Supplemental Information

COVID-19 and Excess All-Cause Mortality in the US and 20 Comparison Countries, June 2021-March 2022

Alyssa Bilinski[1] · Kathryn Thompson · Ezekiel Emanuel

Data sources

We compared the US overall and the 10 most- and least-vaccinated states to 20 Organization for Economic Co-operation and Development (OECD) countries with 2021 population exceeding 5 million (link) and greater than $25,000 per capita gross domestic product in 2021 (link).

For the US, where we required sub-national data for our analyses, we obtained data from CDC files: COVID-19 mortality from "United States COVID-19 Cases and Deaths by State over Time" (link), all-cause mortality data from "Weekly Counts of Deaths by Jurisdiction and Age" (link), and vaccination data from "COVID-19 Vaccination Trends in the United States, National and Jurisdictional" (link). US population data were obtained from the Census (link, link).

For other countries, we obtained data on COVID-19 mortality from the World Health Organization (link), all-cause mortality estimates from OECD.Stat (link), and vaccination data from Our World In Data, which aggregates local estimates (link). We checked data sources by matching US CDC data to WHO and OECD mortality estimates and spot-checking OECD all-cause mortality against country-specific estimates.

COVID-19 mortality was reported daily, and we aggregated weeks beginning on Sunday ("CDC" or "epi weeks"). All-cause mortality was reported weekly. In the US, new weeks began on Sunday ("epi weeks"), but in other countries, weekly data began on Monday ("ISO weeks"). We defined delta and omicron periods based on visual inspection of mortality trends, during summer 2021 for delta and December 2022 for omicron. Because waves were tightly clustered in time, we defined the start of periods based on the earliest mortality turning point for delta and omicron across all locations of interest. Code and additional analyses are available on GitHub.

Calculations in Tables 1 and 2

Potential US deaths averted

Let r_i be the death rate of interest (reported COVID-19 deaths or excess all cause mortality) per 100,000 in country i, and d be US deaths over the period of interest. Let p be the US population in the year of interest. We estimate the difference in deaths or potential US deaths averted:

$$d - (r_i/100,000) * p$$

[1] Contact: alyssa_bilinski@brown.edu

States by Vaccination Rate

Table S1: States by 2-dose vaccination status (as of January 1, 2022). Nine of the 10 most-vaccinated states and 8 of the 10 least-vaccinated states were consistent over the full study period from June 2021 through March 2022; results were robust to including only these states in the top and bottom 10 over the full period (omitting New York from the top 10 and Indiana and North Dakota from the bottom 10) (see GitHub).

State	Rate	Rank	Set
Vermont	79.0	1	Top 10
Rhode Island	77.6	2	Top 10
Maine	77.0	3	Top 10
Connecticut	75.5	4	Top 10
Hawaii	75.3	5	Top 10
Massachusetts	75.1	6	Top 10
New York	72.7	7	Top 10
Maryland	72.0	8	Top 10
New Jersey	71.8	9	Top 10
District of Columbia	71.3	10	Top 10
Indiana	53.4	42	Bottom 10
North Dakota	53.4	42	Bottom 10
Tennessee	52.4	44	Bottom 10
Arkansas	52.3	45	Bottom 10
Georgia	52.2	46	Bottom 10
Idaho	52.0	47	Bottom 10
Louisiana	51.1	48	Bottom 10
Mississippi	49.6	49	Bottom 10
Wyoming	49.1	50	Bottom 10
Alabama	48.9	51	Bottom 10

Excess Mortality Model Selection (Table 2)

We used the following procedure to select models for estimating excess mortality, designed to address concerns about appropriately fitting secular trends raised in prior work (link). In response to this issue, some authors have fit parametric pre-pandemic trends (e.g., link, link) while others have used only 2019 as a benchmark year for counterfactual non-pandemic mortality (e.g., link). To choose functional form, we evaluated 3 possible models on pre-pandemic data: 1) a model fit only on the most recent year of data, 2) models with country-specific trends as fixed effects, and 3) models with country-specific trends random effects.

$$y_{ctk} = W_{ct} + \epsilon_{ctk} \quad \text{fit to only the most-recent year of data} \tag{1}$$

$$y_{ctk} = W_{ct} + \beta_c^F k + \epsilon_{ctk} \tag{2}$$

$$y_{ctk} = W_{ct} + \beta_c^R k + \epsilon_{ctk} \tag{3}$$

where y_{ctk} was mortality per 100,000 population in week t of year k in location c, W_{ct} was a week-location fixed effect, k indicated year, and ϵ_{ctk} was residual error. Country-specific linear trend parameters β_c^F were estimated as fixed effects while β_c^R were estimated as random effects. We considered both models fit on all data and models fit separately on the set of weeks in each period (delta: 26-51 and omicron: 52-53, 1-12).

