Bayesian inference across multiple models suggests a strong increase in lethality of COVID-19 in late 2020 in the UK

Patrick Pietzonka[1,*], Erik Brorson[2], William Bankes[3], Michael E. Cates[1], Robert L. Jack[1,4], and Ronojoy Adhikari[1]

[1]*DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*
[2]*Quantitative Research, JPMorgan Chase & Co., 25 Bank Street, Canary Wharf, London E14 5JP, United Kingdom*
[3]*Applied Machine Learning and Artificial Intelligence, JPMorgan Chase & Co., 25 Bank Street, Canary Wharf, London E14 5JP, United Kingdom*
[4]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
[*]*Corresponding author: pietzonka@damtp.cam.ac.uk*

10 March 2021

We apply Bayesian inference methods to a suite of distinct compartmental models of generalised SEIR type, in which diagnosis and quarantine are included via extra compartments. We investigate the evidence for a change in lethality of COVID-19 in late autumn 2020 in the UK, using age-structured, weekly national aggregate data for cases and mortalities. Models that allow a (step-like or graded) change in infection fatality rate (IFR) have consistently higher model evidence than those without. Moreover, they all infer a close to two-fold increase in IFR. This value lies well above most previously available estimates. However, the same models consistently infer that, most probably, the increase in IFR *preceded* the time window during which new variant B.1.1.7 became the dominant strain in the UK. Therefore, according to our models, the caseload and mortality data do not offer unequivocal evidence for higher lethality *of a new variant*. We compare these results for the UK with similar models for Germany and France, which also show increases in inferred IFR during the same period, despite the even later arrival of new variants in those countries. We argue that while the new variant(s) may be one contributing cause of a large increase in IFR in the UK in autumn 2020, other factors, such as seasonality, or pressure on health services, are likely to also have contributed.

Summary

Model structure. We study a family of non-geographic compartmented models structured by age, and by epidemiological and diagnostic state. They include stages within exposed (E) and infected (I) classes, and separation of symptomatic and asymptomatic infected individuals; E and I classes are subdivided according to whether they have had a positive test.

Disease transmission between compartments is modelled by contact matrices derived from survey data and previous Bayesian inference [1]. Non-pharmaceutical interventions are incorporated via time- and age-dependent transmission rates. This dependence is parameterised by functions that account for lockdowns, seasonal changes, etc., with parameters inferred from the data except where known a priori (e.g. lockdown dates).

The process of diagnosis is modelled explicitly through the subdivided compartments. The targeting of tests towards symptomatic individuals is controlled by a single inferred parameter. Tested individuals have reduced infectivity, as a model for self-isolation. The fraction of asymptomatic cases is presumed fixed. Most other parameters are inferred (including initial conditions in March 2020). Our models are stochastic at all stages, including the infection and the diagnosis processes.

Similar models are considered for Germany and France. All models are analysed with the same Bayesian methodology as outlined below.

Data. We analyse age-structured, weekly data for cases and mortalities for the UK, France and Germany starting in early March 2020. Our explicit modelling of diagnostic tests allows us to directly use data on the number of daily tests performed and their outcomes to inform parameter estimation. For the UK, the data we used are (i) ONS data for COVID deaths; (ii) PHE data for total numbers of tests; (iii) PHE data for numbers of positive tests. The time series analysed extends from March 2020 until January 2021.

Methods. In line with the data used, we treat diagnosed cases and mortalities as the only observables; other compartments are treated as unobserved, with occupancies to be inferred from data. Parameters governing overdispersion for infection and death are likewise inferred.

We perform inference using a likelihood function that includes the stochasticity in all model transitions (infection, disease progression, testing, death). The likelihood of the observed data is computed in an approximation that becomes exact for large population sizes, taking into account all statistical dependencies in this limit. Inference is performed at the level of maximum a posteriori (MAP) estimates. Using these MAP estimates we approximate the evidence for each model, allowing us to compare the credibility of different models given the same set of data.

We use the software package PyRoss to build all the models and run the inference procedure [2, 3].

Model specifications. We compare several distinct models that differ primarily in how interventions and fluctuations are addressed. Some of these variants (Type 1) allow for a time-dependent infection fatality rate (IFR) and some do not (Type 0). Most of our Type 1 models assume a step

change in IFR but a ramped (tanh) variation is also studied. In Type 1 models, the values of IFR before and after the change are parameterised in terms of their log-ratio, whose prior is normal distributed with mean zero and standard deviation log(3). The prior for the time of the change is normal with mean 2020-12-12 and standard deviation 2 weeks.

For both Type 0 and Type 1 models we consider several possibilities for the time-dependence of model parameters other than the IFR. For the UK, these are (A) no dependence; (B) changing infectivity; (C) changing overdispersion: (BC) changing infectivity and overdispersion.

Findings. Type 1 models for the UK consistently attain a significantly higher model evidence than Type 0 models. (Specifically, they show systematically larger values of the log-posterior.) For each Type 1 model, we separately infer that in the UK, the IFR rose by a factor of about two in late October or early November 2020: the MAP values vary between 1.9-2.2 depending on the model chosen. A similar change is also inferred in Germany and France, but is somewhat less pronounced.

Among the France/Germany models, the ratios of inferred IFRs before/after the step are smaller than in UK but still indicate significant changes, around 1.4 for France and 1.8 for Germany.

The IFR changes in UK are inferred to take place significantly before the B.1.1.7 variant became the dominant UK strain. The significant inferred IFR change in Germany is (to our knowledge) not associated with any new variant.

Interpretation. These findings support the hypothesis [4–6] that the lethality of COVID-19 in the UK increased significantly in late autumn 2020. The MAP estimates further suggest that a larger increase than previously suggested [4, 5], of roughly a factor 2, arose with an onset time in early November. However, the fact that this onset time precedes the rise to dominance of new variant B.1.1.7 suggests that this rise is not solely responsible for the change in IFR. Other factors, such as seasonality and/or pressure on health services, may also be important. If these factors came into play significantly before the arrival of B.1.1.7, then the MAP-estimated onset time of early November could represent a compromise between two successive episodes of increasing IFR within models that, as so far implemented, allow only one.

Note that the IFR depends on the true number of infections (not just diagnosed cases as in the case fatality rate) which, in our model, is itself estimated from the data for cases and testing. Underestimation of this number can lead to an overestimate of the IFR. However, to explain the factor-two increase in fitted IFR this way would require implausible assumptions to be made concerning change in testing efficacy in late autumn 2020 (models TT0, TT1, P0).

1. Introduction

The new variant B.1.1.7 of the SARS-CoV-2 virus first emerged in the UK in September 2020. It is now well known to be more infectious than the prior UK strain and for this reason has not only become dominant in the UK itself, but rapidly taken hold in a number of other countries (including the USA) where it is either already the dominant strain or predicted soon to become

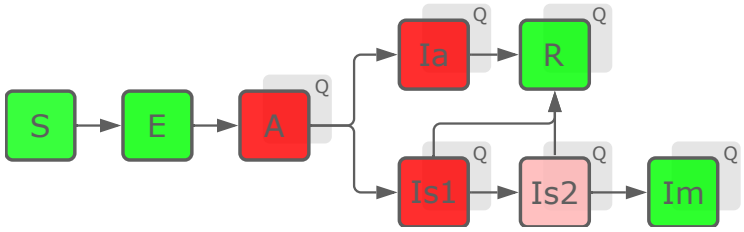

Figure 1: Network structure of the model(s). Every compartment (except S and E) has a quarantined version, transitions into which occur via testing. New infections are transmitted by individuals of the classes coloured red (with the stage Is2 being less infectious). Classes R and Im (mortalities) are no longer infectious.

so [7]. However, it is still uncertain whether B.1.1.7 also leads to more severe cases of COVID-19, resulting in turn in a higher fatality rate [4, 5]. This question has prompted the search for anomalies in the relation between aggregated case and mortality numbers. Simple visual inspections of these data, accounting for the typical time lag between cases and mortality, have suggested that infections with SARS-CoV-2 become more lethal in late 2020 [6].

However, the relation between reported cases and mortalities is not always straightforward to interpret. The time between infection and potential deaths is stochastic, so that a time series of mortality data will tend to show less rapid changes than corresponding case numbers. Moreover, as the available testing capacity changes with time alongside the demand for tests, reported case numbers are not directly representative for the true number of cases.

In order to overcome these challenges while still focusing on (nationally) aggregated data for cases and mortalities, we choose to analyse such data in the context of well-mixed compartmented models, whose complexity is adjustable for this purpose [3, 8]. These models take into account not only the stochastic dynamics of infection and progression of the disease, but also of tests being assigned to individuals of the various compartments at a given overall rate.

Our goal is to establish the evidence for a change in the UK lethality of COVID-19 in a Bayesian fashion, using the reported age-structured data for nationwide cases and mortality. For that purpose, we compare different models that either do or do not allow for a change in the infection fatality rate (IFR). We analyse differences in their posterior probability, optimised for fixed data over a set of model parameters. Generally, one would expect that the likelihood of a model increases as more details are added. In order to judge the significance of changes in the posterior, we compare a whole set of model variants differing in the level of detail, and see whether the ones that do allow for changing IFR perform better than those that do not.

The model specification, simulation, likelihood computation, and optimisation is carried out using the software package PyRoss, which we have developed during the past year [3].

2. Model structure and data

Infection dynamics. We consider a suite of compartmented models (referred to as model variants below) all with $M = 7$ age groups, and 14 classes, as shown in Fig. 1. These classes

are abbreviated as S, E, A, Ia, Is1, Is2, Im, R; along with diagnosed/quarantined versions of all compartments (except S and E), labeled as AQ, IaQ, etc. The age groups are 0-14, 15-29, 30-44, 45-64, 65-74, 75-84, and 85+ (for our UK model variants), referred to by running indices i or j. The transitions from the susceptible (S) to the exposed (E) compartment are induced by infected individuals in the compartments activated (A), asymptomatic (Ia), and the two symptomatic infected stages (Is1, Is2). The corresponding infection rate reads

$$\sum_j \beta_i a_i(t) C_{ij} a_j(t) [A_j + I_j^a + I_j^{s1} + c I_j^{s2}]/N_j, \tag{1}$$

with the infectiousness β_i, the intervention function $a_i(t)$, the contact matrix C_{ij}, the total population per age-group N_i, and a factor c for the reduction of the infectiousness in the second symptomatic stage. The contact matrix we use is based on survey data and previous Bayesian inference [1]. Quarantined individuals are assumed not to infect anyone, so the classes in Eq. (1) are restricted to unquarantined compartments

After a presymptomatic stage (A), individuals become either symptomatically (Is1/Is2) or asymptomatically (Ia) infected, according to an age-dependent fraction α_i of asymptomatic cases. The outcome of the infection is either recovery (R) or death (Im). Progression though all these stages is modelled though linear transition rates, matching the latent and incubation periods, and the typical time from infection until death. The latter is determined by the exit rate γ_s from both stages Is1 and Is2, and is inferred. Otherwise, we fix parameters relating to disease progression to values informed by the literature, see [8] for details.

Contact behaviour. We model changes in the contact behaviour and (potentially) in the infectiousness via the time dependence of $a_i(t)$. These changes can be can be age-dependent, e.g. to model shielding of the elderly or increased contacts at schools and workplaces. To keep the number of control parameters manageable, we separate the time and the age-dependence of $a_i(t)$ as $a_i(t) = a(t)(1 - s(t)h_i)$ and constrain the vector h_i for the age dependence to the form $[0, 0, 0, h_4, h_{5,6}, h_{5,6}, h_7]$, with the largest element set to 1 as a reference. At pre-defined (and occasionally inferred times) the parameters $a(t)$ and $s(t)$ undergo changes and their new values are inferred. We consider the interventions (or other changes) in the UK listed in Tab. 1. See Appendix B for the interventions considered for France and Germany.

