gov/mmwr/volumes/70/wr/mm7014e1.htm?s_cid=mm7014e1_w). (Accessed April 1, 2021)

Table 1. Specific Causes of Death Analyzed, with ICD-10 Codes [a]

Cause of Death (Abbreviation)	ICD-10 Codes
COVID-19 (Covid)	U07.1
Diseases of the Heart (Heart)	I00-I09, Ill I13, I20-I51
Malignant Neoplasms (Cancer)	C00-C97
Cerebrovascular Disease (Stroke)	I60-I69
Alzheimer's Disease	G30
Diabetes Mellitus (Diabetes)	E10-E14
Chronic Lower Respiratory Disease (Respiratory)	J40-J47
Influenza or Pneumonia (Flu/Pneumonia)	J09-J18
Nephritis, Nephrotic Syndrome, Nephrosis (Kidney)	N00-N07, N17-N19, N25-N27
All External Causes (External)	V01-Y89
Unintentional Injuries (Accidents)	V01-X59, Y85-Y86
Drug Overdoses (Drug)	X40-X44, X60-X64, X85, Y10-Y14
Intentional Self-Harm (Suicide)	U03, X60-X84, Y87.0
Motor Vehicle Accidents (Vehicle)	V02-V89.2 with exceptions [b]
Assault (Homicide)	U01-U02, X85-Y09, Y87.1

[a] ICD-10 codes refer to the underlying cause of death, except for COVID-19, where they indicate either an underlying or contributing cause.

[b] The full listing of ICD-10 codes for vehicle fatalities is (V02–V04, V09.0, V09.2,V12–V14,V19.0–V19.2,V19.4–V19.6, V20–V79,V80.3–V80.5, V81.0–V81.1,V82.0–V82.1,V83–V86, V87.0–V87.8,V88.0–V88.8, V89.0,V89.2).

Table 2. Summary of Results for Selected Prior Studies

Study	Time Period	Excess Deaths [a] # (95% CI) {Lower Bound}	Excess Deaths [a] Attributed to COVID-19, % (95% CI), [Upper Bound] [b]	Baseline vs. Actual Deaths 12-Months Earlier: % Δ (95% CI) [Upper Bound] [c]
Sanmarchi et al. (20)	2/26/20 - 12/31/20 [d]	453,521 (421,371 to 485,671)	83.2 (77.7 to 89.6) [e]	0.86 (-0.48 to 2.20)
Rossen et al. (18)	1/26/20 - 2/27/21	660,200 {545,600}	75.4 [88.3]	-1.01 [2.63]
Woolf et al. (17) [f]	3/1/20 - 1/2/21	522,368 (521,413 to 523,332)	72.4 (72.2 to 72.5)	-1.00 (-1.04 to -0.96)
IHME (19)	3/1/20 - 4/24/21 [g]	912,345	63.4	-10.16

[a] Excess deaths calculated in original study.

[b] COVID-19 fatalities include cases where it was identified as the underlying or a contributing cause of death.

[c] Baseline deaths were provided by the authors of some studies and computed for others, as described in Appendix. Deaths 12-months earlier were computed as part of the current analysis in all cases.

[d] Time period for calculating baseline and prior deaths was slightly modified from original study as discussed in Appendix.

[e] Original study provided COVID-19 underlying cause of death numbers. These were converted to multiple cause of death estimates, as described in Appendix.

[f] North Carolina excluded from analysis

[g] Time period was not provided in original study and so was estimated as discussed in the Appendix.

Table 3. Actual, Excess and COVID-19 Deaths, March 2020 to February 2021, All Individuals and Selected Demographic Groups

Group	Actual Deaths # [a]	Excess Deaths, # (95% CI) [b]	Observed-Baseline Death Ratio (95% CI)	COVID-19 Deaths, # [c]	% of Excess Deaths Attributed to COVID-19 (95% CI)
All	3,534,076	646,514 (597,507 to 695,520)	1.22 (1.20 to 1.25)	538,924	83.4% (77.5% to 90.2%)
Male	1,859,870	358,488 (336,331 to 380,645)	1.24 (1.22 to 1.26)	295,195	82.3% (77.6% to 87.8%)
Female	1,674,206	288,221 (259,638 to 316,804)	1.21 (1.18 to 1.23)	243,729	84.6% (76.9% to 93.9%)
White [d]	2,566,117	364,917 (312,500 to 417,334)	1.17 (1.14 to 1.19)	327,320	89.7% (78.4% to 104.7%)
Black [d]	472,671	114,045 (108,858 to 119,231)	1.32 (1.30 to 1.34)	80,573	70.7% (67.6% to 74.0%)
Hispanic	339,054	113,968 (111,937 to 115,999)	1.51 (1.49 to 1.52)	99,552	87.4% (85.8% to 88.9%)
Other [d]	156,234	51,971 (51,038 to 52,903)	1.50 (1.49 to 1.51)	31,479	60.6% (59.5% to 61.7%)
Age: <25	64,872	6,707 (4,493 to 8,920)	1.12 (1.07 to 1.16)	1,023	15.3% (11.5% to 22.8%)
Age: 25-44	185,347	39,253 (34,917 to 43,589)	1.27 (1.23 to 1.31)	12,841	32.7% (29.5% to 36.8%)
Age: 45-64	661,442	125,011 (118,929 to 131,093)	1.23 (1.22 to 1.25)	90,865	72.7% (69.3% to 76.4%)
Age: ≥ 65	2,622,416	452,051 (430,292 to 473,810)	1.21 (1.20 to 1.22)	434,194	96.0% (91.6% to 100.9%)

[a] Actual deaths reported using data from Centers for Disease Control and Prevention, as described in METHODS.

[b] Excess deaths computed as difference between actual and Baseline deaths. Baseline deaths were estimated using data from March 2009 through February 2020 and procedures described in METHODS.

[c] Includes deaths with COVID-19 identified as the underlying or a contributing cause of death.

[d] Excludes Hispanics.

