Methods for estimating the excess mortality associated with the COVID-19 pandemic

World Health Organisation

March 29, 2022

Contents

1 Introduction

The World Health Organisation (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 data do not provide a complete picture of the health burden attributable to COVID-19, nor of how many lives have been lost 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. There have also been variations in the death certification rules countries have applied in regards to COVID-19 (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, Doppen, Black, Hills, and Kearns, 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, Shkolnikov, Smeeth, Magnus, Pechholdová, and Jarvis, 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.

To derive estimates of the excess mortality attributable to COVID-19 in country c, all-cause mortality (ACM) counts in month t for 2020 and 2021 are denoted by $Y_{c,t}$ and, in addition to the 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 are forecasted using historic (prior to the pandemic) deaths data.

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, Phillips, AbouZahr, Setel, Savigny, Lozano, and Lopez, 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 report 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 models used to estimate all-cause mortality under different scenarios of data availability. Section 5 describes the current approach that has been applied to derive preliminary age- and sex-distribution of deaths. Finally section 6 provides some metrics used to assess the model performance.

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.

Table 1 summarises the data that are available for this exercise. Groupings include "Full national" countries, which are countries that have data over all 24 months (January 2020 to December 2021); "Partial national" which have data for less than 24 months; "Mixed data" countries with subnational monthly data for some period (4 countries), national annual data (5 countries) or a combination (China) and then countries without reported ACM that the WHO have had access to. Groupings in the table are according to the 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).

Region	Full National	Partial National	Mixed Data	No Data	Total
AFRO	4	2	0	41	47
AMRO	12	11	4	8	35
EMRO	4	5	0	12	21
EURO	46	5	1	1	53
SEARO	1	1	3	6	11
WPRO	6	3	2	16	27
Global	73	27	10	84	194

Table 1: Country data availability summary for 2020 and 2021

All countries report their official COVID-19 death count, but these do not provide a complete picture of the impact of the pandemic, for many reasons already outlines. 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, Jdanov, and Shkolnikov, 2021) and the World Mortality Dataset (WMD), as described in Karlinsky and Kobak (2021).

The work we report on here is a snapshot of the current state of data availability and over time the situation will improve. As shown in Table 1, 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. 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 64% 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. For all countries we do, however, account for uncertainty in the expected numbers.

2.2 Covariate Data

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 (combines lockdown restrictions and closures), historic 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. A number of these 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, regional values were used for countries with missing data. Details are given in the Supplementary Materials.

3 Deriving expected mortality for years 2020 and 2021

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 $f^{\text{y}}_c(\cdot)$ models the *annual trend*, and $f^{\text{m}}_c(\cdot)$ 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, Rosenbaum, and Quispe, 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 monthly historical data, and so a linear term is used for modeling yearly variation. This model is used to obtain predictions of the expected deaths $\mu^{\text{E}}_{c,t}$ 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 excess expected numbers by month t for $t = 1,\ldots,24$. 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 a multinomial 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,

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

where $\boldsymbol{p}_{c,v} = \{p_{c,v,m}, m = 1,\ldots,12\}$ 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.

To summarize, our strategy for producing expected numbers for countries with annual data only is:

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.

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 acknowledged 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 normal. 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, Martino, and Chopin, 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.

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$(\widehat{\tau}_{c,t'}, \widehat{\tau}_{c,t'}/\widehat{E}_{c,t'})$.

4 Estimating all-cause mortality for years 2020 and 2021

4.1 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 $\delta_{c,t}$.

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.

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 $\theta_{c,t} >/< 1$ 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,

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

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 RW2 prior 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.
- 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, Rue, Riebler, Martins, and Sørbye, 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, Zeger, and Samet, 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.

4.2 Subnational Data Model

For a small number of countries for which national ACM data are not available (e.g., 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, while for China we have subnational monthly and national annual data. 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.

For Turkey we have subnational monthly data over the complete two years of the pandemic, while for Indonesia we have annual 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 (out of 26) over the pandemic period, 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 registration, or by journalists who obtained death registration information through Right To Information requests.

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 $S_t = \{0, K_t\}$. We assume, in month t:

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

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

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

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.

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 $e_t \sim \text{N}(0, \sigma_e^2)$, 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 $\mathrm{E}[Y_{t,2}|p_t] = Y_{t1}(1-p_t)/p_t$, i.e., of the same form as the simple frequentist "obvious" estimator.

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

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

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 annual subnational for 2020 and so we use a national binomial model and then apportion the counts using the multinomial temperature model described in Section 3.2.

4.3 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.

For China, we have annual national data and also subnational monthly data for the first 9 months of each of 2020 and 2021. In the Supplementary Materials we describe a model for combining the two types of data and an MCMC implementation.

5 Methods for deriving sex- and age-pattern

Beyond determining the levels of excess mortality attributable to COVID-19, we intend to disaggregate these deaths by age and sex. For most countries, the sex and age attributes are not identified in the mortality data that are available for years 2020 and 2021. These are necessary inputs when one begins to look at sex and age differentials, impact relative to other causes as well as life tables, and the impact on life expectancy. In particular, these are required for the current year Global Health estimates and projections within the World Population Prospects.