Testing. Testing is modelled as the transfer of individuals from the undiagnosed version to the diagnosed (or "quarantined") version of a compartment. For a given overall rate of testing $\tau(t)$, the transition rate from some compartment X to XQ is

$$\tau_X = \tau(t) \pi_X \phi_X / \mathcal{N} \tag{2}$$

with normalisation $\mathcal{N} = \sum_X \pi_X X$. Assuming perfect tests, which, however, do not detect the exposed stage, we use the true negative rate $\phi_X = 0$ for X being S, E, or R, otherwise the true positive rate $\phi_X = 1$. The factors π_X encode testing priorities for the various compartments. We set $\pi_{Is1} = \pi_{Is2} = 1$ as a reference, and $\pi_{Im} = 20$ to ensure that mortalities get reliably detected[1]. The only fit parameter we infer is π_a, the priority for testing individuals that are not

[1] For the early weeks, when tests were scarce, we set $\pi_{Is2} = 5$ and $\pi_{Im} = 100$.

dates	type	control parameters
before 2020-03-20	before lockdown (reference)	$a(t) = 1$, $s(t) = 0$
2020-03-20 to 2020-03-27	imposition of lockdown	linear decrease of $a(t)$, new value $s(t)$
2020-03-27 to 2020-07-24	easing of / increasing non-compliance with lockdown	linear increase of $a(t)$, linear change of $s(t)$
2020-07-24 to 2020-11-06	lockdown lifted	new values of $a(t)$ and $s(t)$
inferred	increase of contacts / infectiousness in autumn	tanh-shaped increase of $a(t)$, centre and width to be inferred
inferred	several local interventions, summarised as a single one at time to be inferred	new values for $a(t)$ and $s(t)$
2020-11-06 to 2020-12-04	national lockdown (England)	new values for $a(t)$ and $s(t)$
2020-12-04 to 2021-01-08	tiered lockdown	new values for $a(t)$ and $s(t)$
after 2021-01-08	national lockdown	new values for $a(t)$ and $s(t)$

Table 1: Interventions considered in the basic UK model. Dates are always rounded to the closest Friday.

symptomatically infected (i.e. of classes S, A, Ia, R). It interpolates between random testing for $\pi_a = 1$ and very targeted testing for $\pi_a = 0$. The progression through stages in the quarantined compartments is the same as in the non-quarantined ones, but quarantined individuals cause no further infections (effectively assuming perfect self-isolation). Once tested positive, individuals remain in the quarantined compartments. The recovered class RQ therefore includes individuals that have actually left quarantine, but we keep this class separate for the purpose of counting previously diagnosed cases.

Lethality. The lethality of COVID-19 is encoded in the infection fatality rate (IFR), i.e., the probability of any infected to die eventually. As an auxiliary quantity for the specification of our model, we define a *symptomatic* infection fatality rate (SIFR), as the fatality rate for symptomatic infections only. Its numerical value can be expected to be close to reported case fatality rate, provided that testing is exhaustively targeted at symptomatic infections.

We choose both transition rates from Is1 to Is2 and from Is2 to Im as $\gamma_s\sqrt{\text{sifr}_i}$, with the age-dependent symptomatic infection fatality rate sifr_i; recoveries (to class R) happen from both stages at rate $\gamma_s(1 - \sqrt{\text{sifr}_i})$. Since we use a fixed fraction α_i of asymptomatic cases, the IFR follows readily as $\text{ifr}_i = (1-\alpha_i)\text{sifr}_i$. The discussion of relative changes in the IFR applies equally to the SIFR.

Noise. All transitions are modeled as Markov rates, i.e. they are inherently stochastic. In order to account for additional sources of noise or variation that are not present in the well-mixed model, we infer overdispersion parameters that scale up the fluctuations in transitions related to infections, testing, and deaths.

Observables. We consider the reported cumulative case numbers as the sum of all quarantined classes for each age cohort. The reported cumulative mortalities are identified with the numbers of class ImQ[2]. We ensure that no deaths remain unnoticed, by formally assuming that individuals in class Im continue to get tested at high priority. The numbers in all other compartments are considered as hidden, and are implicitly reconstructed by the inference procedure.

Data. We use weekly data from the week beginning 2020-03-07 to the week ending 2021-01-15. For cases, we use the daily numbers reported on the UK government webpage [9], reported by specimen date and in 5-year cohorts. We coarse-grain these data to weekly numbers and to our choice of age groups.

Death numbers by week of reporting have been obtained from the UK Office for National Statistics (ONS) webpage [10]. These data appear to be incomplete after 2020-12-25 and have therefore been padded with mortality data from the government webpage, which are more up-to-date but not age-structured[3].

The daily number of PCR tests performed is available from the government webpage [9]. For early testing data (before 2020-04-21) we use data published by [11, 12]. The data is coarse-grained to weekly numbers, from which we build the testing rate $\tau(t)$ as a stepwise constant function.

For France, we use data for deaths in hospitals [13], and PCR testing [14]. The latter includes the number of tests performed for each age-group, except for early data where instead we use non-age structured test numbers [15].

For Germany, we use data provided by the Robert Koch-Institute for cases and deaths [16], and tests [17].

3. Model variants

We consider several variants of the basic model outlined above, labeled by A0, A1, B0, etc. Some of the variants have an IFR that is constant in time, as indicated by the Type number 0. Type number 1 indicate time-dependent changes of the IFR; Type 2 also indicates this but via a mechanism involving slowed recovery rather than higher death rate (which also ultimately results in more deaths). The Type letters refer to other details of the model variant, *a priori* unrelated to the IFR.

Basic model (A0). The model outlined above, without any further additions.

Step-change in IFR (A1). As model A0, but with a simple step-change in IFR. The size of the change and time of change are inferred (except for the two youngest cohorts, where fatal cases

[2]For mortalities, we merge the age groups 15-29 and 30-44 to match the available data.

[3]More recent data have become available in the meantime, however, we stick to this procedure to ensure comparability between results for the model variants. For the weeks in question, the sum over all age groups in class ImQ is considered as observable.

are extremely rare). The change in IFR is parameterised in terms of the log-ratio of the values, with a prior that is normal distributed with mean zero and standard deviation log(3). The prior for the time of the change is normal with mean 2020-12-12 and standard deviation 2 weeks.

Change in recovery rate (A2). As model A0, but with a step-change of the recovery rate to $r\gamma_s(1-\sqrt{\text{ifr}})$ with some factor r. The rates for progression to Is2 or Im are unchanged. This effectively changes the SIFR (and accordingly the IFR) to

$$\text{sifr}' = \frac{\text{sifr}}{\left(\sqrt{\text{sifr}} + r\left(1-\sqrt{\text{sifr}}\right)\right)^2}. \tag{3}$$

This way of introducing a change in IFR causes the death numbers to evolve more smoothly, without the need for a further fit parameter. For $r < 1$, the IFR rises, and all additional deaths occur *after* the ones that would have occurred had rates remained unchanged. (With a fixed recovery rate, the longer a patient does not recover, the more likely they are to die eventually.)

Model with easing/increase in infectiousness (B0). As model A0, but with a linear increase in $a(t)$ and a linear change in $s(t)$ during the November lockdown. This could model the increasing non-compliance or changes in the infectiousness (new variant of the virus), or both.

Tanh-shaped IFR change (B1). As model B0, but with a tanh-shaped change in IFR. The time around which this change is centred (which we refer to as the onset time), the width, and the amplitude are inferred. (We deem that this increased level of detail is harmonious with the already more detailed model B0.)

Model with change in overdispersion (C0). As model A0, but with a change in the three overdispersion parameters for infections, testing, and deaths. The change is allowed to happen on 2020-10-02, a date chosen to match a potentially new stochastic dynamics as the second wave gains momentum. The new values of the parameters are inferred independently. This reflects potential changes in the testing strategy and in the infection dynamics in the second wave, and can avoid the overestimation of case and death numbers, that is often observed as a side-effect of mismatching overdispersion parameters.

Jump-like change of IFR and overdispersion (C1). As model C0, but with a simple jump-like change in IFR.

Combined model (BC0). A combination of models B0 and C0: It has a change in overdispersion parameters in the second wave and easing/non-compliance (or increasing infectiousness) during the November lockdown.

Combined model with changing IFR (BC1). A combination of models B1 and C1: As model BC0, but with a tanh-shaped change of IFR as in B1.

Test and trace (TT0). When case numbers are low, effective contact tracing is possible. This could mean that more asymptomatic cases are uncovered in summer than at the height of the first and second wave. As a simple model for this effect, building on model BC0, we allow for the inference of the testing priorities π_A and π_{Ia} for pre- and asymptomatic infected individuals different from the priority π_a of classes S and R. This change comes into effect with the beginning of large-scale contact tracing on 28th May 2020. Testing priorities remain unchanged thereafter; however, as long as the testing priorities of A and Ia remain below those of Is, the effect of contact tracing will only become relevant for large test rates and low case numbers, so that the class Is of undiagnosed individuals can be depleted.

Test and trace with changing IFR (TT1). As model TT0, but with a tanh-shaped change in IFR (as in B1).

Time-dependent test priority (P0). As model BC0, but with a time-dependent change in π_a, the only parameter entering our model for testing. The change in π_a is tanh-shaped, with centre (the onset time), width and amplitude to be inferred. This change may reflect changes in the testing strategy, that have happened during the course of the pandemic.

4. Model comparison

Using our software package PyRoss, we can calculate the logarithmic likelihood of the observed data for each of the model variants and for any choice of the model parameters and initial conditions [8]. This computation is based on the inherent stochasticity specified for the model. It employs a multivariate Gaussian approximation of the joint probability of all compartment values at all observed points in time, thus taking into account all correlations across compartments and time. The approximation becomes exact in the limit of large population sizes and is therefore appropriate for our well-mixed models applied at national level.

Given an informed choice of prior distributions for all parameters and initial conditions, we have determined for each model variant the parameters that maximise the posterior probability (MAP). Results are summarised in Tab. 2 and Fig. 2. Remarkably, the model variants with a change in IFR consistently attain higher posteriors. As we have centred the prior for the factor of change in IFR to the value 1 (representing the null-hypothesis that there has been no change), the prior always decreases with a change in IFR, but this reduction of the prior is overbalanced by quite some margin by the increased likelihood derived from data. The inferred change in IFR is always around the factor of two for the UK data. Note that for model A2 the change in the recovery rate amounts even to a factor of 2.35 in IFR for the oldest cohort.

Although closely related, the log-posterior should not be naively equated with model evidence. To calculate the latter, one would also need to consider (sub-exponential) contributions from integrating over the parameter space, and the Occam factor penalising the additional number of fit parameters. Nonetheless, it is remarkable that even though the models A0, B0, C0, and BC0 have different numbers of fit parameters, their variation in the log-posterior is nowhere near as big as the difference to the variants with a changing IFR. We have not calculated the full model

Country	Model	# Params	log-Prior		log-Posterior		IFR change	
			abs	Δ	abs	Δ	Onset	Factor
UK	A0 (base)	67	−354		−4281			
	A1	69		−20		+55	26 Oct	1.967
	A2	69		−9		+49	23 Nov	$r = 0.399$
UK	B0 (base)	69	−350		−4269			
	B1	72		−21		+61	7 Nov	2.135
UK	C0 (base)	70	−365		−4282			
	C1	72		−35		+52	27 Oct	2.017
UK	BC0 (base)	72	−400		−4272			
	BC1	75		−26		+56	9 Nov (±9 days)	2.197(±0.19)
	P0	75		−22		+59	[29 Nov]	[3.000][4]
UK	TT0 (base)	73	−368		−4270			
	TT1	76		−25		+56	9 Nov	2.201
GER	C0 (base)	59	−268		−3438			
	C1	61		+3		+79	26 Nov (±2 days)	1.803(±0.08)
FRA	C0 (base)	75	−254		−4137			
	C1	77		0		+55	3 Nov	1.372

Table 2: Summary of MAP results. We list the country considered along with the model variant, the number of inferred parameters and initial conditions, the logarithmic prior and posterior, and, if applicable, the inferred onset time and factor of a change in IFR. For Type 0 (base) models without any IFR change, absolute values of log-prior and log-posterior are shown, indicated by "abs". For other models, we show values relative to the corresponding base model, indicated by Δ. For models UK-BC1 and GER-C1 we indicate the uncertainty in the inferred parameters for the IFR change as the standard deviation in a Gaussian approximation of the posterior.

evidence [18–20] for each variant, because it is computationally expensive. However, our previous results obtained using Markov Chain Monte Carlo simulations for models calibrated with death numbers have shown that indeed the model evidence is largely dominated by the log-posterior, such that the latter can be used as a meaningful proxy to compare model evidences [8].

