# Integrative modelling of reported case numbers and seroprevalence reveals time-dependent test efficiency and infection rates

Lorenzo Contento[1,*], Noemi Castelletti[2,*], Elba Raimúndez[1,5], Ronan Le Gleut[3,4], Yannik Schälte[3,5], Paul Stapor[3,5], Ludwig Christian Hinske[6], Michael Hölscher[2,7,8], Andreas Wieser[2,8,+], Katja Radon[8,9,10,+], Christiane Fuchs[3,4,5,11,+], Jan Hasenauer[1,3,5,+,†], and the KoCo19 study group[‡]

[1]Faculty of Mathematics and Natural Sciences, University of Bonn, Bonn, Germany
[2]Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, Munich, Germany
[3]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[4]Core Facility Statistical Consulting, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[5]Center for Mathematics, Technische Universität München, Garching, Germany
[6]Institut für medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Munich, Germany
[7]Center for International Health (CIH), University Hospital, LMU Munich, Munich, Germany
[8]German Center for Infection Research (DZIF), partner site Munich, Germany
[9]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Munich, Germany
[10]Comprehensive Pneumology Center (CPC) Munich, German Center for Lung Research (DZL), Munich, Germany
[11]Faculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany
[*]equal contribution first authors
[+]equal contribution last authors
[†]corresponding author: jan.hasenauer@uni-bonn.de
[‡]KoCo19 study group: Emad Alamoudi, Jared Anderson, Valeria Baldassare, Maximilian Baumann, Marc Becker, Marieke Behlen, Jessica Beyerl, Rebecca Böhnlein, Anna Brauer, Vera Britz, Friedrich Caroli, Lorenzo Contento, Alina Czwienzek, Flora Deák, Emma Dech, Laura Dech, Jana Diekmannshemke, Anna Do, Gerhard Dobler, Jürgen Durner, Ute Eberle, Judith Eckstein, Tabea M. Eser, Philine Falk, Volker Fingerle, Stefanie Fischer, Marius Gasser, Sonja Gauder, Otto Geisenberger, Christof Geldmacher, Leonard Gilberg, Kristina Gillig, Philipp Girl, Elias Golschan, Vitus Grauvogl, Celina Halfmann, Tim Haselwarter, Arlett Heiber, Matthias Herrmann, Stefan Hillmann, Christian Hinske, Janna Hoefflin, Tim Hofberger, Michael Höfinger, Larissa Hofmann, Sacha Horn, Kristina Huber, Christian Janke, Ursula Kappl, Charlotte Kiani, Isabel Klugherz, Norah Kreider, Arne Kroidl, Magdalena Lang,

Clemens Lang, Silvan Lange, Ekaterina Lapteva, Michael Laxy, Reiner Leidl, Felix Lindner, Alexander Maczka, Alisa Markgraf, Paula Matcau, Rebecca Mayrhofer, Anna-Maria Mekota, Hannah Müller, Katharina Müller, Leonie Pattard, Claire Pleimelding, Stephan Prückner, Konstantin Pusl, Elba Raimúndez, Julius Raschka, Jakob Reich, Raquel Rubio-Acero, Nicole Schäfer, Paul Schandelmaier, Lara Schneider, Sophie Schultz, Mirjam Schunk, Lars Schwettmann, Heidi Seibold, Paul Stapor, Jeni Tang, Fabian Theis, Sophie Thiesbrummel, Eva Thumser, Niklas Thur, Julian Ullrich, Julia Waibel, Claudia Wallrauch, Julia Wolff, Pia Wullinger, Tobias Würfel, Patrick Wustrow, Houda Yaqine, Sabine Zange, Eleftheria Zeggini, Thorbjörn Zimmer, Thomas Zimmermann, Lea Zuche.

October 1, 2021

## Abstract

Mathematical models have been widely used during the ongoing SARS-CoV-2 pandemic for data interpretation, forecasting, and policy making. However, most models are based on officially reported case numbers, which depend on test availability and test strategies. The time dependence of these factors renders interpretation difficult and might even result in estimation biases.

Here, we present a computational modelling framework that allows for the integration of reported case numbers with seroprevalence estimates obtained from representative population cohorts. To account for the time dependence of infection and testing rates, we embed flexible splines in an epidemiological model. The parameters of these splines are estimated, along with the other parameters, from the available data using a Bayesian approach.

The application of this approach to the official case numbers reported for Munich (Germany) and the seroprevalence reported by the prospective COVID-19 Cohort Munich (KoCo19) provides first estimates for the time dependence of the under-reporting factor. Furthermore, we estimate how the effectiveness of non-pharmaceutical interventions and of the testing strategy evolves over time. Overall, our results show that the integration of temporally highly resolved and representative data is beneficial for accurate epidemiological analyses.

## Introduction

Social distancing, mask wearing, lockdowns and other non-pharmaceutical interventions (NPIs) are used worldwide to slow the spread of SARS-CoV-2 and to avoid overburdening health care systems. Various studies have analysed how such NPIs influence the contact rate (Latsuzbaia et al. 2020; Jarvis et al. 2020), the infection rate (Courtemanche et al. 2020; Hartl, Wälde, and Weber 2020; Li et al. 2020; Lyu and Wehby 2020; Siedner et al. 2020), the reproduction number (Giordano et al. 2020; Zhao and Chen 2020; Liu et al. 2021; Brauner et al. 2021; Sypsa et al. 2021) and related quantities. These studies use a broad spectrum of analysis approaches, including statistical methods (e.g., generalized linear models, generalized estimating equations, machine

learning, . . . ) (Streeck et al. 2020; Latsuzbaia et al. 2020; Siedner et al. 2020; Courtemanche et al. 2020), compartmental models based on ordinary differential equations (ODEs) (Barbarossa et al. 2020; Jarvis et al. 2020), and agent-based models (Lorch et al. 2021), and provide important insights. However, a limitation of most studies is that they are exclusively based on official case numbers.

