Estimating the impact of school closures on COVID-19 disease burden

Supplementary Appendix

——— Add Authors List here ———

Contents

1 Model description

1.1 General approach

We use a semi-mechanistic compartmental model of COVID-19 transmission governed by ordinary differential equations (ODEs). Our model captures important factors of COVID-19 transmission and disease such as age-specific characteristics, heterogeneous mixing, vaccination and the emergence of different variants of concern. The ODE-based model is used to capture only states relevant to transmission, whereas hospitalisations and deaths are estimated through a convolution process based on the ODE-based model's outputs. This process combines the model-estimated disease incidence with statistical distributions modelling the time to hospitalisation, the hospital stay duration and the time to death. This approach presents two main advantages. First it reduces the complexity of the dynamic system relying on numerical solving of ODEs, which is computationally expensive. Second, the convolution approach allows for more flexibility and produces more realistic assumptions regarding the timings of hospitalisations and deaths, compared to what could be achieved with a simple compartmental approach. The following sections describe the model in details and Figure S1 summarises the overall approach used in our analysis.

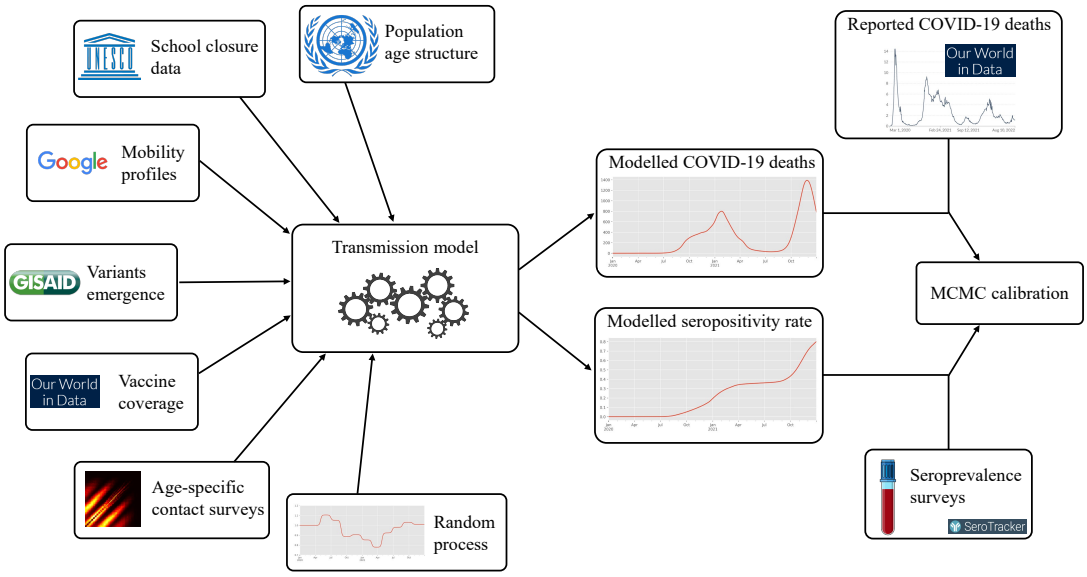

Figure S1: Conceptual approach used to build and calibrate the models

1.2 Transmission model

1.2.1 Compartment types and sequence

Model compartments represent sequential progressions through the processes of infection with, progression through, and recovery from the phases of SARS-CoV-2 infection and COVID-19 disease. The following types of compartments are implemented:

- Susceptible
  - Persons never previously infected with SARS-CoV-2 during or before the model simulation period
- Latent
  - Persons recently infected with SARS-CoV-2, but not in the active phase of the disease yet.
  - These individuals may still be infectious (see details in next paragraph).
- Active

  - Persons with active COVID-19 who are currently infectious.

- Recovered
  - Persons recovered from COVID-19 during the model simulation period
  - Reinfection from these compartments is permitted through exposure to a different strain than the one that most recently infected the individual (see strain stratification section for details).

The base model structure consists of a sequence of one susceptible compartment (S), four latent compartments (E_1, ..., E_4), four active disease compartments (I^1, ..., I^4) and one recovered compartment (R) (Figure S2).

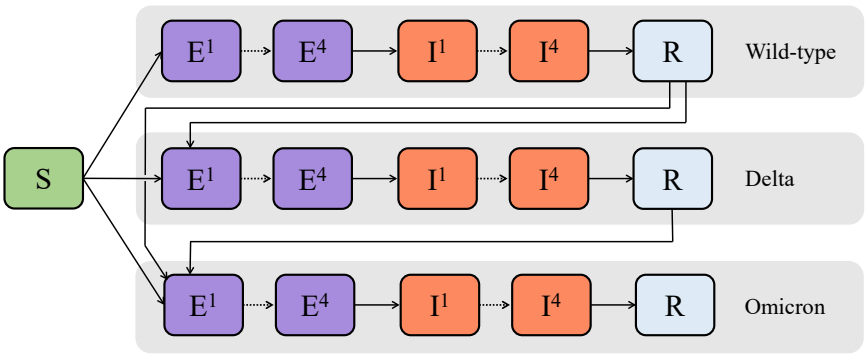

Figure S2: Compartmental model structure. S = Susceptible, E = Exposed / Latent, I = Active disease, R = Recovered. Stratification by age and vaccination status are not shown here.