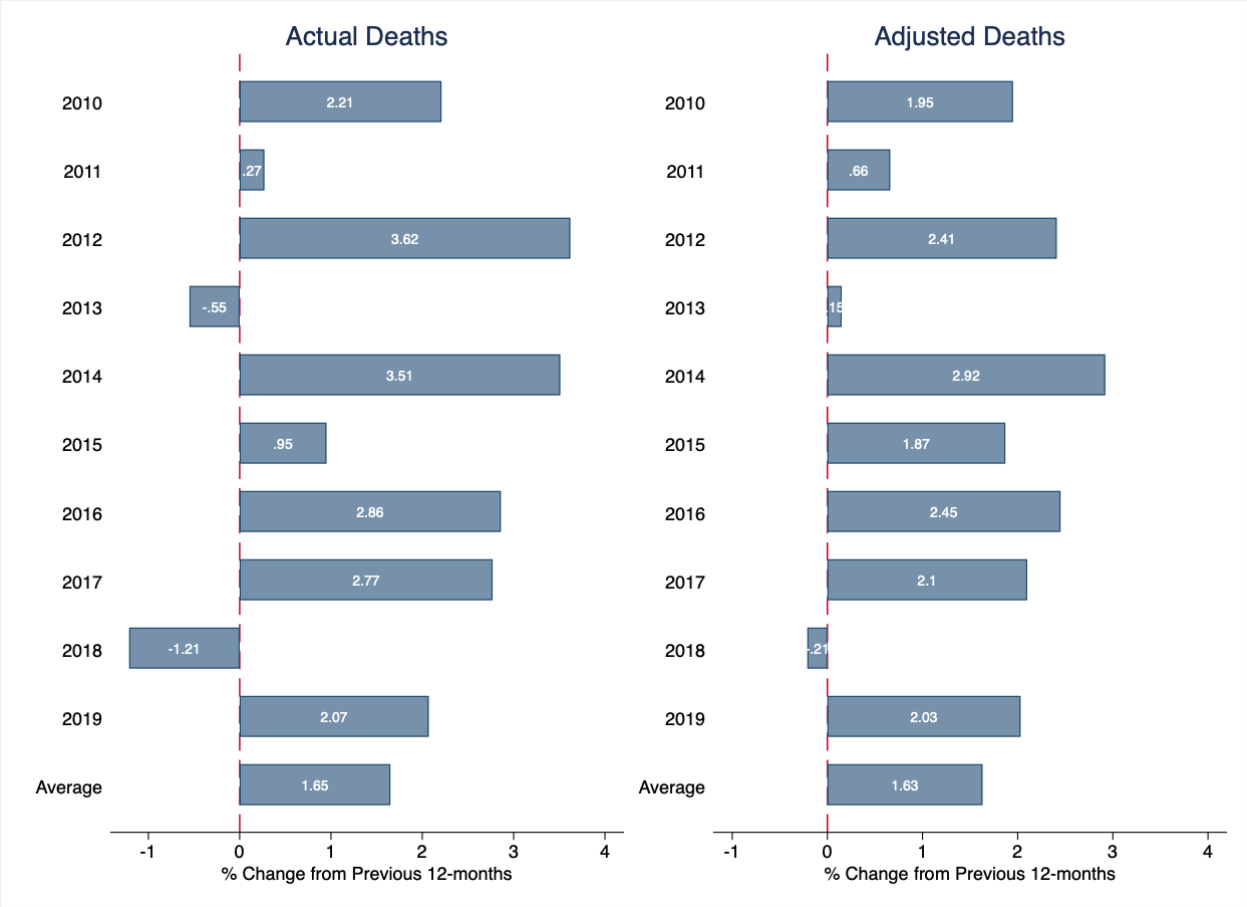
Table 4. Actual, Excess and COVID-19 Deaths, March 1, 2020 to February 28, 2021, Selected Causes

Cause	Actual Deaths, # [a]	Excess Deaths, # (95% CI)[b]	Actual-Baseline Death Ratio (95% CI)
Heart	701,498	33,133 (26,312 to 39,953)	1.05 (1.04 to 1.06)
Cancer	599,273	-1,840 (-4,454 to 775)	1.00 (0.99 to 1.00)
Stroke	161,779	10,482 (7,281 to 13,683)	1.07 (1.05 to 1.09)
Alzheimer's	135,452	11,386 (7,277 to 15,495)	1.09 (1.06 to 1.13)
Diabetes	104,982	14,379 (13,026 to 15,733)	1.16 (1.14 to 1.18)
Lower Respiratory	146,549	-7,314 (-12,215 to -2,414)	0.95 (0.92 to 0.98)
Influenza/Pneumonia	48,479	-2,026 (-7,035 to 2,982)	0.96 (0.87 to 1.07)
Kidney	52,815	511 (-10 to 1,031)	1.01 (1.00 to 1.02)
All External	290,433	31,309 (22,317 to 40,302)	1.12 (1.08 to 1.16)
Accidents	207,145	28,688 (21,101 to 36,275)	1.16 (1.11 to 1.21)
Drugs	95,622	21,431 (16,199 to 26,664)	1.29 (1.20 to 1.39)
Suicide	45,764	-4,357 (-5,237 to -3,477)	0.91 (0.90 to 0.93)
Vehicle	42,993	3,940 (3,142 to 4,739)	1.10 (1.08 to 1.12)
Homicide	25,107	5,310 (3,401 to 7,219)	1.27 (1.16 to 1.40)

[a] Actual deaths reported using data from Centers for Disease Control and Prevention, as described in METHODS.

