Submitted to the Annals of Applied Statistics

ESTIMATING GLOBAL AND COUNTRY-SPECIFIC EXCESS MORTALITY DURING THE COVID-19 PANDEMIC

BY VICTORIA KNUTSON[1], SERGE ALESHIN-GUENDEL[1], ARIEL KARLINSKY[2], WILLIAM MSEMBURI[3], JON WAKEFIELD[1,4]

[1]*Department of Biostatistics, University of Washington, Seattle, USA*

[2]*Hebrew University, Jerusalem, Israel*

[3]*World Health Organization, Geneva, Switzerland*

[4]*Department of Statistics, University of Washington, Seattle, USA*

Abstract: Estimating the true mortality burden of COVID-19 for every country in the world is a difficult, but crucial, public health endeavor. Attributing deaths, direct or indirect, to COVID-19 is problematic. A more attainable target is the "excess deaths", the number of deaths in a particular period, relative to that expected during "normal times", and we estimate this for all countries on a monthly time scale for 2020 and 2021. The excess mortality requires two numbers, the total deaths and the expected deaths, but the former is unavailable for many countries, and so modeling is required for these countries, and the expected deaths are based on historic data and we develop a model for producing expected estimates for all countries. We allow for uncertainty in the modeled expected numbers when calculating the excess. We describe the methods that were developed to produce World Health Organization (WHO) excess death estimates. To achieve both interpretability and transparency we developed a relatively simple overdispersed Poisson count framework, within which the various data types can be modeled. We use data from countries with national monthly data to build a predictive log-linear regression model with time-varying coefficients for countries without data. For a number of countries, subnational data only are available, and we construct a multinomial model for such data, based on the assumption that the fractions of deaths in specific sub-regions remain approximately constant over time. Our inferential approach is Bayesian, with the covariate predictive model being implemented in the fast and accurate `INLA` software. The subnational modeling was carried out using MCMC in `Stan` or in some non-standard data situations, using our own MCMC code. Based on our modeling, the 95% interval estimate for global excess mortality, over 2020–2021, is 13.3–16.6 million.

1. Introduction. The World Health Organization (WHO) has been tracking the impact of COVID-19 as the pandemic has evolved over time. Aggregate case and death numbers are being reported to the WHO and the data have been made publicly available at `https://covid19.who.int/`. For a number of reasons, these reported data neither provide a complete picture of the health burden attributable to COVID-19, nor of how many lives have been lost, directly and indirectly, due to the pandemic. Some deaths that are attributable to COVID-19 have not been certified as such because tests had not been conducted prior to death. Deaths may also be mistakenly certified as COVID-19, though this is less likely. It does not affect our estimates of excess mortality, based on all-cause mortality (ACM) data, however, only causing the resultant ratio of excess mortality to reported COVID-19 deaths to be lower than if such mistaken certification did not occur. There have also been variations in the death certification rules countries have applied in

Keywords and phrases: Bayesian inference, Global health, Poisson framework, Subnational modeling.

regards to COVID-19 (Garcia *et al.*, 2021; Riffe and Acosta, 2021). The impact of the pandemic is far reaching. Beyond the deaths directly attributable to it are those that can be linked to the conditions that have prevailed since the pandemic began and have led to some health systems being overwhelmed or some patients avoiding healthcare. In countries where COVID-19 spread was limited, due to lockdown measures or otherwise, some potential causes of death have decreased, such as those attributable to air pollution, or traffic accidents, or from other communicable diseases such as influenza like illness, resulting in negative excess or deficit deaths (Kung *et al.*, 2020; Karlinsky and Kobak, 2021). In light of the challenges posed by using reported COVID-19 data, excess mortality is considered a more objective and comparable (across countries) measure of the mortality impact of COVID-19 (Leon *et al.*, 2020). The WHO defines excess mortality as, "the mortality above what would be expected based on the non-crisis mortality rate in the population of interest" (`https://www.who.int/hac/about/definitions/en/`). Knowledge of the excess deaths not only paints a clearer picture of the pandemic, but can also aid in implementing public health initiatives.

The excess mortality in country c, ACM counts in month t for 2020 and 2021 are denoted by $Y_{c,t}$. These counts, in addition to the contribution from expected deaths, are assumed to be a result of the direct effects of COVID-19 (i.e., deaths attributable to it) and the indirect knock-on effects on health systems and society, along with deaths that were averted. The choice of a monthly time scale gives sufficient temporal resolution for most public health purposes. The hypothetical or "counterfactual" no-COVID-19 scenario uses the expected death numbers $E_{c,t}$, which have been forecasted to month t, using historic (prior to the pandemic) deaths data, usually over 2015–2019. Excess deaths are defined as:

$$\delta_{c,t} = Y_{c,t} - E_{c,t} \tag{1}$$

for country c where $c = 1, \ldots, 194$, and in month t where $t = 1, \ldots, 24$, represent months in 2020 and 2021.

The exercise of determining excess deaths for all countries is non-trivial, because the required ACM counts $Y_{c,t}$ are currently unavailable for many country/month combinations. Routine mortality data is often received by the WHO a year or more after the year of death. In addition, differential reporting capacity and variable data quality across countries has resulted in many nations lacking the systems to provide good quality routine data even historically (Mikkelsen *et al.*, 2015; Adair and Lopez, 2018; GBD, 2020; UNSD, 2021; Karlinsky, 2021). Correspondingly, these countries lack the capacity required to monitor ACM during the unprecedented COVID-19 pandemic. Hence, a number of countries are unable to contribute to the centralized systematic mortality surveillance that would be needed to measure global, regional and country level excess mortality by the WHO.

In this paper we describe our ongoing methods development to produce the WHO excess mortality estimates. In Section 2 we discuss data sources, before describing models for estimation of the expected numbers in Section 3. Section 4 describes our national covariate model and in Section 5 we outline the models we used for countries with subnational monthly data, national annual data, or a combination. Section 6 provides main results, with more extensive summaries appearing in the Supplementary Materials. Two other sets of global estimates of excess deaths have been produced by The Economist and the Institute for Health Metrics and Evaluation (IHME) with the latter being described in Wang *et al.* (2022). We fully describe and critique these methods in Section 7. The paper concludes with a discussion in Section 8.

2. Data Sources.

2.1. *Mortality Data.* Excess mortality cannot be directly measured for all countries due to many not having the required ACM data. The WHO usually receives routine mortality data on an annual basis in the year after the year of death or perhaps after an even greater lag. Civil registration and vital statistics (CRVS) systems differ greatly across countries with varying timelines and quality control measures for compiling unit record cause-of-death numbers into aggregates identified by cause, age, sex, place, and period of death. In addition, differential reporting coverage, the absence of electronic surveillance systems in some locations and limited investments in CRVS systems has resulted in many nations lacking the structures necessary to provide good quality routine data, even before the COVID-19 pandemic. This lack of capacity and the data required to monitor ACM has been exacerbated during the unprecedented pandemic. Therefore, many countries are unable to contribute to a centralized systematic mortality surveillance that would be needed to measure global, regional and country level excess mortality by the WHO.

Region	Full National	Partial National	Mixed Data	No Data	Total	Proportion Population
AFRO	4	2	0	41	47	0.13
AMRO	11	12	4	9	36	0.90
EMRO	4	5	0	12	21	0.32
EURO	46	5	1	1	53	0.89
SEARO	1	1	3	6	11	0.04
WPRO	6	3	2	16	27	0.18
Global	72	27	10	85	194	0.33

TABLE 1

Country data availability summary for 2020 and 2021. Full national countries have data over all 24 months and partial national have data for less than 24 months; for example, 83 countries have data for at least the first 18 months, and 96 countries have data for at least the first 12 months. Mixed data refers to countries with subnational monthly data for some period (4 countries), national annual data (5 countries) or a combination (China). WHO regions: African Region (AFRO), Region of the Americas (AMRO), Eastern Mediterranean Region (EMRO), European Region (EURO), South-East Asian Region (SEARO), Western Pacific Region (WPRO). The proportion of the population that are observed column is calculated at the country-month level. The Supplementary Materials include a table that lists the type of data available for each country.

All countries report their official COVID-19 death count, but we would not expect this to be accurate, and for many countries we would expect serious underestimation, for the reasons already outlined and for political reasons. However, the official count does provide an interesting summary for comparison with the estimated excess, and the COVID-19 death rate is used as a covariate in our ACM estimation model.

For this study, our main sources of data are reports of ACM as collected and reported by countries' relevant institutions – from national statistics offices, ministries of health, population registries, etc. These have been collected in several repositories such as the data routinely shared with WHO as part of its standing agreement with member states, Eurostat, The Human Mortality Database (HMD) as part of the Short-Term Mortality Fluctuations (STMF) project (Németh *et al.*, 2021) and the World Mortality Dataset (WMD), as described in Karlinsky and Kobak (2021). Monthly data are included after accounting for delayed registration either by adjusting for registration delay (Australia, Brazil, United States) or by not-including highly incomplete months.

In this paper we report the current state of data at our disposal. This project is ongoing and data is added as soon as available. Table 1 shows the breakdown of data availability by WHO region. Just over a half (99) of the 194 countries provide monthly national data

from at least some of the pandemic period, while 10 other countries provide subnational monthly data, national annual data, or a combination of the two (this includes Argentina which has partial national and subnational data, so could fit in with partial or mixed data). It is immediately clear that there is a huge regional imbalance in data availability, with the EURO region being very well represented, the AMRO region having data from 75% of the countries, and other regions being more poorly represented. For example, in the AFRO region we only have data from 6 out of 47 countries. For those countries with data in month t, we assume that the ACM part of the excess $\delta_{c,t}$, as defined in (1), is known exactly. Hence, we do not account for inaccuracies in the reported deaths (beyond the aforementioned accounting for delayed registration). For all countries we do, however, account for uncertainty in the expected numbers.