To generate estimates of excess mortality by age and sex we consider the expected sex and age profile for the years 2020 and 2021 and assess the difference in the expected and the observed in the places with reported data. For a specific location, the observed/reported death numbers for an age-group x, sex s, country c in year y, can be represented by

$$Y_{c,y,s,x}$$

and are assumed to be a result of the direct effects of COVID-19 (deaths attributable to it) and the indirect knock-on effects on health systems and society. The hypothetical or "counterfactual" no-COVID-19 scenario uses the expected death numbers

$$E_{c,y,s,x}$$

which are forecasted using historical data. Excess deaths by sex and age, represented by $\delta_{c,y,s,x}$, can thus be defined as the difference:

$$\delta_{c,y,s,x} = Y_{c,y,s,x} - E_{c,y,s,x}$$

As with the approach taken for deaths over all ages and for both sexes combined, the goal is to determine standard patterns of excess mortality by sex and age for the places with reported data and then generalize them to the other countries without. Simultaneously, we aim to propagate the uncertainty in the overall excess death numbers for the years 2020 and 2021 (predicted using the statistical models for overall mortality) to the predicted sex- and age-patterns. The steps towards accomplishing this are described in more detail in the sections that follow.

5.1 Countries with observed sex-age data for years 2020 and 2021

We consider country- and sex-specific deaths for the year 2020 aggregated to 5-year age-bands $x \in \{0-4, 5-9, \ldots, 85+\}$. Of interest is the location- and year-t-specific death-rate in age interval $[x, x+n)$, represented by ${}_nm_{x,t}$ which is calculated using the counts and the population numbers according to WPP2019. Only a subset of countries has observed data to estimate these

quantities at this level of granularity for 2020 and 2021. Of these, excluding the countries that have experienced conflict, have very small population numbers, incomplete deaths and/or have erratic/implausible age-patterns, the countries with age-sex patterns we have applied in a model framework are listed in Table 2 below:

Table 2: Countries with age and sex data used in model

	ISO3C	WHO_region	Population 2019	Deaths 2020	Deaths 2021
1	ALB	EURO	2877800	39204	43429
2	AUS	WPRO	25499881	162333	168685
3	AUT	EURO	9006400	93476	93175
4	BEL	EURO	11589616	132411	117500
5	BGR	EURO	6948445	124823	148203
6	BRA	AMRO	212559409	1608136	
7	CHE	EURO	8654618	77613	71519
8	CHL	AMRO	19116209	126178	137667
9	COL	AMRO	50882884	301705	
10	CYP	EURO	1207361	7002	7625
11	CZE	EURO	10708982	129289	140564
12	DEU	EURO	83783945	986018	1017531
13	DNK	EURO	5792203	54790	57363
14	ECU	AMRO	17643060	126523	
15	ESP	EURO	46754783	506331	468096
16	EST	EURO	1326539	15864	18665
17	FIN	EURO	5540718	56926	58944
18	FRA	EURO	65273512	654772	643178
19	GBR	EURO	67886004	703504	683116
20	GRC	EURO	10423056	132134	145102
21	HRV	EURO	4105268	57261	63085
22	HUN	EURO	9660350	141092	155418
23	IRL	EURO	4937796	31981	
24	IRN	EMRO	83992953	504872	
25	IRQ	EMRO	40222503	203651	
26	ISR	EURO	8655541	48753	50769
27	ITA	EURO	60461828	756293	717949
28	JPN	WPRO	126476458	1384544	
29	KAZ	EURO	18776707	160872	180287
30	KOR	WPRO	51269183	306226	318282
31	LTU	EURO	2722291	48299	52233
32	MEX	AMRO	128932753	1065357	
33	MUS	AFRO	1271767	11271	
34	NLD	EURO	17134873	170443	172409
35	NOR	EURO	5421242	40867	42046
36	NZL	WPRO	4822233	32745	34878

37	PER	AMRO	32971846	280706	
38	POL	EURO	37846605	478578	519896
39	ROU	EURO	19237682	297122	334239
40	RUS	EURO	145934460	2138586	
41	SGP	WPRO	5850343	27722	30463
42	SRB	EURO	8737370	133273	154700
43	SVK	EURO	5459643	59210	72685
44	SVN	EURO	2078932	24246	23192
45	SWE	EURO	10099270	102022	96755
46	TUN	EMRO	11818618	75058	
47	UKR	EURO	43733759	630859	
48	URY	AMRO	3473727	32638	
49	USA	AMRO	331002647	3504924	3583956
50	ZAF	AFRO	59308690	581618	

5.2 Grouping countries to generalize sex-age patterns

We consider how to group the countries with data in order to extrapolate any of the estimated 2020/2021 impacts to the locations without. A natural grouping would be geographically using some regional identification e.g., the WHO region. However, this poses two dilemmas:

1. Not all regions are represented adequately in the observed data e.g., there are no countries from SEARO region and there is only one country from AFRO.

2. Even within close geographic proximity, the scale of the impact of the pandemic overall and thus potentially by sex and age, varies e.g., Finland shares borders with Russia but the reported excess rates for the countries is on significantly different orders if magnitude.

Instead of using the natural geography to group the data, we allow the data to drive the clusters. We apply the K-means clustering approach (Likas, Vlassis, and Verbeek, 2003). K-means is a method commonly used to automatically partition a data set into K groups. The K-means method uses K centers of clusters, to characterize the data. These centers are determined by minimizing the sum of squared errors,

$$J_{\mathrm{K}} = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - \mathbf{m}_k)^2$$

where $(x_1, \ldots, x_n) = X$ is the data matrix and $\mathbf{m}_k = \sum_{i \in C_k} x_i / n_k$ is the centroid of cluster C_k and n_k is the number of points in C_k.