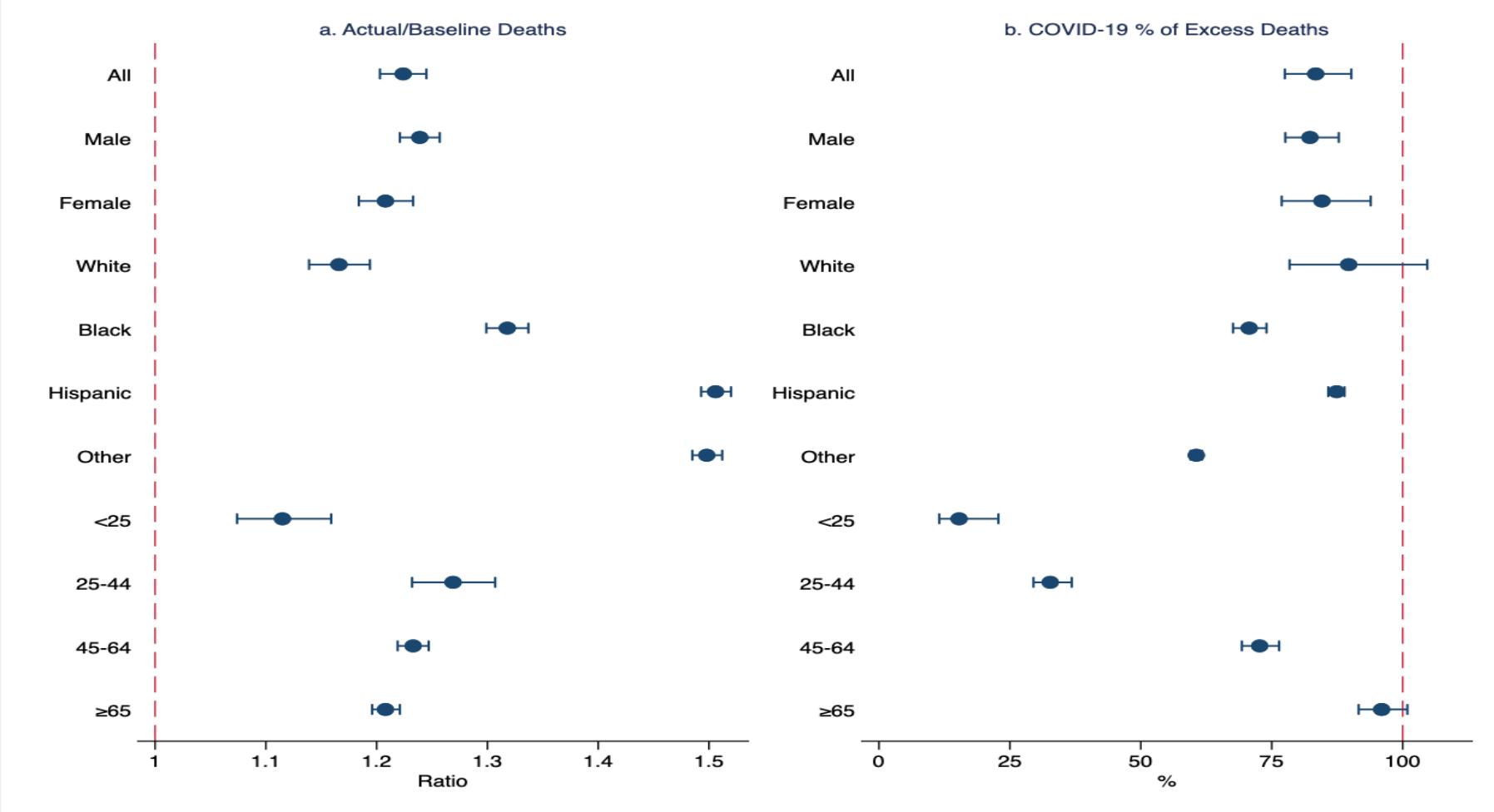
[b] Excess deaths computed as difference between actual and Baseline deaths. Baseline deaths were estimated using data from March 2009 through February 2020 and procedures described in METHODS.