The uncertainty of the inferred parameters around the MAP values for a given model variant can be assessed by considering the local shape of the posterior close to the MAP parameters. For models BC1 in the UK and model C1 in Germany, we have calculated the Hessian of the log-posterior at the MAP value, which yields a multivariate Gaussian approximation of the posterior. Evaluating the standard deviation for the onset time and factor of the IFR change results in error estimates that are in line with the observed variation of those parameters across the different model variants. We expect similar uncertainties for all UK model variants, but refrain from repeating this rather tedious computation for each of them. The smaller uncertainties observed for Germany are indicative for the overall better fit of the model with the German data, but may also be characteristic for the step-like change of IFR in model C1.

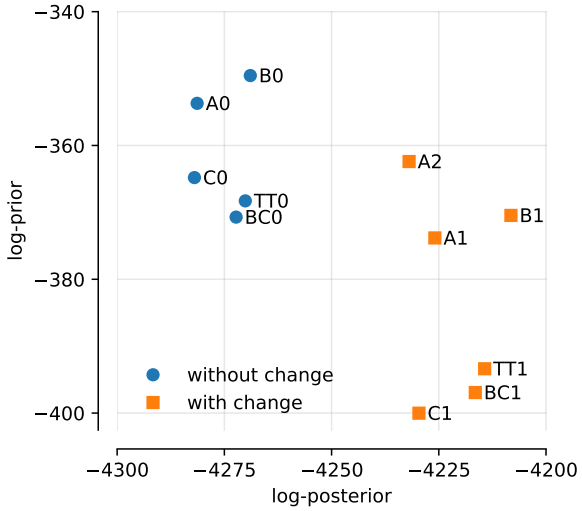

Figure 2: Plot of log-posterior and log-prior. Models without change in IFR in blue, models with change in IFR in orange.

For each model and set of MAP parameters, we can plot a deterministic solution. This is the most likely trajectory, conditional on the inferred initial condition, and the mean of the multivariate Gaussian approximation for all compartment values at all times. Fully detailed plots of the MAP trajectories for each of the model variants are shown in Appendix C.

Unsurprisingly, the results for the most detailed model variants BC0 and BC1 produce mean trajectories for cases and mortalities that fit the data best, as shown in Fig. 3. Among these two solutions, the even better fit of the model variant with change in IFR (BC1) is not only evident through the likelihood computation, but also visible to the naked eye.

In Fig. 4 we show the MAP trajectories for the variants of models A, B and C. On visual inspection, none of these models fit the data as accurately as BC0 and BC1 in Fig 3. However, such visual inspection is not always as reliable an estimate as our Bayesian posterior, which accounts for temporal correlations in the data. For example, we find that the likelihood strongly penalises models where the rate of growth (or decay) of infections does not match the data. Also, the (idealised) step-changes in the intervention function $a(t)$ mean that MAP trajectories may over/under-shoot the data at the change points. In combination, these two factors mean that agreement between expected trajectory and data may be imperfect on visual inspection; however, the posterior is being correctly maximised and the results of inference are robust. This is because deviations of the data from the mean trajectory are consistently taken into accounted for the implicit inference of unobserved compartments. For example, the inferred change in IFR is remarkably consistent between the model variants for the UK, despite considerable differences in the mean trajectory for each model.This observation reassures us that the evidence for a change in IFR would persist in a model that is even more detailed than model BC.

Some differences between models A/B and C are due to a generic feature of our computation

[4]Change of π_a. MAP value attained upper bound set by prior.

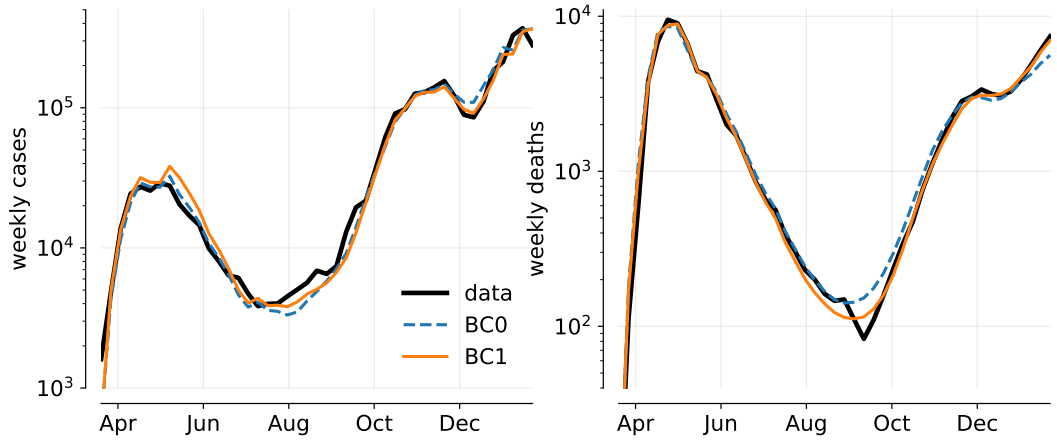

Figure 3: Mean trajectories of weekly cases and deaths for the MAP parameters of models BC0 (dashed blue) and BC1 (solid orange), along with data (black)

of the likelihood. When the estimated overdispersion parameters are too low to account for the observed noise in the data, the optimiser tends to overestimate the expected case and deaths numbers, thereby increasing the variance of weekly changes. Considering the resulting MAP trajectories for models A and B, it seems that the inferred overdispersion parameters, which serve well to fit the first wave, are too small to match the level of noise in the second wave. This leads to an overestimation of both expected case and death numbers. Differences between cases and deaths in this overestimation could negate the perceived change in IFR. This has prompted us to analyse model C, allowing for changes in the overdispersion parameters for the second wave. It is consistent with the results of models A and B, which rules out that the observed changes in the IFR stem from temporal changes in the overdispersion.

The models C0/C1 reproduce the observed height of the second wave, but not the dip between the second and third wave. Note that model C (just like model A) has the November lockdown fixed without easing, leading to a larger reduction in cases and deaths than in reality. Model BC (easing and change of overdispersion) reproduces the short-livedness of this reduction better.

The sudden drop in mortalities (individuals with COVID-19 mentioned on the death certificate) in late August / early September is not reproduced by any of our models. This might be related to changes in the legal definition of such deaths. We do not model this here, but note that adjusting the data for a lasting change in the definition from this time onwards would likely lead to an even larger increase in the IFR than the MAP estimates 1.9-2.2 reported above.

We also did the inference procedure for model C0/C1 with data for France and Germany, using appropriate forms of the intervention function, as detailed in the Appendix B. A similar change in IFR seems to be present there as well, though somewhat less pronounced in France and happening somewhat later in Germany.

In Fig. 5 we show the inferred mean fraction of infected individuals, conditional on the observation of cases and deaths at all points in time. This result is compared to the data from the

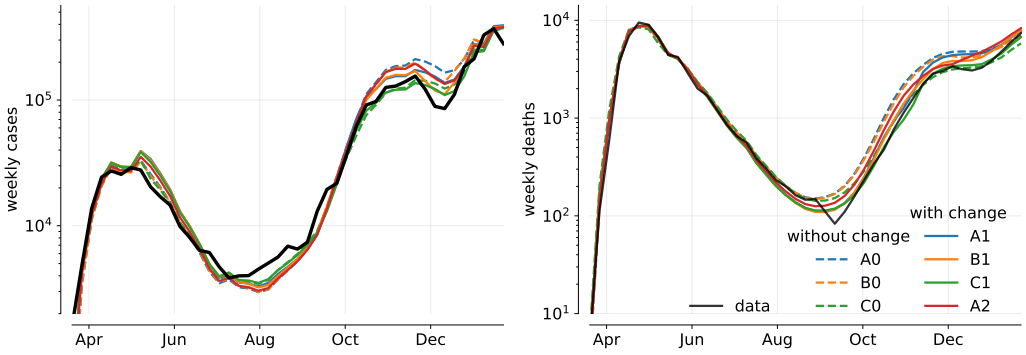

Figure 4: Mean trajectories of weekly cases and deaths for the MAP parameters of the various model variants, along with data (black). Models without change in IFR are shown as dashed, models with change in IFR solid.

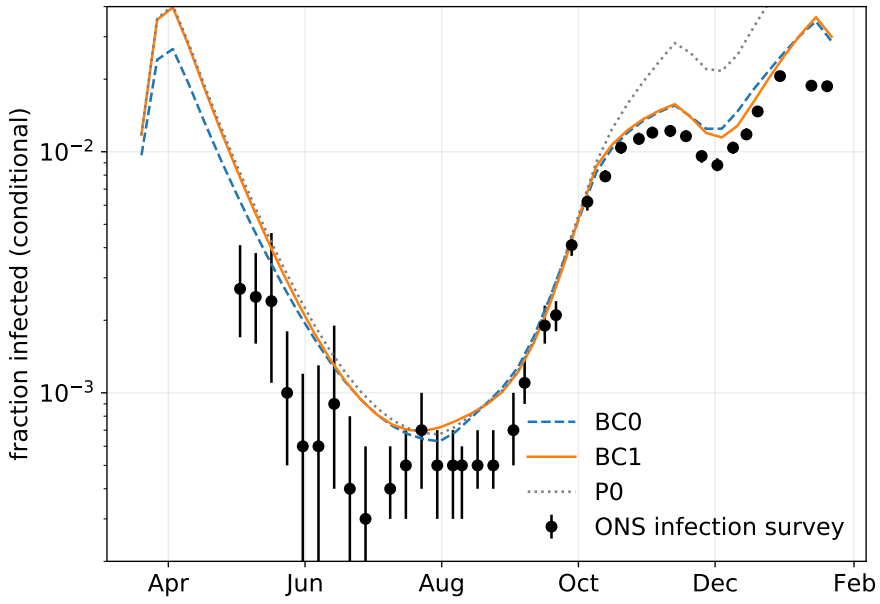

Figure 5: Mean fraction of people infected (symptomatic and asymptomatic) in the total population, conditional on the observation of cases and deaths in models BC0 (blue), BC1 (orange) and P0 (grey). For comparison, we also show the prevalence of infections reported in the ONS infection survey.

ONS infection survey [21], seeking to determine the fraction of the infected population through random asymptomatic testing. We note that numbers in our results are always bigger than in the infection survey, except when they are rapidly rising. This could be an indication that individuals count as infected for longer in our models than they test positive in a PCR test. Also, the fixed estimate we use for the age-dependent fraction of asymptomatic infections may be incorrect (based on early data [22], as described in our paper [8]).

We note that differences in the conditional numbers of infections between models BC0 and BC1 mainly show up in the first wave. For that time, data from the infection survey is not available, and data on testing may be incomplete. Nonetheless, it is remarkable that the inferred numbers of infection largely agree between models BC0 and BC1 from June onwards, encompassing the inferred time of change in IFR. Hence, we can rule out that the change in IFR in model BC1 is merely due to changes in the inferred true number of cases. It is rather that the observed timeline of deaths is more likely in model BC1 than in model BC0, for similar estimated total numbers of infections.