There are many options of features that could be used to create the data matrix $\mathbf{X}$ which is

informing the clusters. Using the human development index (HDI) (Anand and Sen, 1994) for 2019, the age-specific all-cause deaths in 2019 $Y_{x,2019}$, the estimated/reported total deaths in 2020 $\hat{Y}_{2020}$, age-specific population in years 2019 and 2020, $N_{x,2019}$ and $N_{x,2020}$ respectively, and the predicted overall excess deaths for 2020, δ, we consider three features in total (one of which relates to the predicted excess mortality attributable to COVID-19):

(i) For each country, the human development index for year 2019:

$$f_1 = hdi$$

(ii) For each country, the mean age at death in 2019:

$$f_2 = \frac{\sum_x x \times Y_{x,2019}}{\sum_x Y_{x,2019}}$$

(iii) For each country, the crude excess mortality rate for 2020:

$$f_3 = \frac{\delta}{\sum_x N_{x,2020}}$$

These are derived for each of the 194 member states and normalized to derive the $mathbf X$ matrix:

$$\mathbf{X}_{(194,3)} = \begin{bmatrix} j & 1 & 2 & 3 \\ 1 & f_{1,1} & f_{2,1} & f_{3,1} \\ \vdots & \vdots & \vdots & \vdots \\ j & f_{j,1} & f_{j,2} & f_{j,3} \\ \vdots & \vdots & \vdots & \vdots \\ 194 & f_{194,1} & f_{194,2} & f_{194,3} \end{bmatrix}$$

For the subset of countries with observed sex- and age-specific data for 2020, the K-means approach is applied. The number of clusters K is chosen to maximise the variation between clusters and to minimise the variation within clusters. Five clusters are selected. The resultant cluster compositions are shown in Figure 1 with Table 3 listing cluster constituents:

Figure 1: K-means cluster allocation by ISO3C

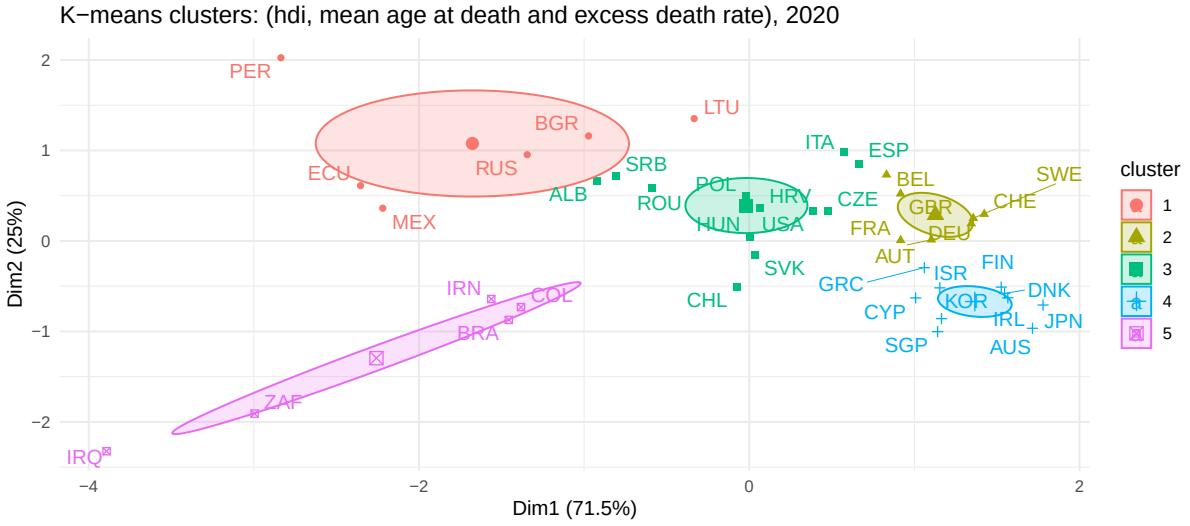


Table 3: K-means cluster allocation by ISO3C

1	2	3	4	5
BGR	AUT	ALB	AUS	BRA
ECU	BEL	CHL	CYP	COL
LTU	CHE	CZE	DNK	IRN
MEX	DEU	ESP	FIN	IRQ
PER	FRA	HRV	GRC	ZAF
RUS	GBR	HUN	IRL	
	SWE	ITA	ISR	
		POL	JPN	
		ROU	KOR	
		SRB	SGP	
		SVK		
		USA		

5.3 Extrapolating cluster groupings to countries without observed data

The K-means clusters provide groupings for the data based on the HDI, the mean age at death, as well as the overall excess mortality rates and porportion of total deaths. Not all countries are included in the original clustering as this was done for a subset to ensure that all countries with data are clustered into optimal bins. However, we require all countries to be assigned to clusters and this is accomplished by mapping each country to the 5 K-mean clusters using the multivariate minkowski distance (Singh, Yadav, and Rana, 2013) between the X matrix values and the cluster averages shown in the Table 4:

Table 4: K-means cluster average normalised values

Cluster	HDI	Mean age at death	Crude excess rate
1	0.517	0.531	1.470
2	1.433	1.205	0.066
3	0.922	0.971	0.611
4	1.409	1.140	-0.579
5	0.296	0.310	0.205

5.4 Extrapolating sex-age impact to countries without observed data

To capture the Covid-19 excess mortality impact for the years 2020 and 2021 for the countries with observed data as well as those without, requires expected death numbers by sex and age. The expected death numbers by age and sex are derived using the GHE2019 age-sex patterns rebalanced to the expected death numbers that are derived for the overall longitudinal mortality model. This enforces consistency in the expected numbers and also minimizes any potential bias that can be introduced in the single year forecast that would otherwise be necessary to derive the expected deaths for 2020. Acknowledging the uncertainty inherent in the GHE2019 estimates and that this is a very short forecast, we assume that changes in the sex- and age-pattern between observed 2019 and expected 2020 are minimal after adjusting for the level through the aggregate projection.

