Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

Excess mortality versus COVID-19 death rates: a spatial analysis of socioeconomic disparities and political allegiance across US states

Janine Aron & John Muellbauer

9th December 2021

INET Oxford Working Paper No. 2021-24

Excess mortality versus COVID-19 death rates: a spatial analysis of socioeconomic disparities and political allegiance across US states

BY JANINE ARON

Department of Economics, University of Oxford

AND

JOHN MUELLBAUER*

Nuffield College, University of Oxford

9-December-2021

Abstract

Excess mortality is a more robust measure than the counts of COVID-19 deaths typically used in epidemiological and spatial studies. Measurement issues around excess mortality, considering data quality and comparability both internationally and within the U.S., are surveyed. This paper is the first state-level spatial analysis of cumulative excess mortality for the U.S. in the first full year of the pandemic. There is strong evidence that, given appropriate controls, states with higher Democrat vote shares experienced lower excess mortality (consistent with county-level studies of COVID-19 deaths). Important demographic and socio-economic controls from a broad set tested were racial composition, age structure, population density, poverty, income, temperature, and timing of arrival of the pandemic. Interaction effects suggest the Democrat vote share effect of reducing mortality was even greater in states where the pandemic arrived early. Omitting political allegiance leads to a significant underestimation of the mortality disparities for minority populations.

JEL Codes: I14, I18, J11.

Keywords: Excess Mortality, COVID-19, Spatial Analysis, US States, Political Polarization.

Janine Aron and John Muellbauer are also affiliated to INET, Oxford Martin School, University of Oxford and the Leverhulme Centre for Demographic Science, Nuffield College Oxford.

Note: We are grateful to Eric Beinhocker and David Hendry of INET for comments, to Lauren Rossen of the US CDC (Centres for Disease Control and Prevention), and to participants at seminar meetings at the Dallas Federal Reserve, Peterson Institute, D.C., particularly Olivier Blanchard and Marcus Noland, and at INET, Oxford University, for discussions.

*Correspondence to: John Muellbauer, Nuffield College, University of Oxford, New Road, Oxford, OX1 1NF, UK (john.muellbauer@nuffield.ox.ac.uk)

1. INTRODUCTION

Excess mortality is a count of deaths from 'all causes' expressed relative to the benchmark of 'normal' deaths. 'Normal' death rates reflect persistent factors such as the age composition of the population, the incidence of smoking and air pollution, the prevalence of obesity, poverty and inequality, and the normal quality of health service delivery. 'Normal' deaths are typically estimated from several years of data on pre-pandemic mortality using methods of varying sophistication. In a pandemic, deaths rise sharply, but causes are often inaccurately recorded, particularly when reliable tests are not widely available. Thus, the death counts[1] attributed to COVID-19 may have been significantly undercounted. Excess mortality data overcome two problems in reporting COVID-19-related deaths. Miscounting from the misdiagnosis or under-reporting of COVID-19-related deaths is avoided. Excess mortality data also include 'collateral damage' from other health conditions, left untreated if the health system is overwhelmed by COVID-19 cases, or by deliberate actions that prioritise patients with COVID-19 over those with other symptoms. Precautionary measures taken by governments and individuals may also influence death rates in a pandemic. Deaths from traffic accidents and deaths from other infectious disease such as influenza may decline; however, suicide rates may rise.[2] Excess mortality captures the *net outcome* of all these factors.

Excess mortality data can be used to draw lessons from cross-country and within-country differences and to analyse the social and economic consequences of the pandemic and of lockdown restrictions. Excess death figures may help to avoid the measurement biases inherent in other data typically used to estimate the virus reproduction rate, R, in epidemiological models[3], crucial for designing and assessing non-pharmaceutical interventions such as lock-downs.

Studies comparing the US to other countries find that in 2020 it ranked amongst the highest in COVID-19 deaths per 100,000 (Bilinski, 2020) and in rates of excess deaths (OECD paper by Morgan et al. (2020), ONS (2021) and earlier versions, and Aron and Muellbauer (2020c)). Woolf et al. (2020, 2021), comparing US mortality from COVID-19 (March-October, 2020) to leading causes of death two years before the pandemic (March-October, 2018), finds that COVID-19 was one of the leading causes of death; in the Spring and late Autumn of that year, it was *the* leading cause of death in the US. The pandemic is likely to exacerbate the decline in life expectancy that has been apparent since 2014 (Koh et al., 2020).

[1] For example, see webpage: COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE), Johns Hopkins University (JHU).

[2] Other examples are increases in self-harm, domestic abuse and other crime; use of tobacco, drugs and alcohol; and anxiety and changed quality of diet from loss of jobs and income, see Kontis et al. (2020).

[3] See the evidence of Prof. John Edmunds to the UK Science and Technology Parliamentary Select Committee on 7th May 2020. He explained that while excess mortality data lag Covid-19 infections, the data are an important check on earlier estimates of the rate of spread of the virus.

Virtually all spatial analyses of mortality in the US are based on case counts or counts of COVID-19 deaths. The only exception is a county-level study of excess mortality by Stokes et al. (2021), of which more below. An indication of the limitations and biases in the data on infections and COVID deaths is given in IHME (2021), who suggest that death counts are a less biased estimate of true COVID-19-related deaths than COVID-19 case counts are of the true number of infections.[4] Yet, Weinberger et al. (2020) find that official tallies likely undercount US deaths due to the virus, with the completeness of the tallies varying markedly between states; they also advocate excess all-cause mortality data as more reliable to estimate the full COVID-19 burden.

In the first 52 weeks of the pandemic, there were around 650,300 excess deaths in the US, compared with COVID-19 deaths of around 499,500, sourced from Coronavirus Resource Center, Johns Hopkins University (JHU), or around 530,000, when sourced from the US Centres for Disease Control and Prevention (CDC). Figure 1 shows the time profile of weekly per capita excess deaths at the national level, and the ratio of the CDC count of COVID-19 deaths to excess deaths. This shows severe under-counting of COVID deaths at the start of the pandemic in the Spring and suggests considerable under-counting in the Summer and early Autumn of 2020. The figure also shows the ratio of JHU-sourced COVID-19 deaths to CDC-sourced COVID-19 deaths; the high ratio suggests an even greater under-counting by the JHU source than the CDC source at the start of the pandemic. Moreover, the divergence between the two measures persists throughout the pandemic and is greatest at the peaks of the waves. Our empirical work on Covid-19 deaths suggests strongly that the CDC-sourced COVID death count is preferable to the JHU data, see Section 5.4. Figure 2 ranks the US states by the cumulated excess deaths per capita for the 52 weeks, comparing with the P-score and the CDC measure of per capita COVID-19 deaths. Comparing the COVID-19 death count to excess deaths across states reveals considerable variation in the degree of under-counting.

Our study focuses on cumulative US *excess mortality* across 51 states (including District of Columbia) in the first 52 weeks of the pandemic. This avoids potential mismeasurement problems in the usual dependent variables, and we compare the results with a model for COVID-19 deaths per capita. One reason for the choice of state comparisons is that the US CDC (Centres for Disease Control and Prevention) does not generate county-level estimates of excess mortality.[5] We have found only two spatial analyses of US COVID-19-related mortality at the state level, IHME (2020) and Doti (2020),[6] both modelling COVID-19 deaths. Thus, our paper is the first state-level spatial analysis of excess

[4] Case count data are affected by differences in treatment-seeking behaviour, testing protocols and access to care, and further compromised by infectious asymptomatic individuals or pre-symptomatic individuals. Testing results may be compromised by accuracy concerns.

[5] Stokes et al. (2021) generated their own excess mortality data covering two-thirds of all US counties. Their estimates of 'normal deaths' are a simple average of the per capita death rates for 2013-2018, adjusted by a national trend factor (see Section 2.3 for a discussion of the estimation issues around 'normal' deaths).

[6] The article has specification errors; for instance, the state intervention mandates are not lagged, introducing endogeneity bias.

mortality, and the first state state-level spatial analysis of mortality that explicitly includes political variables.[7]

There are several advantages to a state-level perspective, apart from a simpler and more easily interpretable spatial model. Using states can be justified by their crucial political role defined by the Constitution, e.g. their equal representation in the Senate and their role in the Electoral College, which elects the President. It is possible to flexibly explore different hypotheses without the significantly greater challenge faced by county studies of properly capturing complex local spatial correlation. Few county studies deal seriously with county spill-over effects. The use of state fixed effects in county models can help address such flaws but they are difficult to interpret, and much of what is of central interest to policy can be thereby 'washed out'. While the state-focus has the obvious cost of the reduced range of spatial variation and fewer degrees of freedom, it provides a useful complement with implications for county-level research.

The heterogeneity across US states in excess deaths linked to COVID-19 in the first 52 weeks was enormous, from 305 per 100,000 in Mississippi, to 64 in Maine and 65 in Washington State, the two lowest on the mainland. Using the right controls for state-level comparisons is crucial to disentangle the effects of political partisanship from other determinants. Fortunately, there have been many studies at much more fine-grained spatial levels, e.g. over 3000 counties, from which the most important controls can be deduced. For the majority of studies (an exception is Stokes et al. (2021)), the dependent variable is a per capita measure of the infection count and/or of COVID-19 deaths, see Table 1. These dependent variables embody measurement bias, although some parameterisations in a dynamic model can reduce the bias subject to simplifying assumptions (e.g. Rubin et al. (2020)). Five examples of cross-sectional spatial studies that include socio-demographic and health determinants but do not include political variables are Stokes et al. (2021), Knittel and Ozaltun (2020), McLaren (2020), Karmakar et al. (2020) at the county level, and Doti (2020) at the state level, who also includes state interventions on social distancing. Considering also the role of partisanship and COVID-19 infections and deaths are Liao and De Maio (2021) and Desmet and Wacziarg (2021).[8] A detailed critical review of these studies can be found in Aron and Muellbauer (2021).

Structural differences between locations had huge effects on mortality outcomes in the pandemic's first year. A first group of baseline population characteristics, affecting the transmission risk of contracting COVID, and vulnerability to the serious health consequences of infection and to non-pharmaceutical interventions by governments, is likely to remain largely unchanged over the

[7] IHME (2020) is a dynamic panel study which uses fixed effects to control for state differences. Doti (2020) does not include political variables.

[8] Of these studies, Karmakar et al. (2020) and Desmet and Wacziarg (20201) also have a dynamic aspect. Other studies introducing dynamics into the spatial analysis are Rubin et al. (2020), Gerritse (2020), IHME (2020), Hamidi et al. (2020), Gollwitzer et al. (2020) and Almagro et al. (2020, 2021).

pandemic.[9] A second set of public health and social care determinants experienced rapid rescaling and reskilling to affect capacity. Over time there was an improved understanding of the disease and how to treat it, and later, vaccines were deployed, and new virus variants encountered. A third group, policies for lockdown and other restrictions, varying widely across states and countries, have been tightened and relaxed at times over the different waves of the pandemic. Finally, compliance with policies and scientific advice may also have altered over the pandemic, affected by the perception of economic trade-offs, and by the media and political role models. Generally, therefore, longitudinal spatial models would be expected to be subject to changing values over time of the coefficients of the last three sets of determinants, but also of the first set, to the extent that the correlations with omitted variables are subject to alteration. In a cross-sectional context, these expected changes suggest testing for potential interaction effects, for example with measures of the timing of first arrival of significant levels of infection.

Political partisanship, e.g. measured by the US electoral vote share, has supplemented the controls in some county studies of pandemic deaths and case counts to proxy private attitudes and compliance. Gollwitzer et al. (2020) summarise studies of partisanship and its measurement, and the link with social judgements and behaviours (e.g. Van Bavel (2018)). Allcott et al. (2020) study partisan differences in Americans' surveyed beliefs concerning their infection risk and the likely severity of the pandemic and find that social distancing behaviours reflected these beliefs. Makridis and Rothwell (2020) use nationally-representative US panel data to demonstrate that the formation of beliefs about the pandemic and social distancing behaviour is driven primarily by political affiliation. Druckman et al. (2021) find a strong association between citizens' levels of partisan animosity and their attitudes about the pandemic, and the actions they take in response to it. Hamel et al. (2021) analyse the results of multiple surveys confirming the role of partisanship in explaining spatial differences in US vaccination rates.

Omitted variables are likely to be the most prominent source of bias if they are correlated with the included regressors. The inclusion of political partisanship adds an important omitted variable to the more typical set of regressors, which are focused on the characteristics affecting transmission risk and vulnerability to infection and the preparedness and capacity of the public health and social care systems. As in other cross-section studies, there may be omitted variables that are correlated with an included regressor but are themselves difficult to measure. Examples are wealth inequality across race and racial discrimination, which may provide channels to explain the widely-found significance of racial and ethnic regressors in the above types of analyses, conditional on inclusion of a set of co-variates.[10]

[9] The baseline variables include demographic and health characteristics differentiated by gender; measures of poverty, income and inequality; racial and ethnic group status; employment status, type of occupation and working conditions; transport measures such as use of public transport, commuting across states and international linkages through airports; and housing density. The relative influence of such baseline variables can, of course, evolve over the course of the pandemic.

[10] See Hardy and Logan (2020) for a comprehensive analysis of the impact of racial and ethnic inequality on COVID-19 mortality and McLaren (2020) for statistical evidence.

Another example is that the quite accidental arrival of the pandemic in certain counties and states early in the pandemic - because of returning travellers from Europe, or crowd events such as New Orleans' Mardi Gras - will have been strongly linked with high subsequent mortality. Omission of relevant controls, such as enplanement measures of numbers of travellers from the most infected foreign origins, can bias the estimated effects for those counties.[11] Alone amongst the above studies in controlling for temperature is Knittel and Ozaltun (2020), a variable that has been found important in historical patterns of mortality, e.g. Kontis et al. (2020).