2.2. *Covariate Data.* For countries with no data, we predict the ACM count using a log-linear covariate model. A range of covariates were considered, including a high income country binary indicator, COVID-19 test positivity rate, COVID-19 death rate, temperature, population density, socio-demographic index (SDI), human development index (HDI), stringency (index for lockdown restrictions and closures, overall government response, economic (including measures such as income support and debt relief), containment (this index combines "lockdown" restrictions and closures with measures such as testing policy and contact tracing, short term investment in healthcare, as well investments in vaccines – it is calculated using all ordinal containment and closure policy indicators and health system policy indicators, for further details see Hale *et al.* (2020)), historic (from 2019) non-communicable disease rates, historic cardiovascular disease rate, historic HIV rate, historic diabetes prevalence, life expectancy, proportion of the population under-15, proportion of the population over-65. Some of the covariates are time-varying (COVID-19 test positivity rate, COVID-19 death rate, temperature, stringency, overall government response, containment), while the remainder are constant over time. A number of the covariates were not available by month for all countries and so their values were imputed. Specifically, (WHO) regional medians were used for countries with missing data. Details are given in the Supplementary Materials.

3. Expected Mortality Modeling. A key component of the excess mortality calculation is the ACM count that would be expected in non-pandemic times, for each country and month. We describe models for two types of countries: those that have historic monthly ACM data, and those that have historic annual ACM data only – 100 countries have historic monthly data and 94 have historic annual data. In terms of the period upon which we base the expected numbers, it is usually 2015–2019 for countries with monthly historical data, and is usually 2000–2019 for countries with annual historical data.

3.1. *Countries with Monthly Data.* We consider first those countries with monthly ACM data over multiple years (usually 2015–2019). For country c, $Y_{c,t}$ represents the ACM count for country c and month t, for $t = 1, \ldots, M_c$, where M_c is the number of historic months for which we have data. We assume the sampling model for $Y_{c,t}$ is,

$$Y_{c,t}|\mu_{c,t} \sim \text{NegBin}(\mu^{\text{E}}_{c,t}, \phi^{\text{E}}_c),$$

parametrized in terms of the mean, $\mu^{\text{E}}_{c,t}$, and the overdispersion parameter, ϕ^{E}_c, such that $\text{var}(Y_{c,t}|\mu^{\text{E}}_{c,t}, \phi^{\text{E}}_c) = \mu^{\text{E}}_{c,t}(1 + \mu^{\text{E}}_{c,t}/\phi^{\text{E}}_c)$, with the Poisson model being recovered as $\phi^{\text{E}}_c \to \infty$. We let $v[t]$ index the year in which month t occurred (for example, labeled $1, \ldots, 5$ when data are available for 2015–2019) and $m[t]$ be the month (labeled $1, \ldots, 12$), so that given v, m we can find t as $t = 12(v-1) + m$. The mean is modeled as,

$$\eta_{c,t} = \log(\mu_{c,t}) = f^{\text{y}}_c(v[t]) + f^{\text{m}}_c(m[t]) \tag{2}$$

where $f_c^{\mathrm{y}}(\cdot)$ models the *annual trend*, and $f_c^{\mathrm{m}}(\cdot)$ is a smooth function of time t which accounts for *within-year* seasonal variation. The yearly trend is modeled with a thin-plate spline and within-year variation with a cyclic cubic spline (Rivera *et al.*, 2020). In both cases we use the `gam` function in the `mgcv` package with generalized cross-validation (Wood, 2017, Section 4.5.3) used to select smoothing parameters. The spline model is fitted separately for each country. Algeria, Iraq and Sri Lanka have less than three years of historical data, and so a linear term is used for modeling yearly variation. This model is used to obtain predictions of the expected deaths $\mu_{c,t}^{\mathrm{E}}$ for all t in 2020 and 2021, with both a point estimate and a standard error being produced.

3.2. *Countries with Annual Data.* For countries with only annual historic data, the goal is to predict expected numbers by month t for $t = 1, \ldots, 24$. We summarize our strategy for producing expected numbers for countries with annual data only:

1. Fit a negative binomial spline model to the countries with annual counts only. Use the spline to predict the total annual ACM for 2020 and 2021, for these countries.
2. In a separate exercise, fit the multinomial model to all of the countries with monthly data, with deaths being attributed via the log-linear temperature model (3). This produces an estimate $\widehat{\beta}$.
3. Combine the spline model with the multinomial model using monthly temperature apportionment to obtain expected numbers for the countries without monthly data.

The annual trend can be estimated for each country using the method we described in the previous section minus the monthly term, i.e., a spline in year. To apportion the yearly totals to the months, we use the fact that a collection of Poisson random variables conditioned on their sum produce a multinomial distribution with within-year variation modeled using temperature, which is acting as a surrogate for seasonality. This relationship is learned from countries with historic monthly data. We use a smooth series of monthly temperatures since 2015. Let $\boldsymbol{Y}_{c,v} = \{Y_{c,v,m}, m = 1, \ldots, 12\}$ be the vector that contains the ACM counts by month in year v, $v = 1, \ldots, 5$. Suppose each of the 12 constituent counts are Poisson with mean $\zeta_{c,v,m}$, for $m = 1, \ldots, 12$. Then, within the year, conditional on the total ACM,

$$\boldsymbol{Y}_{c,v} | Y_{c,v}^{+}, \boldsymbol{p}_{c,v} \sim \text{Multinomial}(Y_{c,v}^{+}, \boldsymbol{p}_{c,v}),$$

where $Y_{c,v}^{+}$ is the national count in country c and year v and $\boldsymbol{p}_{c,v} = \{p_{c,v,m}, m = 1, \ldots, 12\}$ with

$$p_{c,v,m} = \frac{\zeta_{c,v,m}}{\sum_{m'=1}^{12} \zeta_{c,v,m'}},$$

We assume,

$$\log(\zeta_{c,v,m}) = z_{c,v,m}\beta \tag{3}$$

where $z_{c,v,m}$ is the temperature and β is the associated log-linear coefficient. The multinomial model can be fitted in `INLA` using the Poisson trick (Baker, 1994) which involves fitting the Poisson model for the data in country c, month m:

$$Y_{c,v,m} | \lambda_{c,v} \sim \text{Poisson}(\ \lambda_{c,v} \mathrm{e}^{z_{c,v,m}\beta}\ ),$$

where the $\lambda_{c,v}$ parameters are given (improper) priors $\pi(\lambda_{c,v}) \propto 1/\lambda_{c,v}$. Further details may be found in the Supplementary Materials.

The estimated expected counts are shown in blue in Figure 1, for selected countries.

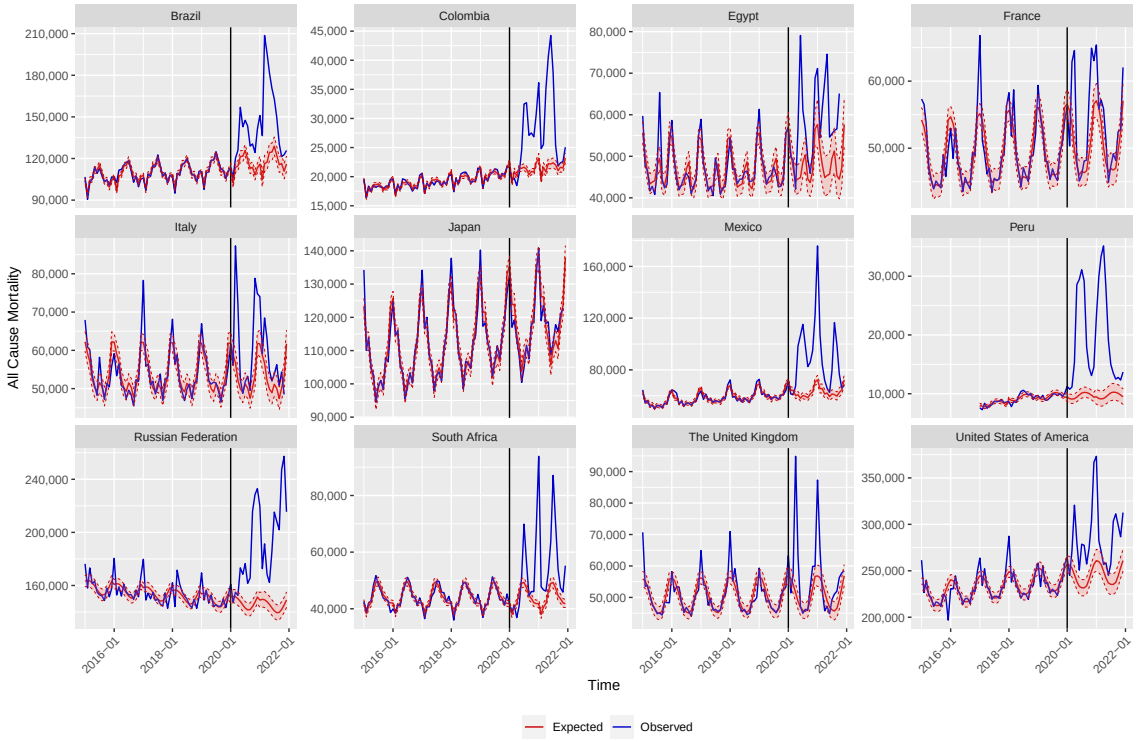

FIG 1. *Monthly time series of all cause mortality: expected counts in red and observed counts in blue, for selected countries. The black vertical line is drawn at the start of 2020. The dashed red bands denote 95% uncertainty intervals for the mean expected numbers. For these countries, ACM counts are available for all months apart from Egypt, for which the last month is missing.*

3.3. *Modeling Uncertainty in the Expected Numbers.* For all countries the expected numbers appear directly in the excess calculation, (1). In addition, for countries with no pandemic ACM data, the Poisson model we adopt for covariate modeling includes the expected number as an offset. For all countries and months, we obtain not just an estimate of the mean expected mortality but also a measure of the uncertainty (due to uncertainty in estimating the spline model) in this estimate. We now describe how the uncertainty in the mean expected count is accounted for in our modeling.