For the countries with data and listed in the table above (we exclude the countries with low death numbers and zero inflated counts), for each sex and by age-group, we look at the scale $r_{c,y,s,x}$:

$$r_{c,y,s,x} = \left(\frac{log(Y_{c,y,s,x}/N_{c,y,s,x})}{log(E_{c,y,s,x}/N_{c,y,s,x})} \right)$$

where $N_{c,y,s,x}$ are the population counts. This quantity contrasts the observed log mortality rate against the expected to capture the sex- and age-specific changes for 2020. The K-means clusters are used for two aspects of the extrapolation to locations without data. Firstly, to summarise these log mortality scalars into cluster specific distributions. And secondly, to derive country-specific estimates of predicted deaths by sex and age, conditional on the clusters the country lies in.

For each cluster k (and by extension, each country j in the cluster k), we generate sex-specific distribution for the $r_{s,x}$ scalars based on the observed data (dropping subscripts c and y for simplicity). The empirical bootstrap distribution is generated by first smoothing the observed series by age for each country in the cluster and then repeatedly sampling from the smooth

series. The range of possible values by age is assumed to be a Gaussian approximate with distribution.

$$\hat{r}^k_{s,x} \sim N(\bar{r}^k_{s,x}, \sigma^k_{s,x})$$

where $\bar{r}^k_{s,x}$ and $\sigma^k_{s,x}$ are the sex s and age x specific mean and standard deviations for cluster k derived using the smoothed draws of the observed data. Following recommendations from the United Nations Interagency Group for Child Mortality Estimation (UNIGME), we do not extrapolate any protection or otherwise to children and young adults. Figures 2 to 4 are of the observed and smoothed series and are filtered by cluster:

Figure 2: Smoothed ratio by age and sex for clusters 1 to 2

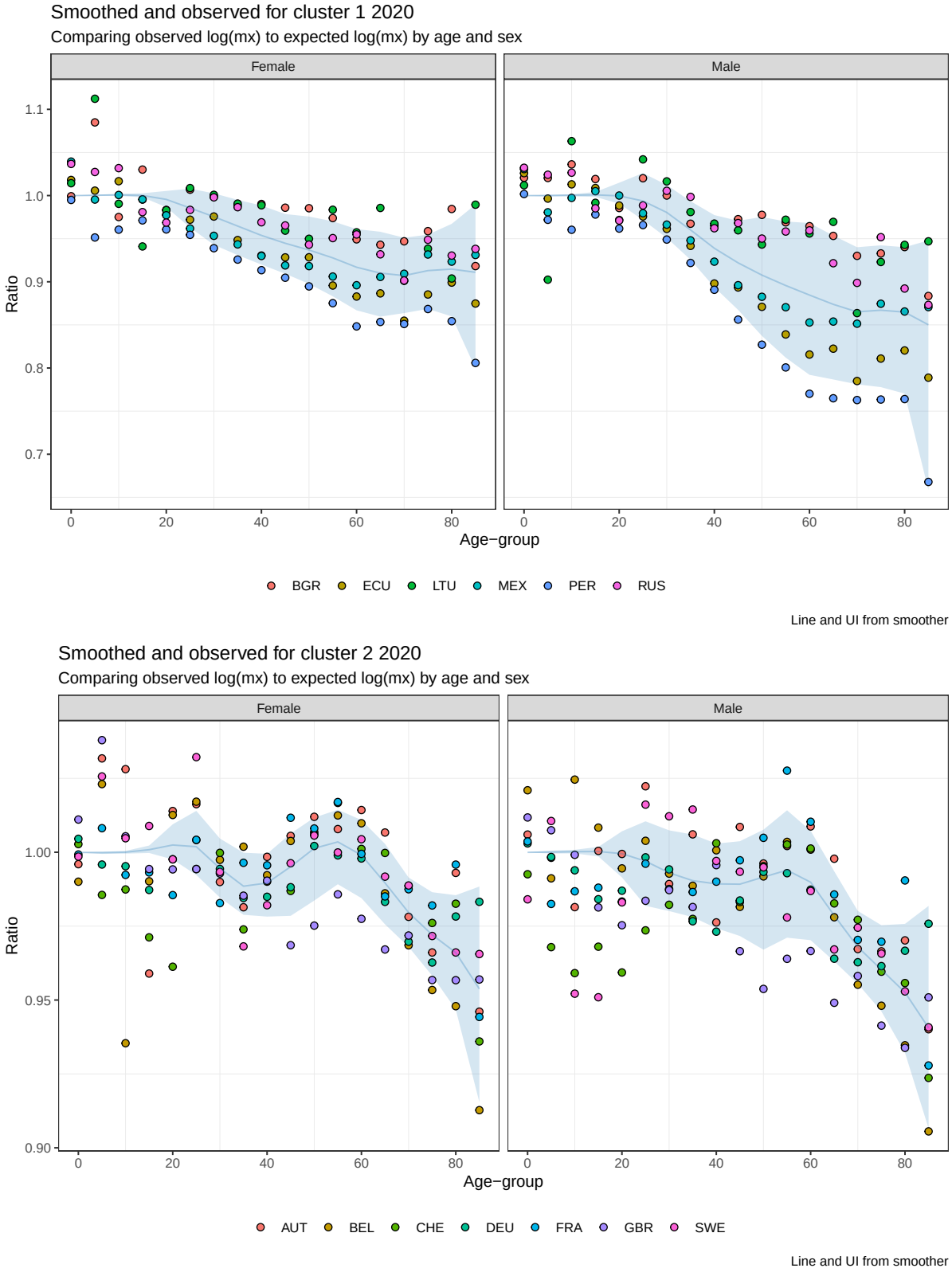


Figure 3: Smoothed ratio by age and sex for clusters 3 to 4

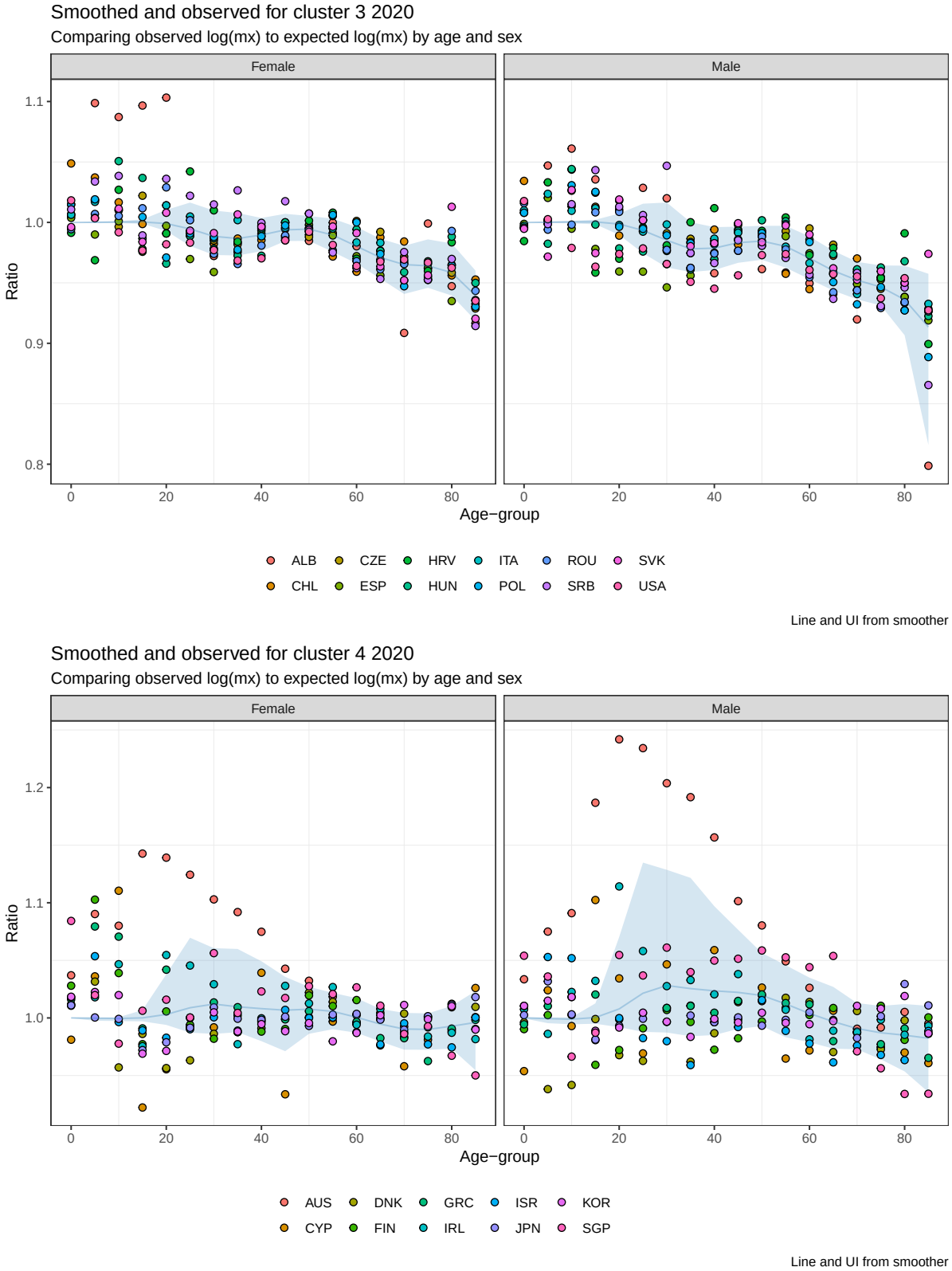
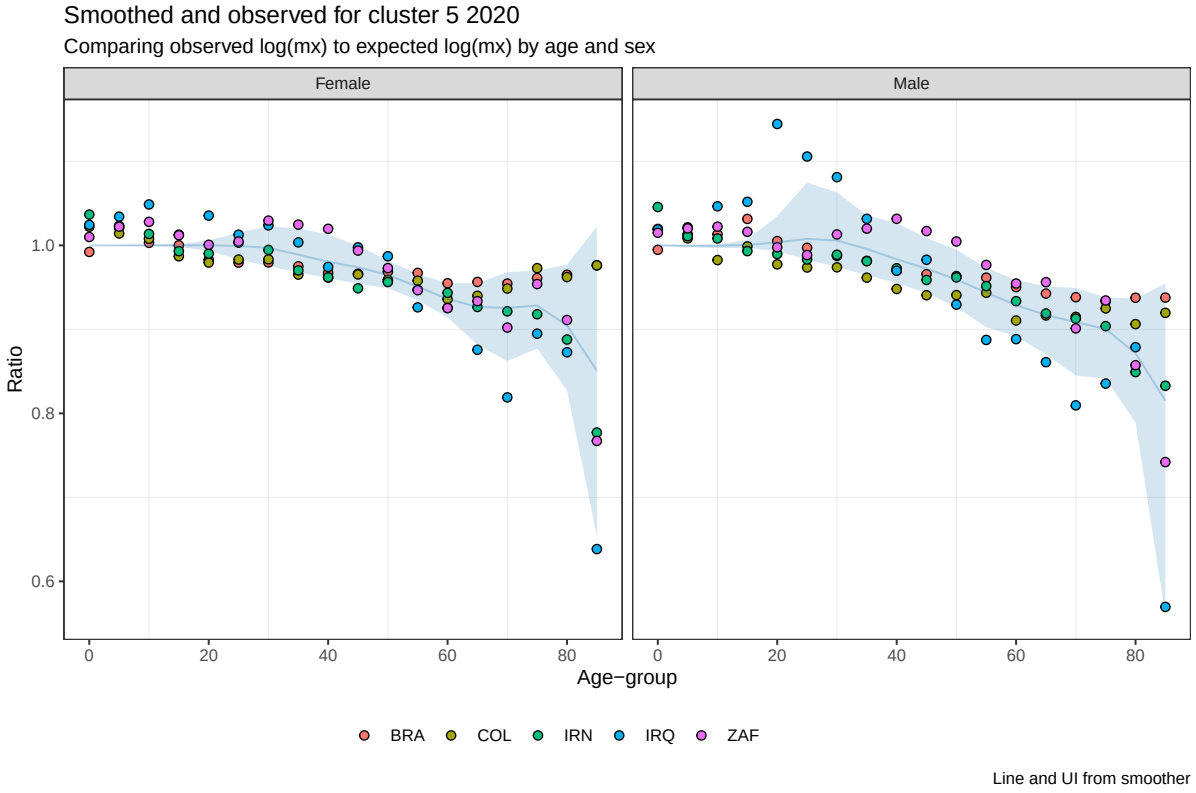