The main rationale for using four serial compartments for both the latent and active states is to achieve an Erlang distribution for the time spent in each of these states. This distribution is more realistic than the exponential distribution that would have been associated with a single compartment, because the Erlang distribution does not have a large density mass around 0 and is not heavy-tailed. Figure S3 illustrates the modelled distributions of the incubation period and the active disease period.

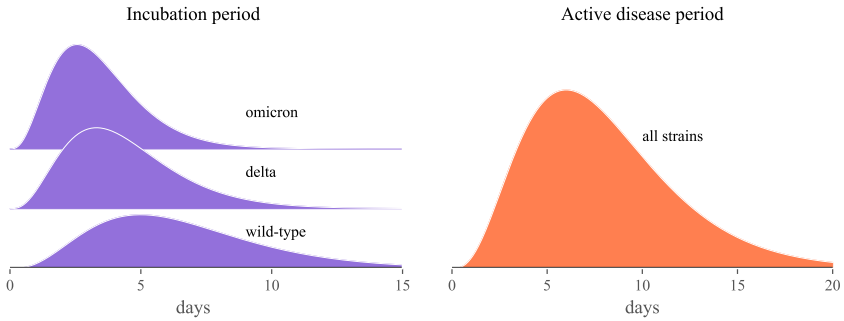

Figure S3: Modelled distributions of the incubation and active disease periods. Active disease period shown for an average duration of 8 days, but this parameter is varied during calibration.

The four active disease compartments all have exactly the same characteristics. However, the last two latent compartments (E^3 and E^4) are infectious whereas the first two (E^1 and E^2) are not. We further assume that the infectious latent compartments are half as infectious as the active disease compartments.

1.2.2 Model stratification by age

All compartments of the base compartmental structure are stratified by age into the following age bands: zero to 14 years / 15 to 24 years / 25 to 49 years / 50 to 69 years / 70 years and above.

The initial population is distributed between the different age bands to reflect the age distributions reported by the United Nations Population Division. Demographic processes, including births, ageing and non-infection-related deaths are not simulated, given the timeframes considered in this simulation.

We assume heterogeneous mixing between the different age groups to account for the assortative nature of social interactions by age (see Section 1.2.5). Age is assumed to affect:

- The susceptibility to infection
- The risk of COVID-19 hospitalisation
- The risk of COVID-19 death
- The vaccination rate (see Section 1.2.3)

1.2.3 Capturing the effects of vaccination

History of vaccination is captured by stratifying all model compartments by vaccination status. Two vaccination strata are included to represent those who have received at least two doses of a COVID-19 vaccine, and those who have not.

We use data from *Our World in Data* to inform the modelled dynamic vaccination coverage. In particular, we use the reported proportion of people "fully vaccinated" to specify the time-variant proportion of vaccinated people in our model. We assume that older individuals are vaccinated first by prioritising the modelled age groups in descending order. That is, the oldest age group receives all available vaccines until a saturation coverage of 80% is reached for this group. Then the next oldest category starts receiving vaccines and we repeat this process until all available vaccines are allocated. Note that in the event that the population-level vaccine coverage exceeds 80%, the saturation coverage is set equal to the population-level coverage.

Let us consider two successive time points t_i and t_{i+1} for which vaccination data is available. Let us denote $r_{a,i}$ and $r_{a,i+1}$ the associated vaccine coverage for age group a. The time-variant and age-specific vaccination rate per capita $w_a(t)$ verifies:

$$1 - r_{a,i+1} = (1 - r_{a,i})e^{-w_a(t)(t_{i+1}-t_i)} \quad ,\forall t \in [t_i, t_{i+1}). \tag{1}$$

Then,

$$w_a(t) = \frac{\ln(1 - r_{a,i}) - \ln(1 - r_{a,i+1})}{t_{i+1} - t_i} \quad ,\forall t \in [t_i, t_{i+1}), \tag{2}$$

where $\ln(x)$ represents the natural logarithm of x.

Figure S4 shows the modelled vaccination coverage over time against the reported data for the analysed countries.

The effect of vaccination on transmission is to partially reduce the rate of infection for all persons at-risk of infection in the vaccinated stratum. This includes both fully susceptible (never previously infected) persons, as well as recovered persons who are at risk of reinfection. The model allows for hybrid immunity in the sense that the vaccination-induced relative reduction of transmission risk is multiplied with that induced by previous infection. Vaccination is also assumed to reduce the risk of hospitalisation and death

Emerging variants of concern (VoCs) may partially escape vaccine-induced (as well as infection-induced) immunity, as described further below (see Section / Table XX).

1.2.4 Modelling multiple viral strains

The model is stratified by "strain" to simulate the emergence of multiple variants of concern (VoC). This approach explicitly represents multiple competing strains, each with a separate force of infection calculation. We assume that VoCs can have different levels of transmissibility, incubation period's duration and disease severity (hospitalisation and death risks) compared to the ancestral COVID-19 strain. In addition, VoCs are assumed to escape immunity partially for both vaccination- and infection-related immunity. The parameters used to represent the VOCs' characteristics are presented in Table XXX.

Seeding of each new strain into the model is achieved through the importation of a small number (10 per million population) of new infectious persons with the relevant strain into the model. The seeding

process is done over a ten-day period and the start of this period is set between 30 days before and 30 days after the emergence date as reported by GISAID Table XXX. The exact start date is automatically calibrated by the MCMC algorithm.

1.2.5 Dynamic social mixing

The model captures changes in social interactions over time through a dynamic age-specific mixing matrix. The following sections describe how this matrix is defined and how it captures the different non-pharmaceutical interventions implemented in the analysed countries, including school closures. The overall approach is also illustrated by Figure S6.