For countries with monthly data, we use the spline model to predict the log of the mean expected number of deaths. Asymptotically, the estimator for the log of the mean expected numbers is normally distributed. Let $\widehat{\eta}_{c,t'}$ and $\widehat{\sigma}^2_{c,t'}$ represent the mean and standard deviation of the prediction for pandemic months, labeled as $t' = 1, \ldots, 24$. We simulate S samples from the asymptotic normal sampling distribution with mean $\widehat{\eta}_{c,t'}$ and standard deviation $\widehat{\sigma}_{c,t'}$; denote these samples by $\eta^{(s)}_{c,t'}$, $s = 1, \ldots, S$. We then transform the samples so that we have samples for the expected numbers $E^{(s)}_{c,t'} = \exp(\eta^{(s)}_{c,t'})$, for $s = 1, \ldots, S$. We then use the method of moments to fit a gamma distribution to these S samples with shape $\tau_{c,t'}$ and rate $\tau_{c,t'}/E_{c,t'}$. In particular, letting $m_{c,t'}$ denote the sample mean, and $V_{c,t'}$ denote the sample variance, we set $\widehat{E}_{c,t'} = m_{c,t'}$ and $\widehat{\tau}_{c,t'} = m^2_{c,t'}/V_{c,t'}$. We approximate the distribution of the expected numbers as gamma, since this is conjugate to the Poisson, and so allows efficient inference with INLA (Rue *et al.*, 2009) using a negative binomial, as we describe in Section 4. Effectively, we are approximating the sampling distribution of the mean expected count by a gamma distribution.

We now consider a generic country c with yearly data only. In pandemic year v', we use the spline model to predict the log of the expected number of deaths. Let $\widehat{\eta}_{c,v'}$ and $\widehat{\sigma}^2_{c,v'}$ represent the mean and standard deviation of the prediction, for $v' = 1, 2$ (the two pandemic years). We then simulate S samples from a normal distribution with mean $\widehat{\eta}_{c,v'}$ and standard deviation $\widehat{\sigma}_{c,v'}$; denote these samples by $\eta^{(s)}_{c,v'}$, $s = 1, \ldots, S$. We then transform the samples so that we have samples for the expected numbers $E^{(s)}_{c,v'} = \exp(\eta^{(s)}_{c,v'})$, for $s = 1, \ldots, S$. We then apply the monthly temperature model to produce predictions of the proportion of deaths in each month in each year, i.e., for a given pandemic month m', we have S samples of the predicted proportion of deaths in month m' of year v', $p^{(s)}_{c,v',m'}$, for $s = 1, \ldots, S$. Converting to pandemic cumulative months $t' = 12(v' - 1) + m'$ we then produce samples of the expected number of deaths in month t', as $E^{(s)}_{c,t'} = E^{(s)}_{c,v'} \times p^{(s)}_{c,v',m'}$. We then use the method of moments to fit a gamma distribution to these S samples as for the countries with monthly data. To summarize, in both cases we have a distribution for $E_{c,t'}$ which is Gamma$(\widehat{\tau}_{c,t'}, \widehat{\tau}_{c,t'}/\widehat{E}_{c,t'})$. The Supplementary Materials provide comparisons of the true distribution of the mean expected counts and the approximating gamma distributions, and illustrates that the latter are accurate. We also experimented with including negative binomial sampling variability in the calculation of the expected numbers, but it made little additional contribution to the intervals for the excess.

In the next section we describe a Bayesian modeling of ACM in the pandemic, for countries without data. Inference for the expected numbers is frequentist, and we sample from the asymptotic normal distribution, but with flat priors, this will approximate a Bayesian analysis, and so when we combine the two components in the excess (1) we view the resultant inference as Bayesian.

We next describe how we model ACM – we have different models for different data scenarios but in each case the starting point is the Poisson distribution.

4. National Mortality Models for Countries with No Data. For countries with observed monthly national ACM data, $Y_{c,t}$, we use these directly in the excess calculation. In the countries with no data we need to estimate the ACM count. We follow a Bayesian approach so that for countries without data we obtain a predictive distribution over this count and this, when combined with the gamma distribution for the expected numbers, gives a distribution for the excess $\delta_{c,t}$.

In Figure 2 we plot the monthly counts for a range of countries with monthly ACM data, along with the reported COVID-19 deaths and the expected numbers. We see very different scenarios in different countries. In all countries but Japan there is a clear large difference between the observed and the expected, though within each country this difference shows large fluctuations over time. In Figure 3, again for countries with monthly ACM data, we plot the excess $\delta_{c,t} = Y_{c,t} - E_{c,t}$, as a function of month t (including uncertainty in the expected numbers), along with the reported COVID-19 deaths. As expected, $\delta_{c,t}$ is greater than the reported overall in general, but not for Japan, and for most countries displayed the difference between the excess and the reported shows a complex temporal pattern.

While complex models that attempt to pick up data nuances are desirable, given the idiosyncrasies of the different data sources described in Section 2, any modeling exercise is fraught with difficulties, and we resort to a relatively simple model in which we build an overdispersed Poisson log-linear regression model for the available monthly ACM data to predict the monthly ACM in those countries with no data. We cannot overemphasize the regional imbalance of the missing ACM data – in the AFRO region in particular, our estimates should be viewed with extreme caution, since they are predicted from data which overwhelmingly is from other regions.

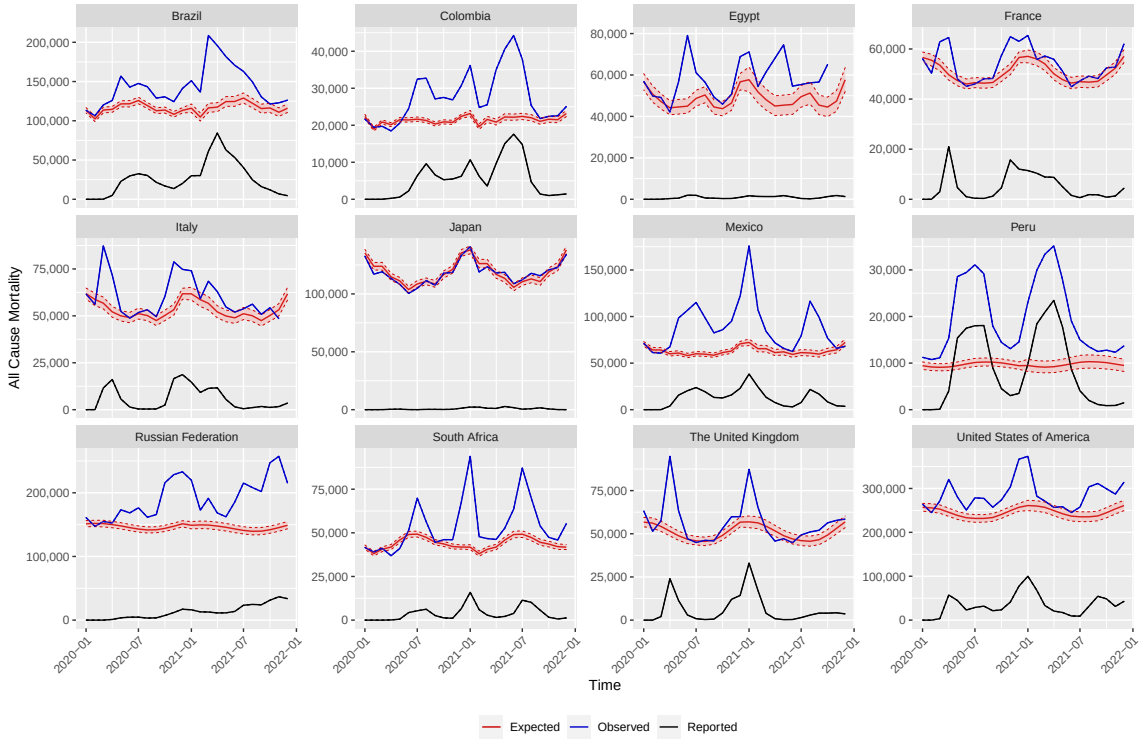

FIG 2. *Monthly time series of ACM counts, expected counts (with 95% interval estimates) and reported COVID-19 mortality counts, for selected countries. ACM counts are available for all months apart from Egypt, for which the last month is missing.*