To minimise the effects of omitted variables it is important to test for a comprehensive set of potential initial controls, an important feature of our own methodology. Approaches amongst the above-cited articles differ in the selection of controls, which is often arbitrary, leaving out key controls such as temperature and population density. However, in a large set, many controls may be collinear with other controls or appear insignificant. At least two approaches have been used in this context. The Lasso (least absolute shrinkage and selection operator) regression analysis method aims to enhance the prediction accuracy and interpretability of the resulting statistical model, by requiring the sum of the absolute value of the regression coefficients to be less than a fixed value, which forces certain coefficients to zero, thereby excluding them. Castle et al. (2020) argue that Lasso struggles with negative correlations,[12] and find better performance, from the 'general to specific' approach 'implemented in the Autometrics' software, which we use to check our regressions.[13]

Our analysis of US state differences in pandemic-related rates of mortality estimates the effects of racial composition, age structure, poverty, population density, care capacity and other structural features, the timing of the pandemic onset, Spring temperatures (°F) and of political allegiance. Across the 51 US states, we find that political allegiance expressed in the way people voted in 2016 had a major effect on mortality outcomes, given the inclusion of the socio-economic and other controls. This is consistent with spatial studies at the county level, linking partisan allegiance with private attitudes, behaviour and COVID-19 deaths. The Desmet and Wacziarg (2021) county-level study of COVID-19 deaths and infection rates in the US established that counties with a high vote-share for the Republicans in 2016 had higher rates of COVID-19 deaths up to the end of November, accounting for population density, racial/ethnic composition and other controls). We confirm this result at the state level for the full year since the arrival of the pandemic when using rates of *excess mortality* as the dependent variable, as well as for COVID-19 death counts per capita. Our controls also include state interaction

[11] Save for Desmet and Waziarg (2021), none of the above studies corrects for the bias from the differential early onset of the pandemic in some states and later onset in others.

[12] This is because negatively correlated variables need to enter jointly as they may not matter much individually. This also proves to be a problem for step-wise regression.

[13] The Autometrics algorithms are available in Doornik and Hendry (2018), see also www.doornik.com, the Excel add-in XLModeler, www.xlmodeler.com, and in R (Pretis et al., 2018).

effects with the timing of first arrival of the pandemic, implying that the effect of partisanship was even greater in states where the pandemic arrived early.

The paper sets out in Section 2 why excess mortality expressed as a rate most accurately captures the impact of the COVID-19 pandemic. Different measures of pandemic outcomes are compared and contrasted, especially in relation to the valid comparability of deaths, case counts, 'normal deaths' excess deaths and excess mortality across regions, states and countries. Data sources and data quality are assessed, and suggestions made for improving the transparency and granularity of excess mortality data. Section 3 lays out the conceptual framework and the drivers of excess mortality, and a reduced form empirical model for analysing cross-state variation in rates of cumulated excess mortality, and Section 4 the data sources, transformations and statistics. In Section 5, the data and empirical results are described for the impact on rates of cumulated excess mortality, and for comparison, of rates of COVID-19 deaths, of state variations in political allegiance and socioeconomic factors in the first 52 weeks of the pandemic. Section 6 concludes.

2. EXCESS MORTALITY – DEFINITION AND MEASUREMENT

For country or state comparisons (where the under-recording of pandemic deaths may differ), a *robust measure* of the count of excess deaths (actual deaths minus 'normal' deaths) expressed relative to the population or relative to the benchmark of 'normal' deaths (which we have named the P-score),[14] is greatly to be preferred to simple counts (including per capita) of COVID-19 death rates and infectious case counts, see Table 1.

This section explains the data quality problems with the raw case and deaths data, it compares and contrasts different measures of excess mortality, and discusses an alternative measure of the toll of the pandemic, quality-adjusted life expectancy.

2.1 *Why use excess mortality?*

Comparisons of excess mortality across regions, states or countries have several purposes. The first is to compare the death toll of the pandemic. The death count of COVID-19, as noted above, suffers from a number of biases, making it an unreliable dependent variable, especially when comparing across countries or states with different definitions of what constitutes a COVID death. Even within the US, we noted significant discrepancies between the CDC and JHU sources for COVID-19 deaths, see Figure 1. Countries with a wide definition for COVID-19 deaths (e.g. Belgium and France) will show that most excess deaths are accounted for by COVID-19, as compared to those with a narrower definition. In the

[14] This terminology has now been adopted more widely, e.g. OWID, the ONS (2020; 2021) and the OECD (2020).

US, as Fineberg (2020) observes, counts of deaths from all causes from the National Vital Statistics System (NVSS)[15] are incomplete for recent weeks, and lags may be as long as eight weeks.[16] COVID-19 deaths tend to be under-reported based on the listed causes of death, which reflect varying uncertainty and the judgment of the certifier. For instance, Woolf et al. (2020) find that mortality rates for Alzheimer disease/dementia and heart disease rose during Spring and Summer pandemic surges, with statistical significance. This could suggest misdiagnosis of a COVID-19 death or that COVID was implicated in these deaths by preventing early treatment. Supporting evidence for the above is from Woolf et al. (2020) who find that COVID-19 deaths were a documented cause of death for "only" 67% of excess deaths in the US (1-March to 1-August 2020). Their table shows great variation in the COVID-19 share of excess deaths across the US states, pointing to varying degrees of mismeasurement across states in COVID-19 implicated mortality, as implied by our Figure 2. Figure 1 provided national evidence on the shifting COVID-19 share of excess deaths over time, reflecting improvements in the understanding of the disease, in testing capacity, in diagnosis and other factors.[17]

A second reason for making comparisons of excess mortality, to evaluate the effectiveness of policy responses, requires one to dig deeper, and even the simple measures above require further interpretation. Countries may differ in the size of the initial source of infection, in their age structure, in the distribution of co-morbidities in the population and the prevalence of dense urban centres, making some countries more vulnerable.

The third motivation for comparisons is the purely objective one of improving the scientific understanding of the dynamics of the spread of infections, their incidence and the death rates of those infected. Key to this last endeavour is the production of granular data, i.e. disaggregation of excess deaths data by age, gender, region, and, where possible, socio-economic categories.

2.2 *Measures, sources, and their variable quality*

Several definitions of the dependent variables capturing pandemic outcomes and used in spatial analyses are summarised and evaluated in Table 1. These are presented in two groups: measures of COVID deaths, COVID-related deaths and COVID-cases; and measures of excess mortality.

To address the measurement problems inherent in the former group, we argued at an early stage of the pandemic that national statistical offices should publish more granular data and excess mortality P-

[15] The US National Center for Health Statistics (NCHS), within the Centers for Disease Control and Prevention (CDC), operates the National Vital Statistics System (NVSS) for the US.
[16] The lags were longer for North Carolina, as it transitioned from a paper-based to a digital system of recording deaths.
[17] Some of the discrepancy between reported COVID-19 deaths and excess deaths could be related to the intensity and timing of increases in testing, and differential guidelines on the recording of deaths that are suspected to be COVID-19 but without a laboratory confirmation; the location of death (hospital, nursing home, or at home) has also affected whether it is recorded as a COVID-19 death, (Weinberger et al., 2020).

scores for states and sub-regions, disaggregated by age, gender and race.[18] The P-score (ratio or percentage of excess deaths relative to 'normal' deaths) is an easily interpretable measure. While many national statistical agencies have published actual weekly deaths and averages of past 'normal' deaths, there were few published benchmarks for more granular or disaggregated data, such as sub-regions or cities. In the U.S., the CDC publishes data on excess deaths and a variant on P-scores (see Table 1), defining excess deaths as deviations from 'normal' deaths plus a margin adjusting for the uncertainty around estimated normal deaths.[19] This variant is a lower bound estimate of excess mortality, since the upper 95 percent confidence interval is an upper bound estimate of normal deaths. The variant has the disadvantage that excess mortality data cannot be cumulated over a number of weeks since the margin of uncertainty will narrow as randomness at the weekly level smooths out. These data include states but not counties, and are also available disaggregated by gender, age and ethnicity. However, to obtain cross-European or cross-global comparisons in 2020 required data collation from individual national agencies to construct these measures.

Early in the pandemic, separate journalistic endeavours engaged in the time-consuming effort of collating and presenting more transparent excess mortality data, see Aron and Muellbauer (2020a, Table 1). In the intervening year, several agencies have geared up to provide underlying data or present the P-score measures. Perhaps the biggest single pitfall for comparability arises from the accuracy of the raw mortality data. An important drawback of the reported numbers concerns lags in recording and reporting deaths. Countries differ in the efficiency of their death registration systems, particularly where those systems are devolved to regional or local administrations. Problems in one location can affect or delay the national data, and sometimes the national recording system can be slow to absorb regional information. Even in countries with the most sophisticated recording systems, reported mortality lags weeks behind the facts. In a pandemic, it can happen that the capacity of systems is temporarily overwhelmed, most of all in hotspots, often in urban areas. Occasionally the recording methods may be so weak overall, that the observers resort to data on burials. These definitional differences need to be highlighted and made transparent across country data providers and international organisations reporting excess mortality statistics. The period over which comparisons are made needs to be specified carefully, as it is likely that reporting lags are far from uniform across countries.

The Human Mortality Database's Short-term Mortality Fluctuations (STMF) project offers high quality national mortality data by week for 38 countries, and access to the exemplary statistical metafile of HMD. Baseline data cover mainly 2015-2019 (2016 for a few countries), back in many cases to 2000, and disaggregation by several age categories and gender. This provides the raw data from which excess

[18] See Aron and Muellbauer (2020a, 2020b, 2020c).

[19] See webpage: "National Center for Health Statistics", Centers for Disease Control and Prevention (CDC), US Government. These estimates use statistical models at the state level incorporating seasonals and trends to define normal deaths.

mortality measures can be constructed. Eurostat[20] produce granular mortality data, cross-classified by sex, five-year age-groups and NUTS 3 regional levels within countries for 26 EU member states, EFTA countries and five non-member neighbouring countries. They also compile monthly estimates of P-scores using normal deaths defined as the monthly average for 2016 to 2019.

The World Mortality Database has the largest set of countries (94) with a mix of weekly and monthly data. Around half of these come from the above sources, and the rest are directly sourced from national authorities, though some data are of questionable quality.[21] Some of the countries covered by WMD publish data with lags as long as 6 months and even those data may be under-recording deaths in the final weeks of the period covered. Data are presented normalising the excess mortality estimates by the population size, though without evaluating the quality of the underlying population data.[22] Their P-scores (for 'all ages' only) use 'normal' deaths based on the previous 4 to 5 years of data, where available, using seasonals and annual time trends in regressions to project 'normal' deaths to 2020 and 2021.[23] This is a simplified version of the methods used for instance by the CDC of the US, which provides 'normal' seasonally-adjusted baselines on its site. It also differs from the method used by Our World in Data (OWID) which sources data from the above three websites and presents excess mortality statistics (P-scores) for 70 countries, using an arithmetic average for 'normal' deaths of the years 2015 (or 2016) -2019. OWID disaggregate by several age categories and by gender, have a discussion of data quality and comparability, and are clearer on the time-frame for their data – they do not use the last few weeks because of recording lags.

2.3 *Issues around the measurement of normal deaths – the case of the CDC.*

Using the arithmetic average of previous years as the baseline for normal deaths has the advantage of simplicity. However, there are differences in underlying trends in deaths which are likely to be dominated by population growth and the changing age structures of the population, and in other health conditions and their treatment. Ignoring such trends can result in over- or under-estimates of 'normal' deaths, and hence in under- or under-estimates of excess deaths in comparisons between countries or regions. The CDC's estimates of weekly normal deaths at the state level[24] implement the Farrington algorithm, see Noufaily et al. (2012), which uses over-dispersed Poisson generalized linear models with

[20] See webpage: Eurostat excess mortality statistics.

[21] The dataset is a mixture of reliable and poor-quality data, without discussion of comparative quality. Monthly data were used for countries where weekly data are not available. Availability of weekly data might be considered as indirect indicator of data quality. It is not always the case and there are some exceptions (e.g. Japan doesn't publish weekly data but has high quality data).

[22] Notably, a few countries with acceptable mortality quality were excluded from the HMD excess mortality statistics (STMF), mainly because of problematic population estimates (HMD publishes rates).

[23] It is not fully clear from the WMD website which countries have data for the full five years for the baseline estimation: 2015-2019. However, if the baseline is estimated for one year of data only, then no trend could be estimated, leading to biased results.

[24] See the CDC website.

spline terms to model trends in counts, and accounts for seasonality. The CDC's approach does not take into account evolving state-level population and its age distribution in previous years in modelling normal deaths. Moreover, the Poisson model, designed for small number count data, makes strong assumptions about the underlying stochastic process, which are contradicted by evidence for larger populations, see Aron and Muellbauer (2020b). Even for the least populous US states, weekly deaths almost never fall below 60, which is not a 'small number' in this context. Hence, a better approximation to the data-generating process is likely to be offered by the more flexible ARIMA models. These more flexible ARIMA models have been used at a *national* level to estimate normal deaths, by Rossen et al. (2021), Faust et al. (2021) and Shiels et al. (2020), among others. These authors apply ARIMA models to estimate trends and seasonals from historic data on per capita deaths for different age groups. Estimates of normal deaths for the pandemic period are then made by projecting these trends and seasonals and multiplying up by the current population data for each age group. The pandemic has reduced the population count, especially of older age groups who have high per capita death rates. This method results in lower estimates of normal deaths and higher estimates of excess deaths than a linear projection of past trends which ignores the changing population and age structure. Applying such an approach at the state level would improve the accuracy of excess mortality estimates.

2.4 *Comparability of the different measures across countries, regions or states*

The different measures of excess mortality are compared and contrasted in Table 1. Assuming that the data definitions for the death counts, such as the definition of the week, type of death count data collected (e.g. registration versus occurrence data) and timeliness of the collection are identical across countries, see Aron and Muellbauer (2020b), we consider the relative comparability of the statistical measures of excess mortality.

