# Global excess deaths associated with the COVID-19 pandemic

**William Msemburi, PhD**
**Technical Officer**

*Open virtual meeting of the IAEG-SDGs, 15 June 2022*

World Health Organization

# Outline

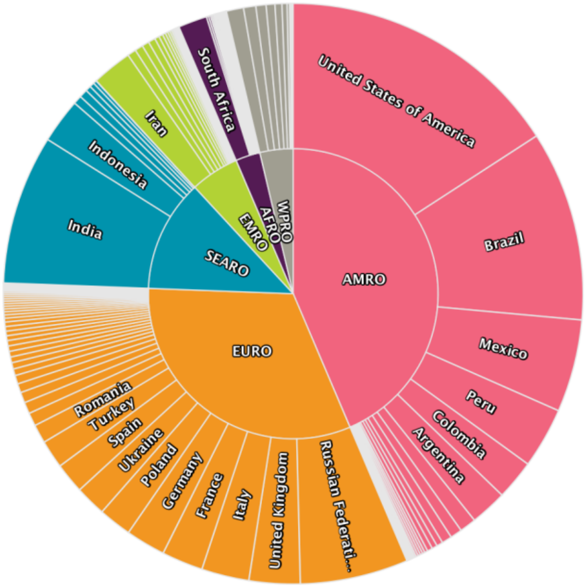
- **Background**
  - Reported COVID-19 death numbers
  - Why excess mortality?
  - Data availability to estimate excess
  - Technical advisory group
- **Model framework**
  - Expected all-cause deaths (if pandemic had not occurred)
  - Predicting all-cause deaths (during pandemic)
- **Overview of results**
- **Discussion**

# Overview of reported COVID-19 numbers

- As of June 10 2022, 532 million COVID-19 cases and 6.3 million deaths have been reported to WHO,
- Deaths concentrated in three regions, 44% Americas, 32% Europe and 13% South-East Asia (~89%)
- Upper middle income and high income economy deaths ~83% of total (44% and 39%, respectively)

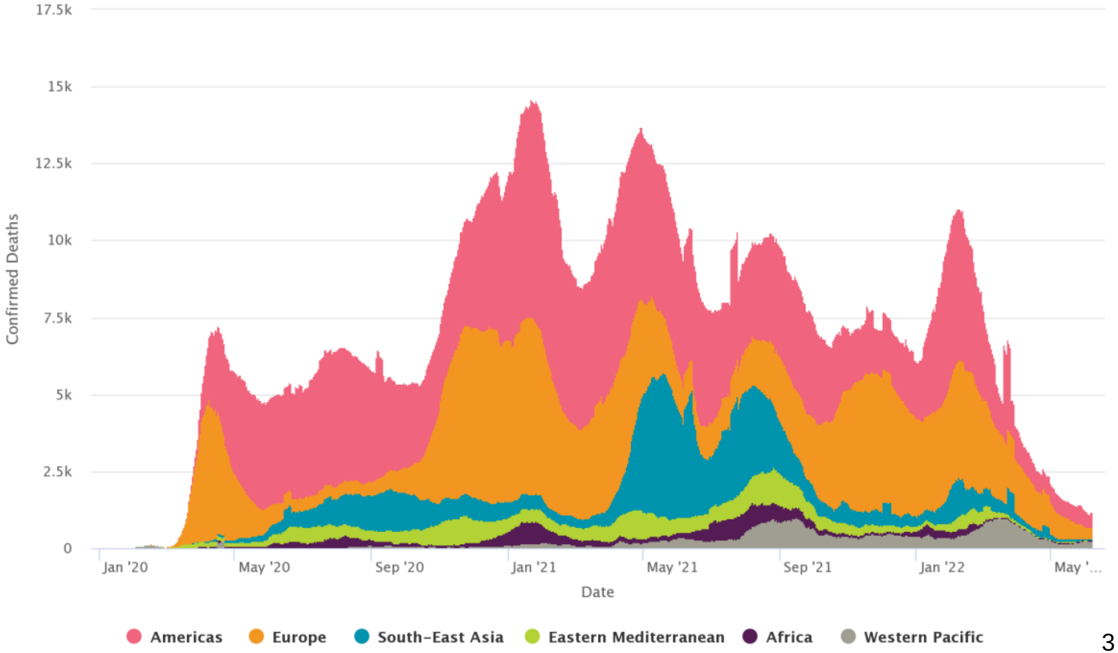
Confirmed COVID-19 deaths by country and region