The basic starting model is

$$Y_{c,t}|E_{c,t},\theta_{c,t} \sim \text{Poisson}(E_{c,t}\theta_{c,t}), \tag{4}$$

so that $\theta_{c,t} > 0$ is a relative rate parameter, with $\theta_{c,t} > 1$ / $\theta_{c,t} < 1$ corresponding to a higher/lower ACM rate than expected, based on historic data. Recall, from Section 3, that we model the distribution of the expected counts $E_{c,t}$ as Gamma$(\widehat{\tau}_{c,t}, \widehat{\tau}_{c,t}/\widehat{E}_{c,t})$. When combined with (4), we obtain the sampling model,

$$Y_{c,t}|\theta_{c,t} \sim \text{NegBin}(\widehat{E}_{c,t}\theta_{c,t}, \widehat{\tau}_{c,t})$$

with known overdispersion parameter $\widehat{\tau}_{c,t}$ to give $\text{var}(Y_{c,t}|\theta_{c,t}) = \widehat{E}_{c,t}\theta_{c,t}(1 + \widehat{E}_{c,t}\theta_{c,t}/\widehat{\tau}_{c,t})$. The mean is $\text{E}[Y_{c,t}|\theta_{c,t}] = \widehat{E}_{c,t}\theta_{c,t}$. The relative rate parameter $\theta_{c,t}$ is modeled as,

$$\log\theta_{c,t} = \alpha + \sum_{b=1}^{B}\beta_{bt}X_{bct} + \sum_{g=1}^{G}\gamma_g Z_{gc} + \epsilon_{c,t}. \tag{5}$$

The model details are:

- The intercept is α and the time-invariant covariates (e.g., SDI, historic diabetes rate) have fixed association parameters γ_g.
- We have B time-varying covariates (e.g., sqrt(C19 death rate), test positivity rate, containment), and we allow the associations for these variables, β_{bt}, to be time-varying via a random walk of order 2 (RW2) prior (Rue and Held, 2005) which has variance σ^2_β. These parameters include a sum-to-zero constraint, since we include a fixed effect for the overall association (across months) – these are included in the G time-invariant part of the model.

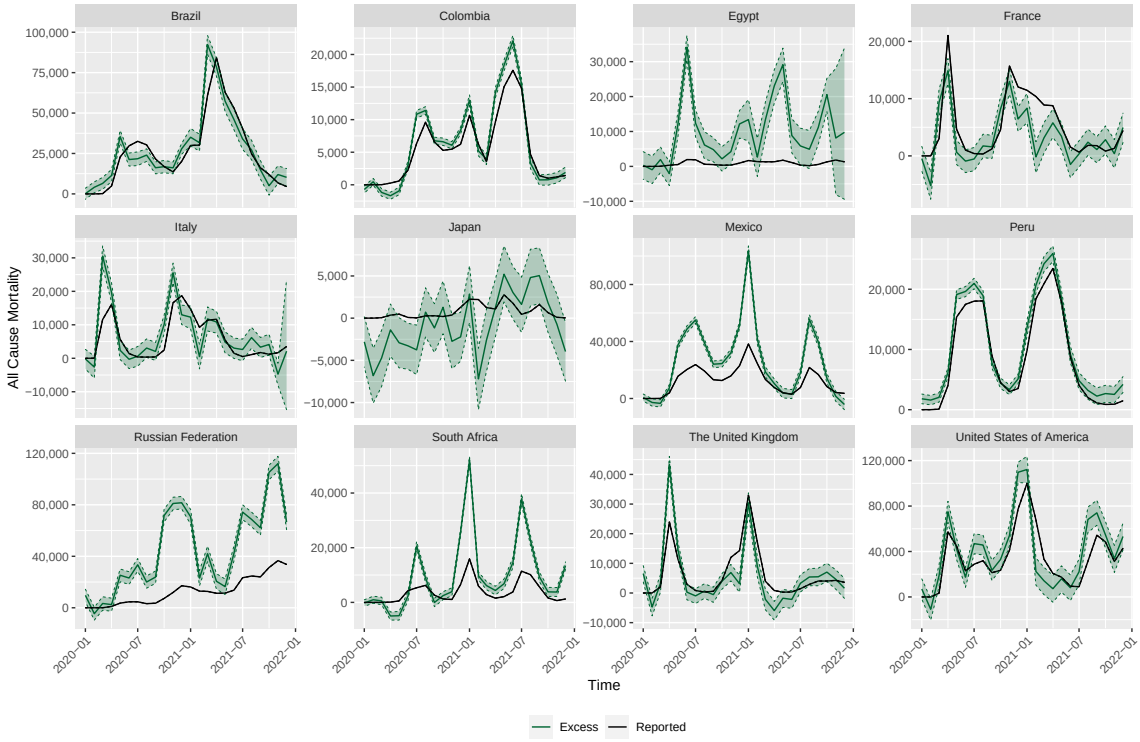

FIG 3. *Monthly time series of excess mortality, along with reported COVID-19 mortality counts. ACM counts are available for all months apart from Egypt, for which the last month is missing. For this month, the covariate prediction model is used for the point and interval estimates.*

- There are two sources of excess-Poisson variation in our model. The negative binomial component, with known $\widehat{\tau}_{c,t}$, arises because of the uncertainty in the expected numbers, while the $\epsilon_{c,t} \sim \mathrm{N}(0, \sigma_\epsilon^2)$ adjustments allow for overdispersion, given a fixed value of the expected numbers.
- The Bayesian model is completed by prior specifications on the regression coefficients of the loglinear model and any hyperparameters. We use default priors (normal with large variance) on the intercept and fixed association parameters, and penalized complexity (PC) priors on the RW2 standard deviations and on σ_ϵ (Simpson *et al.*, 2017). Specifically, letting σ_β denote a generic RW2 standard deviation parameter, the PC priors are such that $\Pr(\sigma_\beta > 1) = 0.01$, and the PC prior on the overdispersion parameter σ_ϵ has $\Pr(\sigma_\epsilon > 1) = 0.01$.

Each country will clearly have its own specific temporally correlated baseline, as a result of unobserved covariates and model misspecification, but we did not include terms to model such a baseline (using a RW2 or a spline, for example), since fits from this model are not being used to estimate the excess for countries with data. Rather, we are using this model to predict the ACM for countries with no data. Hence, we did not use RW2 intercepts as these would dilute the covariate effects, due to confounding by time (Kelsall *et al.*, 1999), and it is these covariate effects that are key to prediction for countries with no data. If we had included a RW2 baseline, then a country-specific RW2 model would give estimated contributions of zero in countries with no data and so would not provide any benefit. This is but one of the model assumptions that are forced upon us by the limited data we have available. The country-level model was fitted using the INLA method (Rue *et al.*, 2009) and accompanying `R` implementation.

For countries with no ACM data, we obtain a predictive distribution by averaging the negative binomial model with respect to the posterior via,

$$\Pr(Y_{c,t}|\boldsymbol{y}) = \int \underbrace{\Pr(Y_{c,t}|\theta_{c,t})}_{\text{Negative Binomial}} \times \underbrace{p(\theta_{c,t}|\boldsymbol{y})}_{\text{Posterior}}\, d\theta_{c,t}.$$

We use INLA to fit the covariate model, and then use the posterior sampling feature to produce samples for the components of (5), which in turn produces samples $\theta_{c,t}^{(s)} \sim p(\theta_{c,t}|\boldsymbol{y})$ from the posterior. and we then then simulate $Y_{c,t}^{(s)}|\theta_{c,t}^{(s)}$ from the negative binomial, for $s = 1, \ldots, S$.

Partial monthly data is available for 27 countries, and for these we require a switch from observed data to the covariate modeled ACM. The naive application of the covariate model will lead to the possibility of unrealistic jumps (up or down) when we switch from the observed data to the covariate model, and to alleviate this problem we benchmark the predictions to the last observed data point. We let $T_c^{(1)}$ represent the number of observed months of data and $T_c^{(2)}$ be the number of months for which there is no data, for country c. For a country with partial data, let $\boldsymbol{y}_c^{(1)} = [y_{c,1}, \ldots, y_{c,T_c^{(1)}}]$ represent the observed partial data. We then wish to predict the ACM counts $\boldsymbol{y}_c^{(2)} = [y_{c,T_c^{(1)}+1}, \ldots, y_{c,T_c^{(1)}+T_c^{(2)}}]$ for the missing period. The model for the missing data period is,

$$y_{c,t}^{(2)}|\boldsymbol{y}_c^{(2)}, \theta_{c,t}, f_c \sim \text{NegBin}(\widehat{E}_{c,t}\theta_{c,t}f_c, \widehat{\tau}_{c,t}), \tag{6}$$

for $t = T_c^{(1)} + 1, \ldots, T_c^{(1)} + T_c^{(2)}$, where $\theta_{c,t}$ is a function of the covariates in the missing data period (specifically given by (5)), and the benchmarking factor is,

$$f_c = f_c\left(\theta_{c,T_c^{(1)}}\right) = \frac{y_{c,T_c^{(1)}}}{\widehat{E}_{c,T_c^{(1)}}\theta_{c,T_c^{(1)}}},$$

where $\theta_{c,T_c^{(1)}}$ is given by equation (5). This factor matches the last observed death count to the covariate model projected back to the last observed count. This factor is applied subsequently to all of the missing data months. To implement the benchmark, samples from the posteriors for $\theta_{c,t}$ and f_c are used in (6), and then negative binomial counts are drawn.

5. Observed Mortality Mixed Data Modeling. For a small number of countries for which national ACM data are not available (Argentina, India, Indonesia and Turkey) we instead have ACM data from subregions, with the number of regions with data potentially changing over time. For other countries we obtain national annual ACM data only. In this section we describe the models we use in these situations. For the subnational scenario we construct a statistical model building on, and expanding, a method previously proposed by Karlinsky (2022) that is based on a proportionality assumption.