In Figure 3, the weekly per capita excess deaths and P-scores for the US as a whole are plotted. The P-scores have the advantage that by normalising relative to 'normal' death counts, they reflect persistent factors affecting normal mortality such as the age composition of the population, the incidence of smoking and air pollution, the prevalence of obesity, poverty and inequality, and the normal quality of health service delivery.[25] A country like Italy, with an older population, will fare somewhat worse in a per capita excess mortality comparison with countries having younger populations than in a P-score comparison. In a multivariate statistical study, the inclusion of comprehensive controls reduces this advantage of the P-score over a per capita measure of excess mortality, though the P-score reduces the risk of potential bias from unobserved heterogeneity in normal health risks. Moreover, while P-scores are less affected than per capita excess deaths by differences in the age-composition of the population, they are not immune. Differences in the age distribution between countries would only be irrelevant if

[25]A possible argument in favour of per capita excess mortality is that total population could be regarded as a rough proxy for the ability of the society to absorb excess deaths.

mortality risk increased in the same proportion for all. This is not the case because children have a far lower *relative* mortality risk in the COVID-19 pandemic than under normal conditions. Moreover, differences in urban structure and in population density have relatively little effect on normal mortality rates but have major effects on the spread of a pandemic. P-scores are therefore far from immune to structural differences between countries and regions. However, for temporal comparisons for the same country, their time profile differs little from per capita excess deaths, see Figure 3.

These themes can be illustrated by comparing rankings of COVID-19 related rates of mortality across US states. Because normal deaths are higher for the elderly and for those with co-morbidities, scaling by normal deaths takes some account of differentials in age composition and socioeconomic characteristics between countries and regions. Indeed, comparing US states, the rankings of states according to the two metrics are notably different, see Figure 2. For example, Mississippi had the highest per capita rate of excess mortality in the US, while California ranked in the middle of the distribution at number 25. However, on the P-score, California has higher mortality in 5th place while Mississippi is in 7th place. Clearly, normal death rates are far higher in Mississippi than in California.

Similar issues affect age-standardised mortality comparisons. The age-standardised mortality rate takes the age-specific mortality rate for each age group, and measures their weighted average using the proportion of the population in the corresponding age groups in a reference population. The same reference population is used in comparing any two countries or regions. While this controls for some of the effects of differences in age structures it neglects the other structural difference affecting pandemic-related mortality in different countries.

An alternative measure is the Z-scores compiled by EuroMOMO[26] for 29 states, see Table 1. The Z-scores standardise data on excess deaths by scaling by the standard deviation of deaths outside periods of notable excess mortality. The expected value of each country's weekly deaths is estimated using data for the previous five years, taking seasonal factors and trends into account, and adjusting for delays in registration. To fit the baseline, normal variability is measured after excluding seasons leading to excess deaths from additional processes (e.g. Winter influenza and Summer heat waves). In contrast to the P-scores, the Z-scores are a less easily interpretable measure. Moreover, if the natural variability of the weekly data is lower in one country compared to another, for example in larger populations compared with smaller ones, then the Z-scores lead to exaggeration of excess mortality compared to the P-scores[27]. Graphic presentation of the Z-scores for different time-periods, countries, and age-groups, with the

[26] EuroMOMO is a European mortality monitoring entity, aiming to detect and measure excess deaths related to seasonal influenza, pandemics and other public health threats. Official national mortality statistics are provided weekly from the 24 European countries and regions in the EuroMOMO collaborative network, supported by the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO).

[27] Given the Poisson model used by EUROMOMO, there should be large differences in Z-scores between countries with different populations even if the P-scores were identical. In practice, because the Poisson is likely to be poor approximation to the stochastic process for the number of deaths, the differences are less pronounced than one would expect, see Appendix in Aron and Muellbauer (2020b).

estimated confidence intervals back to 2015, provides a visual guide to their variability. A further disadvantage of Z-scores, compared to P-scores and per capita excess death measures, is that their cumulation over multiple pandemic weeks is problematic. While excess deaths can be cumulated, the standard deviation of normal deaths cannot. EuroMOMO do not reveal the standard deviations used in their calculations. This makes it hard to obtain a comprehensive comparative summary of the pandemic's impact from the Z-scores.

2.5 *Quality-adjusted life expectancy*

Finally, it should be considered whether excess mortality statistics alone are sufficient to measure the impact of a pandemic. One has to be aware of the limitation of any single measure of comparability between countries. Subsumed within the excess death aggregates are implicit value judgements. For example, crucially in the case of a pandemic, there is an implicit assumption that the toll of an older life lost is the same as that of a younger life. However, when a younger life is lost, many more years of life expectancy are lost, and one might want to attach a larger weight to deaths of the young.

The health economics literature has given attention to Quality Adjusted Life Expectancy (QALY) as a criterion for expenditure on health-improving policies. QALYs measure the number of reasonably healthy years a person might expect to live. The number of QALYs lost could supplement the increased death count resulting from the pandemic as a measure of its impact. However, detailed actuarial and medical information is entailed in the complex estimation of the number of QALYs lost. QALYs and the attachment of monetary values to QALYs have long been controversial, see Loomes and Mackenzie (1989), but the concept of a QALY does focus attention on the relative value (by age group) of expected years lost in a pandemic. The excess mortality of working age adults with a normal life expectancy of 30 years might be weighed against the excess mortality of 85-year olds with a life expectancy of 5 years. Attaching more weight to excess mortality for working age adults will affect comparisons of countries with different age-specific mortality rates. Pifarré i Arolas et al. (2021) estimate years of life lost (YLL) for 81 countries from premature deaths due to COVID-19 based on age-specific life-expectancy tables for each country. For most countries, they based their estimates on COVID-19 death counts, but for a subset of 18 they use excess mortality data. They find that close to half of YLL for all the countries are in the 55 to 75 age group and that only around a quarter of YLL occurred for the over 75s.

To end on a cautionary note that affects all the weekly measures of excess mortality, it is important to examine excess mortality in a longer-term perspective. If, as argued, for example, by British statistician, Spiegelhalter (2020), the main impact of COVID-19 is simply to shift forward the date of death by a few months for those close to death because of underlying poor health, then a peak in weekly

deaths should be followed by a trough in the following months.[28] For the US, Faust et al. (2021) have estimated the impact of so many people dying in the initial wave, that there were fewer vulnerable people as time went on, and proposed a method of adjusting expected counts of deaths downward because of the excess mortality that happened earlier in the year.

3. CONCEPTUAL FRAMEWORK AND MODEL FORMULATION

Our aim in this paper is to analyse the main factors accounting for cross-state variations in cumulated excess mortality after one year of the pandemic in reduced form models. The papers cited in Section 1 have examined socio-economic drivers of recorded COVID-19 cases and deaths using county-level cross-section data. A few also examine political drivers of COVID-19 cases and deaths. A major limitation of such studies is the serious measurement biases in reported infections and COVID-19 attributed deaths, particularly early in the pandemic when testing capacity was often limited, and unequally distributed.

If 'all-cause' death registration data are accurate, then excess mortality will not be subject to these measurement biases;. However, excess mortality includes the other two components discussed above: avoidable deaths due to non-occurrence of treatments for other causes of ill-health and deaths avoided from shifts in behaviour linked with the pandemic. While the peak incidence of COVID-19 deaths occurs 2 to 3 weeks after infection, though with a long tail of later incidence, the timing of the last-mentioned components is likely to be different. The effects of non-treatment of preventable ill-health on mortality include missing early diagnosis and starting cancer treatments later than is advisable, and therefore have mortality consequences likely to materialise months, and in some cases years, later. Similarly, the health damage from the economic disruption caused by the pandemic, especially for lower income people, is likely to affect mortality for years to come.

3.1 *The drivers of spatial variation in excess mortality*

To interpret the large differences in cumulative COVID-19 death rates among states requires consideration of several factors: the average infection rates in preceding weeks, average mortality risk from COVID-19 and constraints on COVID-19-specific health capacity, given the prevailing state of knowledge about treatment.

Turning to the first of the factors, consider differences in infection rates. Compare two states with the same average COVID-19 case fatality risk where 1 percent of all adults are infected in A, while 5 percent are infected in B. Then the rate of excess deaths for adults measured by the P-score will be

[28] Actuary McDonald has disagreed with claims that a majority would have died in the next 3 months, see Edwards and McDonald (2020). Spiegelhalter subsequently admitted over-estimating this 'harvesting' effect, Kelly (2020).

about 5 times as large in B in the weeks following the incidence of the infection. States that locked down early and had effective test, trace and isolate procedures kept down the average infection rate and hence the excess death rate. Transmission and hence rates of infection are also influenced by factors like the nature of social distancing, availability and use of face masks, and cultural differences in the exercise of self-discipline and following of advice. This set of mitigating factors can be influenced by public policies enacted at state and local levels. Other factors impacting infection rates include types of occupation, density of living circumstances and proximity to international or cross-state travellers who might import infection. For example, New York's higher excess mortality was influenced by higher initial imports of infections and a higher virus reproduction number given its high density and hard-to-avoid close physical contact on public transport and at work in New York City. States with a higher fraction of adults in multi-generational families, and in locations or occupations (e.g. health workers or taxi-drivers) where the virus can more easily spread, will tend to have higher excess death rates. The influence of the above factors is likely to evolve over the course of the pandemic as the main sources of infection change and as individual behaviour and public policies respond.

The second of the factors mentioned above is mortality risk for infected adults, and this can differ between and within states. The steep age gradient of COVID-19 mortality implies that states with older populations will have higher per capita COVID-19 mortality, other things being equal. The percentage increase in mortality risk may be greater for some ethnic groups, or for some co-morbidities such as diabetes or pre-existing lung conditions, which are often a function of low income. Then state differences in ethnic composition, the prevalence of obesity and smoking, and poverty, are likely to influence comparative excess mortality.

Lastly, a state's COVID-19 mortality is increased, and potentially amplified, by limited COVID-19-specific health capacity. The death rate among infected adults depends on capacity constraints on hospital beds and staff, particularly of nurses with expertise, on ventilators, PPE and on testing and on logistical failures in delivery, e.g. to care homes. Given similar initial capacities, a state with a higher average infection rate will be more likely to run into these constraints. By the same logic, given the same high infection rate, a state with lower health capacity would have a higher rate of excess mortality. This is why there is such a focus on 'flattening the pandemic curve'. Different capacity constraints can have different implications for different groups. For example, lack of PPE and testing facilities in care homes will have disproportionately larger effects on mortality for the oldest individuals and this could affect state comparisons. However, as these health capacities evolve over time in response to the pandemic, the influence of differences in pre-existing health capacity is likely to decline. Further, the timing of the pandemic's incidence matters also, as medical interventions became more effective with learning about the nature of the virus and its treatment.

The probability of an individual death from COVID-19, $P(D)$, is the product of the probability of being infected, $P(I)$, and the probability of death *given* infection. Thus,

$$P(D) = P(I) * P(D/I), \text{ so that}$$

$$\log P(D) = \log P(I) + \log(D/I) \quad (1)$$

At the state level, *assuming correct measurement of death counts and infection counts,* aggregating the individual probabilities yields population proportions of infections and deaths. The log of the state COVID-19 mortality rate is then the sum of two functions, the log of the (lagged) infection rate and the log of the average case fatality rate (CFR) for the population of that state (that is, the proportion of infected people who die from the virus):

$$\log(mortality\ rate) = \log\ (lagged\ infection\ rate) + \log\ (CFR) \quad (2)$$

Equation (2) justifies a log formulation of the mortality rate. A further reason arises from the highly skewed nature of the levels data, greatly reduced in the log transformation.

The lagged infection rate will be affected by the variables discussed in Section 1, such as population and housing density, the use of public transit, the proportion of occupations exposure to early infections arriving from Europe, lock-down and social distancing measures and private behaviour responding to the risk of infection and to public measures trying to limit the spread of the virus. The average case fatality rate for the population of that state will vary with factors such as age, race and ethnicity, poverty and inequality, access to good medical care and the capacity of the health system. Our study estimates the cumulative effects of these influences both on infection rates and case fatality rates over 52 weeks.

3.2 *An empirical model for the pandemic*

We adopt a two-stage model. In the first stage, the time of arrival of a significant level of infection for each state is modelled. In the second stage, rates of excess mortality measured either per capita or in terms of P-scores are modelled as a function of the time elapsed, from the end of February to the time of arrival of the infection, and of socioeconomic, political, demographic and environmental factors. For comparison, the dependent variable, per capita COVID-19 deaths, is also tested.

A later local onset of the pandemic should have enabled state and local authorities to take advantage of rapidly improving medical knowledge and capacity (the nature of the disease, treatment regimes, testing capacity, and the effectiveness of policies such as social distancing and masks). Private individuals would also have had more time to learn precautionary behaviour. Kaplan et al. (2020) use a logistic function in time - a 'learning function' to capture the effect of this evolution of behaviours, policies and capacities on health outcomes. We adapt the idea to define a 'learning function' that captures the advantage that some states obtained from the later arrival of the virus. The timing of the arrival of the virus in each state is measured by the first day that the 14-day average of daily cases of

infection reached, or exceeded, a threshold of 6 cases per 100,000 persons. Our 'Timing of onset' function is defined as the inverse of days elapsed from the last day of February to the threshold date (signalling the arrival of serious levels of the virus). The 'Timing of onset' function, like the logistic, has the property that the effect is strong at the beginning, but each additional day of delay matters less and less. The inverse function is the dependent variable in the first stage regression estimated in a cross-section regression across states.

Given the probable undercounting of infections in the first wave, it is likely that the dates when the threshold was breached occurred somewhat earlier than indicated in the reported counts. If the bias was uniform across states, it would not matter much. To the extent that the bias varies with socioeconomic differences between states, the interpretation of estimated socioeconomic effects needs to consider the possibility that, in part, these effects may be compensating for measurement bias in the timing measure. If the bias is independent of political allegiance at the state level, it should not affect the estimated effect of political allegiance on excess mortality.

As New York City had the highest initial incidence of the virus, nearness to New York is likely to have been a factor in explaining the timing for other states. Factors such as the degree of urbanisation of the state, density of its metropolitan areas, the use of public transport, and socioeconomic correlates of dense housing conditions are plausible additional candidates for this first stage model of timing.

The second stage consists of a cross-section regression for the 51 US states of the log of cumulated excess mortality on the timing function and on socioeconomic, political, demographic and environmental factors.