We first used 2018 and 2019 as test data. We fit models (1)-(3) on data from 2015-2017, predicting out-of-sample mortality for each country-week in 2018, and 2015-2018, predicting out-of-sample mortality in 2019. We estimated root mean-squared error for each period of weeks (delta: 26-51 and omicron: 52-53, 1-12) ($\mathcal{P}$) and test year (K) at the week-level $\left(RMSE_{\mathcal{P},K} = \sqrt{\frac{1}{N_c}\sum_{c\in\mathcal{C}}\sum_{t\in\mathcal{P}}\left(y_{ctK} - \hat{y}_{ctK}\right)^2}\right)$ and at the period-level $\left(RMSE^{CP}_{\mathcal{P},K} = \sqrt{\frac{1}{N_c}\sum_{c\in\mathcal{C}}\left(\sum_{t\in\mathcal{P}}\left(y_{ctK} - \hat{y}_{ctK}\right)\right)^2}\right)$. We also ranked models by RMSE within each country and evaluated mean model ranks. We found that model (3) (random effects), estimated with all weeks of data (rather than separately for each period), either strictly or weakly dominated all other models over these metrics during these test years and periods.

Second, we fit models for data from 2015-2017, and predicted 2019 to approximate 2021 with only 2019 pre-intervention data. (We did not want to fit trend models on only 2 years of pre-pandemic data and therefore did not use 2018 as a test year.) Model (3) continued to outperform other models overall; for $RMSE^{CP}$ during omicron weeks, model (2) estimated on all pre-intervention data slightly outperformed model (3), but the magnitude of the difference was negligible. We therefore used model (3) estimated over all weeks combined as our main specification. See GitHub for further notes, summaries of model evaluation statistics, and sensitivity analyses.

Statistical Analyses

Table 1

To compare COVID-19 death rates across locations, we let d_{ct} be the number of COVID-19 deaths in location c during period t and p_{ct} be its population at time t. We assumed that $d_{ct} \sim Pois(\lambda_{ct})$ and

$$\mathbb{E}\left[log(\lambda_{ct})\right] = \beta_0 + \sum_{j\in\mathcal{C}, j\neq US} \beta_j \mathbb{I}\left(c = j\right) + log(p_{ct}),$$

where β_j compared the death rate in country j to the United States. We employed a similar model to make comparisons with the 10 most-vaccinated states and 10 least-vaccinated states. We used standard Wald tests to evaluate statistical significance.

Table 2

Per the model-fitting process described above, we modeled excess all-cause mortality:

$$y_{ctk} = \sum_{j\in\mathcal{C}}\sum_{w\in\mathcal{W}}\sum_{\ell\geq 2020} \beta_{jw\ell}\mathbb{I}\{c = j \cap t = w \cap k = \ell\} + W_{ct} + \beta^R_c k + \epsilon_{ctk},$$

where y_{ctk} was mortality per 100,000 population in week t of year k in location c, W_{ct} was a location-week fixed effect, and k indicated year. We estimated β^R_c as a random effect. Week-level treatment effects were saturated to avoid estimating pre-pandemic trends with pandemic data. In this framework, the linear combination of $\sum_{j=Comparator}\sum_{w,\ell\in\mathcal{P}}\beta_{jw\ell} - \sum_{j=US}\sum_{w,\ell\in\mathcal{P}}\beta_{jw\ell}$ compared excess mortality between each comparator location and the US over the set of week-years in period $\mathcal{P}$. We used standard Wald tests to evaluate statistical significance. We made similar comparisons between the 10 most-vaccinated states and 10 least-vaccinated states, and also compared state subgroups to other countries. To estimate the difference between COVID-19 and all-cause mortality, we used as an outcome $y_{ctk} - d_{ctk}$, where d_{ctk} was COVID-19 mortality per 100,000 for location c in week t and year k.