5.1. *Subnational Data Model.* For Turkey we have subnational monthly data over the complete two years of the pandemic, while for Indonesia we have monthly subnational data for 2020 and for the first six month of 2021. Argentina has observed data for 2020 and subnational monthly data for 2021. India has data from up to 17 states and union territories (from now on, states) over the pandemic period (out of 36), but this number varies by month.

We consider the most complex subnational scenario in which the number of regions with monthly data varies by month, using India as an example. For India, we use a variety of sources for registered number of deaths at the state and union-territory level. The information was either reported directly by the states through official reports and automatic vital

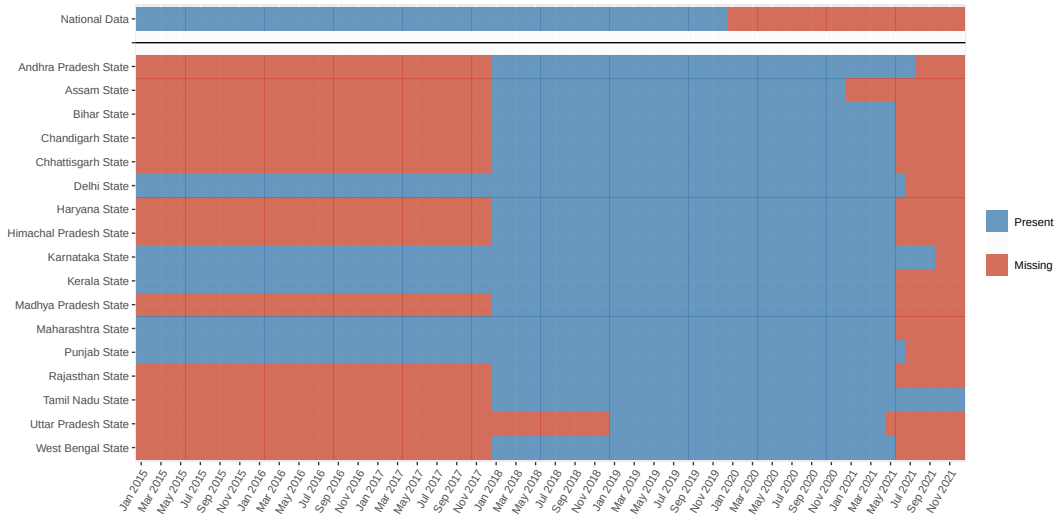

FIG 4. *Plot of missingness in subnational data for India across 2015–2021.*

registration, or by journalists who obtained death registration information through Right To Information requests (see the Supplementary Materials for full details). The available data we have for India is summarized in Figure 4. We assume in total that there are K regions that contribute data at any time. We develop the model for a generic country and hence drop the c subscript. For the historic data in month t we have total deaths counts along with counts over regions, $Y_{t,k}$, $k \in K_t$, so that in period t, $|K_t|$ is the number of regions that provide data with $k \in K_t$ being the indices of these areas from $1, \ldots, K$. We let region 0 denote all other regions, which are not observed in pandemic times, at time t and $S_t = \{0\} \cup K_t$. We assume, in month t:

$$Y_{t,k}|\lambda_{t,k} \sim \text{Poisson}(N_{t,k}\lambda_{t,k}), \qquad k \in S_t,$$

where $N_{t,k}$ is the population size, and $\lambda_{t,k}$ is the rate of mortality. Hence,

$$Y_{t,+}|\lambda_{t,k}, k \in S_t \sim \text{Poisson}\left(\sum_{k \in S_t} N_{t,k}\lambda_{t,k}\right).$$

If we condition on the total deaths, we obtain,

$$\boldsymbol{Y}_t|\boldsymbol{p}_t \sim \text{Multinomial}_{|S_t|}(Y_{t,+}, \boldsymbol{p}_t),$$

with $\boldsymbol{p}_t = \{p_{t,k}, k \in S_t\}$, with

$$p_{t,k} = \Pr(\text{ death in region } k \mid \text{ month } t, \text{ death }) = \frac{N_{t,k}\lambda_{t,k}}{N_{t,+}\lambda_{t,+}},$$

Our method hinges on this ratio being approximately constant over time. If, over all regions, there are significant changes in the proportions of deaths in the regions as compared to the national total, or changes in the populations within the regions over time, then the approach will be imprecise. For India, the fractions of the total ACM by state are shown in Figure 5. There are certainly deviations from constancy for some states, but in general the assumption appears tenable, at least in pre-pandemic periods. Of course, the great unknown is whether the assumption remains reasonable over the pandemic. To address this, we carry out extensive sensitivity and cross-validation analyses.

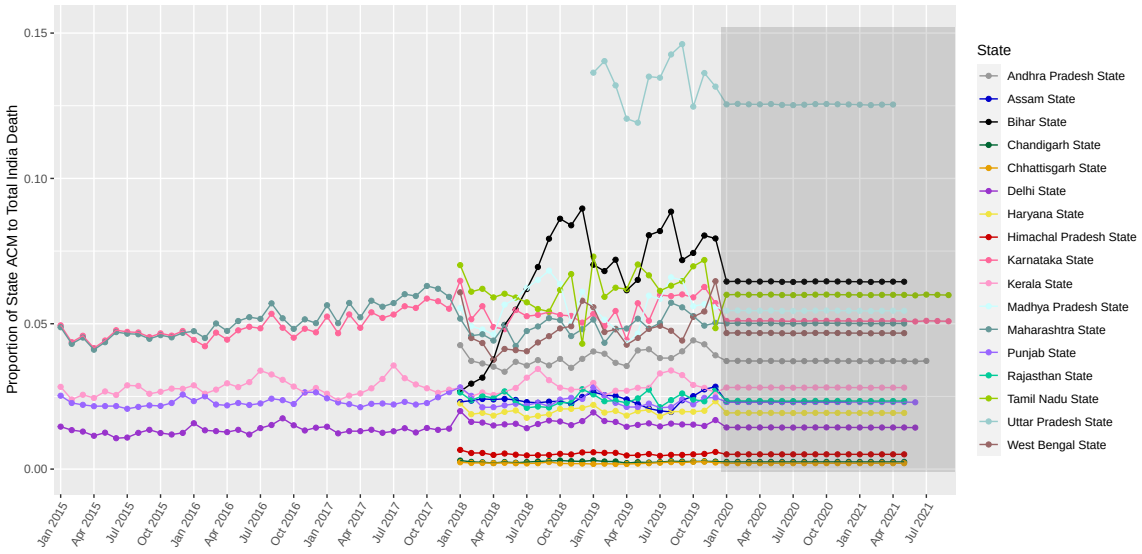

FIG 5. *Plot of estimated proportion of subnational deaths to national deaths in pre-pandemic and pandemic periods. The horizontal flat lines are the point estimates for the fraction for the respective states during the pandemic months.*

We model the monthly probabilities as,

$$\log\left(\frac{p_{t,k}}{p_{t,|K_t|+1}}\right) = \alpha_k + e_t, \qquad k \in S_t, \tag{7}$$

where the α_k parameters are unrestricted and $e_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$, and we examine the size and temporal structure of the error terms e_t, to assess the proportionality assumption, at least over the available pre-pandemic period.

To specify the model, we take a multinomial with a total number of categories that corresponds to all regions that appear in the data, K, and specify the likelihood over all months by exploiting the property that a multinomial collapsed over cells is also multinomial. Hence, in year t we have a multinomial with $|K_t| + 1$ categories with constituent probabilities constructed from the full set of $K + 1$ probabilities.

To derive the predictive distribution, we abuse notation and let $Y_{t,1}$ denote the total number of observed subnational deaths at time t, and $Y_{t,2}$ the total number of unobserved subnational deaths at time t, with $Y_{t,+} = Y_{t,1} + Y_{t,2}$ being the total (national) number of deaths at time t. Hence, at time t, $Y_{t,1}|p_t, Y_{t,+} \sim \text{Binomial}(Y_{t,+}, p_t)$, where $p_t = \sum_{k \in K_t} p_{t,k}$. In order to fit the multinomial model in a Bayesian framework and predict the total number of deaths in 2020–2021, we need to specify a prior for $Y_{t,2}$ or, equivalently, for $Y_{t,+}$, where t indexes months in this period. We will use the prior $p(Y_{t,+}) \propto 1/Y_{t,+}$, which is a common non-informative prior for a binomial sample size (Link, 2013), and has the desirable property that the posterior mean for $Y_{t,2}$, conditional on p_t, is $\mathrm{E}[Y_{t,2}|p_t] = Y_{t,1}(1-p_t)/p_t$, i.e., of the same form as the simple frequentist "obvious" estimator, which leads to the naive estimate of the ACM, $Y_{t,1} + \widehat{Y}_{t,2} = Y_{t,1}/p_t$.

To give more details for implementation we will use a general result. Suppose

$$Y_{t,1}|Y_{t,+}, p_t \sim \text{Binomial}(Y_{t,+}, p_t)$$
$$p(Y_{t,+}) \propto 1/Y_{t,+},$$

so that, in particular, the marginal distribution of Y_{+t} does not depend on p_t. Then the posterior for the missing ACM count, conditional on p_t, is

$$Y_{t,+}|Y_{t,1}, p_t \sim Y_{t,1} + \text{NegBin}(Y_{t,+}, 1 - p_t),$$

or, equivalently,

$$Y_{t,+} - Y_{t,1}|Y_{t,1}, p_t \sim \text{NegBin}(Y_{t,1}, 1 - p_t).$$

This links to one of the usual motivations for a negative binomial (number of trials until we observe a certain fixed number of events) — making inference for the number of total deaths it takes to produce $Y_{t,1}$ deaths in the sub-regions. We implement this model in `Stan`. In the Appendix we detail a simulation study that validates the method in the situation in which the missing data follow the assumed form.