As the literature review on more granular spatial differences indicated, pre-pandemic socioeconomic controls at the state level should include at least the population proportions who are of African American, Hispanic or Asian origin, in the 65+ age group, population density, a measure of health capacity, income and a measure of the incidence of poverty. To these we add the Spring and Autumn temperatures (averaged over March, April and May, and over October to December, respectively) in each state. For excess mortality, very cold weather is likely to induce more influenza and other deaths, as well as increasing COVID-19 deaths by forcing people indoors, where lack of social distancing and of adequate ventilation may increase virus transmission rates. Separating the above into factors affecting the rate of infection vs. those affecting the case fatality rate is typically not possible. For example, if African Americans are more likely to live in crowded housing conditions and work in occupations involving more face-to-face contact, they suffer higher infection rates. In addition, they are likely to suffer higher case fatality rates, for example, because of pre-existing co-morbidities. Similarly, Spring and Autumn temperatures probably affect both the rates of infection and case fatality.

The two-equation model for the 52-week pandemic period may be represented thus:

$$Timing\ of\ onset = g(Z_1, Z_2, \ldots . Z_r) \tag{3}$$

$$\log EMR_{52\,weeks} = F(Timing\ of\ onset, X_1, X_2, \ldots . X_k) \quad (4)$$

where EMR is the cumulative excess mortality rate. In equation (3), the inverse function of days elapsed, is explained by a vector of r pre-pandemic structural variables, denoted by *Z*, where state subscripts have been suppressed. In equation (4), the log of the cumulative rate of excess mortality for the pandemic, $EMR_{52\,weeks}$ is explained by $Timing\ of\ onset$ and a second vector of k pre-pandemic structural variables, denoted by X. There can be overlap between the variables in the vectors *Z* and *X*, but it is crucial for identification that the *Z* vector includes some variables not included in *X*.

The list of relevant variables is by no means exclusive, though there are strong priors based on the evidence from county-level studies. Model selection methods, starting with more general specifications including up to 30 regressors, were used to check for the relevance of the other explanatory variables. Since variation across 51 states is much more limited than across over 3000 counties, sign priors on relevant variables, as well as statistical significance, can help the variable selection process.

For the analysis of cumulative rates of excess mortality in the first 52 weeks of the pandemic, no attempt is made to control for differences in non-pharmaceutical interventions (NPIs) at the state level. State NPIs are endogenous, likely to be switched on when case-counts and COVID-19 deaths rise strongly. The positive correlation induced would bias estimates of the beneficial effects of NPIs on subsequent excess mortality. In order to measure such effects, excess mortality would need to be considered over shorter intervals, and the measures of NPIs lagged to avoid endogeneity bias.

4. DATA

4.1 Dependent variables: excess mortality and COVID-19 deaths

Estimates of excess deaths - defined as the number of persons who have died from all causes, in excess of the expected number of deaths for a given place and time - are from the CDC's National Center for Health Statistics (NCHS), see discussion in section 2.3. Successive vintages of these estimates reveal surprisingly large revisions in estimates of normal deaths and hence excess deaths. One reason is a switch from historical data for 2016-2019 to data for 2017-2019 in late January 2021, to estimate normal deaths.[29] The longer historical sample is likely to result in less noisy estimates at the state level. We therefore used the CDC estimates of 'normal' deaths based on 2016-2019 up to week 3 of 2021. For weeks 4 to 8 of 2021, the CDC estimates of 'normal' deaths in February 2021 based on 2017-2019 were used. We used the weekly count of excess deaths calculated as observed deaths for that week minus the 'normal' (average expected) number of deaths and cumulate over 52 weeks. For weeks where excess deaths are estimated to be negative, we followed the CDC and use a count of zero. The percentage

[29] Private communication from Lauren Rossen of the CDC.

excess deaths (*the P-score*) are excess deaths divided by the expected number of deaths. To calculate *excess mortality per capita*, the excess deaths are divided by 2019 state population (US Census).

Observed death counts are weighted by the CDC to account for incomplete reporting by 51 state jurisdictions in the most recent weeks, and weights are based on completeness of provisional data in the past year as mortality data are recorded with a lag. As we use observed deaths as recorded over 9 months after the end of the period analysed, this is not a significant issue. In the first weeks of the pandemic, our data on the sum of state-level excess deaths are marginally higher than the national data from the CDC. In the rare cases where measured weekly excess deaths are negative, we replace such state-level values by zeroes. However, at the national level, there were no negative weekly excess deaths, in our sample (week 9, 2020 to week 8, 2021).

We compared two sources of COVID-19 death counts, sourced from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and the US Centres for Disease Control and Prevention (CDC), see Table 2.[30]

4.2 Time and learning functions

Studies that capture time variation in the infection and mortality rates note that a later arrival of the virus reduces cumulative COVID-19 attributed mortality. As discussed above, the effect of learning and adaptation gradually fades with time, implying a non-linear function of time elapsed. In place of the logistic function of Kaplan et al. (2020), we use a simpler function with similar properties: the inverse of the number of days elapsed between the end of February 2020 and the day at which a given case-count threshold was breached. The chosen threshold is the day the 14-day average of new infections exceeded 6 per 100,000. To reduce measurement error, we average case infections from two sources: the CDC and The COVID Tracking Project. The latter, widely-used by other researchers, has a more comprehensive data collection, often giving a higher case count. The inverse days measure is normalised by dividing by its mean.

Except for Desmet and Wacziarg (2021), none of the studies cited in Section 1 adequately addresses the bias created by arrival of the virus in some states before others, initially largely by the accident of international travel. Dynamic panel studies with the case count as a variable will in principle control for this, as the case count will reflect early incidence. However, this models deaths conditional on infections but does not explain what drives the infections. The case count is endogenous, and when modelled separately, e.g. in a SEM framework, there ought to be a control such as the enplanement measure of Desmet and Wacziarg (2020) linked with travel from high-severity countries, or a learning function as above. Desmet and Wacziarg (2021) use both *calendar date* and *synchronised* studies. Greater weight

[30] The weekly CDC state-level data record missing values for COVID-19 death counts of between 1 and 9. For states with small populations, there are a number of these low counts. Regressions of the CDC-counts on weekly JHU state-level data are used to fill in the missing values.

should be accorded to the calendar year results because the synchronised sample results suffer from two problems: sample selection and the mixing up of effects that are likely to vary with time.[31] Simply including the number of days elapsed since the first case (e.g. Liao and Maio (2021)) fails to capture the non-linear learning aspect.

4.3 Temperature

In the public health domain, the effects of cold weather on the spread or the severity of the corona virus have been widely discussed,[32] though less so in the scientific literature. Medical research suggests the virus is more stable at low temperatures. In a study of hospital patients, Kifer et al. (2021) find an association between cold weather and mortality. Even if there were no direct link between cold weather and the virus, cold weather drives people indoors, where aerosol spread is a greater risk factor.

Only one of the studies reviewed, Karmakar et al. (2020), includes temperature as a co-variate.[33] Its omission potentially creates an omitted variable bias since cross-state temperature variations are correlated with other characteristics, for example, the Democrat vote share.

We included Spring and Autumn temperatures in our regressions using data from monthly reports on the larger cities in each state from of the National Oceanic and Atmospheric Association (NOAA), National Climate Report. The temperature in °F and the 1981-2020 average temperature in °F were averaged to the state level, and the state-level Spring and Autumn temperatures and deviations from the average were tested in regressions, see Table 2. Spring is defined to include the months from March to May. Autumn covers October to December.

4.4 Characteristics of individuals and communities affecting transmission and vulnerability

The first set of potential determinants, see Section 1, includes characteristics of demography, ethnicity and race, health, poverty, income and inequality, education, employment and occupation, commuting and density. With one exception, all covariates in this group retain their original scale and units to assist understanding of the regression coefficients; but the log of median household income is defined as the deviation around the mean value across states.

Since the higher mortality rates for older people and for 'Blacks and African Americans' and 'Hispanics and Latinos' have been obvious from early in the pandemic, *controls for age and ethnicity*

[31] Many states had not yet reached the '225 days since onset' criterion that defines the synchronised sample by 30th November, and these states are likely to be systematically different from the others. To illustrate the second issue, a cross-section for the synchronised sample will mix counties at quite different points in the calendar year, so that a like-for-like comparison of the effect of differences in the use of public transit, for example, cannot be made. Transit options in the early days of the pandemic differed, since multiple adaptations of transport use occurred subsequently.

[32] Examples are, for the UK, the ONS guidance in ONS (2020), and for the US, the MIT Technology Review.

[33] Rubin et al. (2020) in a dynamic study of COVID-19 cases and deaths find important temperature effects.

are common to most (but not all) studies. Following McLaren (2020), we abbreviate the above two racial categories to 'African American' and 'Hispanic'. We also include the proportions reported as 'Asian', and 'American Indian and Alaska Native'. The age distribution (including proportions of the population aged 0-18 years, and older than 65 years) and proportion of the population in racial and ethnic categories were sourced from the United States Census, American Community Survey (ACS) for 2019, see Table 2. Our general specifications also included the share of multi-generational households, and average family size, from the ACS (2019).

Several measures of *co-morbidities* sourced from the Kaiser Family Foundation (KFF) were tested in the general specifications of our regressions: adults who report smoking, or that they are obese, all in 2019. We also tested uninsured rates for the nonelderly. Categories of *vulnerable persons,* also from KFF, include numbers of residential nursing home residents as a fraction of the over 65s, and the proportion of incarcerated adults in 2019.

Economic variables included: total Gross State Product in 2018 (in millions of current dollars), sourced from U.S. Bureau of Economic Analysis (BEA) via the KFF, and deflated by the 2019 state population; median annual household income from the KFF (2019); the poverty rate from KFF (2019);[34] the US unemployment rate in January 2020 from the KFF; the 2020 St. Louis Fed index of occupations sensitive to the virus; and the 2018 proportion of the population who are below twice the federal poverty income level.

Travel measures included in general specifications of our regressions were the percentage of workers 16 years and over who travelled to work by public transportation (excluding taxicab), and the percent of those commuting alone (by car, van or truck), from the 2018 ACS and enplanements in the top 5 airports in each state.[35] *Educational* variables included the percentage of those over 25 with high school or higher, and also of those over 25 with bachelor's degree or higher, from the ACS (2019).

Various *proximity, density and urbanisation* variables were examined. To capture closeness to the epicentre of the early outbreak in Wave 1, a weighted New York contiguity dummy was constructed for contiguous states, see Table 2. This is the product of a dummy equal to 1 for contiguous states, weighted by the log ratio of the New York State's population to the contiguous state's population, since smaller contiguous states are more likely to be disproportionately affected by their populous neighbour. A dummy was included for remote states defined as Hawaii, Alaska, Maine and Washington State. We calculated a standard measure of population density, defined as the 2019 state population per state area in square km, and used the fraction of each state's population living in large cities and a measure of urbanisation defined as the fraction of each state's population living in urban areas (2010), both sourced

[34] The Gini coefficient, common in several studies, was not used here. The Gini gives a large weight to variations at the top of the distribution, whereas weights at the bottom with vulnerable groups should matter more. Including both poverty and median income should capture inequality.

[35] From 2018 ACS, Tables R0804 and R0802 respectively; and enplanements per state for 2019 from the Federal Aviation Administration website.

from the US census. A more sophisticated measure of urban density using 2010 Census data is the per square km density of urban areas, see Table 2 and Cox (2016).

Several authors have emphasised spill-over effects from commuting in dense Metropolitan Areas, spanning states. We calculated a weighted Metropolitan Statistical Area (MSA) density measure that takes some account of population density in populous overlapping MSAs as follows. Using the 2010 Census state population figures to match the 2010 Census MSA population figures, we calculated first, the actual population of the MSA as a share of the state population. Second, we calculated the average MSA density as the MSA actual population divided by the MSA *occupied* land area. The product of these two is the density of the MSA weighted by the share of MSA population in the state, and it was scaled by 1000. We use a cut-off point for MSAs of populations over 1.5m in 2010. The MSA occupied land area is approximated by multiplying the total MSA land area by the MSA share of state population. This was an elaborate exercise as some MSAs are shared with other states, so that it is required to apportion the part of each shared MSA that belongs to each state. The measure is zero for states in which no MSA's population exceeded 1.5 million.

4.5 Measuring the 'preparedness, resilience and agility' of the public health and social care systems

A second set of potential determinants concern health care capacity, reflected in the availability of PPE, numbers of ICU beds and ventilators, preventive and pre-hospital care, numbers of doctors and critical care nurses, laboratory networks and testing and contact tracing infrastructure. Several measures were sourced from the KFF including the numbers of ICU beds per 10,000 population, of hospital beds per 1000 population, and of critical care nurses per 10,000 adults.

4.6 Political measures

Recent literature adds political partisanship in the US to the subset of drivers of pandemic mortality, which helps to capture private attitudes and behaviour, see Section 1. The hypothesis is that partisanship influences 'compliance' with state-level safety measures that mitigate transmission of infection, coupled with voluntary behaviour to reduce vulnerability. Our measure of partisanship is the Democratic share of the popular vote received in each State in the 2016 Presidential General Election, sourced from the Federal Election Commission of the US, Federal Election Commission (2017), Appendix A. We also included the political affiliation of the Governorship for each state as at 2020, sourced from KFF.

4.7 Interaction effects

Interaction effects were defined between the 'Timing of onset', and the Democrat vote share and log median household income, all taken as deviations from their means, see Section 5 for discussion.

5. POLITICAL ALLEGIANCE AND SOCIOECONOMIC FACTORS IN THE PANDEMIC

5.1 The two-stage model and the role of partisanship

The two-equation model of Section 3.2, represented in equations (3) and (4), was applied across 51 US states (including Washington D.C) using two-stage least squares (2SLS) and OLS. Table 2 provides definitions and sources for the data. The 'Timing of onset' function corresponding to equation (3) was estimated in a first stage, see Section 4.2 and Table 2 for the definition of the dependent variable. The chosen specification is the result of the reduction from a more general to a parsimonious formulation, on plausible correlates of early arrival of infections. The fitted value was used as an instrument in estimating the second-stage regression of the equation for the log of cumulative per capita excess mortality. This helps address the probable endogeneity of the timing of the pandemic's arrival in each state.