Figure 4: Smoothed ratio by age and sex for cluster 5



The samples of the scalars are combined with the sex s and age x specific expected death rates by country and the sex-age-specific population numbers to generate predicted deaths by age and sex:

$$\hat{Y}_{s,x,2020} = \exp({}_n\mu^e_{s,x,2020} \times \hat{r}^k_{s,x}) \times N_{s,x,2020}$$

where ${}_n\mu^e_{s,x} = log({}_nm^e_{s,x})$. The $\hat{Y_{s,x}}$ are rescaled to correspond to a random realization from the Poisson count model $\hat{Y}$ to produce the final $\tilde{Y}_{s,x}$ i.e.

$$\tilde{Y}_{s,x} = \frac{\hat{Y}_{s,x}}{\sum_s \sum_x \hat{Y}_{s,x}} \times \hat{Y}$$

This process is repeated 1,000 times for each country, drawing unique samples of $\hat{r}_{s,x}$ and all-cause mortality $\hat{Y} \sim N(\bar{Y}, \sigma_Y)$ each time. These are used to generate country-specific distributions (and uncertainty intervals) of deaths by sex and age. The uncertainty shown is the propagation of the uncertainty from the K-means cluster smoothed draws and the Poisson count model draws but should not be interpreted as being parametric or containing a hypothetical "true" value. Rather it shows a range of plausible values conditional on the distribution of the total predicted deaths, the expected deaths and the cluster identified for the country.

6 Model assessment

The sampling model we assume is,

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

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

We wish to assess whether the covariate model provides a good fit to the data. To this end we perform a number of model checks:

1. Plot fitted values $\widehat{y}_{c,t}$ versus observed values $y_{c,t}$. The fitted values are given by $\widehat{Y}_{c,t} = \widehat{E}_{c,t}\widehat{\theta}_{c,t}$ where $\widehat{\theta}_{c,t}$ is the posterior median. We color code the points by region.

   These plots are created both for in-sample, and out-of-sample via cross-validation in which data from either a complete country or a complete month are removed.

2. Plot standardized residuals versus time, color coded by region. Standardized residuals are:

$$r_{c,t} = \frac{y_{c,t} - \widehat{y}_{c,t}}{\sqrt{\widehat{E}_{c,t}\widehat{\theta}_{c,t}\left(1 + \widehat{E}_{c,t}\widehat{\theta}_{c,t}/\widehat{\tau}_{c,t}\right)}}$$

3. To compare models we use log CPO.

   The measure we use for assessing models in the CV exercise is the log conditional predictive ordinate (LCPO). For country c let M_c be the set of months for which ACM data are observed. An overall measure of the fit of a model for the country CV scheme is:

$$\text{LCPO-C} = \frac{1}{\sum_c |M_c|} \sum_c \sum_{t \in M_c} \log \Pr(y_{c,t}|\boldsymbol{y}_{-c}),$$

   i.e., the log of the predictive distribution obtained from all data with country c left out, and evaluated at $y_{c,t}$. Similarly, for the MCV scheme:

$$\text{LCPO-M} = \frac{1}{\sum_c |M_c|} \sum_c \sum_{t \in M_c} \log \Pr(y_{c,t}|\boldsymbol{y}_{-ct}).$$

   In each case, the predictive is the sampling model for the data (negative binomial in our case), averaged over the posterior, given the retained data. For example, with country c left out:

$$\Pr(y_{c,t}|\boldsymbol{y}_{-c}) = \int_\theta \int_\phi \Pr(y_{c,t}|\boldsymbol{\theta}, \phi) \times p(\boldsymbol{\theta}, \phi|\boldsymbol{y}_{-c})\, d\boldsymbol{\theta} d\phi. \tag{8}$$

4. We assess the errors in our model, also using CV, over the countries with ACM data. Let $r_{c,t} = Y_{c,t}/N_{c,t}$ be the observed rate and $\widehat{r}_{c,t} = \widehat{Y}_{c,t}/N_{c,t}$ where $\widehat{Y}_{c,t} = \text{PostMedian}(Y_{c,t}|\boldsymbol{y}_{-ct})$ be

the estimated rate. We report the absolute and relative biases of the ACM rate:

$$\frac{1}{\sum_c |M_c|} \sum_c \sum_{t \in M_c} \frac{\widehat{r_{c,t}} - r_{c,t}}{r_{c,t}}. \tag{9}$$

and

$$\frac{1}{\sum_c |M_c|} \sum_c \sum_{t \in M_c} \frac{|\widehat{r_{c,t}} - r_{c,t}|}{r_{c,t}}. \tag{10}$$

These measures can be calculated where the estimated rates are based on data with a complete country or a complete month's data are left out.