For the other countries with subnational data, the number of subregions is constant over time, and so in the above formulation the multinomial is replaced by a binomial. Details for these countries are in the Supplementary Materials. For Indonesia we have subnational data from only Jakarta at the monthly level and historic national ACM at the annual level. Hence, we fit a binomial subnational model to the annual historic data, summing the monthly subnational historic data to the annual level, and then predict the monthly national ACM for 2020–2021 using the p_t fit on the historic annual data.

5.2. *Mixed Data Models and Special Cases.* We have annual national ACM counts for Viet Nam, Grenada, Sri Lanka, Saint Kitts and Nevis, and Saint Vincent and the Grenadines. For these countries we estimate the monthly counts using a multinomial model in which the loglinear covariate model (5) is used to apportion the total count to months.

6. Results. In this section we summarize the excess mortality results, further results are available in the Supplementary Materials, and a ShinyApp is available (`https://msemburi.shinyapps.io/excessvis/`) at which the results may be fully examined. The aim is to build a covariate prediction model for the countries with no ACM data, using (5). The covariate model choice exercise was carried out in an empirical fashion. In an ideal world, we would have had region-specific models, but the paucity of data in many of the regions (as summarized in Table 1) did not allow this. Instead, for all of the time-varying covariates (COVID-19 test positivity rate, COVID-19 death rate, temperature, stringency, overall government response, containment) we added an interaction with the binary country-level variable, low/middle or high income. We examined plots of the covariates by availability in the ACM observed/unobserved countries, and discarded a number of covariates (historic HIV rate, and over-65 and under-15 proportions of the population) that had little overlap over countries with/without ACM data (meaning, for example, that the countries with high HIV rates tended to be those without observed ACM data, making extrapolation hazardous). On a contextual basis we then formed a covariate model with time-varying covariates: containment, square root COVID-19 death rate (the square root transforms helps in preventing the association being driven by a few countries), temperature and COVID-19 positivity rate. The constant covariates are: historic diabetes rate and historic cardiovascular rate. We took this model as our starting point and added and removed variables to examine the sensitivity of the predictions. We evaluated the models using cross-validation and various metrics that are described in the Supplementary Materials. We found that the predictions were quite robust to covariate models and so only report the results for the model described above.

In the Supplementary Materials we describe our approaches to model assessment and model comparison. We assessed the frequentist coverage of our procedure using cross-validation. In particular, we performed two experiments: in one we left out all data from

a country, and in the other we left out all data from one month (systematically going through all countries and all months, respectively, in the two schemes). The model was fitted to the remaining data and was used to produce predictive intervals for the left out data, which can then be compared with the left out data. The empirical coverage at levels, 50%, 80%, 95%, was calculated by summarizing across all left out data. For the leave-one-country out analysis the coverages were 59.3%, 82.7%, 91.6%, and for the leave-one-month out analysis 57.8%, 83.7%, 92.9%. From these summaries, we would conclude that the model is reasonably well calibrated. Using the same cross-validation strategies we also evaluated the relative and absolute bias of the ACM rate. The relative biases from the country and monthly leave out strategies were 1.98% and 1.84%, respectively. The absolute biases from the country and monthly leave out strategies were 10.08% and 10.18%, respectively. The absolute relative bias tells us that the point estimates are reasonable overall, though for any one country are typically off by 10%. It is interesting that leaving out countries or complete months give very similar results. The Supplementary Materials contain comparisons of fitted versus observed, both in-sample and out-of-sample, along with residual plots over time.

Our point estimate for the excess mortality over 2020–2021 is 14.9 million with a 95% credible interval of (13.3, 16.6) million. In Figure 6 we plot global and regional estimates. The excess estimates based purely on countries with observed data are also plotted, with uncertainty, which is due to the expected numbers, as grey rectangles. Note that we do not include subnational data in the rectangles. Globally, and with respect to our estimate, roughly half of the contribution to the excess is from observed data, and half from modeling (though, in addition, India contributes around a third to the total excess, and our estimate is based on extensive subnational data). This further emphasizes which region's estimates are based primarily on observed data (EURO and AMRO) and those that are not. It is interesting that the IHME estimates for EURO and AMRO are relatively higher than the rectangles, even though the excess is observed for the majority of country-month combinations. Our global estimate is the lowest of the three. In general, The Economist uncertainty intervals are widest and those of IHME are the narrowest. As we discuss in Section 8, in terms of the procedures used, the IHME intervals are on the shakiest ground theoretically, even if all model assumptions are satisfied. The SEARO interval is particularly striking, given the uncertainty over India's excess mortality.

Figure 7 gives the cumulative estimated by month and by region. The impact of the surge of deaths in India (which is in the SEARO region) in May 2021 is apparent. The WPRO region has a number of countries with negative excess (because of strong lockdown policies leading to the avoidance of certain types of death), and in this region, the mortality impact of the pandemic was smallest according to our analysis up to the end of 2021.

Figure 8 shows the global excess death rates, where countries without data are highlighted with hatching. The paucity of data in AFRO and SEARO in particular is apparent. India has the largest excess death count (though not the largest excess death rate) and so we briefly discuss our results. The countries with the highest estimated excess yearly death rates (per 100,000 population, and 95% credible intervals) are: Peru with 437 (431, 442), Bulgaria with 415 (399, 432) and Bolivia with 375 (370, 379). Rankings of countries in terms of the excess death rate are included in the Supplementary Materials. Countries with negative excess estimates are Australia, China, Japan, South Korea, Vietnam and New Zealand. Figure 9 maps the ratio of excess deaths to reported COVID-19 deaths. There is a huge range of this excess, with many countries in the AFRO region having high ratios, and countries in Western Europe having ratios closer to 1. Globally, over January 2020–December 2021, there were 542,0534 reported COVID-19 deaths, and according to our estimates, the ratio of excess to reported COVID-19 deaths is 2.75, with a 95% interval estimate of (2.46, 3.07), which is a huge discrepancy.

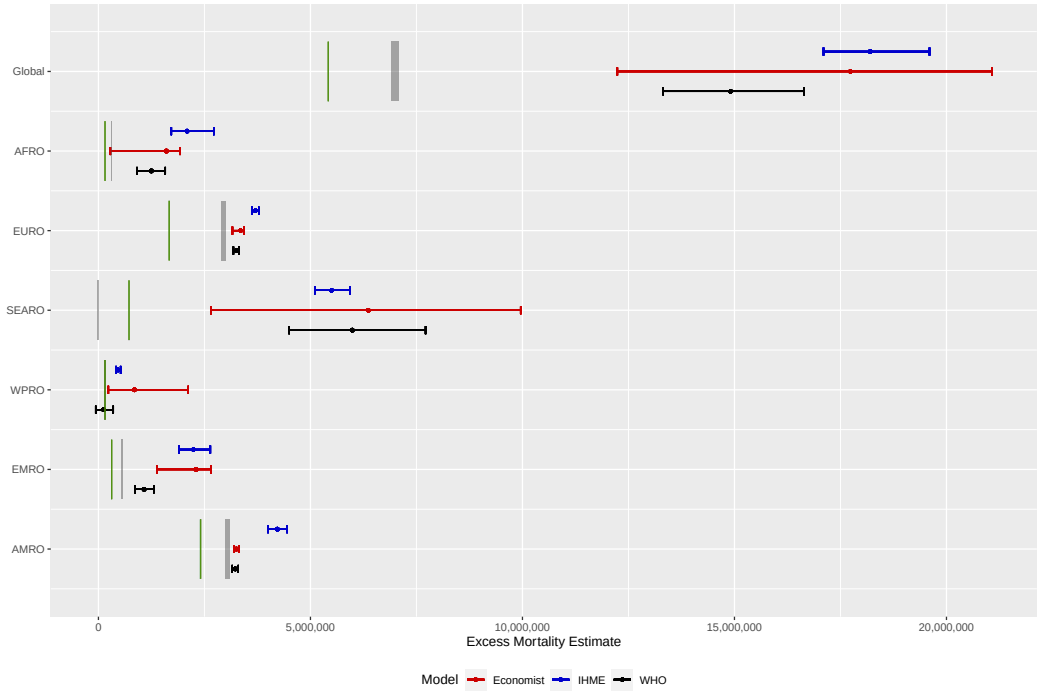

FIG 6. *Global and regional point excess mortality estimates and 95% intervals from WHO, The Economist and IHME. The grey vertical thin rectangles correspond to the excess from those countries with observed ACM death, so the only uncertainty comes from the expected numbers (the width of these rectangles reflects this uncertainty). The green vertical lines show the reported COVID-19 deaths.*