The 'Timing of onset' function has its highest value for New York, clearly the first state to be seriously affected, followed by New Jersey, Michigan, Vermont, Louisiana, Massachusetts, and Connecticut. Those states hit early had a double disadvantage: a longer period for deaths due to the pandemic to cumulate and less time to benefit from learning about appropriate public and private behavioural and medical responses.

The estimated first-stage equation is shown in Table 3. The early arrival of the pandemic is explained by three geographical measures, and by the percentage of the population who are African American, by median household income and by the Spring temperature. A lower median income and a lower Spring temperature are associated with the case-count threshold being breached earlier. The geographical measures are a measure of nearness to New York for the contiguous states (zero for the non-contiguous states), a measure of population density for the metropolitan areas in each state and an index of urbanisation.

The dependent variable for the second equation is the log of the per capita cumulative excess mortality rate, *EMR*, for 52 weeks. Similar models are estimated for the log P-score and log per capita COVID-19 deaths, see Table 4. The first column of Table 4 shows the crude correlation, controlling only of the remoteness dummy, between *log EMR* and the Democrat vote share. The estimated second-stage equation for *log EMR,* using two-stage least squares, is shown in column 2, followed by the OLS estimates in column 3. The estimates in these two columns are fairly close, despite probable endogeneity bias. Columns 5 and 7 show 2SLS estimates for, respectively, the log P-score and log per capita COVID-19 deaths as the dependent variables (the corresponding crude correlations are shown in columns 4 and 6).

Several controls are common to the majority of studies cited in Section 1: measures of density and urban structure, measures of race and ethnicity, the age structure, poverty and income Given the widespread discussion of temperature and our prior that states where the pandemic arrived first suffered

a serious disadvantage, this suggested a basic set of 13 controls plus an intercept, including three geographic measures: remoteness, state population density and urban density. We also controlled for two interaction effects, the first between 'Timing of onset' and the Democrat vote share, and the second between the 'Timing of onset' and log median household income. The former effect would capture more cautious behaviour by Democrat voters mattering more for mortality when the risks were particularly pronounced, as was the case in those states hit hardest early on. Given the pandemic was seeded by the arrival of fairly affluent travellers from Europe, the latter interaction effect would suggest a positive link with higher income states. Desmet and Wazciarg (2021) find that the early positive correlation between COVID-19 mortality and income switches to negative as the pandemic progressed. This might suggest that early arrival states, where the 'Timing of onset' is above average, would experience a positive income effect, while late arrival states would have a negative income effect. Other controls were discussed in Section 4, and included the proportion of workers using public transit, the proportion of those aged under 65 without health insurance, the ratio of nursing home residents to the population aged 65 or above, and 20 other variables.

The Autometrics software of Doornik and Hendry (2018) has the option of searching over a broad set of other controls in a general-to-specific reduction, given the retention of a basic set of key controls. The software was used to confirm that none of these other controls was statistically relevant, resulting in the parsimonious specification shown in columns 2 and 3. A non-nested test, see Aneuryn-Evans and Deaton (1980), strongly supports the log version of the dependent variable versus the linear alternative: the log of the fitted value from the linear version of the equation is insignificant when added to the log specification as shown in columns 2 or 3. However, adding the exponential of the fitted value from the log version to the linear version gives a highly significant result, implying that the linear version is seriously mis-specified. Replacing the 2016 Democrat vote share by the equivalent 2020 vote share, makes little difference to the results, with a slightly lower (negative) coefficient on the Democrat vote share.

5.2 Robustness checks

The robustness of the findings for log per capita excess mortality is demonstrated in Table 5, in turn, dropping the first 10 observations, the second ten, and so on, to the last ten observations. This demonstrates the relative stability of the coefficients on the Democrat vote-share and its interaction with the 'Timing of onset', on the Democrat Governor dummy, and on the proportions of African Americans and Hispanics. All the other parameter estimates (not shown) easily cover the full sample estimates within a 95 percent confidence interval. The implication is that the results are clearly not driven by outliers concentrated in a few states and are fairly insensitive to the exclusion of particular states.

5.3 Comparing results for the P-score and per capita measures of excess mortality

Comparing coefficients in Table 4 for the log P-score measure with those for the specification with log per capita excess mortality shows an only slightly lower (negative) coefficient on the Democrat vote share, somewhat lower coefficients on the percentage of poor residents, population density and on Spring temperature. It is striking that the effect of age composition disappears entirely (as the t-ratio is 0.2, the variable is omitted). As the P-score measures excess deaths relatively to 'normal' deaths, it already captures some differences in mortality due to pre-existing co-morbidities, of which age is the most important. The effects of race and ethnicity are broadly similar for the per capita and P-score measures. By the same token, this suggests that the effects of race and ethnicity are *not* related to the higher, pre-pandemic mortality rates of minority populations.

These findings have implications when comparing the plain P-scores across states and countries. While for basic comparisons this is probably the best measure, and preferable to per capita measures, even for P-scores, structural socioeconomic and environmental differences need to be taken into account. In other words, P-scores do not fully capture the differences in racial and ethnic composition, and in poverty and urban density, despite being normalised against normal deaths. *Unqualified* comparisons not just of per capita excess deaths, but even of the preferred P-score measure, should not be used to assess the relative performance of public policy in different locations.

5.4 Comparing results with the COVID-19 per capita death rate

Given alternative sources of COVID-19 death counts in the US, a comparison was made to select the more robust measure on the basis of whether there is mis-measurement against the excess deaths measure. In time series regressions of aggregate US data of log per capita COVID-19 deaths on log per capita excess deaths, the R-squared is higher and the standard error lower for CDC data than for JHU data, whether or not the first few weeks are included. In cross-state regressions of the 52-week cumulative per capita data, the same conclusion is reached. Even though excess deaths also include spill-overs in deaths from conditions untreated because health systems were overwhelmed, over a 52-week period and cross-state variation, one would not expect such spill-overs to substantially bias the relationship between true COVID-19 death counts and excess deaths. We therefore concluded that the CDC COVID-19 death count is less inaccurate than the JHU data.

There are striking differences in the state rankings by per capita excess mortality versus the rankings by per capita COVID-19 deaths, see Figure 2 and Section 2. Thus, it is somewhat surprising that the estimates in columns 2 and 7 are not more different. For the per capita COVID-19 deaths measure, the effects of the Democrat vote share and the Democrat Governor effect are, respectively, a little stronger, and weaker; the timing effect is slightly stronger; and the proportions of African Americans and Hispanics have somewhat stronger effects, though for Asians, prove less significant. The interaction effects with timing of the pandemic are even stronger for the COVID-19 measure than for the two

excess mortality rate measures. However, consistent with substantial measurement errors, the fit for the COVID-19 specification is much worse, with the equation standard errors twice as high.

5.5 The interpretation of the effects of the controls

The literature cited in Section 1 on the role of partisanship in the pandemic has explored the links between the rates of COVID-19 infections and deaths and political attitudes and beliefs, reflected in private behaviours (such as mask-wearing and social distancing) and compliance with official advice and mandates. The Democrat vote share can be interpreted as a proxy for compliance and informed private behaviours, when controlling for both the differential onset across states of severe outbreaks and the different risk groups. This interpretation accords well with the findings at county-level of Desmet and Wacziarg (2002) and Gollwitzer et al. (2020).

As explained in Section 3, the cross-sectional equations presented in Table 4 are reduced-form equations which mix the effects governing infection rates and those governing mortality (given infection), as well as the pandemic's indirect effects on other types of deaths. For example, the coefficient on the proportion of African Americans in the population may be connected with higher infection rates in states with higher proportions of African Americans, as well as with their higher case-fatality rate. On the face of it, the estimated coefficient of 2.11 in Table 4 column 2, implies that a 1 percent shift in the population from White to African American results in a 2.11 percent increase in excess mortality. However, this cannot be given a strict interpretation of individual mortality risk faced by an African American, even with the other controls in our regression (including poverty, political allegiance, population density and the age distribution). It might be that states with high proportions of African Americans have other characteristics, not controlled for, raising mortality risk. No studies of which we are aware control for differences in wealth between African American and other households, and, as Hardy and Logan (2020) point out, wealth inequality between African Americans and Whites is far greater than earnings inequality. It is plausible that accurate controls for wealth, educational quality, family composition and discrimination (e.g. in labour, housing and credit markets), would greatly reduce and perhaps eliminate racial differences in excess mortality rates.

Our racial-ethnic estimates are broadly in line with those of county-level studies of COVID-19 mortality rates. County-level measures for the effects of variations in the proportion of African Americans, with Whites as the reference group, typically vary in a range from about 1.5 to 3, according to other controls included and the period covered, e.g. McClaren (2020). Similarly, the effect of variations in the proportion of Hispanics, at somewhat over half of the effect for African Americans, is

also not far from county-level estimates. The coefficient on the proportion of Asians is similar to that for African Americans but much less precisely estimated.[36]

It is important to note the important role played by partisanship in these estimates of racial and ethnic disparities. The Democrat vote share effect is highly significant and robust to the exclusion of ten states at a time from the cross-section regressions for COVID-19 related mortality. As racial and ethnic minorities tend to vote disproportionately for the Democratic Party, their population shares are strongly positively correlated with the Democrat vote share, which has a negative effect on excess mortality. Therefore, if the Democrat vote share was omitted from the cross-state regression, this would result in a downward omitted variable bias on the coefficients for the population shares of Afro-Americans and Hispanics. Indeed, the omission almost halves the estimated coefficients for Afro-Americans and Hispanics, with a substantial loss of precision (these results are not reported in Table 4).

The coefficient of 4.4 on the percentage of residents aged 65 or more is consistent with the steep age gradient of COVID-19 mortality and the fact that hardly any deaths occur for those under 18. The estimated coefficient of 7.0 on the percent classified as poor, though broadly consistent with studies showing strong links between economic deprivation and COVID-19 mortality, cannot be taken too literally. On the face of it, it implies that a 1 percent of population increase in those below the poverty line, implying a 1 percent decrease in those above, results in 7.0 percent increase in excess mortality. The figure is surprisingly high given that the percentages of African American and Hispanic residents are also being controlled for, and poverty rates for these groups are above average. It is likely that being classified as poor is associated with other unobserved characteristics that raise mortality risk.[37] The positive interaction effect between timing of onset and median income in a cross-state regression, given controls for race, ethnicity and poverty, likely reflects the fact that many of those who first seeded the infection in the US were affluent travellers returning from Europe. It implies a negative effect of higher incomes on mortality in late onset states. This could be related to the ability of the more affluent to afford good medical care and to avoid close contacts that raise infection risk.

Differences in state population density (measured as population per square km) and in urban density have the expected effects, consistent with the great majority of granular studies cited in Section 1. Through the 'Timing of onset' function, there is an additional effect from density measured for the MSAs to which each state belongs as well as a measure of urbanisation and a control for bordering on New York state. The estimated effect for Spring temperature, measured in degrees Fahrenheit, suggests that a one-degree higher average temperature is associated with a 2 percent lower rate of excess

[36] Rossen et al. (2021) estimate 'normal' deaths by age and racial group at the national level. They report disparities in excess mortality incidence rates in 2020 for different age groups and races. The rate per 100,000 in the 65+ age group for Afro-Americans and Hispanics is just over double that for Whites; for the 25-64 age group, the Afro-American rate is 2.6 times that of Whites, and for Hispanics it is 1.9 times that of Whites. For those of Asian descent, the rates are similar to those of Whites.

[37] Examples are co-morbidities, working in a meat packing plant or in seasonal agriculture without health facilities.

mortality for the full period of 52 weeks. Even if there were no direct link between cold weather and the virus, the fact that cold weather drives people indoors, where aerosol spread is a risk factor, is widely suspected of association with excess mortality. Some studies of historical patterns of mortality, e.g. Kontis et al. (2020), find significant temperature effects with low Spring temperatures and high Summer temperatures associated with higher death rates. As the CDC does not use temperature controls to estimate 'normal' death rates, part of what our temperature effect captures could be due to higher mortality that would have occurred even without the pandemic.

5.6 Summary

For the full 52 weeks of the pandemic analysed, the bilateral correlation is close to zero between any of the three COVID-19 related mortality measures and the 2016 Democrat vote share. Given the inclusion of a set of plausible controls, however, states with higher Democrat vote shares, experienced lower COVID-19 related mortality on all three mortality measures. This finding parallels the evidence at a county-level for data to the end of November 2020 from Desmet and Wacziarg (2021). The finding is consistent with the more cautious and better-informed behaviour by Democrat voters in the 2016 election. Moreover, the interaction effects suggest the negative Democrat vote share effect on mortality was even greater in states where the infection arrived early. If the Democrat vote share is omitted, this results in an under-estimation of the estimated disparities in excess mortality suffered by Afro-Americans and Hispanics.

6. CONCLUSIONS

This paper is the first state-level, spatial analysis of excess mortality across the 51 US states, showing for the full year since the arrival of the pandemic in the US, the effects of racial composition, age structure, poverty, income, the timing of the pandemic onset, temperature, population density and other structural features, and political partisanship. We have focused on two excess mortality measures in a log formulation: per capita excess mortality and the P-score (excess deaths relative to normal deaths). Analysing the drivers of excess mortality measures, rather than counts of COVID-19 deaths as typically used in epidemiological studies, avoids the well-documented mismeasurement biases from under-reported pandemic-related cases and deaths. Our paper clarified definitions and data measurement issues around excess mortality, considering data quality and comparability both internationally and within the US.