5. We also calculate the root mean squared error (RMSE) of the fit:

$$\sqrt{\frac{1}{\sum_c |M_c|} \sum_c \sum_{t \in M_c} (\widehat{r_{c,t}} - r_{c,t})^2}$$

again using the two cross-validation schemes (by month and by country).

6. Coverage of predictive intervals from cross-validation exercises.

For India, we remove one state at a time and examine the sensitivity of the results, in terms of both the monthly time series of excess, and the cumulative by month. We also remove one state at a time, and then estimate the number of deaths we would see in this state using the estimated fraction of the total deaths in that State and the predictive distribution of the national ACM.

	Final Model
R-RATE-C	1.98
A-RATE-C	10.08
RMSE-C (×1000)	1.25
LOO-C Coverage - 50% Interval	59.3
LOO-C Coverage - 80% Interval	82.7
LOO-C Coverage - 95% Interval	91.6
R-RATE-M	1.84
A-RATE-M	10.18
RMSE-M (×1000)	1.24
LOO-M Coverage - 50% Interval	57.8
LOO-M Coverage - 80% Interval	83.7
LOO-M Coverage - 95% Interval	92.9

Table 5: Leave one country and month out model assessment measures. All rates are expressed as percentages.

We consider the excess $\delta_{c,t} = Y_{c,t} - E_{c,t}$ and assess the model performance using the mean relative error of predicted excess mortality rate, as used in Wang, Paulson, Pease, Watson, Comfort, Zheng, Aravkin, Bisignano, Barber, Alam, *et al.* (2022):

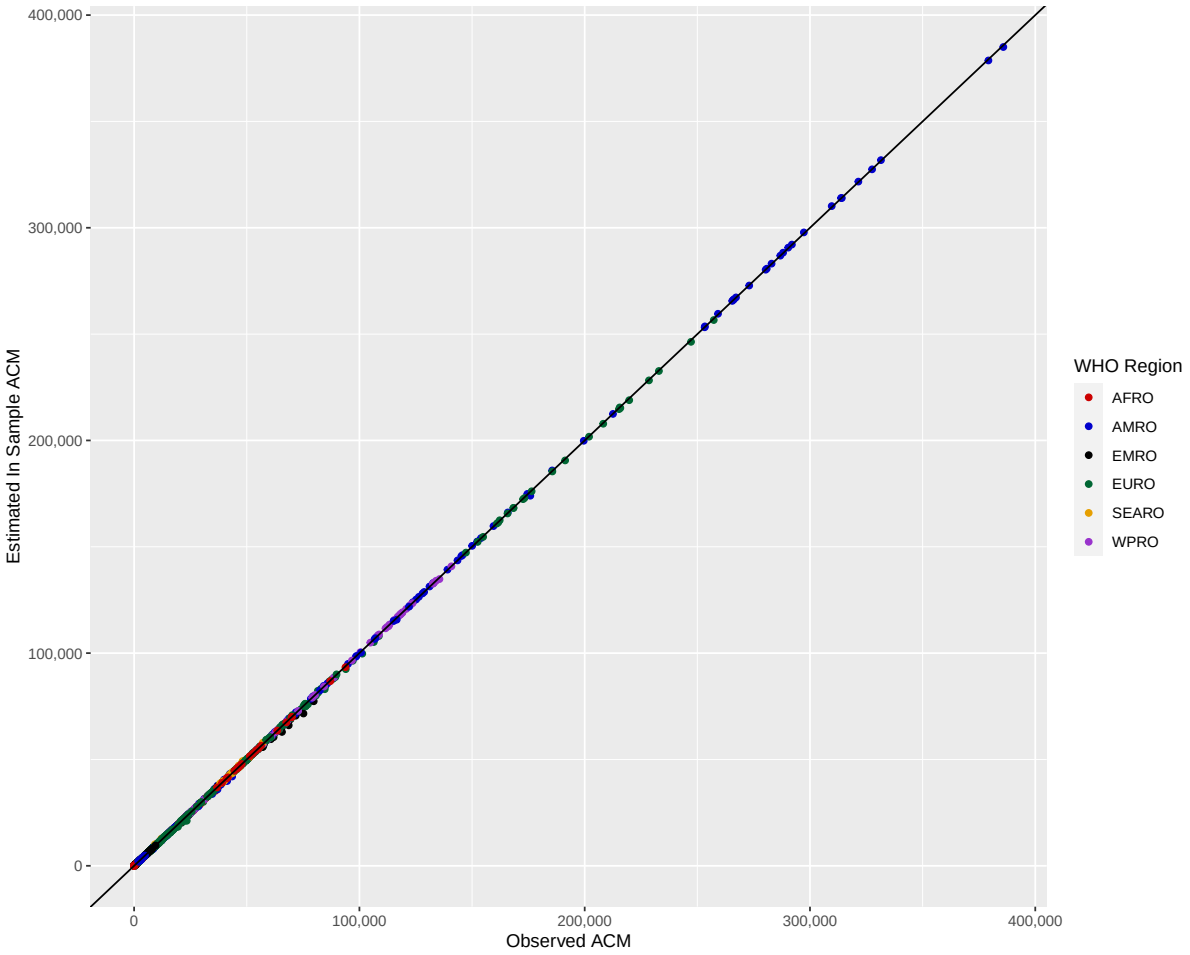


Figure 5: In-sample observed versus predicted, color-coded by region.