1.2.5.1 Reference mixing matrices

We extracted country-specific contact matrices using the *conmat* R package which derives social mixing matrices from contact survey data. These matrices provide the average numbers of contacts per day between different age groups, disaggregated by the following locations: home, school, work, other locations.

The overall contact matrix (before adjustments for mobility changes) results from the summation of the four location-specific contact matrices: $C_0 = C_H + m_S C_S + C_W + C_L$, where C_H, C_S, C_W and C_L are the age-specific contact matrices associated with households, schools, workplaces and other locations, respectively. Note that the school contribution C_S is multiplied by the factor m_S that is varied during model calibration (see Section 2), in order to include uncertainty around the relative contribution of school contacts to overall mixing.

1.2.5.2 Modifications of contact rates over time

To capture mobility changes over time, the contributions of the matrices C_S, C_W and C_L vary with time such that the input contact matrix can be written as:

$$C(t) = h(t)^2 C_H + s(t)^2 m_S C_S + w(t)^2 C_W + l(t)^2 C_L \tag{3}$$

The modifying functions h (for households), s (for schools), w (for work) and l (for other-locations) are each squared to capture the effect of the mobility changes on both the infector and the infectee in any given interaction that could potentially result in transmission.

School closure/re-opening

Reduced attendance at schools is represented through the function s, which represents the proportion of all school students currently attending on-site teaching. If schools are fully closed at time t, $s(t) = 0$ and C_S does not contribute to the overall mixing matrix $C(t)$. The function s is derived from the UNESCO database on school closures since the start of the COVID-19 pandemic (ADD REF HERE). This database provides school opening status over time as a categorical variable taking the following values: "Fully open", "Partially open", "Academic break", "Closed due to COVID-19". Table S1 indicates how the different categorical values are converted into the numerical function s.

UNESCO category	**Assumed proportion of students on-site ($s(t)$)**
Fully open	100%
Partially open	10-30%
Academic break	0%
Closed due to COVID-19	0%

Table S1: Assumed percentage of students on-site for the different UNESCO school closure categories.

We included uncertainty around the value associated with the partial closure category, as there is no quantitative data available to inform this parameter. The partial closure periods are likely to be periods where only a small fraction of students such as children of "essential workers" were attending school. We assume that between 10 and 30% of students attended on-site learning during these periods.

To model the counterfactual "no school closure" scenario, we assume that the schools were "Fully open" during the periods reported as "Partially open" or "Closed due to COVID-19".

Figure S5 summarises the UNESCO data on school closure for the analysed countries.

Dynamic mobility outside of schools and homes

Changes to people's mobility in places other than schools and homes are modelled using Google Mobility data. We use the "Workplace" category of the Google data to scale the work-related matrix contribution C_W to overall mixing over time, using the adjusting function w. The "other locations" matrix C_L is scaled through the adjusting function l which is defined as the average of the Google mobility indicators across the following Google categories: "Retail and recreation", "Grocery and pharmacy" and "Transit stations".

Household contacts

In the base case analysis, the contribution of household contacts to the overall mixing matrix is fixed over time (i.e. $h(t) = 1$ in Equation 3). Although Google provides mobility estimates for residential contacts, the nature of these data is different from that of each of the other Google mobility types. They represent the time spent in that location, as opposed to other categories, which measure a change in total visitors rather than the duration. The daily frequency with which people attend their residence is likely to be close to one and we considered that household members likely have a daily opportunity for infection with each other household member regardless of the background level of mobility. Therefore, we did not implement a function to scale the contribution of household contacts to the mixing matrix with time.

1.2.5.3 Sensitivity analyses around dynamic social mixing

In addition to the Base Case analysis described above, we performed two sensitivity analyses considering different assumptions for the modelled social mixing.

SA1: No Google mobility data In a first sensitivity analysis (SA1), we remove the contribution of the Google mobility data to the modelled social mixing (see Figure S6). In this configuration, the calibrated random process ($W(t)$) is expected to capture mobility changes implicitly.

SA2: School closures increase household contact rates In another sensitivity analysis (SA2), we consider an alternative assumption where the effective contact rates within households are increased during periods of school closure. In that case, the household component of the mixing matrix is modified by the following function:

$$h(t) = 1 + 0.20(1 - s(t)) \quad ,$$

where s is the function modifying school contacts as introduced in Section 1.2.5.2. This is equivalent to assuming that each individual has 20% more household contact potential when schools are fully closed.

1.2.6 Random transmission adjustment

The risk of SARS-CoV2 transmission per contact is adjusted by a time-variant random process, making the model semi-mechanistic. This random process reflects the fact that all the variations observed in the transmission risk in the real world cannot be explained solely by the factors that are included explicitly in the model such as vaccination, dynamic mobility or new variants' emergence. We therefore allow for random perturbations to the risk of transmission over time, although these perturbations are highly auto-correlated to avoid unrealistic changes over a short period of time.

We use a random walk with gaussian update defined by:

$$\begin{aligned} W(0) &= 0 \\ W(t+1) &\sim \mathcal{N}(W(t), \epsilon) \quad , \end{aligned} \tag{4}$$

where $\mathcal{N}$ denotes the normal distribution and where the standard deviation ϵ is automatically calibrated by the MCMC. The random process W is updated every two months and is transformed using the exponential function before being applied to the risk of transmission per contact (see Equation 6).