N = 6.3 million




Confirmed COVID-19 deaths by date and region

N = 6.3 million (seven day rolling mean)

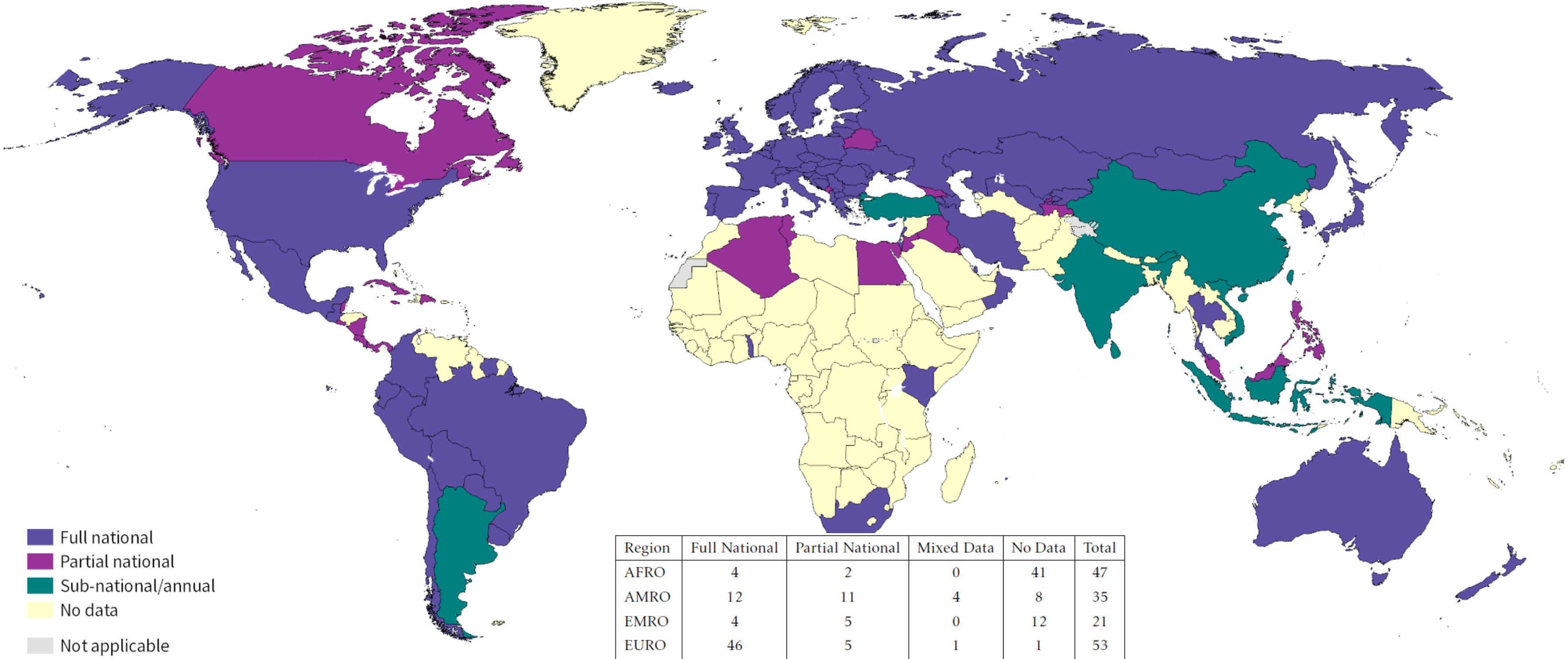



Data source: https://covid19.who.int/ accessed on June 10 2022

# Reported numbers do not provide a complete picture

- **Reported COVID-19 deaths under-estimate lives lost due to pandemic:**
  - miss unreported COVID deaths due to variations in testing access, diagnostic capacity and how COVID-19 deaths are defined,
  - miss increases in other deaths linked to conditions prevailing since the pandemic began, e.g., overwhelmed health systems/patients avoiding care.
- **Excess mortality is crucial for comprehensively quantifying impact:**
  - defined as change in all-cause mortality (ACM) for specified location and time period during a crisis, compared to expected,
  - encompasses deaths from all causes (direct, indirect and other), gives measure of ‘whole system’ impact.
- **Timely, reliable and complete data for estimating excess are limited:**
  - only a subset of countries have complete electronic medical certification and fully functioning CRVs,
  - observed lag of 12-18 months after the end of a calendar for countries to submit COD data to WHO,