A reduced form empirical specification was derived from the theoretical link between the mortality rate and lagged infection rates and average case fatality rates. A log-linear formulation captured a mixture of the influences on infection rates and case fatality rates with co-variates common to granular

studies of COVID-19 per capita death and infection counts. Unlike in most cross-sectional studies, the selection of relevant regressors was not ad hoc, or based on bilateral correlations, but checked against a general to specific econometric analysis from a wide range of initial controls. This set included important socioeconomic regressors, temperature, the timing of the onset of the pandemic, and interaction effects to capture plausible non-linearities, each rarely included in published studies. Our two-stage approach modelled first, the timing of the pandemic across states, and then using two-stage least squares, second stage models for log excess mortality rates. This helped avoid the endemic problem found in almost all the studies we have cited (save for Desmet and Waziarg (2021) of a serious omitted variable bias from the differential arrival in time of pandemic cases across states. Non-nested tests confirmed that the log formulation is far superior to the additive linear formulation used by many studies to model per capita COVID-19 deaths. The latter formulation is a serious mis-specification given that the theory also supports an additive formulation in logs. In general, our study has tried to avoid empirical shortcomings from inappropriate choice of functional form, the exclusion of key controls, and types of selection and measurement biases.

The inclusion of political partisanship adds an important omitted variable to the more typical set of regressors, which are focused on the characteristics affecting transmission risk and vulnerability to infection and the preparedness and capacity of the public health and social care systems. Our evidence is that states with higher Democrat vote shares experienced lower excess mortality rates, controlling for a broad set of the underlying risk factors. This implies more cautious and better-informed behaviours by those who voted Democrat in the 2016 election. These findings, linking partisan differences to mortality outcomes in the pandemic, are strongly consistent with recent studies that clarify the impact of partisanship on actual behaviour. Moreover, the interaction effects suggest that the negative Democrat vote share effect on mortality was even greater in states where the infection arrived early. While our finding parallels the evidence at a county-level for data to the end of November 2020 from Desmet and Wacziarg (2021), interaction effects have not been considered in county-level cross-sectional studies of COVID-19 deaths. Mostly such studies have also not taken Spring temperatures into account. Low Spring temperatures increased COVID-19 related mortality. The absence of interaction effects and the fact that Spring temperatures tend to be lower in states with larger Democrat votes shares, may also suggest that previous estimates of the effect of partisanship on COVID-19 deaths have under-estimated the mortality-reducing effect of the Democrat vote share.

A striking implication of our findings is that the failure in many spatial county-level or state-level studies to control for the effect of political partisanship on COVID-19 related mortality likely resulted in a downward omitted variable bias of the disparities associated with being Afro-American and Hispanic and hence an under-estimation of the effects of race. This is the consequence of a positive correlation between minority population shares and the Democrat votes share, but a negative correlation between the Democrat votes share and COVID-19 related mortality.

No attempt was made to control for differences in non-pharmaceutical interventions (NPIs) at the state level for cumulative rates of excess mortality in the 52-week period as NPIs are likely to be switched on when case-counts and COVID-19 deaths rise strongly. To measure such effects, excess mortality would need to be considered over shorter intervals, and the measures of NPIs lagged to avoid endogeneity bias.

The robustness of our analysis was demonstrated in Section 5.5. We also compared models for the two dependent excess mortality variables (i.e., per capita excess deaths and the P-score). The rankings of US states according to the per capita and P-score measures of excess mortality are notably different, see Section 2. Despite the differences, the cross-section models of state differences for the two excess mortality measures find similar strong effects for partisanship and broadly similar interpretations for the socioeconomic variables. The P-score is the preferred measure for simple cross-country comparisons since it is scaled by 'normal deaths' (taking some account of differentials in age composition and socioeconomic characteristics), but inclusion of comprehensive controls in a multivariate statistical study reduces this advantage over the per capita measure of excess mortality. As might be expected, age drops out in models for the P-score, but it is an important control in models for per capita excess mortality. However, it is striking that there are equally strong racial and ethnic effects for the P-score. These go beyond what is captured in the pre-pandemic 'normal deaths', suggesting levels of discrimination and disadvantage during the pandemic well above those previously prevailing.

Repeating the analysis with the log COVID-19 deaths per capita measure as dependent variable finds a similarly strong political effect, and similar socioeconomic controls mattering, but the equation fit is substantially worse than for excess deaths per capita (the fit is worse still when using the JHU-sourced COVID-19 death count).

Our findings have implications for further research on more granular data. Currently, the US CDC does not produce estimates of weekly excess deaths down to the county level. Such data can be very noisy for counties with small populations. Moving to a monthly or even quarterly frequency would ameliorate this problem and make more granular analysis possible. We also suggest that, at the state level, the CDC control for changes in population and age composition for improved estimates of 'normal' and hence excess deaths.

For making useful comparisons of pandemic related rates of mortality across countries and states, in order to evaluate public policy choices, our findings suggest that while the P-score measure is preferable to per capita excess mortality, it is far from immune to structural differences between countries. The timing of the pandemic, poverty, racial and ethnic composition, occupational structure and the nature of urban density all need to be taken into account in gauging the success or otherwise of public policies in different locations. It would be highly desirable for parallel studies of excess deaths to be carried out. International comparability is harder in these dimensions given difficulties in standardising categories in measures of deprivation, occupational classification (sometimes not

recorded on death certificates, but recoverable from census records) and missing data for some countries on the sensitive issue of ethnicity. The international NUTS classification of regions[38] provides a possible comparable frame for international comparisons. As regions differ in their urban/rural structure, comparing regional data can give important insights into risk factors for death rates. Moreover, as the incidence of the pandemic differs in timing and intensity, regional comparisons can throw light on the dynamics of the spread of infections.

REFERENCES

Allcott, H., L. Boxell, J. Conway, M. Gentzkow, M. Thaler and D. Yang, "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic", *Journal of Public Economics*, 191, 104254, 2020.

Almagro, M., J. Coven, A. Gupta and A. Orane-Hutchinson, "Racial Disparities in Frontline Workers and Housing Crowding during COVID-19: Evidence from Geolocation Data", Opportunity and Inclusive Growth Institute Working Papers 37, Federal Reserve Bank of Minneapolis, 2021. DOI: 10.21034/iwp.37. February 2021

Almagro, M., J. Coven, and A. Orane-Hutchinson, "The determinants of the differential exposure to COVID-19 in New York City and their evolution over time", *Covid Economics,*13, 4, May 2020.

Altman, D., "Understanding the US failure on coronavirus", *British Medical Journal,* 14 Sept. 2020.

Aneuryn-Evans, G. and A. Deaton, "Testing Linear *versus* Logarithmic Regression Models", *The Review of Economic Studies*, 47, 1, 275–291, 1980.

Aron, J. and J. Muellbauer, "A selected survey of US spatial empirical research on the determinants of mortality in the COVID-19 pandemic." MS. Department of Economics, Oxford University, 2021.

Aron, J. and J. Muellbauer, "Measuring excess mortality: England is the European outlier in the COVID-19 pandemic", VoxEU.org, 18 May, 2020a. [See also "Measuring excess mortality: the case of England during the Covid-19 Pandemic", INET Oxford Working Paper No. 2020-11.]

Aron, J. and J. Muellbauer, alongside C. Giattino and H. Ritchie, "A pandemic primer on excess mortality statistics and their comparability across countries", Our World in Data, University of Oxford, 29 June, 2020b.

Aron, J. and J. Muellbauer, "The US excess mortality rate from COVID-19 is substantially worse than Europe's", VoxEU.org, 29 Sept., 2020c. [See also "Transatlantic excess mortality comparisons in the pandemic", INET Oxford Working Paper No. 2020-11, 25 August, 2020.]

Davies, B., B.L Parkes, J. Bennett, D. Fecht, M. Blangiardo, M. Ezzati and P. Elliott, "Community factors and excess mortality in first wave of the COVID-19 pandemic", 2020. doi: https://doi.org/10.1101/2020.11.19.20234849

Bilinski, A., and E.J. Emanuel, "COVID-19 and Excess All-Cause Mortality in the US and 18 Comparison Countries", *JAMA,* 324(20), 2100–2102, 2020.

Castle, J.L., J.A. Doornik and D.F. Hendry, "Robust Discovery of Regression Models", ms., Economics Department, Oxford, 2020.

Chen Y., Glymour, M.M., Catalano, R., *et al*., "Excess Mortality in California During the Coronavirus Disease 2019 Pandemic, March to August 2020", *JAMA Intern Med.*, 2020. doi:10.1001/jamainternmed.2020.7578

Cox, W., "America's Most Urban States", New Geography, July, 2016

Desmet, K and R. Wacziarg, "Understanding spatial variation in COVID-19 across the United States", *Journal of Urban Economics*, 2021. https://doi.org/10.1016/j.jue.2021.103332

Doornik, J. A. and D. F. Hendry, *OxMetrics: An Interface to Empirical Modelling* (8th ed.), London: Timberlake Consultants Press, 2018.

Doti, J.L., "Examining the impact of socioeconomic variables on COVID-19 death rates at the state level", J Bioecon, 2021.

Edwards, M. and S. McDonald, "The co-morbidity question", The Actuary, Institute and Faculty of Actuaries, 7th May 2020.

[38] Eurostat has expanded regional mortality data by the NUTS classification, which should greatly aid research.

Druckman, J.N., S. Klar, Y. Krupnikov, M. Levendusky and J.B. Ryan, "Affective polarization, local contexts and public opinion in America", *Nat Hum Behav,* 5, 28–38, 2021.
Farrington, C.P., N.J Andrews, A.D. Beale and M.A. Catchpole, "A statistical algorithm for the early detection of outbreaks of infectious disease", *Journal of the Royal Statistical Society A* 159: 547-563, 1996.
Faust, J.S., C. Du, S-X Li, Z. Lin, H. M. Krumholz, "Correcting excess mortality for pandemic-associated population decreases", *medRxiv* 2021. doi: https://doi.org/10.1101/2021.02.10.21251461
Federal Election Commission, "Election Results for the U.S. President, the U.S. Senate and the U.S. House of Representatives, Federal Elections 2016", Washington, D.C., Federal Election Commission, December 2017.
Fineberg, H.V., "The Toll of COVID-19", *JAMA* 324(15), 1502-1503, 2020.
Gerritse, M., "Cities and COVID-19 infections: population density, transmission speeds and sheltering responses", *Covid Economics* 37: 1-26, Centre for Economic Policy Research, 2020.
Gollwitzer, A., Martel, C., Brady, W.J. *et al.,* "Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic", *Nat Hum Behav*, 4, 1186–1197, 2020.
Hale, T., J. Anania, N. Angrist, T. Boby et al., "Variation in government responses to COVID-19", BSG-WP-2020/032, Version 11.0, Blavatnic School of Government, University of Oxford, March 2021.
Hamidi, S., S. Sabouri and R. Ewing, "Does Density Aggravate the COVID-19 Pandemic?", *Journal of the American Planning Association*, 86:4, 495-509, 2020. DOI: 10.1080/01944363.2020.1777891
Hardy, B. and T. D. Logan, "Racial economic inequality amid the COVID-19 crisis", Brookings Report, 13 August 2020.
IHME COVID-19 Forecasting Team., Reiner, R.C., R.M. Barber et al., "Modeling COVID-19 scenarios for the United States", *Nat Med*, 27, 94–105, 2021. https://doi.org/10.1038/s41591-020-1132-9
Kaplan, G., B Moll and G. L. Violante, "The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S.", NBER Working Paper Series, Sep. 2020. DOI 10.3386/w27794.
Karmakar, M., P.M. Lantz, and R. Tipirneni, "Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US", *JAMA* 4(1), e2036462, 2021.
Kelly, J., "Spiegelhalter says majority of Covid deaths would not have occurred in coming year", Financial Times, 22 May 2020.
Kifer, D., D. Bugada, J. Villar-Garcia et al., "Effects of Environmental Factors on Severity and Mortality of COVID-19", *Frontiers in Medicine*, 7, 1099, 2021.
Kinsella, M. G. Fowler, J. Boland and D.I. Weiner, "Trump administration abuses thwart US pandemic response", Brennan Center for Justice, 29 September, 2020.
Knittel, C.R., and B. Ozaltun, "What does and does not Correlate with Covid-19 Death Rates", Working Paper 27391 NBER Working Paper Series, June 2020. http://www.nber.org/papers/w27391
Koh, H.K., A.C. Geller and T.J. VanderWeele, "Deaths From COVID-19", *JAMA* 325(2),133–134, 2021.
Kontis, V., J.E. Bennett, T. Rashid, R.M. Parks, J. Pearson-Stuttard, M. Guillot, P. Asaria, B. Zhou, M. Battaglini, G. Corsetti, M. McKee, M. Di Cesare, C.D. Mathers and M. Ezzati, "Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries", *Nature Medicine*, 26, 12, 1919-1928, 2020.
Liao T.F. and F. De Maio, "Association of Social and Economic Inequality With Coronavirus Disease 2019 Incidence and Mortality Across US Counties", *JAMA,* 4(1), e2034578, 2021.
Loomes G. and L. McKenzie, "The use of QALYs in health care decision making", *Soc Sci Med*., 28(4), 299-308, 1989. doi: 10.1016/0277-9536(89)90030-0. PMID: 2649989.
Makridis, C.A and J.T. Rothwell, "The Real Cost of Political Polarization: Evidence from the COVID-19 Pandemic", *Covid Economics,* 34, 47-58, 2020.
McLaren, J., "Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data", NBER Working Paper No. 27407, 2020. See also "Racial disparity in COVID-19 deaths: Seeking economic roots in census data" VoxEU.org, 11 August 2020.
Morgan, D., J. Ino, G. Di Paolantonio and F. Murtin, "Excess mortality: Measuring the direct and indirect impact of COVID-19", OECD Health Working Papers No. 122, 2020. https://dx.doi.org/10.1787/c5dc0c50-en
National Audit Office, "Readying the NHS and adult social care in England for COVID-19", Department of Health and Social Care, June 12, 2020. ISBN: 9781786043191.
National Center for Health Statistics, "Excess deaths associated with COVID-19", Atlanta, GA: US Department of Health and Human Services, CDC, National Center for Health Statistics; 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm

Noufaily A., D.G. Enki, P. Farrington, P. Garthwaite, N. Andrews and A. Charlett, "An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems". *Statistics in Medicine*, 32(7):1206-1222, 2012.

ONS 2020a "Updating ethnic contrasts in deaths involving the coronavirus (COVID-19), England and Wales: deaths occurring 2 March to 28 July 2020", Office for National Statistics, October 2020a.