Based on the remaining discrepancies between the our inferred infection numbers and the ONS survey, one could still argue that we overestimate the true case numbers early on and/or underestimate them later, leading to an apparent increase in the IFR. The model variants TT0/TT1 and P0 serve to address this possibility.

The goal of contact tracing is to detect and isolate asymptomatic cases of COVID-19 and ideally to also detect cases early on in the presymptomatic stage. However, due to limited capacity, the test and trace system is only effective when case numbers are low and there is sufficient testing capacity, such as in the summer months. This could mean that with increasing testing numbers and the large-scale test and trace system being put in place in late May, the reported cases after the first wave are closer to the true cases than expected by the model variants considered so far. The results for the pair of variants TT0 and TT1, with their (albeit rudimentary) realisation of contact tracing, give no indication that this might explain an apparent change in the IFR. Quite to the contrary, TT1 infers a somewhat larger change in IFR than the Type 1 models already considered. The posterior of models TT0 and TT1 change only marginally compared to BC0 and BC1, respectively.

In principle, it is possible that the inferred factor-two change in IFR in early November could instead be explained by large changes in the numbers of undiagnosed cases. We illustrate this fact using model P0. It allows for a time-dependent change of the testing priority π_a, and we deliberately set a loose prior on the timing and amplitude of this change. This model attains a posterior probability that is comparable to that of the models with a change in the IFR. (Coincidentally, prior and posterior are almost identical to those of TT1, therefore P0 is not shown in Fig. 2.) The prior would have allowed to place the change in π_a as early as July, yet it is inferred as late November (four weeks *after* the mean of the Gaussian prior distribution). The amplitude of the IFR change is inferred as a factor 3, unexpectedly saturating an upper bound we set on this parameter. This would mean that, to explain the data in terms of a changed testing regime rather than an actual IFR change, at the height of the second wave, tests must have become at least three times *less* targeted at symptomatic cases of COVID-19. It seems to us highly unlikely that this change could be explained, for example, by a rise in cases of flu, and we would expect evidence for this scenario to rapidly diminish under a more realistic choice of prior. We also note that, as

shown in Fig. 5, the inferred true number of infections for P0 matches the data from the infection survey much worse than for models BC0 and BC1.

5. Discussion

In this paper, we have reported evidence for an increase in lethality of COVID-19 in the UK in late autumn 2020. Bayesian inference provides clear and consistent evidence for such an increase, across the suite of models we considered. This finding complements similar conclusions based on the visual inspection of nationally aggregated case and mortality data [6].

We would not generally expect aggregated case and mortality data, analysed with a well mixed compartmental model (or suite thereof), to identify clear or definitive causes for an increase of this kind. So our conclusion is (in common with [6]) that something happened to increase lethality, but the data does not clearly say what it is that happened.

It is natural to speculate that the increase is related to the emergence of one or more new virus strains, whose potentially increased lethality has been the subject of several cohort-based studies summarised previously in a UK Government publication [5]. Those works estimate a lethality increase for B.1.1.7 by factors ranging up to 1.7 [23]. In contrast, a study by Davies *et al.* using a compartmented model including two virus strains [24] did not produce clear evidence for any change in severity (only a large increase transmissibility). From our suite of compartmented models, we infer not only a larger increase in IFR than so far attributed to B.1.1.7 severity alone, but also a most likely onset date for this increase that *precedes* the widespread emergence of B.1.1.7 in the UK. Moreover, we infer a very similar change in Germany (alongside a more modest one in France) despite even lower prevalence there, at that time, of B.1.1.7 or any comparable new strain. We argue on these grounds that the arrival of B.1.1.7 is unlikely to be the sole cause of the dramatic UK increase in IFR in late 2020.

This suggests that factors such as seasonality and/or pressure on health services may have contributed to the change. Of these, seasonality more credibly would have had similar effects in the UK and Germany (and perhaps weaker effects in France, but this is far from clear). This conclusion may also be supported by a study from Israel addressing changes of in-hospital mortality rates [25]. If such factors started to increase IFR in early autumn, then the MAP estimates of the onset time, for either a single step-change or a single ramped increase, might represent a compromise between two separate episodes of increasing IFR, with only the latter linked to a new variant. This could allow the fitted onset time to precede arrival of B.1.1.7, with the latter still contributing strongly to the total increase in fatality.

Such arguments remain, for now, speculative. To further investigate potential connections between new variants and the observed change in IFR, one should explicitly represent the dominant mutant strain though additional model compartments, with increased infectiousness and possibly increased lethality. Such a model could be calibrated using data for the prevalence of mutant strains.

The increase in lethality might also be generic for the peaks of waves of infection, when hospitals are under severe strain. We have not yet analysed the possibility for *two* changes in IFR: a reduction in spring and an increase in autumn. This might also reflect seasonal variations in

the resilience of patients. Figs. 3 and 4 show that the inferred most likely trajectories actually overestimate the cases in the first wave somewhat in order to get the number of deaths right, and even more so for the models with change in IFR. The possibility of a high IFR in the first wave, falling back in late spring before rising again in autumn, might resolve this discrepancy.

It will be interesting to see whether the evidence for a change in IFR persists as models become calibrated with more recent data and additional types of observations. This might include data for hospital admissions, antibody testing, or random asymptomatic testing (as already considered *a posteriori* in this report). More data would also enable more detailed versions of our models to be constructed. For example, we have neglected so far the possibility of reinfection (loss of immunity). It would be also possible to represent more accurately the timing of the progression through stages of infection [26]. Large-scale vaccination campaigns may also be accounted for in future variants of these models.

As an additional caveat, we emphasise that these well-mixed models cannot describe regional variability, such as situations where the epidemic is shrinking in one region and growing in another. Moreover, the evidence for the UK IFR change is partly based on the relatively poor fits achieved by models with fixed IFR. It follows that if substantially improved fits were obtained for models that include regional variability (with fixed IFR), our conclusions might have to be re-evaluated. On the other hand, we are not aware of a specific mechanism by which regional variability would generate the discrepancies observed here between the models and the data. These possibilities might be also tested by future work.

Finally, we would like to advocate the advantages of creating a suite of models within a single platform to which consistently identical inference procedures can be applied. It is of course valuable to have independent modelling teams doing their own preferred type of data analysis and then comparing the results of these studies to see if a consensus emerges. However, the integrated model-building and inference machinery gathered within the PyRoss platform has in this case allowed rapid implementation of a number of purpose-built models and their comparison, in a fully Bayesian way, in the month immediately following the appearance of [6]. (In fact [6] was posted on medRxiv on 22 January; a complete draft of the current paper was circulated to relevant members of the UK Government advisory group SPI-M on 24 February.)

The PyRoss platform [2] is open source, and freely available to all users.

Acknowledgments. We thank Graeme Ackland, Daan Frenkel, and Julia Gog for helpful discussions. We also thank William Peak and Andrew Ng from JPMorgan Chase & Co. and all Pyross contributors (see [3]), in particular Julian Kappler, Yuting I. Li, Paul B. Rohrbach, Rajesh Singh, and Günther Turk. This work was undertaken as a contribution to the Rapid Assistance in Modelling the Pandemic (RAMP) initiative, coordinated by the Royal Society. This work was funded in part by the European Research Council under the Horizon 2020 Programme, ERC grant 740269, and by the Royal Society grant RP17002. We thank Microsoft Corporation for a Microsoft Research Award for the project "Building an open platform for pandemic modelling".

References

[1] L. Fumanelli, M. Ajelli, P. Manfredi, A. Vespignani, and S. Merler, "Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread", *PLoS Computational Biology* **8**, e1002673 (2012).

[2] "PyRoss: inference, forecasts, and optimised control for epidemiological models in Python", https://github.com/rajeshrinet/pyross.

[3] R. Adhikari, A. Bolitho, F. Caballero, M. E. Cates, J. Dolezal, T. Ekeh, J. Guioth, R. L. Jack, J. Kappler, L. Kikuchi, H. Kobayashi, Y. I. Li, J. D. Peterson, P. Pietzonka, B. Remez, P. B. Rohrbach, R. Singh, and G. Turk, "Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: the pyross library", *arXiv:2005.09625* (2020).

[4] P. Horby, C. Huntley, N. Davies, J. Edmunds, N. Ferguson, G. Medley, and C. Semple, "NERVTAG paper on COVID-19 variant of concern B.1.1.7", https://www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117 (2021).

[5] "NERVTAG: Update note on B.1.1.7 severity, 11 February 2021", https://www.gov.uk/government/publications/nervtag-update-note-on-b117-severity-11-february-2021 (2021).

[6] D. J. Wallace and G. J. Ackland, "Abrupt increase in the UK coronavirus death-case ratio in December 2020", *medRxiv, doi:10.1101/2021.01.21.21250264* (2021).

[7] S. E. Galloway, P. Paul, D. R. MacCannell, M. A. Johansson, J. T. Brooks, A. MacNeil, R. B. Slayton, S. Tong, B. J. Silk, G. L. Armstrong, M. Biggerstaff, and V. G. Dugan, "Emergence of SARS-CoV-2 B.1.1.7 lineage — United States, December 29, 2020–January 12, 2021", *MMWR. Morbidity and Mortality Weekly Report* **70**, 95 (2021).

[8] Y. I. Li, G. Turk, P. B. Rohrbach, P. Pietzonka, J. Kappler, R. Singh, J. Dolezal, T. Ekeh, L. Kikuchi, J. D. Peterson, H. Kobayashi, M. E. Cates, R. Adhikari, and R. L. Jack, "Efficient Bayesian inference of fully stochastic epidemiological models with applications to COVID-19", *arXiv:2010.11783* (2020).

[9] "Coronavirus (COVID-19) in the UK", https://coronavirus.data.gov.uk/details/download.

[10] "Deaths registered weekly in England and Wales, provisional", https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales (2021).

[11] https://ourworldindata.org/coronavirus-testing, accessed 2020-04-16 and 2021-01-01.

[12] J. Hasell, E. Mathieu, D. Beltekian, B. Macdonald, C. Giattino, E. Ortiz-Ospina, M. Roser, and H. Ritchie, "A cross-country database of COVID-19 testing", *Sci. Data* **7** (2020).

[13] "Données hospitalières relatives á l'épidémie de COVID-19", https://www.data.gouv.fr/en/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/.

[14] "Données relatives aux résultats des tests virologiques COVID-19", https://www.data.gouv.fr/en/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19.

[15] "Data on testing for COVID-19 by week and country", https://www.ecdc.europa.eu/en/publications-data/covid-19-testing.

[16] "Robert Koch-Institut: COVID-19-Dashboard", https://www.arcgis.com/home/item.html?id=f10774f1c63e40168479a1feb6c7ca74.

[17] "Tabellen zu Testzahlen, Testkapazitäten und Probenrückstau pro Woche", https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Testzahlen-gesamt.html.

[18] D. J. C. MacKay, "Bayesian interpolation", *Neural Comput.* **4**, 415 (1992).

[19] R. E. Kass and A. E. Raftery, "Bayes factors", *J. Am. Stat. Assoc.* **90**, 773 (1995).

[20] A. E. Raftery, "Bayesian model selection in social research", *Sociol. Methodol.* **25**, 111 (1995).

[21] "Coronavirus (COVID-19) infection survey", https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata (2021).

[22] F. Riccardo, M. Ajelli, X. D. Andrianou, A. Bella, M. D. Manso, M. Fabiani, S. Bellino, S. Boros, A. M. Urdiales, V. Marziano, M. C. Rota, A. Filia, F. D'Ancona, A. Siddu, O. Punzo, F. Trentini, G. Guzzetta, P. Poletti, P. Stefanelli, M. R. Castrucci, A. Ciervo, C. D. Benedetto, M. Tallon, A. Piccioli, S. Brusaferro, G. Rezza, S. Merler, and P. Pezzotti, "Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020", *Euro Surveill.* **25**, 2000790 (2020).