1.2.7 Ordinary differential equations

Let us first introduce some new notations. The different age groups are indicated by the subscript a and $\mathcal{A}$ represents the set of all modelled age groups (see Section 1.2.2). The vaccination status is represented

by the subscript v and $\mathcal{V}$ is the set of vaccination statuses (i.e. $\mathcal{V} = \{\text{“0”}, \text{“1”}\}$ where “0” represents unvaccinated people and “1” represents vaccinated people). The subscript s is used to represent the different viral strains and $\mathcal{S}$ is the set of all strains (i.e. $\mathcal{S} = \{\text{“wild-type”}, \text{“delta”}, \text{“omicron”}\}$). The average incubation period duration associated with strain s is denoted q_s and the average duration of active disease is denoted w. The relative susceptibility to infection of individuals aged a with vaccination status v is denoted $\sigma_{a,v}$. The term $b_{a,v,s}(t)$ is used to designate the imported individuals of age a and with vaccination status v that are infected with strain s (infection seeding). Vaccination is characterised by the age-specific and time-variant per-capita vaccination rate w_a. Finally, $\chi_{s,\sigma}$ represents the relative susceptibility to infection with strain σ for individuals whose most recent infection episode was with strain s. Using these new notations combined with those previously introduced, we can describe the model with the following set of ordinary differential equations:

$$
\begin{aligned}
\frac{dS_{a,v}}{dt} &= -\sum_{s\in\mathcal{S}} \lambda_{a,s}(t)\sigma_{a,v}S_{a,v} + \Phi_v\omega_a(t)S_{a,v=0} \quad , \\
\frac{dE^1_{a,v,s}}{dt} &= \lambda_{a,s}(t)\sigma_{a,v}\Big(S_{a,v} + \sum_{\sigma\in\mathcal{S}} \chi_{\sigma,s}R_{a,v,\sigma}\Big) - \frac{4}{q_s}E^1_{a,v,s} + \Phi_v\omega_a(t)E^1_{a,v=0,s} \quad , \\
\frac{dE^k_{a,v,s}}{dt} &= \frac{4}{q_s}E^{k-1}_{a,v,s} - \frac{4}{q_s}E^k_{a,v,s} + \Phi_v\omega_a(t)E^k_{a,v=0,s} \quad , \forall k \in \{2,3,4\}, \\
\frac{dI^1_{a,v,s}}{dt} &= \frac{4}{q_s}E^4_{a,v,s} - \frac{4}{w}I^1_{a,v,s} + b_{a,v,s}(t) + \Phi_v\omega_a(t)I^1_{a,v=0,s} \quad , \\
\frac{dI^k_{a,v,s}}{dt} &= \frac{4}{w}I^{k-1}_{a,v,s} - \frac{4}{w}I^k_{a,v,s} + \Phi_v\omega_a(t)I^k_{a,v=0,s} \quad , \forall k \in \{2,3,4\}, \\
\frac{dR_{a,v,s}}{dt} &= \frac{4}{w}I^4_{a,v,s} - \sum_{\sigma\in\mathcal{S}} \lambda_{a,\sigma}(t)\sigma_{a,v}\chi_{s,\sigma}R_{a,v,s} + \Phi_v\omega_a(t)R_{a,v=0,s} \quad ,
\end{aligned}
\tag{5}
$$

where $\lambda_{a,s}$ represents the force of infection of strain s affecting individuals of age a. The quantity Φ_v is a binary variable used to switch between plus and minus signs depending on the vaccination status. It is equal to 1 when $v =$“1” and -1 when $v =$“0”. In other words, $\Phi_v = 2\mathbb{1}_{v=\text{“1”}} - 1$.

The force of infection is calculated as:

$$
\lambda_{a,s}(t) = \beta e^{W(t)}\rho_s \sum_{\alpha\in\mathcal{A}}\sum_{v\in\mathcal{V}} c_{a,\alpha}(t)\Big(0.5\sum_{k=3}^{4} E^k_{\alpha,v,s} + \sum_{k=1}^{4} I^k_{\alpha,v,s}\Big) \quad . \tag{6}
$$

In the previous equation, β represents the unadjusted risk of transmission per contact and ρ_s is the relative transmissibility of strain s.

1.3 Estimation of COVID-19-related hospital pressure and deaths

The transmission model described in Section 1.2 provides estimates of COVID-19 incidence over time, disaggregated by age, vaccination status and strain. We combine these incidence estimates with the age-, vaccination- and strain-specific risks of hospitalisation and deaths as well as statistical distributions of time to events to compute COVID-19-related hospital pressure and deaths over time.

1.3.1 COVID-19-related hospital pressure

The risk of hospitalisation given infection is expected to vary dramatically by setting. For example, different countries may have different criteria for whether or not a COVID-19 patient should be admitted to a hospital. This makes it difficult to provide accurate estimates of hospitalisation rates for multiple countries.

This is why we introduce a universal indicator named “hospital pressure” in our analysis. This indicator is obtained by considering the age-specific risk of hospitalisation given infection observed in the first year of the pandemic in the Netherlands, adjusted for vaccination status and for the infecting strain (See Table XX). The “hospital pressure” indicator can therefore be interpreted as the level of hospital occupancy that would be observed in the analysed country if the rates of hospitalisation given infection in

this country were the same as in the Netherlands. This indicator is expected to be roughly proportional to the actual hospital occupancy level of the studied country. Note that this indicator is used in order to make comparisons between scenarios such that one should interpret the relative differences between scenarios rather than the absolute values of the indicator.