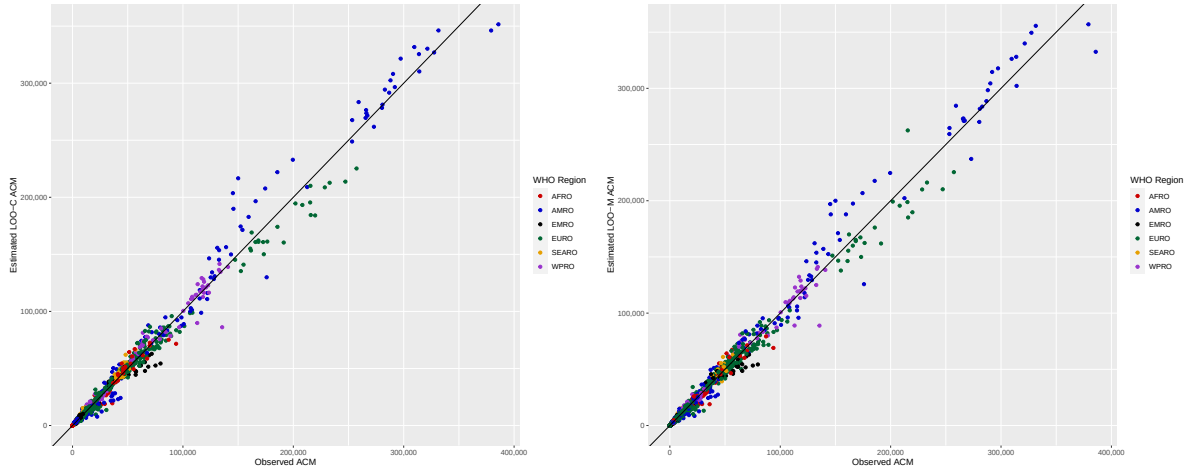


Figure 6: Out-of-sample observed versus predicted: Left: country removed. Right: month removed. Color-coded by region.

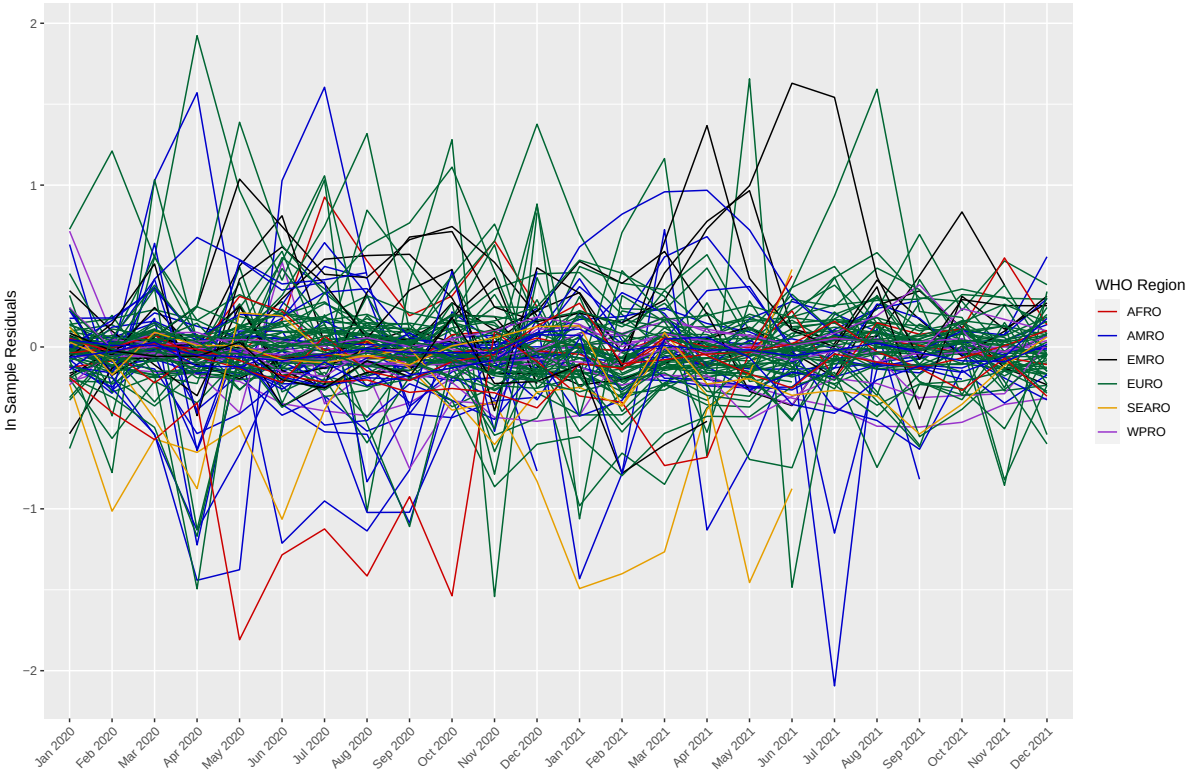


Figure 7: In-Sample standardized residuals over time, color-coded by region.

References

Adair, T. and Lopez, A. D. (2018). Estimating the completeness of death registration: An empirical method. *PLOS ONE*, **13**(5), e0197047.

Anand, S. and Sen, A. (1994). Human Development Index: Methodology and Measurement.

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

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**(10258), 1204–1222.

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**(10234), e81.

Likas, A., Vlassis, N., and Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, **36**(2), 451–461.

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. 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**(10001), 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., 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.

Singh, A., Yadav, A., and Rana, A. (2013). K-means with Three different Distance Metrics. *International Journal of Computer Applications*, **67**(10).

UNSD (2021). Demographic yearbook 2021.

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*. CRC press, second edition.