Figure 1. Changes in U.S. Deaths for 12-Month Periods Starting on March 1



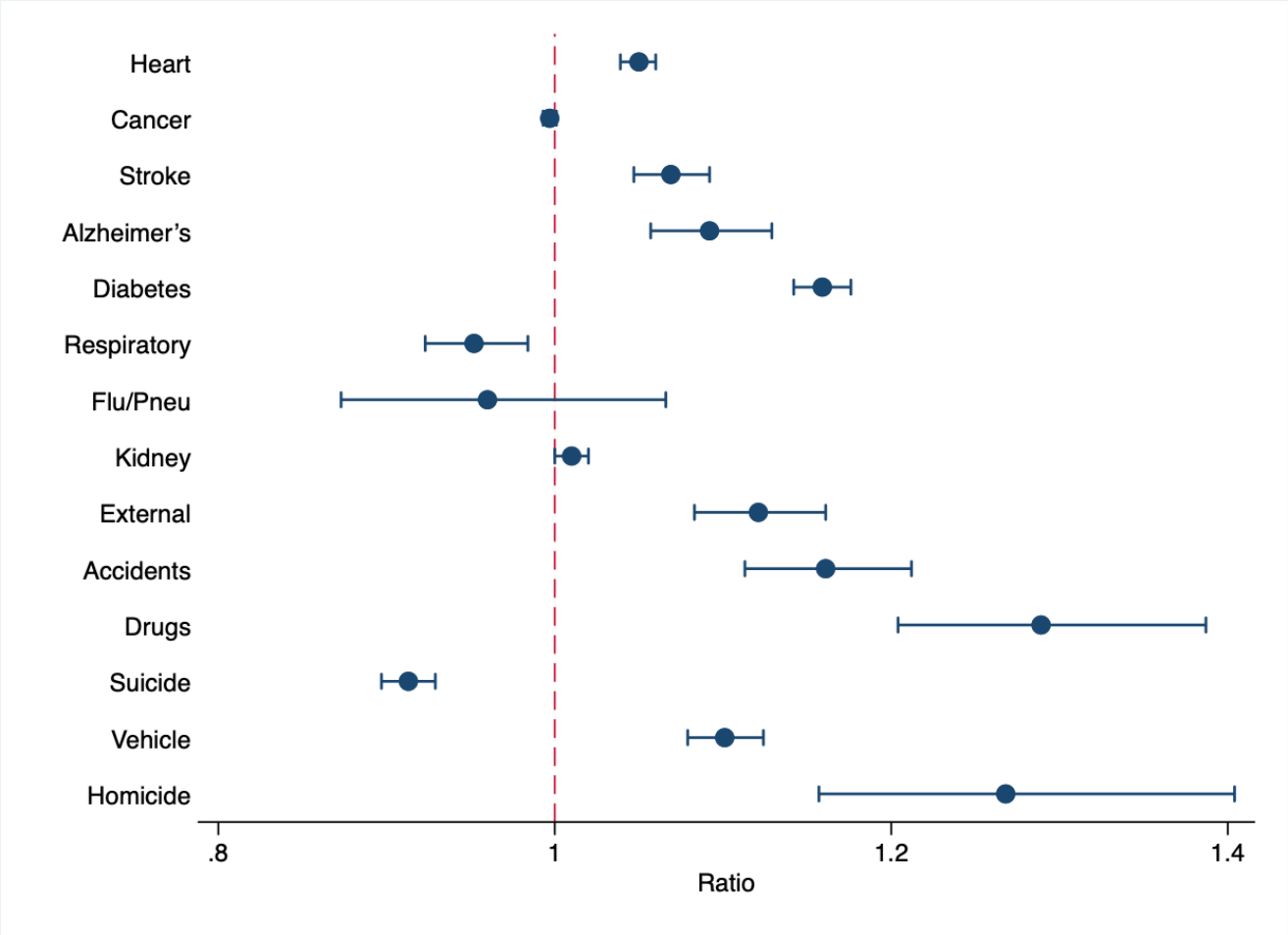
Note: Figure shows percentage changes in deaths for 12-month periods beginning in March 1 of specified year, relative to deaths in the prior 12 months. Deaths were computed using the mortality data, with adjustments to the counts made for February deaths in leap years. Adjusted deaths were calculated by taking actual deaths for 11 months of the (March through February) year, with January fatalities excluded, then multiplying this by the average ratio of all deaths to the 11-month counts over the entire, 2009-2019 year period, as described in METHODS. The standard deviation in the annual change in deaths is 1.70% using the actual data and 1.05% with the adjusted data.

Figure 2. Ratio of Actual vs. Baseline Deaths and Percent of Excess Deaths Involving COVID-19, by Population Group



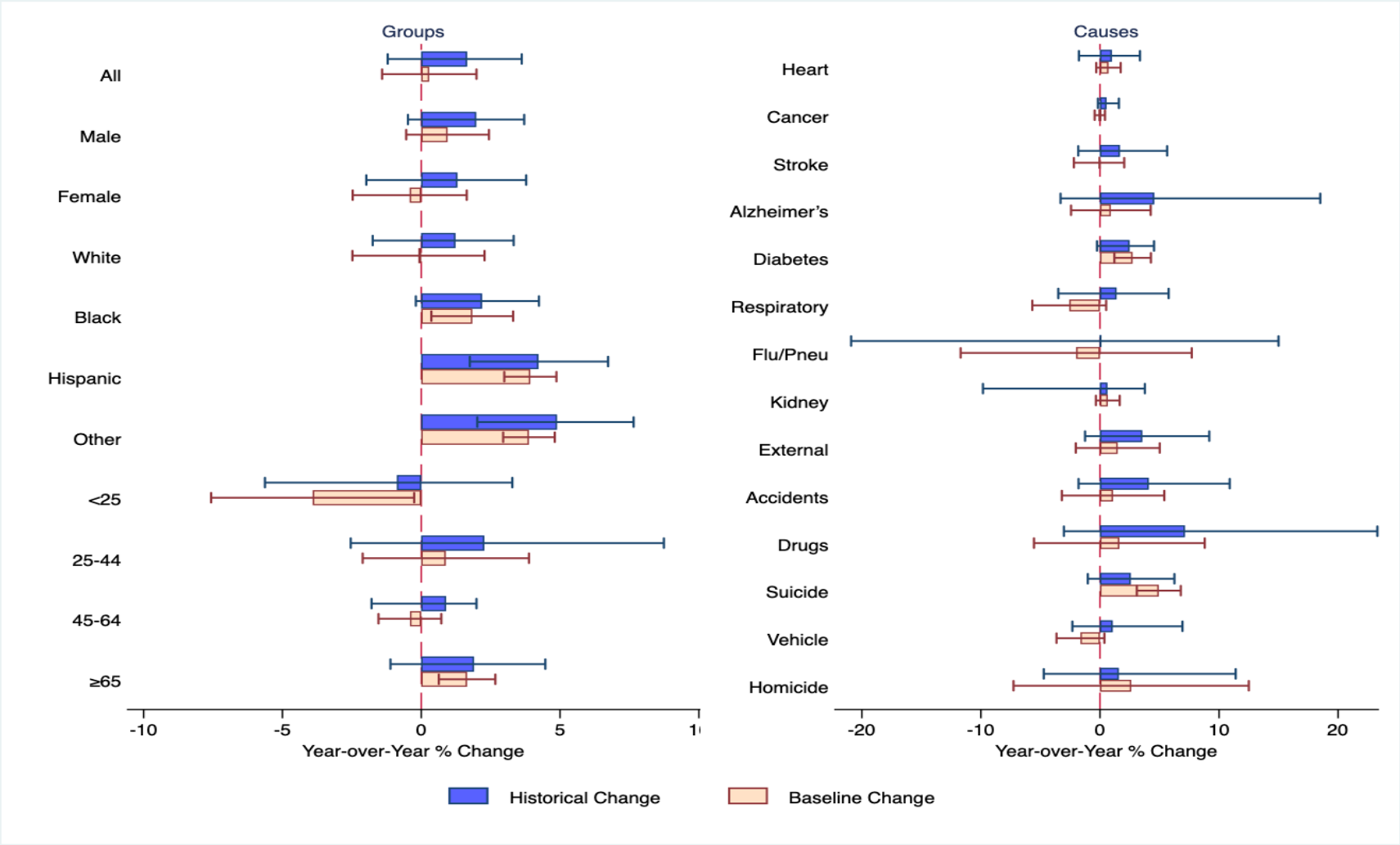
Note: Actual deaths were obtained from the Centers for Disease Control and Prevention Provisional Data for the period March 2020 through February 2021; baseline deaths were estimated using data from March 2009 through February 2020, as described in METHODS. Whiskers show 95% confidence intervals.