Let us denote $i_{a,v,s}(t)$ the number of new disease episodes estimated to start at time t for people aged a with vaccination status v and infected with strain s. The number of new hospital admissions occurring at time t is calculated using the following convolution product:

$$\eta(t) = \sum_{a,v,s} \kappa_{a,v,s} \int_{u \geq 0} i_{a,v,s}(t-u) g_h(u) du \quad , \tag{7}$$

where $\kappa_{a,v,s}$ is the risk of hospitalisation given infection for age a, vaccination status v and strain s based on the Netherlands data, and g_h is the probability density function of the statistical distribution chosen to represent the time from symptom onset to hospitalisation (See Table XX).

We then compute the "hospital pressure" quantity h, which is an indicator of hospital occupancy level, by combining the number of new hospital admissions η with the statistical distribution used to model hospital stay duration:

$$h(t) = \int_{u \geq 0} \eta(t-u)(1-\tau(u)) du \quad , \tag{8}$$

where τ is the cumulative density function of the statistical distribution chosen to represent the hospital stay duration (See Table XX).

1.3.2 COVID-19 deaths

We estimate the number of COVID-19 deaths over time using a similar approach as for the hospital pressure indicator. We use the age-specific infection fatality rates reported in ODriscoll et al. [2], adjusted for vaccination status and for the infecting strain to estimate COVID-19 mortality. Using the same notations as in Section 1.3.1, the number of COVID-19 deaths observed at time t is obtained by:

$$\mu(t) = m_C \sum_{a,v,s} ifr_{a,v,s} \int_{u \geq 0} i_{a,v,s}(t-u) g_d(u) du \quad , \tag{9}$$

where $ifr_{a,v,s}$ is the risk of death given infection for age a, vaccination status v and strain s, and g_d is the probability density function of the statistical distribution chosen to represent the time from symptom onset to death (See Table XX). We use a country-specific adjuster m_C to capture the fact that the infection fatality ratio is expected to vary by country, in part due to differences in COVID-19 death definition and reporting standards. This adjustment is automatically calibrated by the MCMC (Section 2).

2 Model calibration and uncertainty propagation

REVISE THE SECTION BELOW WHEN FINALISED The model was calibrated using a Bayesian approach. In particular, we used the adaptive Metropolis algorithm introduced by Haario *et al.* to sample parameters from their posterior distributions [1]. For each country, we ran 8 independent Metropolis chains initialised using Latin Hypercube Sampling based on the parameter priors. We ran simulations for 1 hour per chain yielding at least 60,000 iterations per chain. We discarded the first 30,000 iterations of each chain as burn-in and combined the samples of the 8 chains to project epidemic trajectories over time. For each modelled country, the epidemic projections presented in our analysis are associated with 1000 parameter sets randomly sampled from the posterior distributions obtained from MCMC sampling. The definitions of the prior distributions and the likelihood are detailed in the following sections.

2.1 Parameters varied during calibration

The parameters varied during calibration along with their associated prior distributions are listed in Table (add ref to table here) and indicated with the superscript ^{c}. We used uniform prior distributions

for all calibrated parameters. The primary parameters varied during calibration are the unadjusted risk of transmission per contact (β), the IFR multiplier (m_C), the infection seeding times of each strain and the proportion of students on-site during "Partially open" periods. Note that the values of the random process W_t described in Section 1.2.6 are also treated like calibrated parameters by the MCMC. The Gaussian auto-regressive component described in Equation 4 is incorporated in the posterior likelihood computation (Section 2.3).

2.2 Calibration targets

For each country model calibration is achieved by targeting two COVID-19 burden indicators: the reported number of COVID-19 deaths over time and country-level seroprevalence estimates.

We used the daily number of COVID-19 deaths reported by Our World in Data and applied a 7-day moving average to the observed data. We used the online platform SeroTracker to extract country-specific seroprevalence estimates. Use Angus' description here.

2.3 Likelihood definition

Let d_w denote the rounded average daily number of COVID-19 deaths during week w, and $\hat{d_w}^{\theta}$ the associated predicted number of deaths according to the model with parameter set θ.

Let us denote m the sample size associated with the seroprevalence survey selected from SeroTracker (Section 2.2), and k the number of seropositive individuals observed in the survey. Let $\hat{\pi}^{\theta}$ denote the modelled proportion ever infected by the time the survey was conducted (using midpoint date) associated with the parameter set θ. The likelihood was defined as follows:

$$\mathcal{L}(\theta) := f_{m,\hat{\pi}^{\theta}}(k) \times \prod_{w} g_r(d_w | \hat{d_w}^{\theta}) \quad , \tag{10}$$

where $f_{n,p}(.)$ is the probability mass function of a binomial distribution $\mathcal{B}(n,p)$, and $g_r(.|\mu)$ is the probability mass function of a negative binomial distribution with mean μ and overdispersion parameter r. The parameter r is automatically estimated by the MCMC algorithm.