[23] R. Challen, E. Brooks-Pollock, J. M. Read, L. Dyson, K. Tsaneva-Atanasova, and L. Danon, "Increased hazard of mortality in cases compatible with SARS-CoV-2 variant of concern 202012/1 - a matched cohort study", *medRxiv, doi:10.1101/2021.02.09.21250937* (2021).

[24] N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, and W. J. Edmunds, "Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England", *medRxiv, doi:10.1101/2020.12.24.20248822* (2020).

[25] H. Rossman, T. Meir, J. Somer, S. Shilo, R. Gutman, A. B. Arie, E. Segal, U. Shalit, and M. Gorfine, "Hospital load and increased COVID-19 related mortality - a nationwide study in Israel", *medRxiv, doi:10.1101/2021.01.11.21249526* (2021).

[26] J. D. Peterson and R. Adhikari, "Efficient and flexible methods for time since infection models", *arXiv:2010.10955* (2020).

A. Constitutive equations for the basic model

The differential equations determining the deterministic evolution of the mean occupation numbers of each compartment read

$$\dot{S}_i = -\lambda_i(t) S_i \tag{4a}$$
$$\dot{E}_i = -\gamma_E E_i + \lambda_i(t) S_i \tag{4b}$$
$$\dot{A}_i = -\gamma_A A_i + \gamma_E E_i - \tau(t)\pi_a A_i/\mathcal{N}(t) \tag{4c}$$
$$\dot{A}_i^Q = -\gamma_A A_i^Q + \tau(t)\pi_a A_i \tag{4d}$$
$$\dot{I}_i^a = \gamma_A \alpha_i A_i - \gamma_a I_i^a - \tau(t)\pi_a I_i^a/\mathcal{N}(t) \tag{4e}$$
$$\dot{I}_i^{a,Q} = \gamma_A \alpha_i A_i^Q - \gamma_a I_i^{a,Q} + \tau(t)\pi_a I_i^a/\mathcal{N}(t) \tag{4f}$$
$$\dot{I}_i^{s1} = \gamma_A(1-\alpha_i) A_i - \gamma_s I_i^{s1} - \tau(t)\pi_{s1} I_i^{s1}/\mathcal{N}(t) \tag{4g}$$
$$\dot{I}_i^{s1,Q} = \gamma_A(1-\alpha_i) A_i^Q - \gamma_s I_i^{s1,Q} + \tau(t)\pi_{s1} I_i^{s1}/\mathcal{N}(t) \tag{4h}$$
$$\dot{I}_i^{s2} = \gamma_s \sqrt{\mathrm{sifr}_i(t)} I_i^{s1} - \gamma_s I_i^{s2} - \tau(t)\pi_{s2} I_i^{s2}/\mathcal{N}(t) \tag{4i}$$
$$\dot{I}_i^{s2,Q} = \gamma_s \sqrt{\mathrm{sifr}_i(t)} I_i^{s1,Q} - \gamma_s I_i^{s2,Q} + \tau(t)\pi_{s2} I_i^{s2}/\mathcal{N}(t) \tag{4j}$$
$$\dot{I}_i^{m} = \gamma_s \sqrt{\mathrm{sifr}_i(t)} I_i^{s2} - \tau(t)\pi_{m} I_i^{m}/\mathcal{N}(t) \tag{4k}$$
$$\dot{I}_i^{m,Q} = \gamma_s \sqrt{\mathrm{sifr}_i(t)} I_i^{s2,Q} + \tau(t)\pi_{m} I_i^{m}/\mathcal{N}(t) \tag{4l}$$
$$\dot{R}_i = \gamma_a I_i^a + \gamma_s(1-\sqrt{\mathrm{sifr}_i(t)})(I_i^{s1} + I_i^{s2}) - \tau(t)\pi_a R_i/\mathcal{N}(t) \tag{4m}$$
$$\dot{R}_i^Q = \gamma_a I_i^{a,Q} + \gamma_s(1-\sqrt{\mathrm{sifr}_i(t)})(I_i^{s1,Q} + I_i^{s2,Q}) + \tau(t)\pi_a R_i/\mathcal{N}(t) \tag{4n}$$

with

$$\lambda_i(t) = \sum_j \beta_i C_{ij}(t)(A_j + I_j^a + I_j^{s1} + c I_j^{s2})/N_j \tag{5}$$

and

$$\mathcal{N}(t) = \sum_i \left[\pi_a(S_i + E_i + A_i + I_i^a + R_i) + \pi_{s1} I_i^{s1} + \pi_{s2} I_i^{s2} + \pi_m I_i^m\right]. \tag{6}$$

The stochastic differential equations underlying our computation of the likelihood follow in a linear noise approximation of the corresponding master equation. Denoting the stochastic compartment numbers also as $S_i, A_i, \ldots$ (for notational simplicity), a stochastic noise term is

added to each term in Eq. (4). This leads to

$$\dot{S}_i = -\lambda_i(t)S_i - \sqrt{\eta_{\text{infect}}\lambda_i(t)S_i}\zeta_i^{(0)}(t) \tag{7a}$$

$$\dot{E}_i = -\gamma_E E_i + \lambda_i(t)S_i - \sqrt{\gamma_E E_i}\zeta_i^{(1)}(t) + \sqrt{\eta_{\text{infect}}\lambda_i(t)S_i}\zeta_i^{(0)}(t) \tag{7b}$$

$$\dot{A}_i = -\gamma_A A_i + \gamma_E E_i - \tau(t)\pi_a A_i/\mathcal{N}(t) - \sqrt{\gamma_A A_i}\zeta_i^{(2)}(t) + \sqrt{\gamma_E E_i}\zeta_i^{(1)}(t) - \sqrt{\eta_{\text{test}}\tau(t)\pi_a A_i/\mathcal{N}(t)}\zeta_i^{(3)}(t) \tag{7c}$$

$$\dot{A}_i^Q = -\gamma_A A_i^Q + \tau(t)\pi_a A_i - \sqrt{\gamma_A A_i^Q}\zeta_i^{(4)}(t) + \sqrt{\eta_{\text{test}}\tau(t)\pi_a A_i/\mathcal{N}(t)}\zeta_i^{(3)}(t) \tag{7d}$$

etc.

with white noise processes satisfying $\langle\zeta_i^{(\mu)}(t)\rangle = 0$ and $\langle\zeta_i^{(\mu)}(t)\zeta_j^{(\nu)}(t')\rangle = \delta_{ij}\delta_{\mu\nu}\delta(t-t')$. In order to account for sources of noise otherwise not resolved in the model, we include overdispersion factors in the noise terms relating to infections (η_{infect}), to testing (η_{test}), and to deaths (η_{death}). The latter is included in transitions from $I^{s2,Q}$ to $I^{m,Q}$ and from I^m to $I^{m,Q}$.

B. Intervention functions for France and Germany

dates	type	control parameters
before 2020-03-08	before lockdown (reference)	$a(t) = 1$, $sh(t) = 1$, $s(t) = 1$
2020-03-18 to 2020-05-13	first lockdown	new values for $a(t)$, $sh(t)$, $s(t)$
2020-05-13 to 2020-06-14	easing of lockdown	linear change for $a(t)$, $sh(t)$, $s(t)$
2020-07-03	start of school holiday	change in $s(t)$
2020-06-14 to 2020-07-11	restrictions are lifted	linear change $a(t)$, $s(t)$
2020-07-11 to 2020-11-01	more easing of restrictions, schools are reopened	new values for $a(t)$, $s(t)$
inferred	autumn	tanh-shaped increase for $a(t)$
2020-11-27	effect before lockdown is imposed	change in $a(t)$
2020-11-01 to 2020-11-27	second lockdown, primary schools never close	change in $a(t)$, $sh(t)$
2020-11-27 to 2020-12-15	easing of second lockdown	new values for $a(t)$, $sh(t)$
after 2020-12-15	introduction of nightly curfew	new values for $a(t)$, $sh(t)$

Table 3: Interventions that are considered in model FRA-C. The $a(t)$ refers to scaling of contacts for non-school contacts, $s(t)$ refers to the scaling of school contacts, and $sh(t)$ refers to the scaling of the shielding vector.

dates	type	control parameters
before 2020-03-09	before lockdown (reference)	$a(t) = 1$, $s(t) = 0$
2020-03-09 to 2020-03-23	imposition of lockdown	linear decrease of $a(t)$, new value $s(t)$
2020-03-23 to 2020-06-22	easing of / increasing non-compliance with lockdown	linear increase of $a(t)$, linear change of $s(t)$
2020-06-22 to 2020-11-02	lockdown lifted	new values of $a(t)$ and $s(t)$
inferred	increase of contacts / infectiousness in autumn	tanh-shaped increase of $a(t)$ and change of $s(t)$, centre and width to be inferred
2020-11-02 to 2020-12-14	lockdown "light", local interventions	new values for $a(t)$ and $s(t)$
after 2021-12-14	national lockdown	new values for $a(t)$ and $s(t)$

Table 4: Interventions considered in model GER-C. Dates are always rounded to the closest Monday.

C. Detailed plots of the MAP trajectories

C.1. UK

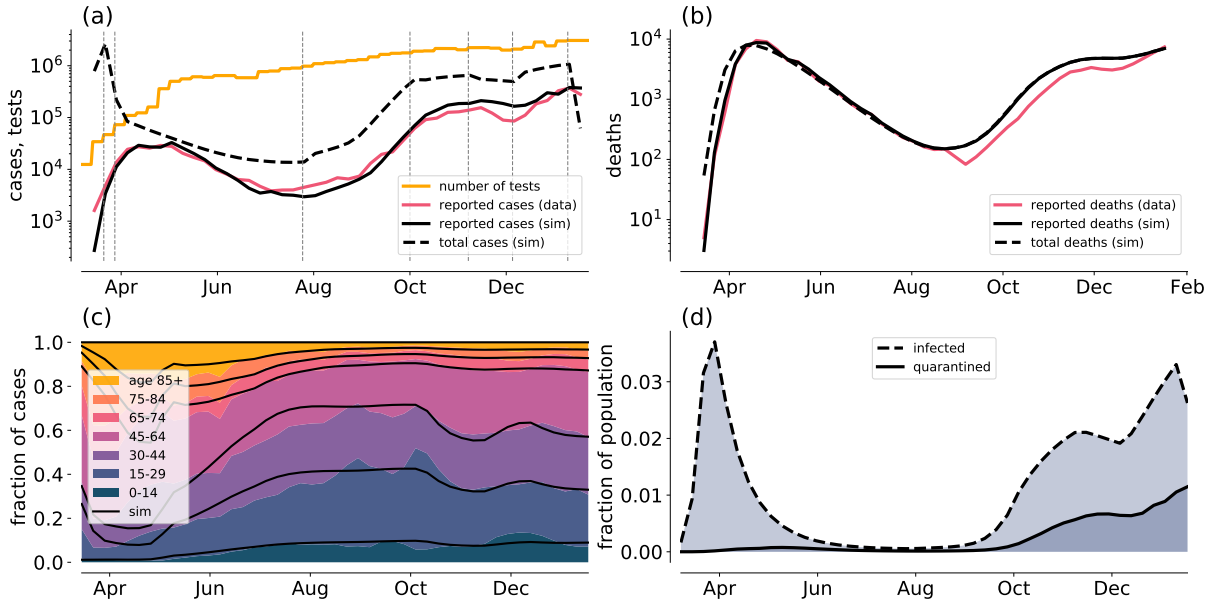

Figure 6: Plots of the MAP trajectories for model UK-A0. Expected trajectories are shown in black and labelled as 'sim' (solid for observable quantities, dashed for hidden quantities). Data are shown in colour. Panel (a) shows weekly diagnosed case numbers along with the total number of new infections and the number of tests performed. The vertical lines indicate times where interventions change. Panel (b) shows weekly deaths. Panel (c) shows the distribution of ages among the weekly new cases. Data are shown in colour, stacked and scaled to add up to 1. Analogously, boundaries between age groups in the simulation are shown in black. Panel (d) shows the prevalence as a fraction of infected (dashed) and quarantined (solid) individuals in the total population.