Figure 3. Ratio of Actual vs. Baseline Deaths, by Cause



Note: Actual deaths were obtained from the Centers for Disease Control and Prevention Provisional Data for the period March 2020 through February 2021; baseline deaths were estimated using data from March 2009 through February 2020, as described in METHODS. Whiskers show 95% confidence intervals.

Figure 4. Historical Average Annual Percent Change in Deaths (2009-2019) and Implied Change in Baseline 2020 vs. Actual 2019 Deaths



Note: Figure shows historical average year-over-year percentage changes in deaths from 2009-2019 and implied percentage change in 2020 baseline deaths versus actual deaths in 2019. Whiskers show the range of year-over-year historical changes in actual changes (left panel) and 95% CI on estimated baseline changes (right panel).

Appendix: Supplemental Methods and Results

Original Analysis of Excess Deaths

Data

Actual and Baseline Death Counts

Testing Adequacy of the Prediction Methods

Excess Death Estimates from Prior Studies

Table A1. Year-Over-Year Changes in Monthly Deaths and in January Deaths by Cause

Table A2. Analysis Period and Model Specification Minimizing Root Mean Square Error (RMSE)

Table A3. Baseline Deaths by Group and Cause of Death

Table A4. Average Annual Historical Change in Deaths (2009-2019) and Implied Change in Baseline 2020 vs. Actual 2019 Deaths

Original Analysis of Excess Deaths

Data

Data on deaths from 2009-2019 came from the *Multiple Cause of Death* files which provided information from death certificates on a single underlying cause of death (UCD), up to 20 contributory causes, and demographic variables (22). Data were utilized on cause of death, using four-digit *International Classification of Diseases, Tenth Revision* (ICD-10) codes, age, race/ethnicity, gender, and month and year of death. The data covered the universe of fatalities occurring in the United States.

Final death counts in 2020 and 2021 were not available at the time of this study. Therefore, mortality data for January 2020 through February 2021 were obtained from files produced by the Centers for Disease Control and Prevention, as part of their *Provisional Death Counts for Coronavirus Disease (COVID-19)* series (23). Specifically, separate data sets supplied information on monthly death counts from various causes (24); weekly counts by sex and age (25); and weekly counts by race and Hispanic origin (26). All three datasets were updated on November 3, 2021.

Using these data, total mortality counts were computed as were those for selected causes and by sex, race/ethnicity, and age. For brevity, non-Hispanic whites, blacks, and other races are referred to below as whites, blacks, and other races. The following causes of death were examined: COVID-19; diseases of the heart (heart); malignant neoplasms (cancer); cerebrovascular disease (stroke); Alzheimer's disease; diabetes mellitus (diabetes); chronic lower respiratory disease (respiratory); influenza or pneumonia (flu/pneumonia); nephritis, nephrotic syndrome or nephrosis (kidney); all external causes (external); unintentional injuries (accidents); motor vehicle accidents (vehicle); drug overdoses (drug); intentional self-harm (suicide); and assault (homicide). Mortality from other causes was not analyzed because they were not included in the provisional data files or were either relatively rare or not descriptively classified. The latter included deaths from septicemia; other diseases of the respiratory system; and symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified.

The provisional death data contained a number of limitations. First, the counts may have been incomplete and were reported with delays. These delays were generally likely to have had fairly minor effects, since the data sets were updated on November 3, 2021 and the most recent deaths analyzed were from February 2021. However, accident, drug, suicide, vehicle, and homicide mortality were reported with a six-month lag, raising potential concern about under-reporting at the end of the sample period for these deaths. Second,

deaths were reported by location of occurrence rather than residence of the individual and approximately 0.2% of fatalities occurred among non-US residents (15). Third, information was only available for certain causes of death and these were sometimes overlapping. Fourth, coronavirus disease deaths were reported if COVID-19 contributed to the death, whether or not it was the underlying cause, but this classification was probably incomplete. Fifth, while it was straight-forward to aggregate the monthly death totals to obtain the counts for January and February of 2020 (which were combined with other data for the computing counts in the 12-months preceding the pandemic) and for March 2020 through February 2021, the procedures for computing annual deaths using weekly data were more complicated. To calculate mortality in January and February of 2020, death counts for the weeks ending 1/4/20 through 2/29/20 were used, with 4/7 of fatalities in the first of these weeks attributed to 2020. For the March 2020 through February 2021 counts, data were used for the weeks ending 3/7/20 through 2/27/21, with the number of deaths for the last of these weeks multiplied by 8/7. Adjustments were also made for the extra day in February during the leap years of 2012, 2016, and 2020.

Actual and Baseline Death Counts

For simplicity, throughout the remainder of the discussion, each March through February period will often be denoted by the year with the majority of these months. For example, 2012 refers to March 2012 through February 2013 and 2020 indicates March 2020 to February 2021.

Observed fatality counts for all deaths or for the specified group or cause were calculated for each month between March 2009 and February 2021 and then aggregated over the 12-month March to February periods to provide observed mortality counts, D_{it}, for all deaths or specified group or cause i in year t. Year-over-year percentage changes were computed as:

$$\Delta D_{it} = \frac{D_{it} - D_{it-1}}{D_{it-1}} \times 100\%, \tag{A.1}$$

for $t = 2010 - 2019$.