  - Attributes such as sex and age are missing from data from many countries,

# Many countries lack the data required to track excess mortality directly:




| 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 |

# WHO & UN DESA established a TAG to tackle the challenges of counting COVID-19 deaths:

- https://www.who.int/data/technical-advisory-group/covid-19--mortality-assessment

- 33 Members and 30+ Observers

- Global and multidisciplinary expert representation: epidemiology, data science and analytics, statistics and biostatistics, demography, national governments, academia, policy makers.

- Establishing globally **standardized methods** and **statistical models** for estimating COVID-19 excess deaths for period January 2020 – December 2021

# Expected ACM for countries with historical time-series

- For each country $c$, let $Y\_ct$, represent the all-cause mortality (ACM) count for country $c$ month $t$. We assume:

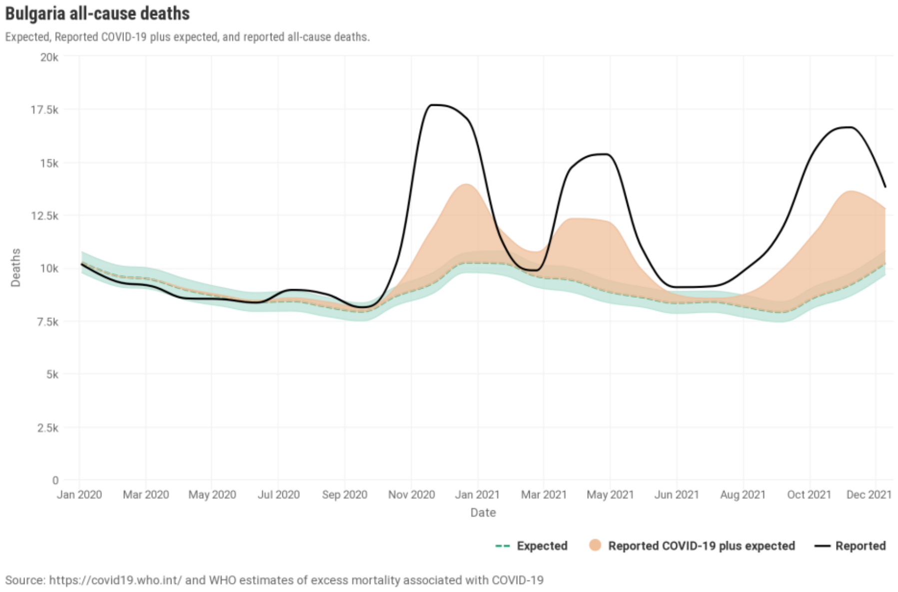
$$Y_{ct}|\mu_{ct} \sim \text{NegBin}(\mu_{ct}, \phi_c)$$

with mean $\mu_{ct}$ , variance $\mu_{ct}(1 + \mu_{ct}/\phi_c)$ and scale $\phi_c$

$$\log(\mu_{ct}) = f_{yc}^{Y}(v[t]) + f_c^{m}(m[t])$$

where $f_{yc}^{Y}(\cdot)$ models annual trend and $f_c^{m}(\cdot)$ accounts for within-year seasonal variation.