ONS, "Guidance: Health matters: cold weather and COVID-19", Office for National Statistics, November 2020b.

ONS, "Monthly mortality analysis, England and Wales: March 2021", Office for National Statistics, March, 2021a. (Regularly updated).

ONS, "Comparisons of all-cause mortality between European countries and regions: 2020: Mortality patterns of selected European countries and regions, week ending 3 January 2020 (Week 1) to week ending 1 January 2021 (Week 53) and up to week ending 12 February 2021 (Week 6) for the UK", Office for National Statistics, March, 2021b. (Regularly updated).

Pifarré i Arolas, H., E. Acosta, López-Casasnovas et al., "Years of life lost to COVID-19 in 81 countries", *Sci Rep*, 11, 3504, 18 February 2021. https://doi.org/10.1038/s41598-021-83040-3

Pretis, F., J. Reade, and G. Sucarrat, "Automated General-to-Specific (GETS) Regression Modeling and Indicator Saturation for Outliers and Structural Breaks", *Journal of Statistical Software*, 2018.

Rossen L.M., F.B. Ahmad, R.N. Anderson, et al., "Disparities in Excess Mortality Associated with COVID-19 — United States, 2020." Morbidity and Mortality Weekly Report (MMWR) Rep 70: 1114–1119, Centers for Disease Control and Prevention (CDC), 2021. DOI: http://dx.doi.org/10.15585/mmwr.mm7033a2

Rubin D, J. Huang, B. T. Fisher et al., "Association of Social Distancing, Population Density, and Temperature with the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States", *JAMA Netw Open*, 3(7), e2016099, 2020.

Spiegelhalter, D., "How much 'normal' risk does Covid represent?", Winton Centre for Risk and Evidence Communication, Cambridge, 21 March, 2020.

Shiels M.S., J.S. Almeida, M. García-Closas, P.S. Albert, N.D. Freedman, and A.B. de González, "Impact of Population Growth and Aging on Estimates of Excess U.S. Deaths During the COVID-19 Pandemic, March to August 2020", *Ann Intern Med.*, 2020. pmid:33316174.

Stokes A.C., Lundberg D.J., Elo I.T., Hempstead K., Bor J., Preston S.H., "COVID-19 and excess mortality in the United States: A county-level analysis." PLoS Med 18(5): e1003571. https://doi.org/10.1371/journal.pmed.1003571, 2021.

Van Bavel, J. J. and A. Pereira, "The Partisan Brain: An Identity-Based Model of Political Belief", *Trends in Cognitive Sciences*, 22, 3, 213-224, 2018.

Weinberger, D.M., J. Chen and T. Cohen, *et al*. "Estimation of Excess Deaths Associated with the COVID-19 Pandemic in the United States, March to May 2020", *JAMA Intern Med.* 180(10), 1336–1344, 2020.

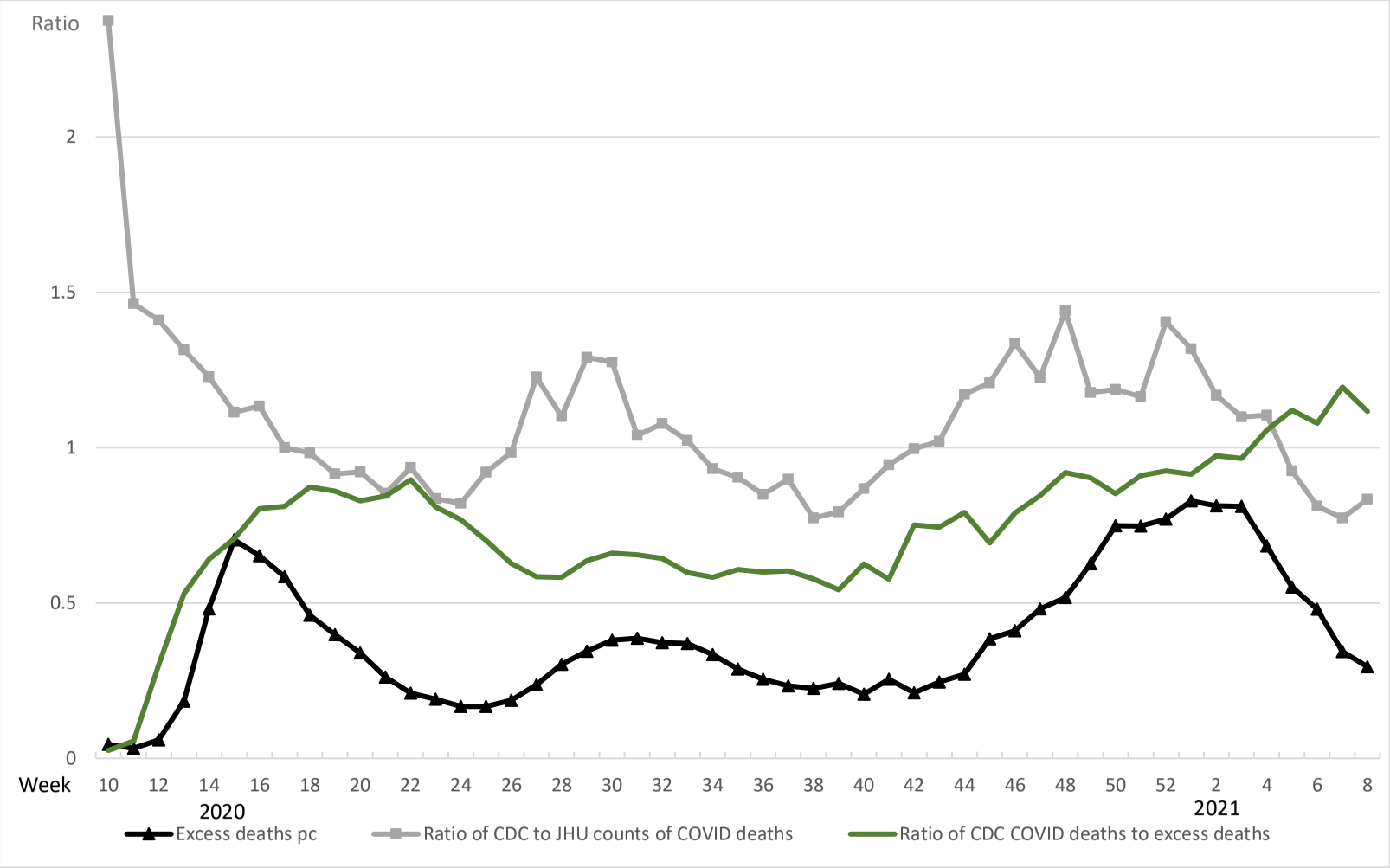
Wieler, L., U. Rexroth and R. Gottschalk, "Emerging COVID-19 success story: Germany's strong enabling environment", Guest Post, Our World in Data, June 2020.

Wieler, L., U. Rexroth and R. Gottschalk, "Emerging COVID-19 success story: Germany's push to maintain progress", Guest Post, Our World in Data, March 2021.

Woolf, S.H., D.A. Chapman, and J.H. Lee, "COVID-19 as the Leading Cause of Death in the United States", *JAMA* 325(2), 123–124, 2021.

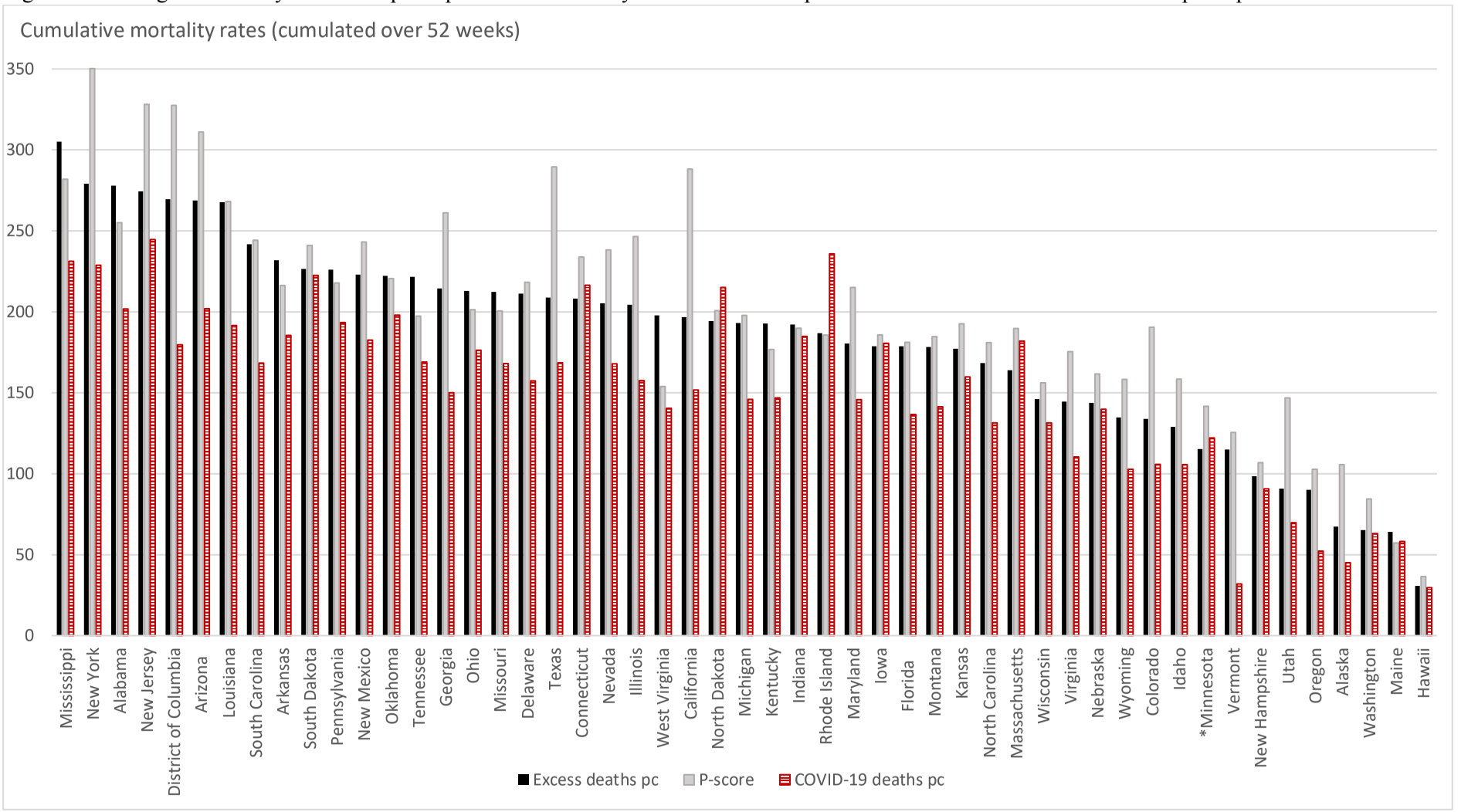
Woolf, S.H., D.A. Chapman, R.T. Sabo, D.M. Weinberger, L. Hill and D.D.H. Taylor, "Excess Deaths From COVID-19 and Other Causes", *JAMA* 324(15), 1562–1564, 2020.

Figure 1: Weekly US per capita excess deaths, the ratio of CDC-sourced to JHU-sourced COVID-19 deaths, and the ratio of CDC-recorded COVID deaths to excess deaths



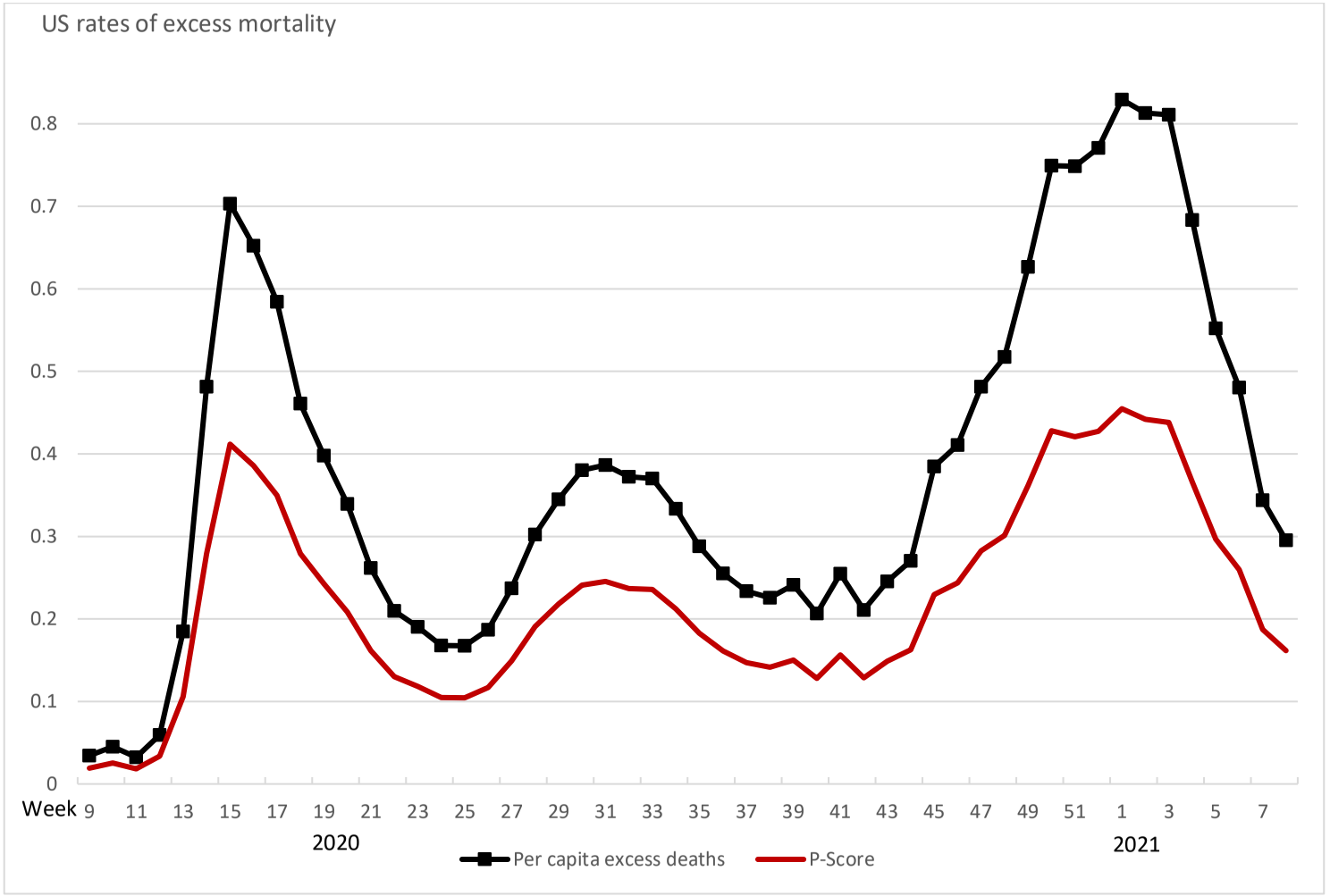
Notes: Calculations by the authors using data from the US Centres for Disease Control and Prevention (CDC) and the Coronavirus Resource Center, Johns Hopkins University, see Tables 1 and 2. Weekly per capita excess deaths are expressed as per 100,000.