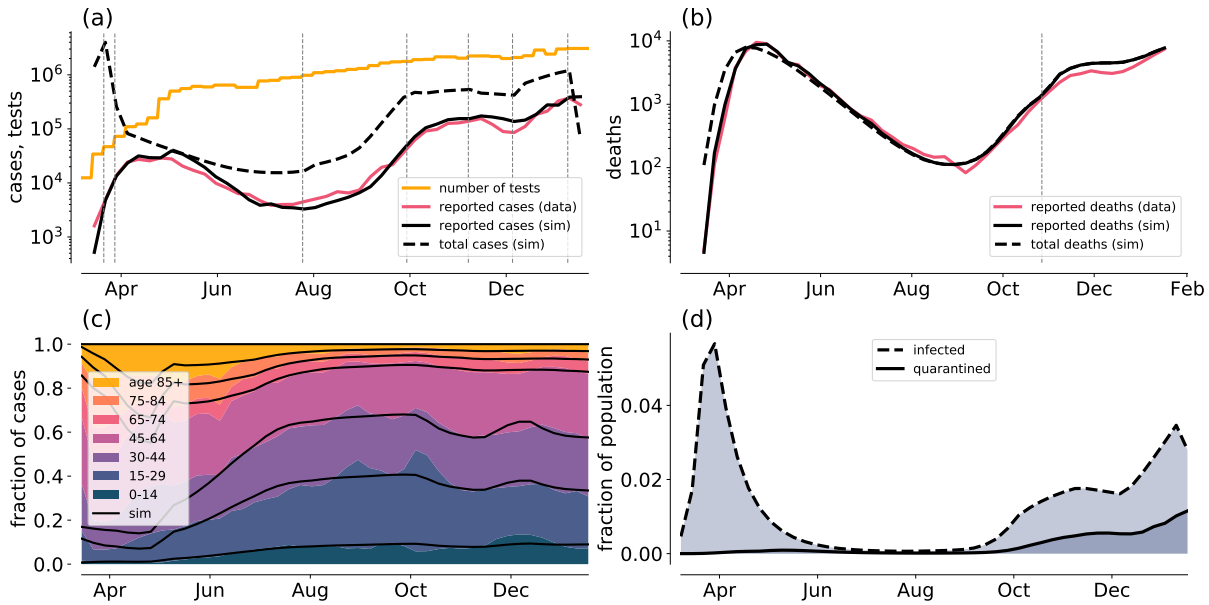

Figure 7: MAP trajectories for model UK-A1, as above. Additionally, the inferred time of change in IFR is indicated as a vertical line in panel (b).