- The model is used to predict expected deaths $\mu_{ct}$ for all $t$ in years 2015 to 2021. Approximate simulations from the Bayesian posterior density are used to generate 95% credibility intervals of the expected deaths.



# Predicting ACM during pandemic

- For country $c$, we let:
  - $N_c$ be the population size in 2020/2021.
  - $Y_{ct}$ be the total deaths in month $t$ (observed in a subset of countries and assumed to be complete with good coverage), $t = 1, \ldots, 24$.
  - $E_{ct}$ be (modeled) deaths that were expected in 2020/2021, in the absence of Covid-19 (observed for all countries).
- We fit overdispersed Poisson Models of the form:

$$Y_{ct} | \theta_{ct} \sim \text{Poisson}(E_{ct}\theta_{ct})$$

which has $\text{E}[Y_{ct}|\theta_{ct}] = E_{ct}\theta_{ct}$ ,so that $\theta_{ct} > 0$ is a relative rate parameter.

  - If this parameter is $< 1$/$> 1$ then the death rate in country $c$ and in month $t$ is less than/greater than that expected, based on historical data.
- For a country $c$ that includes Poisson variation:

$$\underbrace{\Pr(Y_{ct}|\mathbf{y})}_{Predictive} = \int \underbrace{\Pr(Y_{ct}|\boldsymbol{\beta}, \epsilon_{ct})}_{Poisson} \times \underbrace{p(\boldsymbol{\beta}, \epsilon_{ct}|\mathbf{y})}_{Posterior} \, d\boldsymbol{\beta} d\epsilon_{ct}$$

# Relative rate regression model

– We have $\mathrm{E}[Y_{ct}|\theta_{ct}] = E_{ct}\theta_{ct}$ and model the log relative rate as:

$$\beta_0 + \sum_{b=1}^{B} \beta_{bt}^{V} X_{ctb} + \sum_{g=1}^{G} \beta_g^{C} Z_{cg} + \epsilon_{ct}$$

- $\beta_0$ is the intercept
- $\epsilon_{ct} \sim_{iid} \mathrm{N}(0, \sigma_\epsilon^2)$ is a random effect which allows for excess-Poisson variation around the covariate model.
- We have $B$ time-varying covariates, e.g.,~sqrt(C19 death rate), test positivity rate, stringency and containment measures, and we allow the associations with these variables, $\beta_{bt}^{V}$ to be time-varying, via a random walk of order 2 (RW2) prior.
- The $G$ time-invariant covariates, e.g.,~high income or not, human development index (HDI), diabetes prevalence, cardiovascular death rates, proportion of population younger than 15, proportion of population over 65, which have constant association parameters $\beta_g^C$
- We fit the model using INLA

# Subnational regression model

– Assuming $K$ regions contribute data in month $t$,

we have death counts $Y_{tk}$ ,$k = 1, \ldots, K_t$ ,we assume the model for month $t$ is,

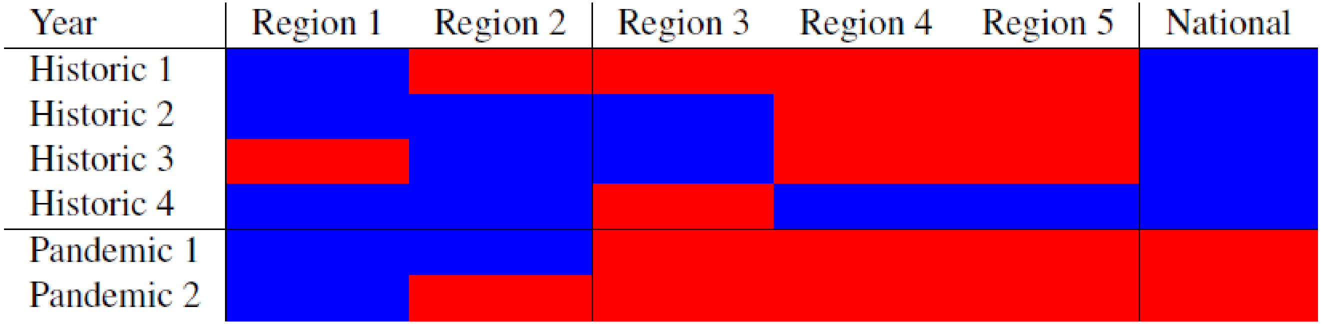
$$\mathbf{Y}_t | \mathbf{p}_t \sim \text{Multinomial}_{K_t+1}(Y_{t+}, \mathbf{p}_t),$$