The likelihood described above represents the goodness of fit of a particular model parameterisation with regards to the targeted data. This quantity needs to be adjusted for the prior likelihood of the parameter set in order to compute the MCMC acceptance quantity $\mathcal{Q}(\theta)$. As we used uniform priors for all the parameters, the inclusion of the individual parameters' priors in the acceptance quantity is not necessary. Indeed, their respective contributions would cancel out as the same quantity would appear in the numerator and the denominator of the MCMC acceptance quantity ratio. However, the auto-regressive relationship described in Equation 4 must be accounted for as part of the combined prior likelihood of a parameter set. This is what prevents unrealistic fluctuations of the random process. If W^{θ} represents the random process associated with the parameter set θ, the overall MCMC acceptance quantity is obtained by:

$$\begin{aligned} \mathcal{Q}(\theta) &= \mathcal{L}(\theta) \times \prod_{i=1}^{n} z_{W_{i-1}^{\theta},\sigma}(W_i^{\theta}) \\ &= f_{m,\hat{\pi}^{\theta}}(k) \times \prod_{w} g_r(d_w | \hat{d_w}^{\theta}) \times \prod_{i=1}^{n} z_{W_{i-1}^{\theta},\sigma}(W_i^{\theta}) \quad , \end{aligned} \tag{11}$$

where $z_{\mu,\sigma}(.)$ represents the probability density function of the normal distribution $\mathcal{N}(\mu,\sigma)$, and n is the number of random process updates. The standard deviation σ is automatically estimated by the MCMC.

2.4 Software implementation

We use the summer2 Python package to implement the model. This is a domain specific library for compartmental epidemiological models that addresses a few of the key concerns as follows.

2.4.1 Application Programming Interface (API)

The model specification is done via a simple yet expressive Python API, while the numerical implementation is largely autogenerated by the summer2 package at runtime. This specification is composable via stratification classes and other reusable components (maybe list some specific summer features with link to code), thus the complexity of the software is kept to a minimum, reducing cognitive overhead for the modeller, and greatly reducing the possibility for error.

2.4.2 Optimising compiler

The summer2 package uses the jax library as its computational backend, meaning that while the specification of models is done largely in Python, the model execution itself is transformed via an optimising compiler into fast native code. This brings the model runtime from several seconds (for a naive implementation) to under 50ms per iteration.

References

[1] Heikki Haario, Eero Saksman, and Johanna Tamminen. "An adaptive Metropolis algorithm". In: *Bernoulli* 7.2 (2001).

[2] M. O'Driscoll et al. "Age-specific mortality and immunity patterns of SARS-CoV-2". In: *Nature* 590.7844 (2021), pp. 140–145. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2918-0. URL: https://www.ncbi.nlm.nih.gov/pubmed/33137809.

Figure S4: Modelled vaccine coverage (lines) against data (dots).

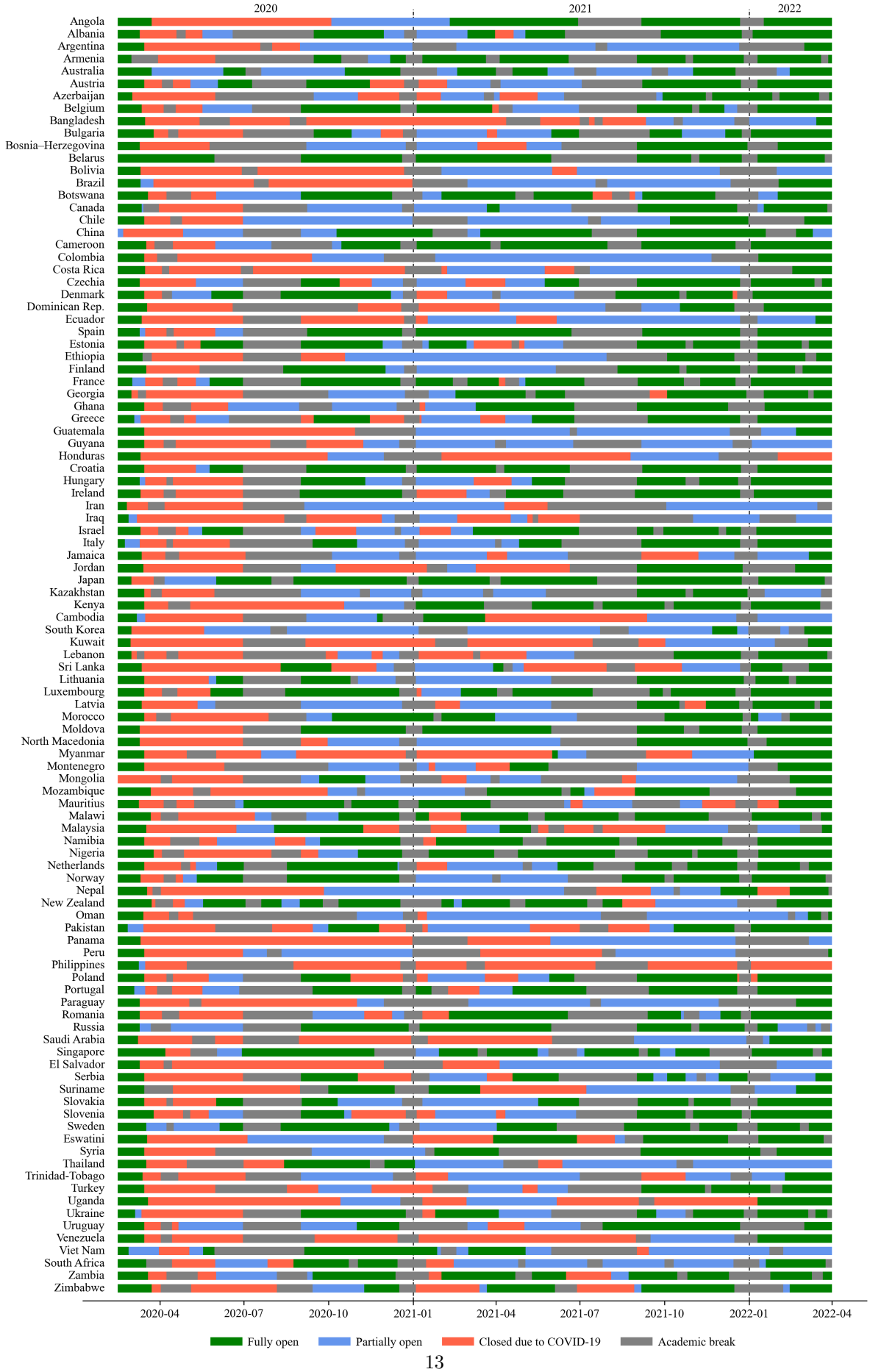

Figure S5: UNESCO data on school closures during the COVID-19 pandemic.