Changes in monthly total death counts, compared to the same month in the prior year, were also calculated and these revealed substantial excess idiosyncratic volatility in January, relative to other months. For instance, the standard deviation in the year-over-year January change was more than twice as large as the next highest month and almost five times as big as for the average of the eleven non-January months (see Table A1). This was partly due to the higher average number of deaths in January than other months but mostly for

other reasons. This can be seen by noting that the coefficient of variation of the year-over-year January change is 4.3 times the average for the other eleven months and 1.9 times that of December, the month with the next highest coefficient of variation. This excess volatility, which preliminary work indicated was due to a variety of causes including, importantly but not exclusively, the severity of seasonal influenza/pneumonia epidemics, increased the difficulty in forecasting baseline deaths. Therefore, these were calculated based on an adjustment to mortality counts described next.

Define deaths in March through December plus February (i.e. annual counts without January deaths) in year t by D'_{it} and the ratio of average annual death counts from 2009 - 2019 compared to corresponding average fatalities over this 11-month period as:

$$Ratio_i = \left(\frac{\sum_{t=2009}^{2019} D_{it}}{\sum_{2009}^{2019} D'_{it}} \right) / 11. \tag{A.2}$$

The adjusted mortality count, D_{it}^{adj}, was then calculated as:

$$D_{it}^{adj} = D'_{it} \times Ratio_i. \tag{A.3}$$

Thus, the adjusted number of deaths was the actual 11-month (March through December and February) count for the specified year multiplied by the ratio of annual to 11-month (without January) fatality counts over the entire 11-years. This preserves the overall number of deaths from 2009-2019 but removes the volatility in January year-over-year changes.

The baseline number of 2020 year deaths, D_i^B, was estimated from regressions of adjusted annual mortality counts on a linear or quadratic time trend:

$$D_{it}^{adj} = \alpha_i TR_t + \epsilon_i, \tag{A.4a}$$

or

$$D_{it}^{adj} = \alpha_i TR_t + \beta_i TRSQ_t + \epsilon_i, \tag{A.4b}$$

where TR is the linear time trend, $TRSQ = TR \times TR$, the last year of the regression model is 2019 and the first year varies between from 2010 and 2016. The use of Poisson models, which had theoretical appeal since the dependent variables were counts, was tested for in preliminary analyses. However, there was no evidence that these improved model fit and so the somewhat easier to interpret results for regression specifications were focused upon.

The decision of whether to use a linear or quadratic trend, as well as the time-period of analysis, was determined by estimating separate models with both types of trends and for all possible starting years between 2009 and 2016, with 2019 always the ending year. The model with the smallest root mean squared error ($RMSE$) for the specified group or cause (or total mortality) was used, with $RMSE_j$ for group/cause i and the j^{th} specification defined as:

$$RMSE_{ij} = \left(\frac{SSE_{ij}}{n_{ij}-p_{ij}}\right)^{0.5}, \tag{A.5}$$

where n was the number of years in the analysis period (ranging from 5 to 11) and p was the number of parameters estimated (two for the linear models and three for the quadratic specifications). This formulation of $RMSE$ penalized shorter panels and quadratic (relative to linear) trend specifications, due to the degrees of freedom correction.

Table A2 shows the models used for all deaths, the 10 population groups, and the 14 causes. The starting analysis year was 2009, 2010, 2013, 2014, 2015, and 2016 in 1, 1, 6, 5, 2, and 10 cases respectively. Linear models were used in 14 cases and quadratic specifications in the other 11.

Baseline deaths in 2020 were then estimated from the regression equation with TR and $TRSQ$, where applicable, set to their 2020-year values. Ninety-five percent confidence intervals (95% CI's) were computed using the standard error of the regression predicted value.

Testing Adequacy of the Prediction Methods

One way to examine the plausibility of the baseline estimates was to check whether the implied change in 2020 death counts, relative to the period 12-months earlier, were in the range of observed prior year-over-year actual death counts. These are often referred to as "baseline changes" and "historical changes" for brevity. To implement this test, historical changes, from 2009-2019, were calculated overall and for each group and cause. The average historical change was computed as was the range between the lowest and highest year-over-year changes. Next the baseline change, and the associated 95% CI, were determined from results of the regression procedure discussed above. Close accordance between the baseline and historical annual changes provided evidence that the estimation models were plausible. Conversely, discordant patterns, such as negative estimated baseline changes, despite a common historical pattern of mortality growth, raise questions about the adequacy of the estimation models.

Excess Death Estimates from Prior Studies

The methods used in prior research have often underestimated the number of counterfactual baseline deaths (expected in the absence of COVID-19) and, as a result, overstated both excess deaths and the share of them attributed to causes other than COVID-19. To demonstrate this, implied changes in baseline deaths were calculated for four studies (17–20). The baseline changes were then compared to historical changes from 2009-2019. Here, deaths in 2019 and 2020 were calculated using CDC Provisional COVID-19 datasets, providing information on weekly deaths in 2019 and 2020 (27,28). For all cases, the 12-month lag period was set to start and end on the same day of the week as in the study analysis timespan. For example, when considering excess deaths from March 1, 2020 through January 2, 2021, the 12-month lag was from March 3, 2019 to January 4, 2020, so that both periods contained the same number of days, started on the first Sunday in March and ended on the first Saturday in January. This procedure ensured that the counts were not affected by either day-of-the week or seasonal differences in death rates.