Figure 2: Ranking US states by cumulated per capita excess mortality for 52 weeks: comparisons with P-scores and CDC-sourced per capita COVID-19 deaths



Notes: Calculations by the authors using data from the US Centres for Disease Control and Prevention (CDC), see Tables 1 and 2. Cumulative per capita excess and COVID-19 deaths are expressed per 100,000. The P-scores have been multiplied by 1000.

Figure 3: Weekly excess mortality per capita and P-score for the US



Notes: Calculations by the authors using data from the US Centres for Disease Control and Prevention (CDC), see Tables 1 and 2. Weekly per capita excess deaths are expressed as per 100,000.

Table 1: Measures of pandemic incidence, deaths and excess mortality used in spatial studies

Measure	Definition	Sources	*Comparability* across regions, states and countries (with reference to measurement and data quality)	Spatial studies using this measure
Measures of COVID deaths and COVID-related deaths and COVID-cases				
Case count	$COVID\ case\ count$	National authorities, e.g. Office for National Statistics (ONS) in the UK; National Center of Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC) in the US.	Poor comparability due to differential measurement biases by location, from missed diagnosis and constraints on testing capacity. Has improved over time with better capacity; highly variable across countries.	Not used.
Per capita case count	$\frac{COVID\ case\ count}{population}$		As above, but further compromised by poor population statistics in some cases.	Widely-used.
COVID deaths	$COVID\ deaths$ (as attributed by country definitions)		Poor comparability due to measurement errors. Some countries have poor systems for recording deaths.	Not used.
Per capita COVID deaths	$\frac{COVID\ deaths}{population}$		As above, but further compromised by poor population statistics in some cases.	Widely-used.
Age-standardised COVID deaths	$\sum_i(w_i\ p_i)$ where w_i is the fraction of the reference population in age group i, and p_i is the age-specific *COVID death rate* for age group i.	e.g. ONS (2021a) regularly updated article.	Poor comparability due to measurement errors.	ONS (2020a)
Measures of excess mortality				
Excess deaths	$X - \bar{X}$ Denote by X: the number of *per period* deaths. Denote by $\bar{X}$: expected value of X for the population (i.e. 'normal' deaths).	e.g. Eurostat (Europe), CDC (US), The Human Mortality database (HMD) for 38 countries; World Mortality Database (WMD), WHO Mortality database.	Requires great care. Some countries have poor systems for recording deaths. Almost everywhere there are significant lags in recording deaths. Techniques differ in the estimation of 'normal' deaths; sometimes historical data are absent. Comparative data quality is discussed in Section 2.	Not used.
Age-standardised excess deaths	$\sum_i(w_i\ p_i)$ where w_i is the fraction of the reference population in age group i, and p_i is the age-specific *excess death rate* for age group i.	e.g. ONS (2021a) regularly updated article.	Good comparability though still affected by socioeconomic differences between countries or regions.	ONS (2020b); Morgan et al. (2020)
Per capita excess deaths	$\frac{X - \bar{X}}{population}$	e.g. Kontis et al (2020), Woolf et al. (2020).	Reasonable comparability but sensitive to the age distribution, as well as to socioeconomic differences between countries or regions.	Used in this paper. Used in Chen et al. (2020)
The P-score	$\frac{X - \bar{X}}{\bar{X}}$	e.g. Our World in Data website; ONS (2021b) regularly updated article.	Good comparability, though still affected by socioeconomic differences between countries or regions	Used in this paper.

	P-score: (X minus the expected value of X for the population), divided by the expected value of X for the population.			
Variant P-score	$\frac{X - upper\ threshold\ \bar{X}}{upper\ threshold\ \bar{X}}$ *Upper threshold:* the upper 95% confidence interval for this expected value. Takes into account uncertainty created by the natural variability of X.	e.g. U.S. National Center of Health Statistics.	As above for the P-score.	Not used to the best of our knowledge.
The Z-score	$\frac{X - \bar{X}}{std\ deviation\ of\ the\ population}$ *Z-score:* (X minus the expected value of X for the population), divided by the standard deviation for the population of X around its expected value in normal times. **	EuroMOMO, webpage: "Methods".	Not comparable where the standard deviations differ	Not used.

Notes: ** Assumes a Poisson distribution, adjusted for excess dispersion to approximate the underlying probability distribution of weekly deaths. The Poisson is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate and independently of the time since the last event. The calculation is described in Farrington et al. (1996).

Table 2: Data definitions and sources

Variable	Definition	Mean	Std. deviation	Maximum	Minimum	Data source
52 weeks: week 9, 2020 to week 8, 2021						
DEPENDENT VARIABLES						
Cumulated excess mortality per 100,000	Weekly excess deaths summed from week 9, 2020 to week 8, 2021, divided by state population; in logs; negative weekly values set to zero.	5.13	0.448	3.43	5.72	Calculated by the authors using data from CDC's National Center for Health Statistics (NCHS), November 2021 vintage data for observed deaths; January and February 2021 vintage data for normal deaths, see Section 4.1. State population for 2019 from the US Census Bureau.
Cumulated P-scores	Weekly excess deaths summed from week 9, 2020 to week 8, 2021, divided by corresponding sum of normal deaths; in logs; negative weekly values set to zero.	-1.68	0.420	-3.31	-1.05	
Cumulated COVID-19 Deaths per 100,000	Cumulated COVID-19 death count to end of week 8, 2021, divided by state population; in logs.	4.92	0.494	3.39	5.50	CDC, November 2021 vintage.
Cumulated COVID-19 Deaths per 100,000	Cumulated COVID-19 death count to end of week 8, 2021, divided by state population; in logs.	4.85	0.512	3.41	5.55	Coronavirus Resource Center, Johns Hopkins University, 1 March 2021 vintage.
INDEPENDENT VARIABLES						
2020 data						
Learning function or Timing of pandemic onset	Inverse of the number of days elapsed between the end of February and the day a given case-count threshold was breached, the threshold being the day the 14-day average of new infections exceeded 6 per 100,000. Scaled by mean of inverse days.	1	0.561	0.228	2.48	Constructed by authors, see Section 4.2.
Spring temperature	Temperature in °F, State average for main cities. Spring is defined as March to May.	54.1	9.46	28.0	75.9	Constructed by authors, see Section 4.3, using the National Oceanic and Atmospheric Association (NOAA), National Climate Report: Spring report and months to the end of 2020.
Pre-pandemic data						
Political vote share	The Democrat share of the popular vote in the 2016 Presidential General Election.	0.447	0.122	0.219	0.909	Federal Election Commission (2017)
Democratic Governor	Political affiliation of Governor.	0.490	0.505	0	1	Kaiser Family Foundation.
African American	Proportion of the population who are Black or African American	0.128	0.108	0.0116	0.474	United States Census, American Community Survey (ACS), ACS Demographic And Housing Estimates.
Hispanic	Proportion of the population who are Hispanic or Latino.	0.112	0.105	0.004	0.487	
Asian	Proportion of the population reporting as Asian	0.0149	0.0289	0.001	0.151	
Proportion aged 65+	Proportion of the population aged 65 years and over.	0.171	0.0202	0.115	0.215	

Remoteness dummy	Dummy=1 for Alaska, Hawaii, Maine and Washington State.	0.0784	0.272	0	1	Constructed by authors, see Section 4.4.
Population density	Defined as the 2019 state population divided by the area of the state in square km, in logs.	3.60	1.49	-0.857	8.29	US Census Bureau.
Urban density	The per square km density of urban areas, in logs.	6.71	0.361	6.17	8.24	US Census Bureau.
Poverty rate	The proportion of households below the poverty line	0.122	0.0263	0.075	0.196	Kaiser Family Foundation.
Interactions with learning function						
Interaction with Democrat vote share	Interaction effect between the 'Timing of onset', taken as the deviation from the mean, and the Democrat vote share.	0.0319	0.286	-0.399	0.870	As above.
Interaction with log median household income	Interaction effect between the 'Timing of onset', taken as the deviation from the mean, and log median household income.	0.0291	0.0937	-0.204	0.303	As above.
Additional variables for first stage equation for learning function						
Median household income	Median annual household income; in logs.	0	0.168	-0.344	0.356	Kaiser Family Foundation.
Nearness to New York	Dummy=1 for contiguous states; =0 for the non-contiguous states. Weighted by log ratio of state population to NY state population.	0.145	0.557	0	3.44	Constructed by authors, see Section 4.4.
Metropolitan Area population density	For each state, the density of large Metropolitan Areas occupied in each state, weighted by the 2010 share of MSA population in the state, and scaled by 1000.	0.119	0.178	0	0.826	Constructed by authors, see details in Section 4.4, using the US Census Bureau data.
Index of urbanisation.	2010 fraction of the state population living in urban areas; in logs.	4.28	0.220	3.66	4.61	US Census Bureau.

Notes: Several other variables were tried in general initial sets, adopting a general-to-specific approach as a diagnostic tool, see Section 4.

Table 3: The equation for the timing of the pandemic onset across US states

Dependent variable (over 52 weeks): Timing of pandemic onset	*Coefficient*
Constant	7.8*
Proportion African American	2.06***
Spring temperature	0.020***
MSA density	1.40***
Log fraction of urban population	0.77**
New York contiguity dummy	0.50***
Log median income	-0.85*
Equation standard error	0.346
Adjusted R-squared	0.62

Notes: Stars indicate significance levels: *** p-value lower than 0.01, ** p-value between 0.01 and 0.05, * p-value between 0.05 and 0.1. All variables are defined in Table 2. MSA stands for Metropolitan Statistical Area, and for the density measure, see Section 4.4.

Table 4: Comparing models with interaction effects for different measures of mortality

Dependent variables (cumulated, 52 weeks)	*log per capita excess mortality*			*log P-score*		*log per capita COVID-19 deaths*	
Variables	*eq. 1*	*eq. 2 2SLS*	*eq. 3 OLS*	*eq. 4*	*eq. 5*	*eq. 6*	*eq. 7*
Constant	5.16***	2.53	2.72	-1.81***	-3.32***	5.02***	2.00
Timing of pandemic onset		0.228***	0.178***		0.285***		0.075
Spring temperature (°F)		-0.0202***	-0.0206***		-0.0150***		-0.0338***
Proportion voting Democrat	0.170	-2.08***	-1.92***	0.489	-1.68***	-0.0002	-3.61***
Democrat Governor		-0.108*	-0.101*		-0.093		-0.045
Remoteness	-1.24***	-0.595***	-0.630***	-1.15***	-0.561***	-1.17***	-0.497***
Log of population density		0.106**	0.101**		0.0527		0.255***
Log of urban density		0.310***	0.293***		0.280***		0.527***
Proportion African American population		2.11***	2.13***		1.99***		2.56***
Proportion Hispanic population		1.46***	1.41***		1.82***		1.79***
Proportion Asian		2.00*	1.96*		1.96		3.73**
Proportion of population aged 65+		4.39**	4.05**		-		5.32*
Poverty		7.01***	7.01***		3.90***		5.67**
Interaction: Proportion voting Democrat × Timing of pandemic onset		-1.79***	-1.87***		-1.34**		-4.34***
Interaction: log median household income × Timing of pandemic onset		0.98**	1.16***		0.76		2.10***
Equation standard error	0.305	0.143	0.140	0.288	0.152	0.388	0.232
Adjusted R-squared	0.538	0.898	0.902	0.531	0.869	0.385	0.780

Notes: Stars indicate significance levels: *** p-value lower than 0.01, ** p-value between 0.01 and 0.05, * p-value between 0.05 and 0.1. In the interaction effects, variables are expressed as a deviation from their means. All variables are defined in Table 2.

Table 5: Robustness tests for sub-samples of equation fit

Dependent variable (cumulated over 52 weeks: log per capita excess mortality	*Full sample of 51 states*	*Variations in sample: omit 10 states*				
		first 10 states	*second 10*	*third 10*	*fourth 10*	*final 10*
Proportion voting Democrat	-2.08***	-3.00***	-2.17***	-1.90***	-1.02	-2.34***
Interaction: Proportion voting Democrat × Timing of pandemic onset	-1.79***	-2.30**	-2.01**	-2.30***	-1.49**	-1.85
Democratic Governor	-0.108*	-0.084	-0.121	-0.174***	-0.045	-0.091
Proportion African American population	2.11***	2.22***	1.94***	2.40***	1.70***	2.16***
Proportion Hispanic population	1.46***	1.73**	1.26**	1.59***	1.48***	1.48**
Equation standard error	0.143	0.169	0.162	0.135	0.122	0.152
Adjusted R-squared	0.898	0.864	0.817	0.920	0.932	0.887

Notes: Only selected coefficients are shown (see Table 4 for the full set of variables included in the regressions). Stars indicate significance levels: *** p-value lower than 0.01, ** p-value between 0.01 and 0.05, * p-value between 0.05 and 0.1. In the interaction effects, variables are expressed as a deviation from their means. All variables are defined in Table 2.