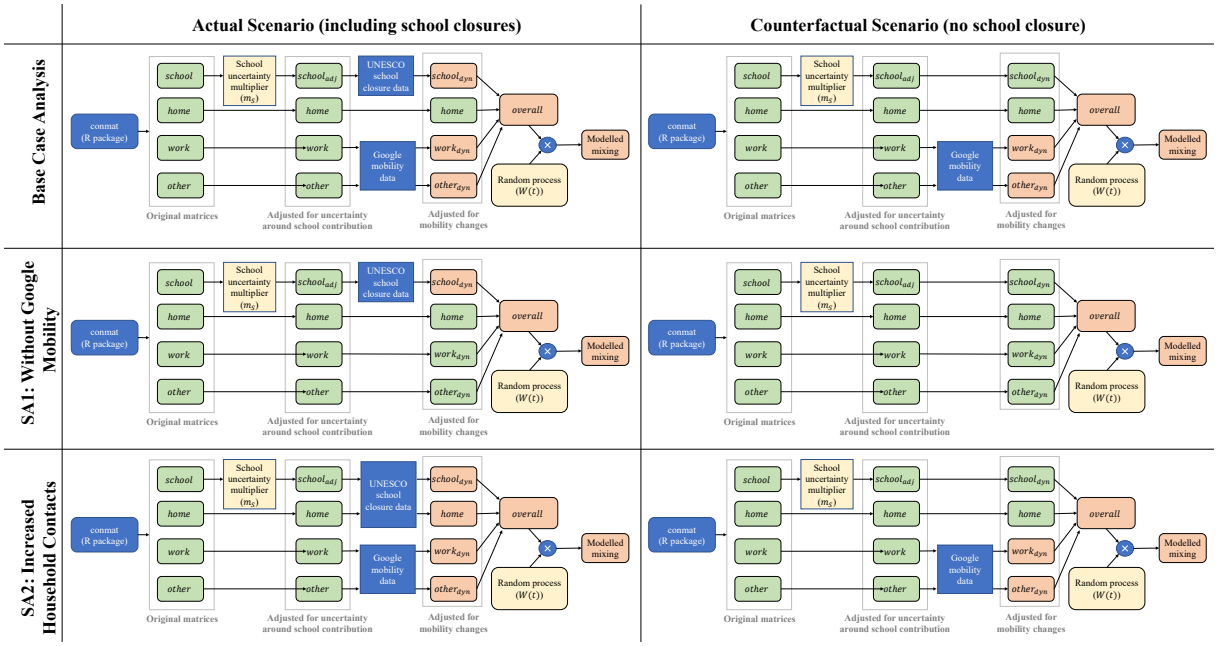

Figure S6: Modelled social mixing under Base Case and Sensitivity Analyses.