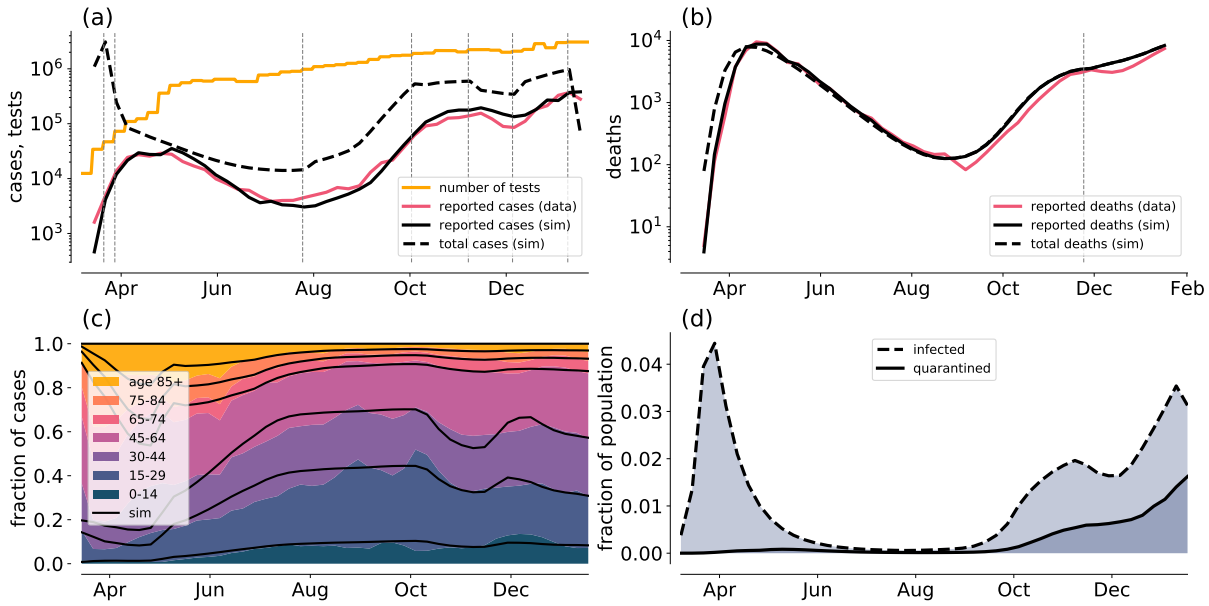

Figure 8: Model UK-A2

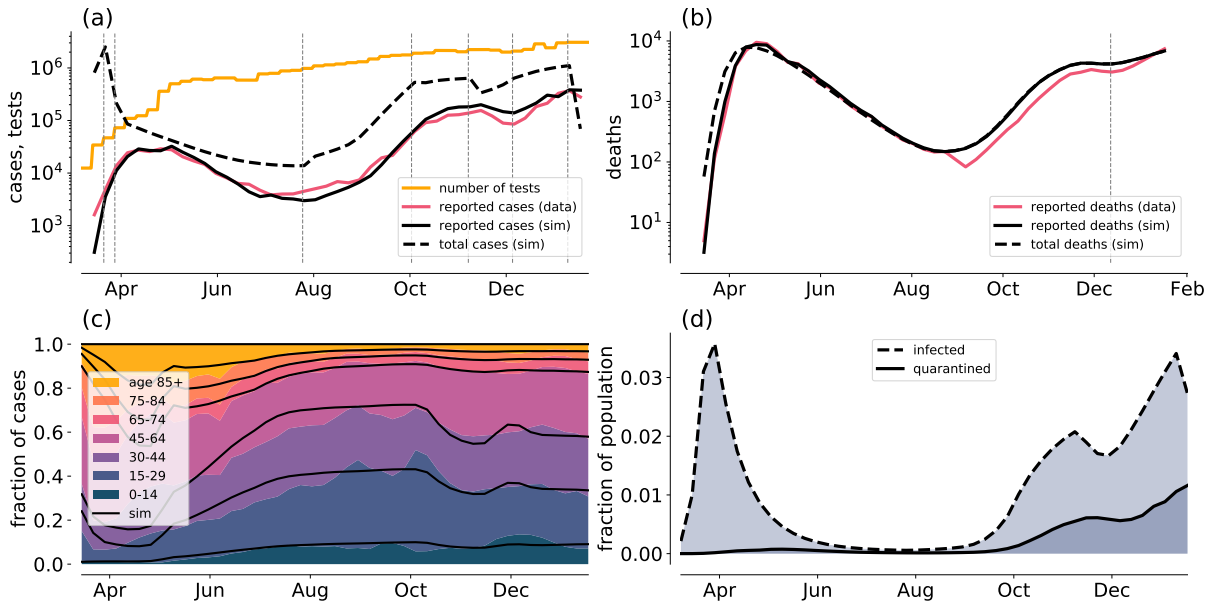

Figure 9: Model UK-B0

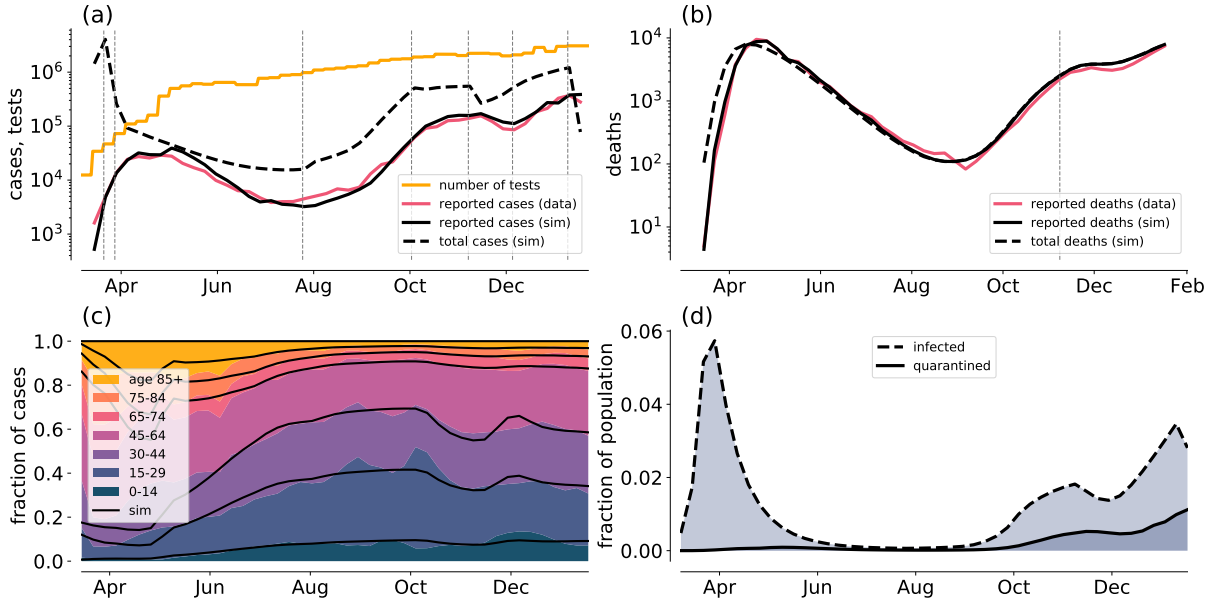

Figure 10: Model UK-B1

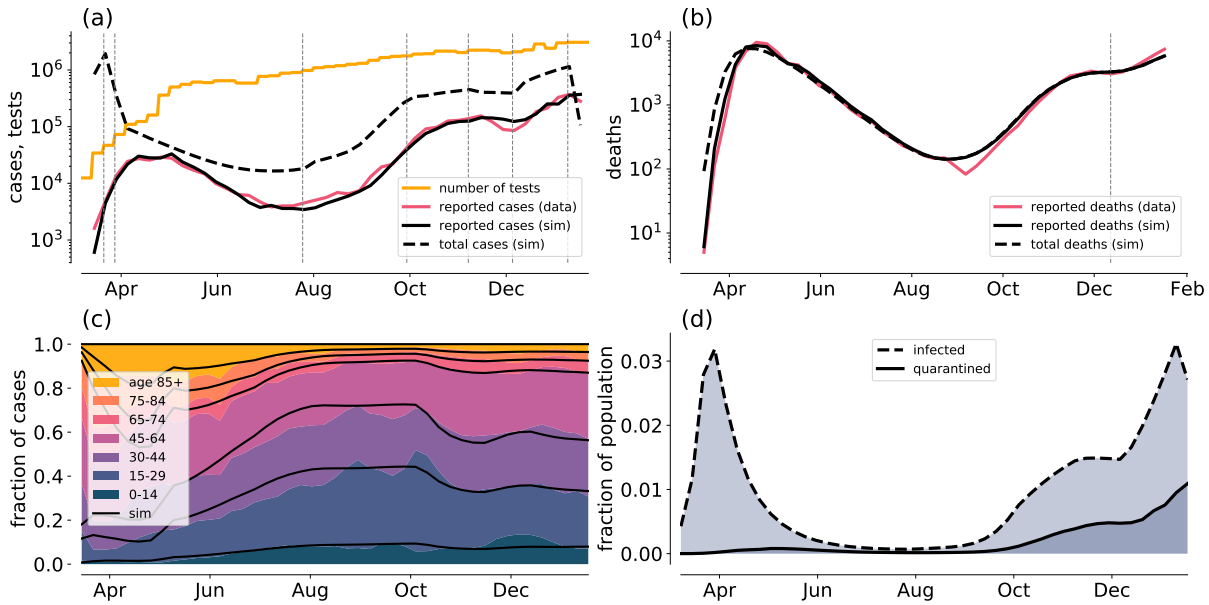

Figure 11: Model UK-C0

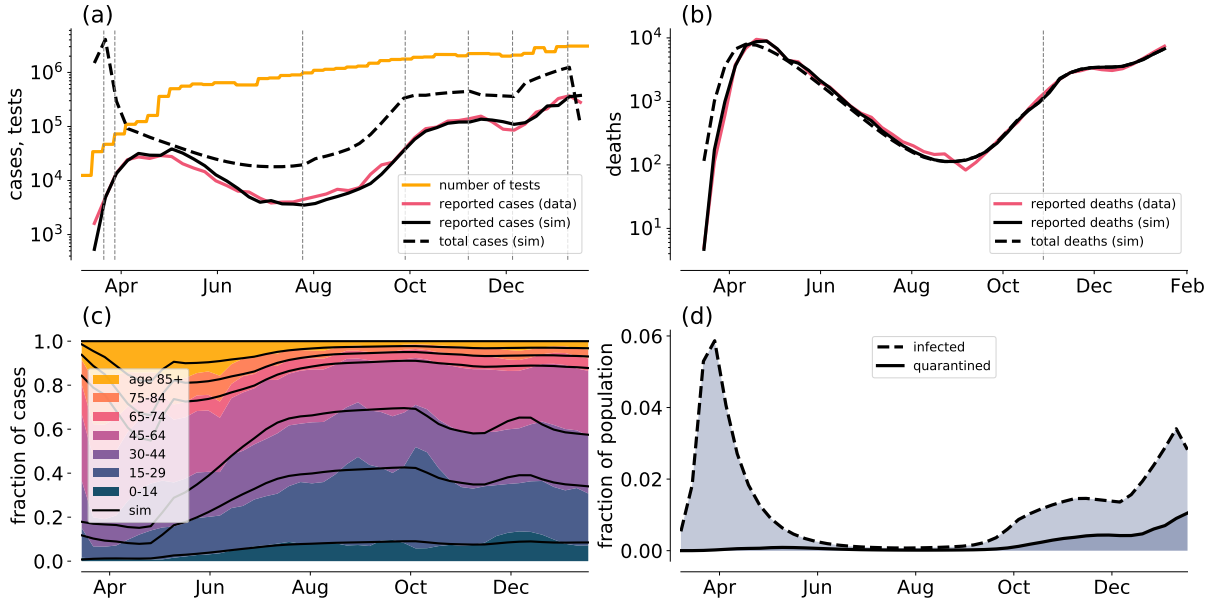

Figure 12: Model UK-C1

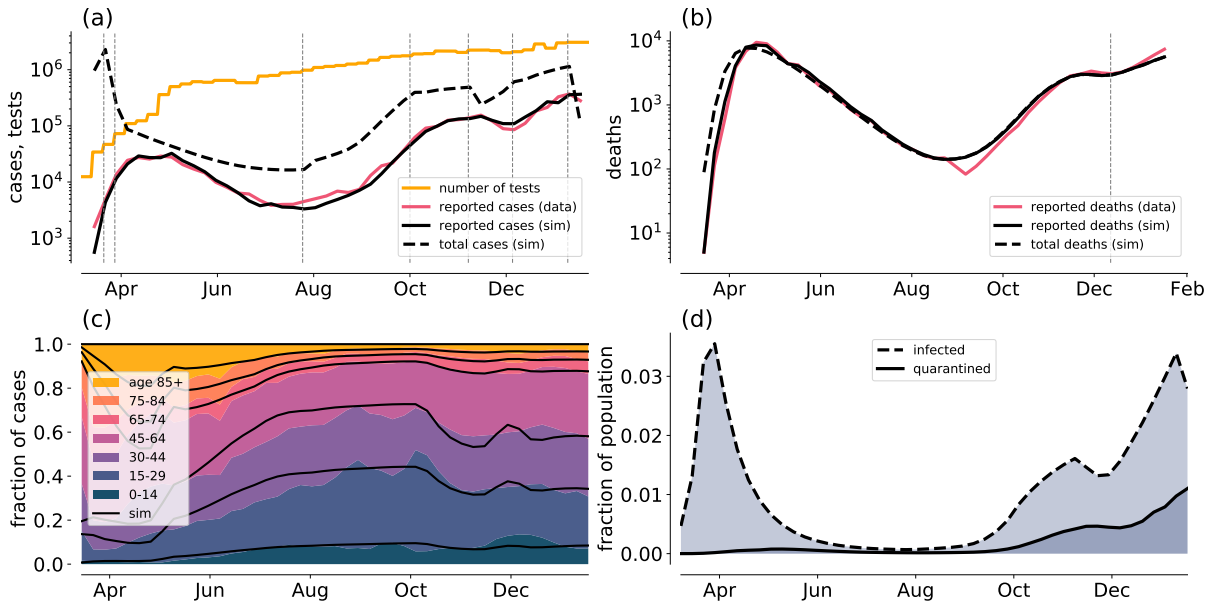

Figure 13: Model UK-BC0

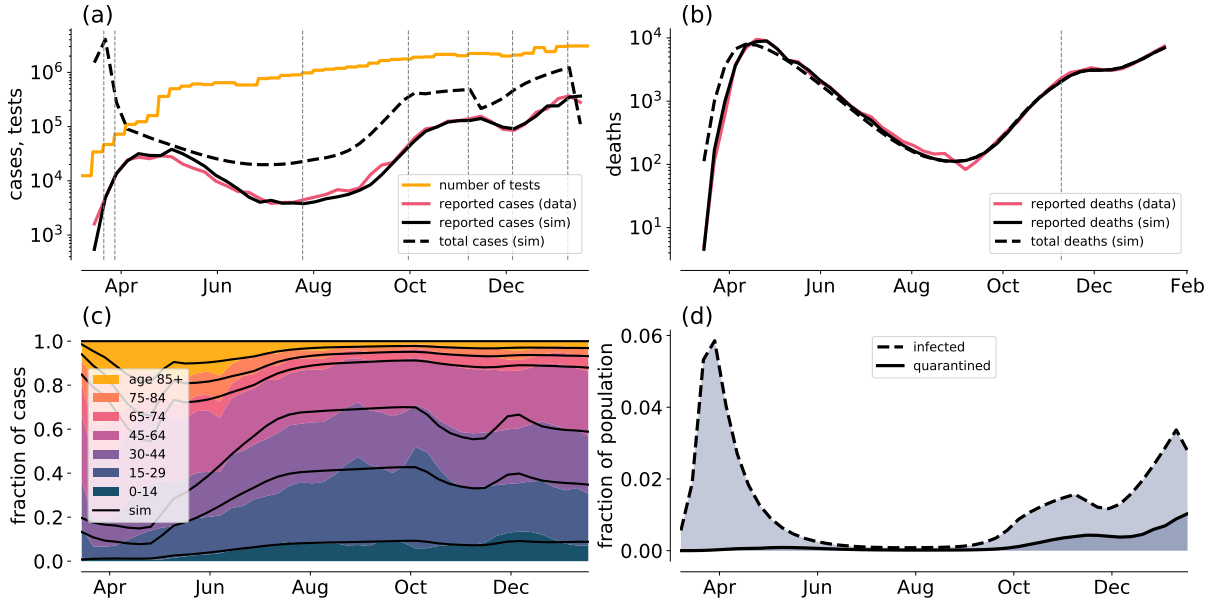

Figure 14: Model UK-BC1