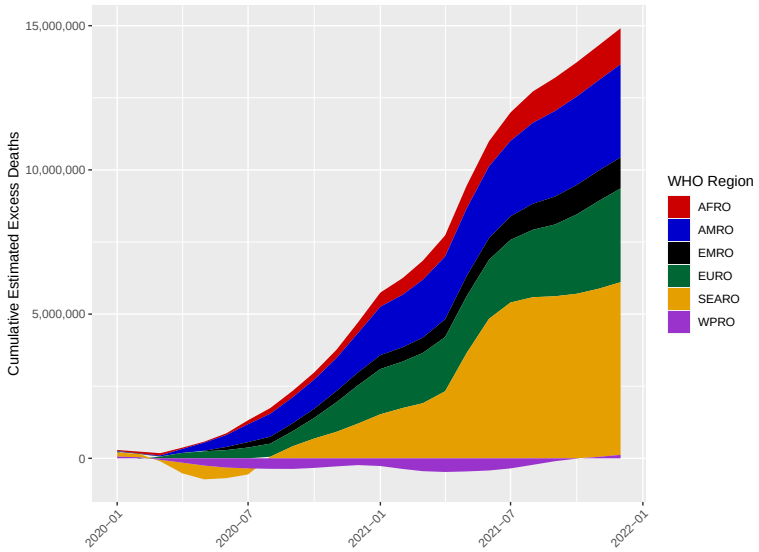

FIG 7. *Cumulative excess deaths over 2020–2021 for all countries, by region.*

Figure 10 shows the predicted ACM for India, based on the state level data, along with the expected deaths, both with uncertainty estimates. We estimate that India has the highest cumulative excess of 4.7 million deaths, with a 95% interval of (1.4, 9.3) million. For the

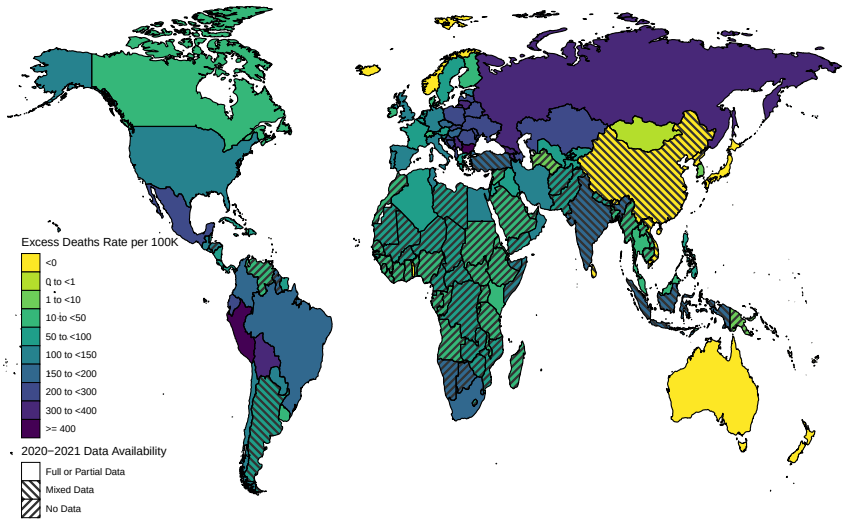

FIG 8. *Excess death rate, per 100,000 by country. Countries with no hatching have monthly observed data, and the two hatching types indicate other data types.*

final 3 months of 2021 there is a single state only (Tamil Nadu) available, and for these 3 months the counts appear high (perhaps due to late registration), and so we do not use these data and instead use a simple predictive model. Specifically, we model $\log(Y_t/E_t)$ (using the estimated Y_t for the first 21 months and weighting by the variance of the estimate) using an autoregressive order 1 (AR1) model, in INLA and then predict the final 3 months. Recall that these estimates are based on subnational data, and hinge on the assumption that at any month, the sum of the available states proportions are close to those observed historically. We cannot check this assumption and so we interpret our results with caution. The choice is between using the global covariate model, or the subnational data, and we went with the latter. The Supplementary Materials contain a sensitivity analysis in which we remove data from different states and examine the excess mortality estimates, and also provide a comparison between our estimates and those of different groups, which shows our estimates are consistent with previous studies.

7. Comparison to Alternative Methods. The Economist and IHME also produce country and global excess mortality estimates and The Economist update their estimates daily (which is not our objective). The Economist method is the more transparent and defensible of the two. The Economist estimates excess deaths for all countries (Economist and Solstad, 2021b) using methods described at Economist and Solstad (2021a). The Economist is not a peer-reviewed publication. From the start of the work, The Economist's methods and code have been freely available. The response is taken as excess deaths per 100k population, per day and the regression approach is gradient boosting (Friedman, 2001), with regression trees applied to a very large collection of variables (144 in total) at the 7-day average level, when available. Since the excess is modeled, negative excess is possible; as we describe shortly, the IHME approach models log excess, so that negative values are not possible. A weighting of log population is taken in The Economist approach, though this choice is arbitrary, beside having the desirable property of having weights that increase with increasing population size. The weights are reduced by 50% for subnational data sources, which is also arbitrary. The loss function is taken as mean squared error. An alternative would be to take the negative log likelihood of a Poisson as the loss function as described, for example, in Section 7.2 of

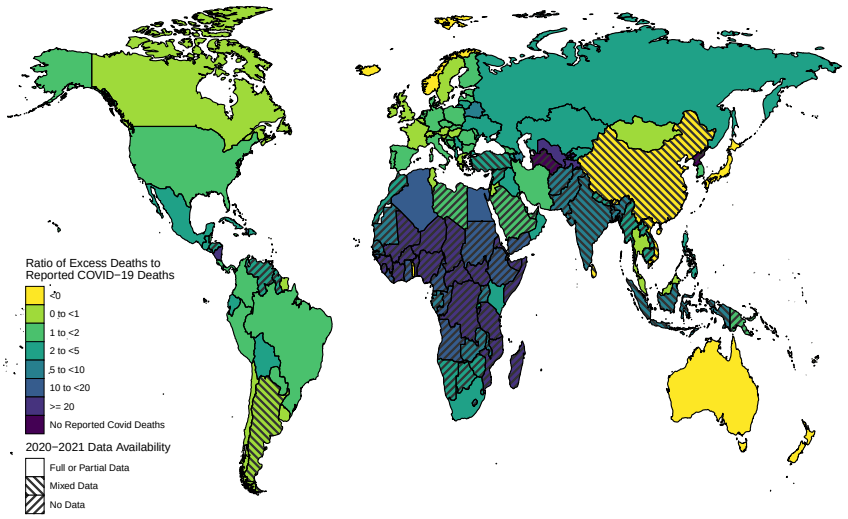

FIG 9. *Ratio of excess death rate to reported COVID-19 death rate, per 100,000 by country. Countries with no hatching have monthly observed data, and the two hatching types indicate other data types.*

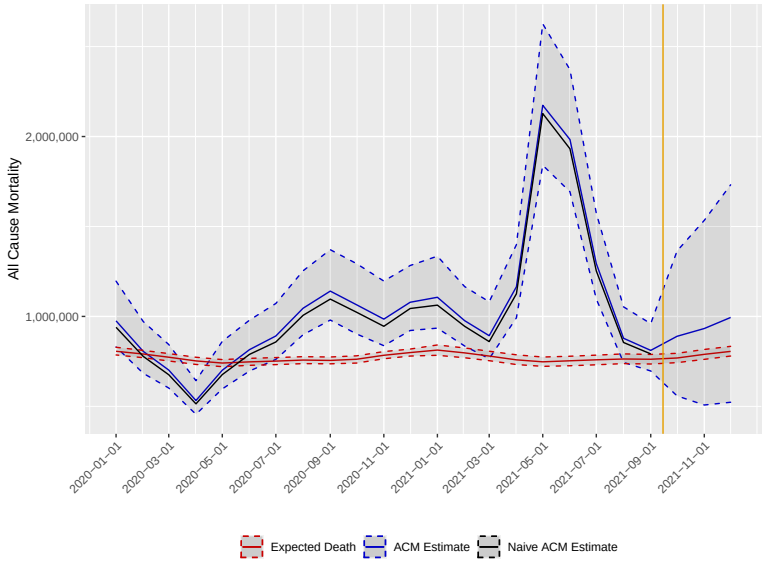

FIG 10. *Estimates with 95% uncertainty for India. The final 3 months are based on an AR1 model.*

Bühlmann and Hothorn (2007). Model assessment is based on 10-fold cross-validation and a non-parametric bootstrap is used to assess (frequentist) uncertainty, based on 200 datasets sampled with replacement from the full data (with random sampling of countries first, and then observations within the sampled country). The expected numbers are modeled using the method described in Karlinsky and Kobak (2021). Specifically, the number of deaths is modeled as linear in year, with weekly (or monthly or quarterly, if weekly data not available) intercepts, using data from 2015–2019. These expected numbers are used directly in the calculation of the excess, when ACM data are observed. For countries without ACM data the excess (i.e., $\delta_{c,t}$) is predicted directly, though the 2019 WHO ACM counts are used as one of the covariates. Uncertainty in the expected numbers in the overall uncertainty for the excess

mortality is not accounted for. The Economist's estimates update daily and are freely available. The models themselves also update daily, with two new models trained on the latest data every morning, replacing old models and then used for improved central estimates and estimates of uncertainty.

IHME also produce estimates of excess mortality with methods described in the Appendix of Wang *et al.* (2022). Expected mortality is estimated using an ensemble approach in which six different models are used to model the expected numbers. The expected ACM is only calculated for time periods not affected by late registration, which if not accounted for, would lead to underestimation of excess mortality rate. An out of sample prediction is then carried out for each of the models, and then the final predicted expected number is a weighted combination of the six models, with weights proportional to the mean squared error of prediction, as estimated from a leave-out exercise. While superficially this approach has elements in common with the super learner prediction algorithm (Van der Laan *et al.*, 2007), it differs in key elements and does not share the optimality properties of super learner.