Two of the studies supplied direct information on baseline deaths (17,20), which could be combined with the lagged mortality counts to provide an implied percentage baseline death change. These two investigations also supplied 95% CI's for baseline mortality, or for excess deaths from which those on baseline deaths could be calculated. The other two studies (18,19), provided information on excess deaths but did not indicate either actual or baseline fatality counts. In addition, the Institute for Health Metrics Evaluation (19) analysis did not precisely indicate the period studied. The publication claimed to measure excess deaths from March 2020 through May 10, 2021; however, since it was published on May 13, 2021, data on deaths during the final weeks of this period would not have yet been available. Here, the actual analysis period was assumed to be March 1, 2020 through April 24, 2021. The rationale for this was that data on weekly deaths in the provisional COVID-19 deaths datasets first became available with a 10-14 day lag, so that data through the week ending April 24, 2021 would be the latest available for an analysis conducted approximately 10-days before the final results were published. For these two studies, 2020-2021 deaths counts were calculated using previously released versions of the data set mentioned above (28), designed to roughly correspond to the information available to the authors at the time of their analyses. Specifically, for Rossen et al. (18) the data set was updated on April 21, 2021 and for IHME (19) it was updated on May 5, 2021. If the authors had more recent data (e.g. updated through November 3, as in this paper's original analysis of excess deaths), they would have obtained higher counts of both total and excess deaths, but it is not obvious how baseline death counts would have changed.

Rossen et al. (18) did not supply confidence intervals but did provide lower-bounds on the number of excess deaths which, when combined with the data on actual lagged deaths, permitted calculation of an upper-bound on the change in baseline deaths compared to the 12-month lag. IHME (19) supplied only a point estimate on the number of excess deaths and so no confidence intervals or bounds could be placed on the implied change in baseline deaths compared to actual deaths one-year earlier.

There was an extra complication when evaluating the IHME (19) estimates. Since their analysis period covered (approximately) March 1, 2020 through April 24, 2021, the 12-month lag would include roughly March 1, 2019 through April 24, 2020. However, since substantial numbers of COVID-19 deaths occurred in March and April of 2020, including deaths from these two months in the calculation of lagged counts would have incorporated some effects of the COVID-19 pandemic. As an alternative, lagged counts were constructed using the following procedure. First, actual deaths were determined for March 2019 through February 2020, reflecting lagged effects for the first 12 months of the analysis period (March 2020 through February 2021). Second, for March 1 through April 24 of 2021, lagged deaths were constructed by taking the number of actual fatalities from the same dates in 2019 and then multiplying these by 1.0165. The multiplication factor reflected the average annual historical percentage increase in deaths and was included since this portion of then lag period was 24 rather than 12 months prior to the timespan of the original analysis.

Additional adjustments were also needed for the Sanmarchi et al. (20) investigation. Their stated analysis period was February 26, 2020 through December 31, 2020. However, the number of total and COVID-19 deaths claimed for this timespan did not match those from the CDC provisional data (probably because they used different data sources). After some investigation, I determined that the best match using the CDC data was for 2/27/20 – 12/30/20. Over this period, the provisional death data indicated that there were 2,873,335 deaths, compared to the 2,870,292 reported in the original study, a difference of just 0.1%. Reported deaths with COVID-19 as an underlying cause were also similar (346,036 versus 344,730 in the original study).

A second complication was that Sanmarchi et al. (20) reported cases where COVID-19 was the underlying cause of deaths, whereas most other prior analyses, and the original estimates in this paper, allowed COVID-19 to be either an underlying or a contributory cause of death. To make the Sanmarchi et al. estimates comparable, the underlying cause of death estimates were converted those expected using the broader definition of COVID-19 fatalities. This was done by inflating their original numbers by 10.47%, which reflected the difference between the multiple cause and underlying cause number of COVID-19 deaths during their analysis period in the CDC data.

The implied percentage changes in baseline versus actual lagged deaths were also computed for three studies (12–14) that preceded but were closely related to Woolf et al. (17) and covered earlier and shorter periods. Using similar methods to those described above, the estimated baseline death rates in these investigations were between 2 and 5 percent lower than actual deaths one year earlier, suggesting an even larger understatement of baseline mortality and greater overstatement of excess deaths in those analyses.

Table A1. Year-Over-Year Changes in Monthly Deaths and in January Deaths by Cause [a]

Month/Cause	Mean	Standard Deviation	Coefficient of Variation	Minimum	Maximum
Year-Over-Year Changes by Calendar Month					
1	4,142	21,718	5.24	-29,209	35,715
2	3,581	9,528	2.66	-9,997	15,888
3	3,888	6,387	1.64	-4,359	14,302
4	3,251	2,997	0.92	-677	8,884
5	3,664	3,147	0.86	-822	8,134
6	3,474	2,179	0.63	1,156	6,573
7	3,493	1,947	0.56	588	7,694
8	3,497	1,840	0.53	-56	6,518
9	3,325	2,976	0.90	-2,095	6,540
10	3,058	2,793	0.91	-1,833	6,484
11	3,791	4,105	1.08	-2,371	9,552
12	3,842	10,581	2.75	-9,611	16,610
Year-Over-Year Changes in January by Cause of Death					
All	4,142	21,718	5.24	-29,209	35,715
Heart	555	5,101	9.19	-6,357	8,317
Cancer	311	868	2.79	-788	1,850
Stroke	216	915	4.25	-1,217	1,513
Alzheimer's	316	1,666	5.28	-2,450	3,332
Diabetes	175	617	3.52	-662	1,059
Lower Respiratory	259	2,379	9.19	-2,953	4,208
Influenza/Pneumonia	130	3,712	28.60	-6,460	5,257
Kidney	21	387	18.79	-606	533
All External	671	938	1.40	-915	2,472
Accidents	507	712	1.40	-772	1,994
Drugs	326	604	1.85	-577	1,633
Suicide	95	166	1.74	-207	295
Vehicle	38	163	4.25	-135	315
Homicide	36	127	3.50	-165	227

[a] Table shows summary statistics on year-over-year changes in total monthly deaths and in January deaths by cause for the period 2009-2019.