country	sampling start date	sampling end date	age min	age max	denom. value	serum pos prevalence	estimate grade	overall risk of bias
Australia	2020-11-03	2021-03-12		19	1685	0.0023	National	High
Austria	2020-06-05	2020-12-04	18	72	20228	0.025	National	High
Bangladesh	2020-10-15	2021-02-15	10		3220	0.673	National	Moderate
Belgium	2020-05-18	2020-05-25		101	3242	0.062	National	Low
Brazil	2020-05-14	2020-06-23			89362	0.023	National	Moderate
Burkina Faso	2020-08-22	2020-11-19	5		998	0.032	National	High
Canada	2020-05-09	2020-07-21	17		74642	0.007	National	Moderate
Chile	2020-09-25	2020-11-25	7	94	2493	0.104	National	Low
Colombia	2020-09-21	2020-12-11	5	80	17863	0.3253	National	Moderate
Croatia	2020-12-15	2021-02-15			1436	0.251	National	High
Czechia	2021-02-01	2021-03-31	18		19548	0.51	National	High
Denmark	2020-08-15	2020-09-30	12		11478	0.02	National	Low
Ecuador	2020-10-12	2020-10-19			1250	0.1168	National	Moderate
Estonia	2021-02-08	2021-03-25			2517	0.201	National	High
Ethiopia	2020-06-24	2020-07-08	15		16932	0.035	National	Low
France	2020-05-11	2020-05-17			3592	0.0493	National	Low
Germany	2020-10-06	2021-02-28	18	99	14781	0.013	National	Moderate
Honduras	2020-06-16	2020-06-23	5		792	0.062	National	Moderate
Hungary	2020-05-01	2020-05-16	14		10474	0.0068	National	Low
India	2020-08-18	2020-09-20	10		29082	0.066	National	Low
Iran	2020-08-03	2020-10-31	6	109	11256	0.142	National	Low
Israel	2020-06-28	2020-09-14			54357	0.046	National	Moderate
Italy	2020-05-25	2020-07-15			64660	0.025	National	Moderate
Japan	2020-06-01	2020-06-07	20		7950	0.001	National	Moderate
Jordan	2020-10-01	2020-10-31			5470	0.07	National	Moderate
Kenya	2021-01-03	2021-03-15	16	64	3018	0.485	National	Moderate
Lebanon	2020-12-07	2021-01-15			2058	0.185	National	Low
Lithuania	2020-08-10	2020-09-10	18	92	3089	0.014	National	Moderate
Malawi	2020-10-14	2020-12-08			4261	0.078	National	Low
Mexico	2020-08-15	2020-11-15	3	12	944	0.187	National	Low
Mongolia	2020-10-13	2020-12-04			5000	0.0136	National	Low
Nepal	2020-10-09	2020-10-22			3040	0.144	National	Low
Norway	2020-12-27	2021-02-13			1912	0.032	National	Moderate
Oman	2020-09-13	2020-09-24	5		4780	0.164	National	Moderate
Pakistan	2020-07-15	2020-07-31			15390	0.424	National	Moderate
Portugal	2021-03-01	2021-03-17			2435	0.173	National	Low
Qatar	2020-05-12	2020-09-09			112941	0.148	National	Moderate
Senegal	2020-06-15	2020-10-15			3231	0.204	National	High
Sierra Leone	2021-03-01	2021-03-31			1893	0.028	National	Low
Singapore	2020-11-15	2020-12-15	23	83	937	0.0016	National	High
Slovenia	2020-10-17	2020-11-10		99	1211	0.0429	National	Low
South Africa	2020-11-15	2021-04-15			7577	0.4523	National	Moderate
Rep. of Korea	2020-09-24	2020-12-09	18	86	4085	0.0039	National	Moderate
Spain	2020-06-08	2020-06-22			62167	0.052	National	Low
Sweden	2020-11-23	2020-12-04			3183	0.07	National	Moderate
Switzerland	2020-10-18	2021-02-15	18	64	2393	0.173	National	Moderate
USA	2020-07-01	2020-07-31	18		28503	0.093	National	Low
UK	2020-06-01	2020-06-28	17	69	10318	0.072	National	Low
Uzbekistan	2020-08-27	2020-09-11			86879	0.231	National	Moderate
Zambia	2020-07-04	2020-07-27			2704	0.021	National	Low

Table S2: Seroprevalence data extracted from SeroTracker (national surveys).

country	sampling start date	sampling end date	age min	age max	denom. value	serum pos prevalence	estimate grade	overall risk of bias
Albania	2020-12-21	2020-12-28	20	70	815	0.482	Local	Moderate
Argentina	2020-12-15	2021-01-15	2		3225	0.1668	Local	Moderate
Benin	2021-03-03	2021-03-15			1405	0.2977	Local	Low
Bosnia and Herzegovina	2020-12-01	2021-01-15	1	88	1855	0.4027	Regional	Moderate
Cameroon	2020-10-14	2020-11-26	5	80	971	0.292	Local	Low
China	2020-05-25	2020-07-08	20		14765	0.0005	Local	Moderate
Congo (DRC)	2021-03-06	2021-03-14			2476	0.722	Local	Low
Egypt	2020-05-05	2020-10-31			2927	0.2982	Local	High
Gabon	2020-07-15	2020-10-15		78	1495	0.362	Local	High
Ghana	2020-11-01	2020-11-30			1346	0.396	Local	Low
Greece	2021-03-01	2021-03-22			764	0.116	Local	High
Indonesia	2020-06-15	2020-12-15	16		1819	0.114	Regional	Moderate
Iraq	2021-01-10	2021-01-30	16		743	0.6258	Local	Moderate
Ireland	2020-06-22	2020-07-16	12	69	913	0.0312	Local	Low
Kazakhstan	2020-10-24	2021-01-11	5		4461	0.57	Local	Low
Malaysia	2020-12-15	2021-04-15			653	0.03	Regional	Moderate
Maldives	2020-10-15	2020-11-16			1940	0.134	Local	Moderate
Mali	2020-09-01	2020-09-30	1	100	1327	0.164	Local	Low
Mozambique	2020-11-02	2020-11-17			9756	0.014	Local	Moderate
Netherlands	2020-06-24	2020-10-09	18	70	2483	0.0906	Local	Moderate
Nigeria	2020-12-08	2020-12-15	1		3142	0.1607	Regional	Moderate
North Macedonia	2020-11-15	2021-03-15			9773	0.25	Local	High
Palestinian territories	2020-06-15	2020-06-30	15		1355	0.0	Local	Low
Paraguay	2020-12-22	2021-03-15			1699	0.269	Local	Moderate
Peru	2020-06-28	2020-07-09			3212	0.208	Local	Low
Philippines	2020-05-29	2021-03-23	1		615	0.3935	Local	High
Poland	2020-10-15	2020-11-15	1	94	1167	0.114	Local	Moderate
Romania	2020-07-08	2020-09-01	18	65	2115	0.0151	Local	Moderate
Russian Federation	2020-10-12	2020-12-06			1322	0.229	Local	Low
Saudi Arabia	2020-08-01	2020-08-31			5629	0.502	Local	Moderate
South Sudan	2020-08-10	2020-09-11			2214	0.383	Local	Low
United Arab Emirates	2020-07-19	2020-08-14			8831	0.104	Regional	Moderate
Zimbabwe	2021-02-10	2021-04-17		100	1530	0.53	Local	Moderate

Table S3: Seroprevalence data extracted from SeroTracker (subnational surveys).