An unweighted analysis is used, with response the log excess cumulative mortality rate:

$$Z_c = \log[(Y_c - E_c)/N_c]$$

where Y_c, E_c and N_c are the observed cumulative counts, expected cumulative counts and population size respectively, for data in country c, all over the relevant observed period. The modeling of this difference does not seem as natural as modeling the log of observed over expected mortality which would be an approximation to the response we have used (though we model over time also). The uncertainty in the true rate of excess is highly dependent on the population size, but this information is not used, since the model implicitly assumes each data point has the same uncertainty attached to COVID-19. If we assume that $\mathrm{E}[Y_c] = N_c\phi_c$ and $\mathrm{var}(Y_c) = \kappa \mathrm{E}[Y_c]$ then, the delta method gives $\mathrm{var}(Z_c) \approx \kappa N_c\phi_c/(Y_c - E_c)^2$, which would give weights approximately proportional to N_c, illustrating the inadequacy of the constant variance assumption. The covariates are also included based on the expected direction of the association, but this expected direction is presumably with respect to univariate models, and in a predictive model with multiple covariates it seems overly restrictive. Covariates are selected in an initial step using the log cumulative excess, as defined above, and the lasso (Tibshirani, 1996). Since cumulative rates are used, a weighted average (e.g., using population) of the covariates is taken.

The uncertainty in this initial covariate selection phase is not accounted for, so that we would expect, all else being equal, the final predictive intervals to be too narrow. With the selected covariates (16 are listed in Section 4.2.2 of the Appendix of Wang *et al.* (2022)), the log of the excess rate is modeled (using the expected ACM rate from the ensemble step and the observed ACM rate). We might also expect the modeling of the log excess to in some cases push estimates of the excess rate that are close to zero upwards. At this stage, Global Burden of Disease (GBD) defined regional and super regional residuals (GBD, 2020) are generated, and their mean is added to the prediction – it is not clear why fixed (or random) effects are not added to the log excess rate model directly. This would make the calculation of uncertainty measures more straightforward.

We describe the estimation of the excess rate for four different data scenarios:

- For countries with observed ACM data over the whole 2-year period, the only uncertainty arises from the modeling of the expected numbers – the uncertainty in this step comes from parameter uncertainty, and not Poisson variation. For each of the six constituent models 100 draws are taken from the asymptotic normal distribution of the estimators, and then a weighted combination of the resultant predicted expected numbers is taken. This is similar to the approach we take, albeit with a different model and a different calculation of uncertainty for the expected numbers.

- For countries with no ACM data, similar to The Economist method, the expected numbers are not calculated, but instead the model directly predicts the excess rate using the estimated regression coefficients of the model. The uncertainty here comes from the random covariates and from the expected numbers modeling, not from any parameter uncertainty for any one fit. But 100 fits are carried out with 100 different expected numbers. There is also no sampling uncertainty, analogous to our negative binomial uncertainty for ACM. This, combined with the lasso pre-selection of covariates, would indicate that the interval estimates would be too narrow, perhaps substantially so.
- For countries with partial data, the cumulative excess rate over the missing (customized to each country) period is obtained from the regression model, adjusted by the residuals (as described above), and then taking random covariates for the missing period.
- It is not possible to obtain negative estimates from the log excess rate model, and so the only way for negative excess to arise is from countries with observed ACM data (Iceland, Australia, Singapore, New Zealand, and Taiwan). The rationale is that there are few locations with a cumulative negative excess rate, and so they wish to avoid making predictions of negative excess.

The overall approach (which has not been peer-reviewed in the statistical literature) is more algorithmic than statistical in nature, and it would be impossible to determine its operating characteristics. In particular, the uncertainty estimates are unlikely to be well-calibrated – we saw they were relatively narrow in Figure 6.

8. Discussion. The estimation of excess mortality during the COVID-19 pandemic is hamstrung by the lack of ACM data for almost half the countries of the world, with EURO and AMRO being well-represented in the databases, but other regions very poorly. We have presented a relatively simple Poisson modeling framework, as we wanted to strive for transparency and leverage a well-understood Bayesian hierarchical structure. We stress that, within the Poisson framework, though we have different models for countries with different data types, the estimates for each country are comparable, and so side-by-side comparisons can be made (with the caveat that the range of uncertainty in the estimates for different countries varies considerably). We deliberately avoid breaking down excess mortality into that directly attributable to COVID-19 and that not, since we believe the information required to do this accurately is unavailable.

We did not adjust the observed ACM on the basis of heatwaves, as done by Karlinsky and Kobak (2021) and Wang *et al.* (2022), and neither did we adjust for conflicts (The Economist adjusts for conflict by excluding ACM data from places which entered large conflicts in the period). Another inadequacy of our modeling is that we are missing covariates in some countries, and regional values are used instead, we do not account for this uncertainty in our modeling. Some of the covariates are modeled, but we do not account for this aspect, which is considered (albeit in an ad hoc, unvalidated procedure) by Wang *et al.* (2022).

Estimating excess mortality by month over the pandemic is a dynamic process and the results we have shown are a snapshot, given the current version of the model and the currently available data. As new data become available we will continue to both update our estimates, and refine our model. Another aspect we will explore is the use of a spatial model, though we approach this with hesitancy.

To reiterate: *the biggest limitation to our study is the lack of any observed mortality data in around half of the countries of the world, which requires us to predict these counts based on a model built with data from countries which are not representative of the missing countries.* In Section 6, we reported coverage estimates, calculated via cross-validation, that were reasonably close to the nominal. However, given the aforementioned regional imbalance in countries for which we have data, we would not expect the coverage to be as accurate for

the missing countries. Improvements in death registration systems is vital to understand and react to pandemics in a timely manner, and obviate the need to carry out such modeling.

Acknowledgments. The authors would like to thank the WHO Technical Advisory Group on COVID-19 Mortality Assessment, for helpful feedback during the model development, and also Sondre Ulvund Solstad and Haidong Wang for generously sharing details on The Economist and IHME approaches, respectively.

REFERENCES

Adair, T. and Lopez, A. D. (2018). Estimating the completeness of death registration: An empirical method. *PLOS ONE*, **13**, e0197047.

Baker, S. G. (1994). The multinomial-Poisson transformation. *Journal of the Royal Statistical Society: Series D*, **43**, 495–504.

Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, **22**, 477–505.

Economist, T. and Solstad, S. (2021a). https://www.economist.com/graphic-detail/coronavirus-excess-deaths-estimates.

Economist, T. and Solstad, S. (2021b). The pandemic's true death toll. https://www.economist.com/graphic-detail/2021/05/13/how-we-estimated-the-true-death-toll-of-the-pandemic.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, **29**, 1189–1232.

Garcia, J., Torres, C., Barbieri, M., Camarda, C. G., Cambois, E., Caporali, A., Meslé, F., Poniakina, S., and Robine, J.-M. (2021). Differences in COVID-19 mortality: Implications of imperfect and diverse data collection systems. *Population*, **76**, 35–72.

GBD (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, **396**, 1204–1222.

Hale, T., Angrist, N., Kira, B., Petherick, A., Phillips, T., and Webster, S. (2020). Variation in government responses to COVID-19. Technical report, University of Oxford.

Karlinsky, A. (2021). International completeness of death registration 2015-2019. *medRxiv*.

Karlinsky, A. (2022). Estimating national excess mortality from subnational data: application to Argentina. *Revista Panamericana de Salud Publica*.

Karlinsky, A. and Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife*, **10**, e69336.

Kelsall, J. E., Zeger, S. L., and Samet, J. M. (1999). Frequency domain log-linear models; air pollution and mortality. *Journal of the Royal Statistical Society: Series C*, **48**, 331–344.

Kung, S., Doppen, M., Black, M., Hills, T., and Kearns, N. (2020). Reduced mortality in New Zealand during the COVID-19 pandemic. *The Lancet*.

Leon, D. A., Shkolnikov, V. M., Smeeth, L., Magnus, P., Pechholdová, M., and Jarvis, C. I. (2020). Covid-19: a need for real-time monitoring of weekly excess deaths. *The Lancet*, **395**, e81.

Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial N. *Ecology*, **94**, 2173–2179.

Mikkelsen, L., Phillips, D. E., AbouZahr, C., Setel, P. W., Savigny, D., Lozano, R., and Lopez, A. D. (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, **386**, 1395–1406.

Németh, L., Jdanov, D. A., and Shkolnikov, V. M. (2021). An open-sourced, web-based application to analyze weekly excess mortality based on the Short-term Mortality Fluctuations data series. *PloS ONE*, **16**, e0246663.

Riffe, T. and Acosta, E. (2021). Data resource profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. *International Journal of Epidemiology*, **50**, 390–390f.

Rivera, R., Rosenbaum, J. E., and Quispe, W. (2020). Excess mortality in the United States during the first three months of the covid-19 pandemic. *Epidemiology and Infection*, **148**.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Application*. Chapman and Hall/CRC Press, Boca Raton.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.

Simpson, D., Rue, H., Riebler, A., Martins, T., and Sørbye, S. (2017). Penalising model component complexity: A principled, practical approach to constructing priors (with discussion). *Statistical Science*, **32**, 1–28.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267–288.

UNSD (2021). Demographic Yearbook 2021. Technical report, United Nations Statistics Division.

Van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, **6**.

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., Aravkin, A. Y., Bisignano, C., Barber, R. M., Alam, T., *et al.* (2022). Estimating excess mortality due to the covid-19 pandemic: a systematic analysis of covid-19-related mortality, 2020–21. *The Lancet*.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. CRC Press.