Table A2. Analysis Period and Model Specification Minimizing Root Mean Square Error (RMSE) [a]

Group/Cause	First Year [b]	Linear/Quadratic [c]
All	2013	Quadratic
Male	2013	Quadratic
Female	2013	Quadratic
White	2014	Quadratic
Black	2016	Linear
Hispanic	2013	Linear
Other	2015	Linear
<25	2016	Linear
25-44	2016	Linear
45-64	2016	Linear
≥65	2014	Linear
Heart	2014	Quadratic
Cancer	2014	Quadratic
Stroke	2013	Quadratic
Alzheimer's	2016	Linear
Diabetes	2013	Quadratic
Lower Respiratory	2014	Quadratic
Influenza/Pneumonia	2010	Quadratic
Kidney	2015	Linear
All External	2016	Linear
Accidents	2016	Linear
Drugs	2016	Linear
Suicide	2009	Linear
Vehicle	2016	Linear
Homicide	2016	Quadratic

[a] RMSE = $[SSE/(n\text{-}p)]^{0.5}$, where SSE is the sum of squared errors, n is the number of years (ranging between 5 and 11) and p is the number of parameters estimated (2 in the linear models and 3 in the quadratic specifications).

[b] First year indicates the starting year of the analysis, which ranges from 2009 to 2016. The end year is always 2019.

[c] Models include either a linear trend variable or a quadratic trend, as noted.

Table A3. Baseline Deaths By Group and Cause of Death

Group	Baseline Deaths (95% CI) [a]	Cause	Baseline Deaths (95% CI) [a]
All	2,887,563 (2,838,556 to 2,936,569)	Heart	668,365 (661,545 to 675,186)
Male	1,501,382 (1,479,225 to 1,523,539)	Cancer	601,113 (598,498 to 603,727)
Female	1,385,985 (1,357,402 to 1,414,569)	Stroke	151,297 (148,096 to 154,498)
White [b]	2,201,200 (2,148,783 to 2,253,617)	Alzheimer's	124,066 (119,957 to 128,175)
Black [b]	358,626 (353,440 to 363,813)	Diabetes	90,603 (89,249 to 91,956)
Hispanic	225,086 (223,055 to 227,117)	Lower Respiratory	153,863 (148,963 to 158,764)
Other [b]	104,263 (103,331 to 105,196)	Influenza/Pneumonia	50,505 (45,497 to 55,514)
Age: <25	58,165 (55,952 to 60,379)	Kidney	52,304 (51,784 to 52,825)
Age: 25-44	146,094 (141,758 to 150,430)	All External	259,124 (250,131 to 268,116)
Age: 45-64	536,431 (530,349 to 542,513)	Accidents	178,457 (170,870 to 186,044)
Age: ≥ 65	2,170,365 (2,148,606 to 2,192,125)	Drugs	74,191 (68,958 to 79,423)
		Suicide	50,121 (49,241 to 51,001)
		Vehicle	39,053 (38,254 to 39,851)
		Homicide	19,797 (17,888 to 21,706)

[a] Estimate of deaths from March 2020 – February 2021 that would have occurred in the absence of the COVID-19 pandemic, calculated as described in METHODS.

[b] Excludes Hispanics.

Table A4. Average Annual Historical Change in Deaths (2009-2019) and Implied Change in Baseline 2020 vs. Actual 2019 Deaths

Group/Cause	Average % Change in Deaths Vs. Prior Year, 2009-2019 (minimum & maximum change) [a]	Change in 2020 Baseline Deaths vs. Actual 2019 Deaths (95% CI) [b]
All	1.7% (-1.2% to 3.6%)	0.3% (-1.4% to 2.0%)
Male	2.0% (-0.5% to 3.7%)	0.9% (-0.5% to 2.4%)
Female	1.3% (-2.0% to 3.8%)	-0.4% (-2.5% to 1.6%)
White	1.2% (-1.8% to 3.3%)	-0.1% (-2.5% to 2.3%)
Black	2.2% (-0.2% to 4.2%)	1.8% (0.4% to 3.3%)
Hispanic	4.2% (1.8% to 6.7%)	3.9% (3.0% to 4.9%)
Other	4.9% (2.0% to 7.7%)	3.9% (3.0% to 4.8%)
<25	-0.9% (-5.6% to 3.3%)	-3.9% (-7.6% to -0.3%)
25-44	2.3% (-2.5% to 8.7%)	0.9% (-2.1% to 3.9%)
45-64	0.9% (-1.8% to 2.0%)	-0.4% (-1.5% to 0.7%)
≥65	1.9% (-1.1% to 4.5%)	1.7% (0.6% to 2.7%)
Heart	1.0% (-1.8% to 3.4%)	0.7% (-0.3% to 1.7%)
Cancer	0.6% (-0.2% to 1.6%)	0.0% (-0.4% to 0.4%)
Stroke	1.7% (-1.8% to 5.6%)	-0.1% (-2.2% to 2.0%)
Alzheimer's	4.6% (-3.3% to 18.5%)	0.9% (-2.4% to 4.3%)
Diabetes	2.5% (-0.2% to 4.5%)	2.7% (1.2% to 4.3%)
Lower Respiratory	1.4% (-3.5% to 5.8%)	-2.6% (-5.7% to 0.5%)
Influenza/Pneumonia	0.1% (-20.9% to 15.0%)	-2.0% (-11.7% to 7.7%)
Kidney	0.6% (-9.8% to 3.8%)	0.7% (-0.3% to 1.7%)
All External	3.6% (-1.3% to 9.2%)	1.5% (-2.0% to 5.0%)
Accidents	4.1% (-1.8% to 10.9%)	1.1% (-3.2% to 5.4%)
Drugs	7.1% (-3.0% to 23.3%)	1.6% (-5.5% to 8.8%)
Suicide	2.6% (-1.0% to 6.3%)	4.9% (3.1% to 6.8%)
Vehicle	1.1% (-2.3% to 6.9%)	-1.6% (-3.7% to 0.4%)
Homicide	1.6% (-4.7% to 11.4%)	2.6% (-7.3% to 12.5%)

[a] Actual deaths refer to March 2009 – February 2019 period, using data from Centers for Disease Control and Prevention, as described in METHODS.

[b] Baseline 2020 (March 2020 through February 2021) deaths, calculated as described in METHODS.