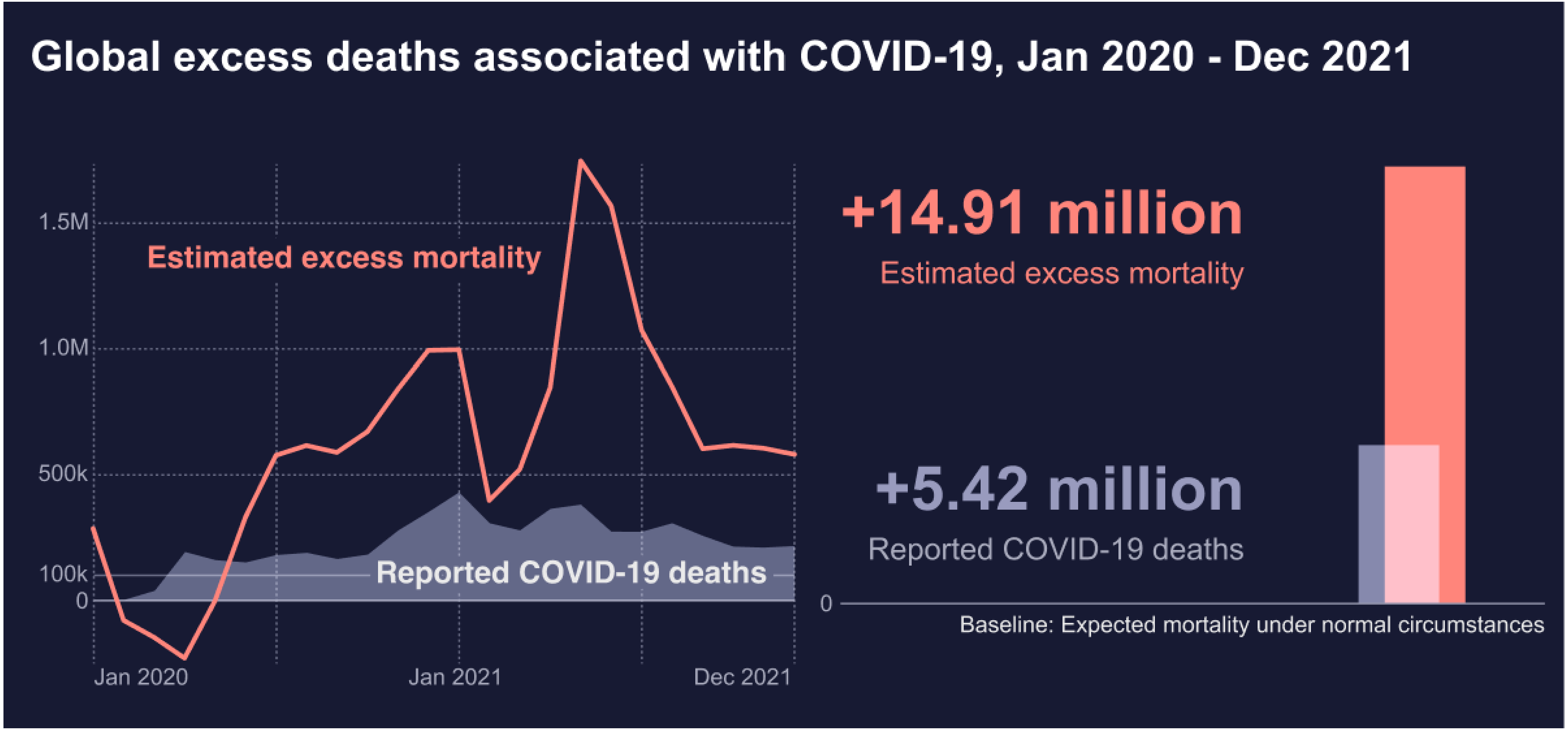
where

$$p_{t,k} = \text{Pr( death in region } k \text{ | period } t\text{, total)},$$

and

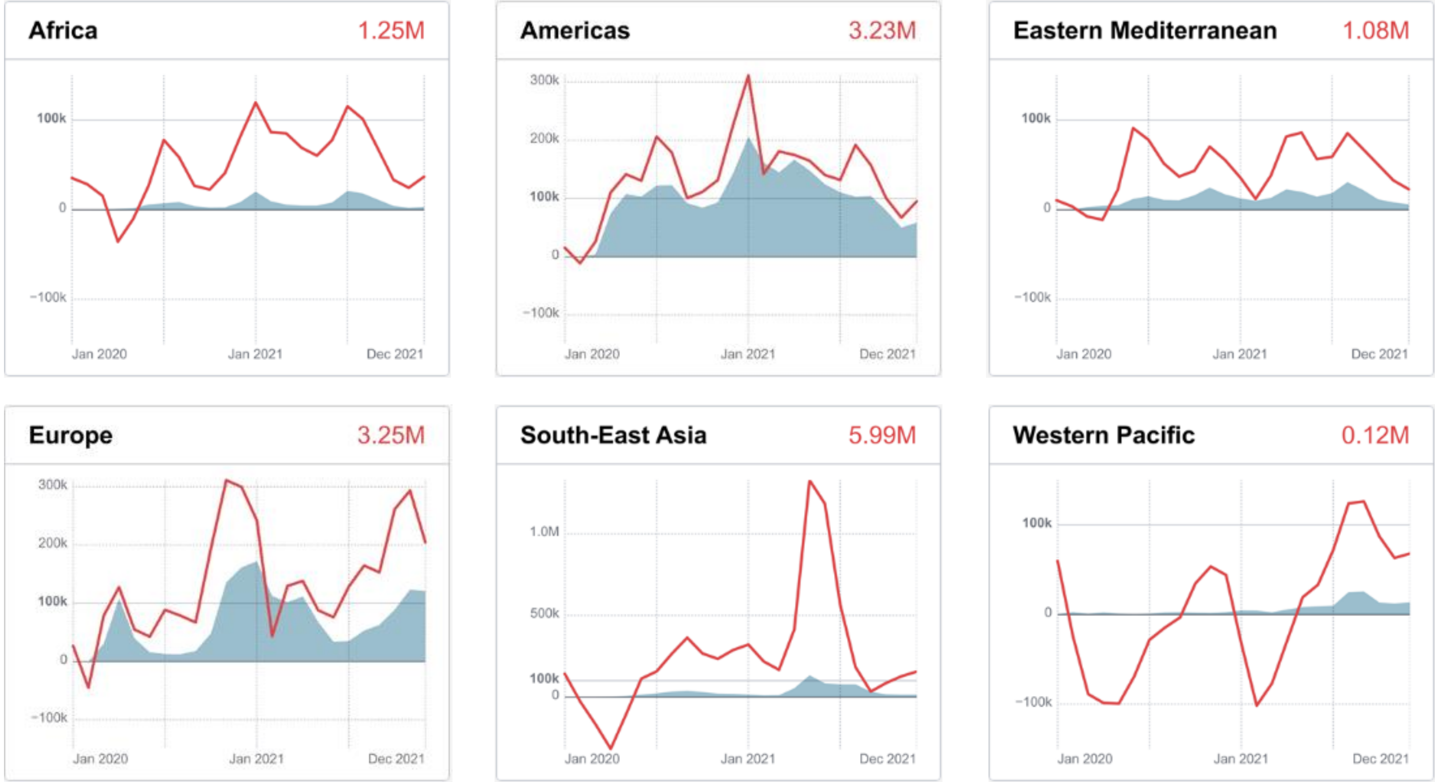
$$\log\left(\frac{p_{t,k}}{p_{t,K_t+1}}\right) = \alpha_k + \epsilon_t, \text{ with } \epsilon_t \sim \mathbf{N}(0, \sigma_\epsilon^2)$$