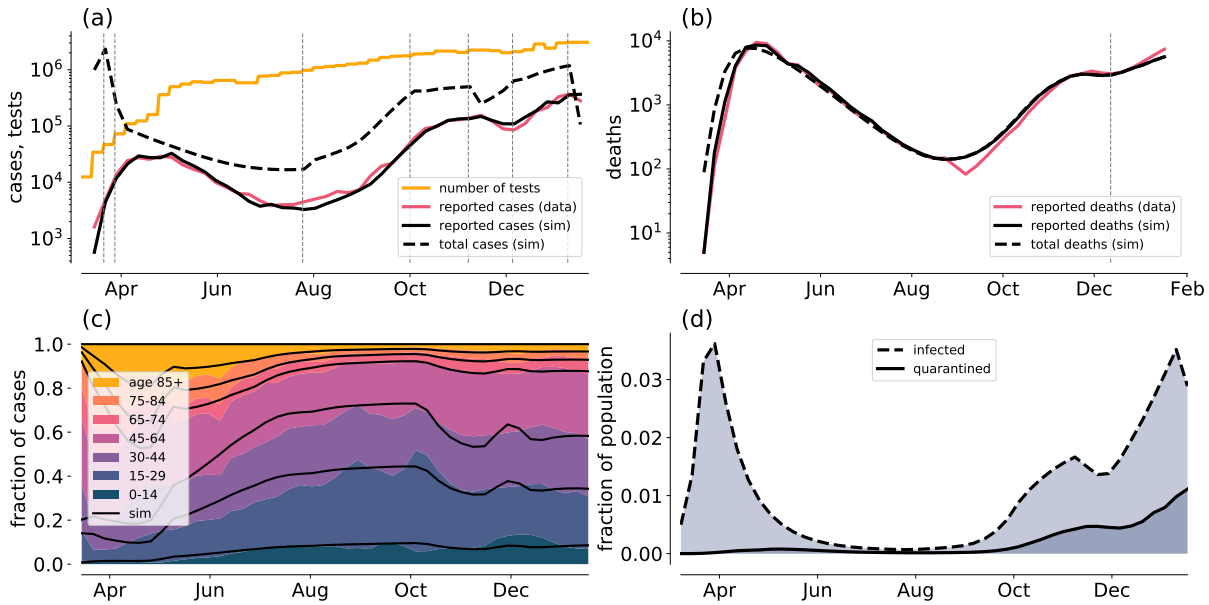

Figure 15: Model UK-TT0

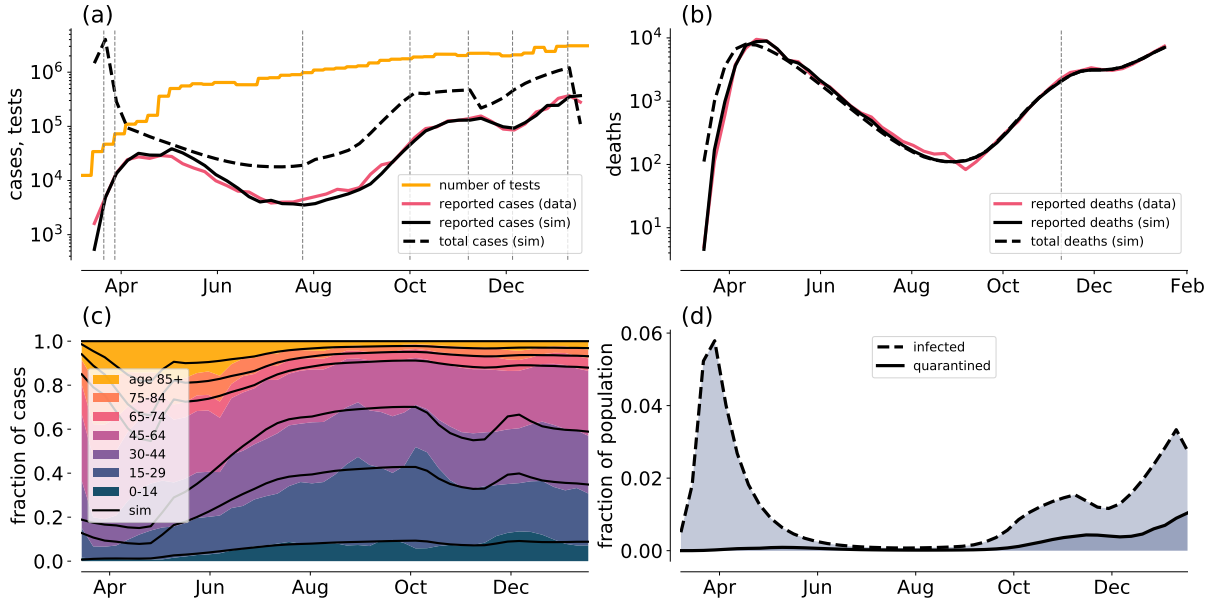

Figure 16: Model UK-TT1

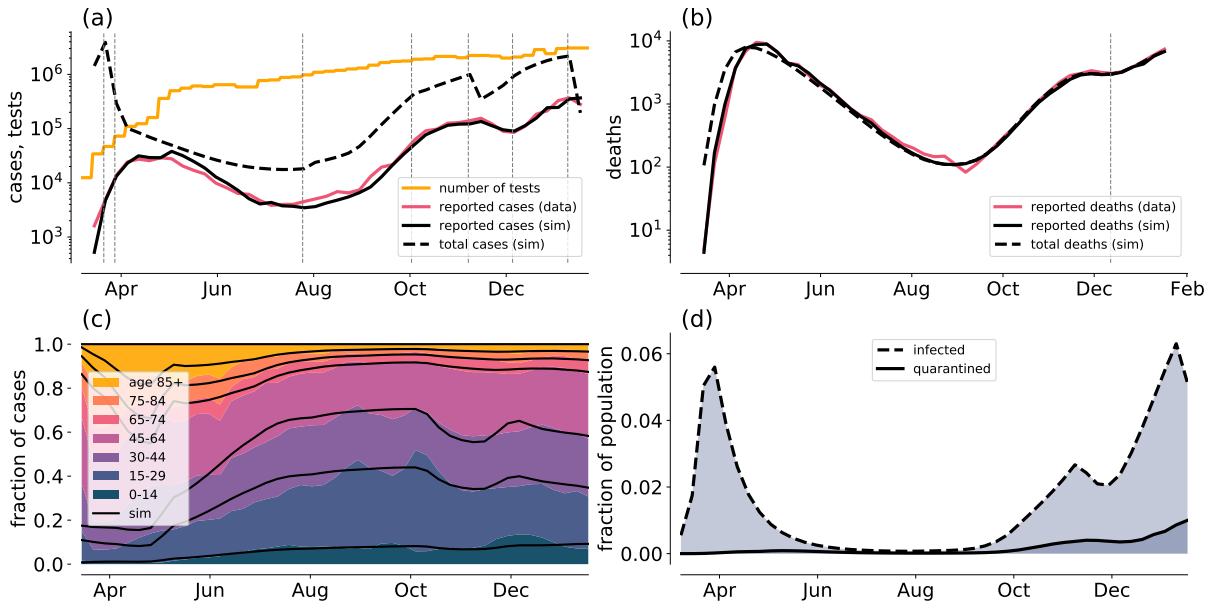

Figure 17: Model UK-P0

C.2. France

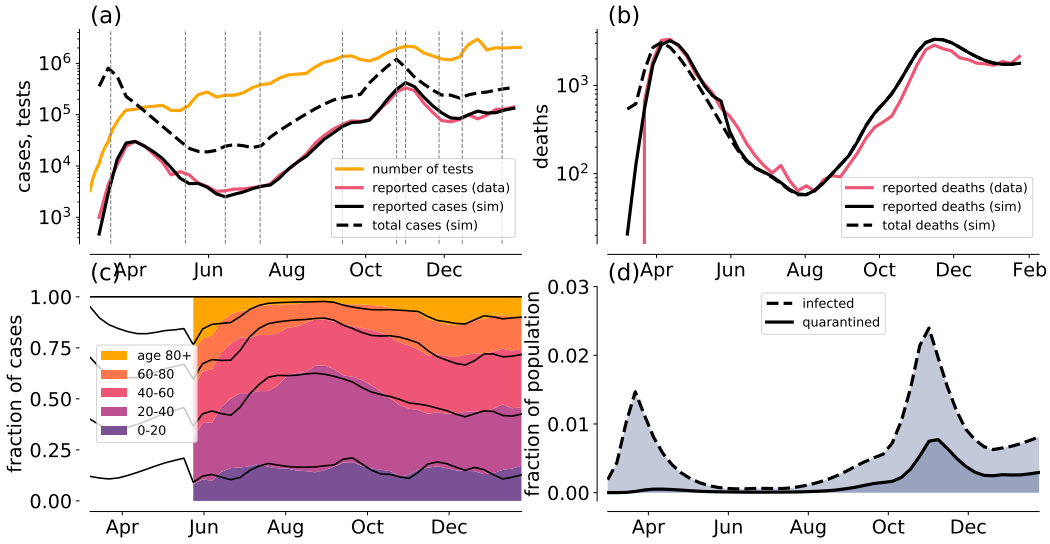

Figure 18: Model FRA C0.

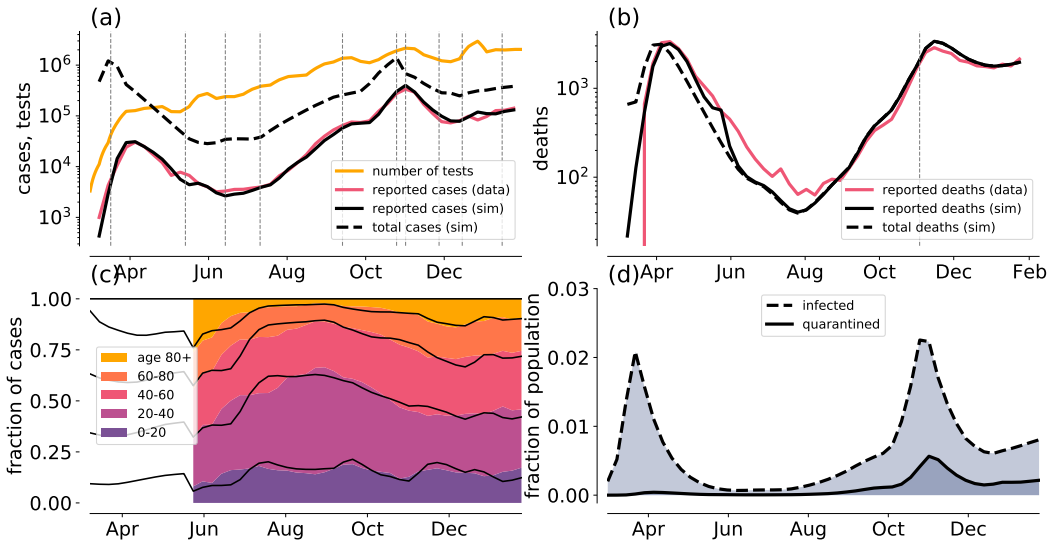

Figure 19: Model FRA C1.

C.3. Germany

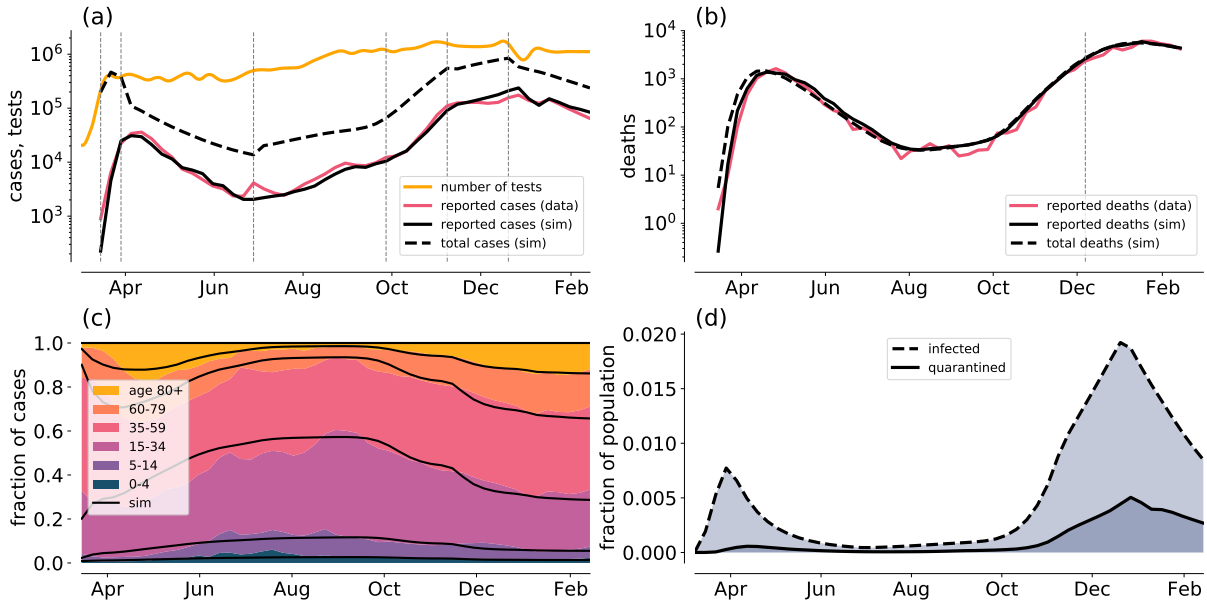

Figure 20: Model GER-C0

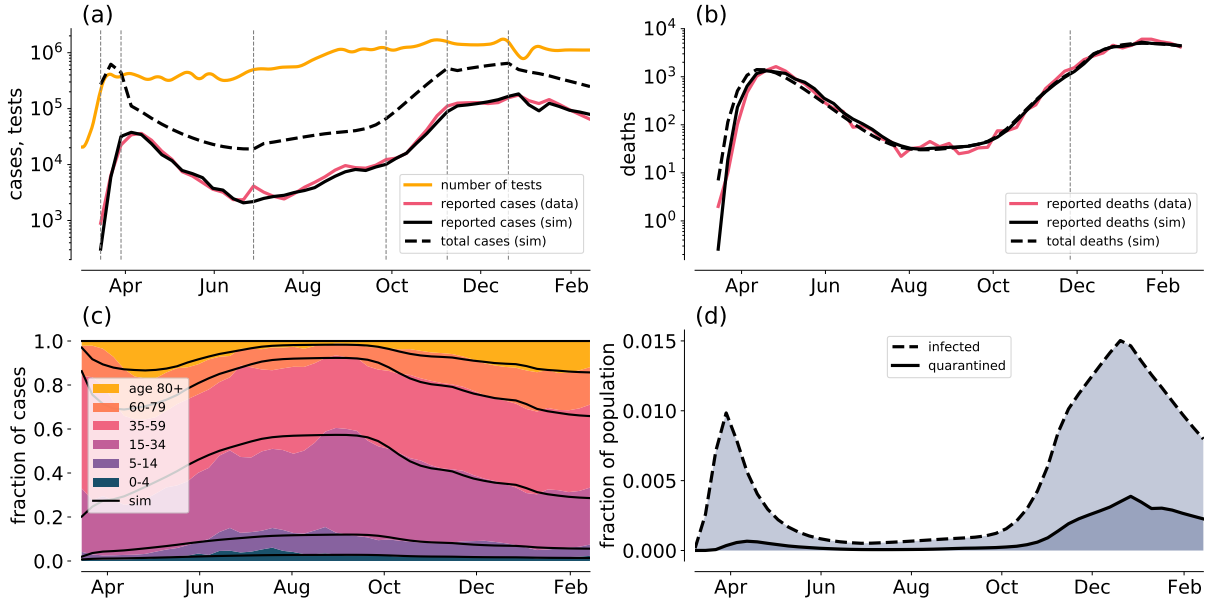

Figure 21: Model GER-C1