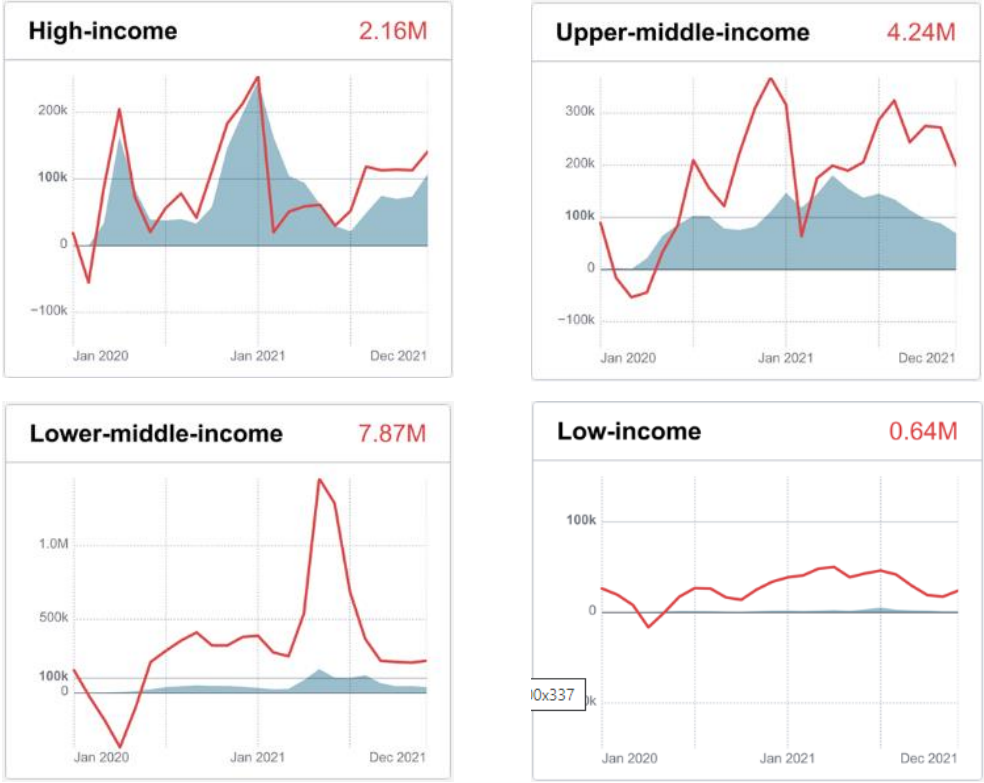



https://www.who.int/data/stories/global-excess-deaths-associated-with-covid-19-january-2020-december-2021

# Difference between reported deaths and excess mortality by WHO region



# Difference between reported deaths and excess mortality by income group



# Discussion

- Estimates can be viewed at https://worldhealthorg.shinyapps.io/covid19excess/

- This work is iterative, we will continue to update periodically as data become available. Future updates include extended assessment of impact by age and sex.

- Not a "one-size-fits-all" model. Uses all available data and balances comprehensiveness and comprehensibility i.e.,
  - Step 1: primary data used when available
  - Step 2: subnational representative primary data used to derive national estimates
  - Step 3: predict overall mortality conditional on expected mortality and contextually relevant variables

- Adherence to the **Guidelines for Accurate and Transparent Health Estimates Reporting** (GATHER)
  - Making available all input data, code and methods to enables Member States to replicate estimates as well as use country-specific input data for estimates

- There are limitations. Every model is an approximation of reality. And conditional on quality of inputs (accuracy of expected, completeness of ACM and consistency of covariates), generalizability of effects (in space and time) and sensitivity of periodicity.