The officially reported case numbers provide in most countries information about the number of positive tests for viral load registered on a specific day. Such tests can either be based on the polymerase chain reaction (PCR) or on antigen detection; for brevity's sake, in the following we will refer to them indiscriminately as PCR tests. However, there are several well-known issues with these numbers (Raimúndez et al. 2021). Besides reporting delays, the most important problems are that case numbers depend on the availability of tests and on the test strategy. Clearly, the number of performed tests and the selection criteria have changed over time (e.g., due to the introduction of antigen tests). As an alternative to the officially reported case numbers, the officially reported death numbers, which are generally considered as more reliable (Radon, Bakuli, et al. 2021; Pritsch et al. 2021), can be used. However, there the effects of NPIs are smoothed over time and only visible after a substantial delay. Furthermore, the observation can be confounded by the quality of medical care, in particular if the number of beds in intensive-care units becomes a limiting factor. In summary, while case and death numbers provide information on the progression of an epidemic, the interpretation is often difficult.

The ideal data-source for the analysis of NPIs as well as the efficiency of test strategies would be a thorough monitoring of a large representative population cohort. Yet this is rather time- and resource-consuming, and in most cases not realistic. Cross-sectional studies based on PCR tests with a high time resolution would provide a comprehensive picture of the spread within populations, but the number of required tests would be very high. For a prevalence of 100 in 100,000 individuals, on average 1000 tests have to be performed to observe a single positive individual. For tight confidence intervals, tens of thousands of PCR tests would be necessary per time point. To make things worse, PCR tests are only positive for a short period after exposure to the virus.

An alternative to PCR tests that allows for the monitoring of epidemics is testing for the presence of antibodies, which assesses seroprevalence . The antibody response is rather stable and can usually be detected even months after the initial infection (Radon, Bakuli, et al. 2021; Olbrich et al. 2021; Isho et al. 2020). Accordingly, antibody tests do not only provide a snapshot of the current situation as PCR tests do, but inform about previous exposure and hence the past history of the epidemic up to the current point. However, to provide an assessment of NPIs and test strategies with a high temporal resolution, a large cohort and immense resources would be required in this case too. This means that the official case numbers (which are anyhow collected for a large population) and representative population studies need to be combined.

Here, we report an analysis of the time-dependent infection rate and the time-dependent efficiency of the test strategy for the spread of SARS-CoV-2 in Munich. We use a compartment model for integrating officially reported case and death numbers, as well as seroprevalence data from the population-based prospective COVID-19 cohort study KoCo19 in Munich (Radon, Saathoff, et al. 2020; Pritsch et al. 2021). The time-dependent rates of infectious contacts under different NPIs, of the test efficiency, and of the reporting delay, are parametrized using splines

and inferred from the data. We show that this approach provides a more stable assessment of the progression dynamics than standard approaches solely based on case numbers.

## Results

### Compartmental model for the COVID-19 epidemic

We developed a compartmental model for the dynamics of the COVID-19 epidemic that allows for the integration of the officially reported numbers of positive PCR tests and COVID-19 related deaths, hospital bed usage, and seroprevalence in representative cohort studies (Figure 1). The model describes the state of individuals: susceptible, exposed, infectious, hospitalized, recovered and deceased. Several of these generic states can be further refined (Figure 1A) by distinguishing infectious individuals with and without symptoms, recovered individuals with and without detectable virus, and hospitalized individuals in ward and intensive care unit (ICU).

Realistic distributions of transition times of individuals between states are achieved by subdividing the illness-related states into multiple sub-states (Figure 1B). This is often referred to as the Gamma Chain Trick or Linear Chain Trick (Smith 2011). Many important transition times related to the COVID-19 disease cannot be reasonably modelled by an exponential distribution (single sub-state case) — a fact that is disregarded by many studies. For example, the incubation time has been shown to be approximately log-normally or Weibull-distributed (Linton et al. 2020; McAloon et al. 2020) and can be reasonably approximated by an Erlang distribution with shape parameter 6 (Lauer et al. 2020), suggesting a split into six sub-states.

The testing process is modelled by splitting illness-related states into two sub-states: infectious individuals that have not been detected, and infectious individuals that have been detected by means of a positive PCR test. The rates at which individuals transition from the undetected to the detected branch reflect the efficacy of the testing system set up by the healthcare authorities. As individuals with symptoms are more likely to get tested, we assume that these detection rates depend on the illness phase. In particular, the detection rate for asymptomatic and presymptomatic individuals are especially important since they can be considered a measure of contact tracing effectiveness. Hospitalized individuals are assumed to be immediately detected. The number of reported positive PCR tests is then equal to the sum of all fluxes from undetected to detected sub-states.

Since antibodies become detectable only about two weeks after an exposed individual has become infectious (Long et al. 2020), it is not possible to obtain the current number of seropositive individuals from the model state. Instead, we compute the number of individuals who will be seropositive in two weeks' time, which is equal to total population size minus the number of susceptible, exposed (but not yet infectious) and deceased individuals.

The rate at which susceptible individuals are infected is proportional to the sum of the number of asymptomatic individuals, presymptomatic individuals, symptomatic individuals, recovered individuals with detectable virus levels, and hospitalized individuals in ward (ICU patients are considered to be not infectious). The elements of this sum must be weighted by degree of infectiousness, which is specific of each compartment. In particular, this infectiousness level depends both on biological factors, such as the stage to which the illness has progressed , and

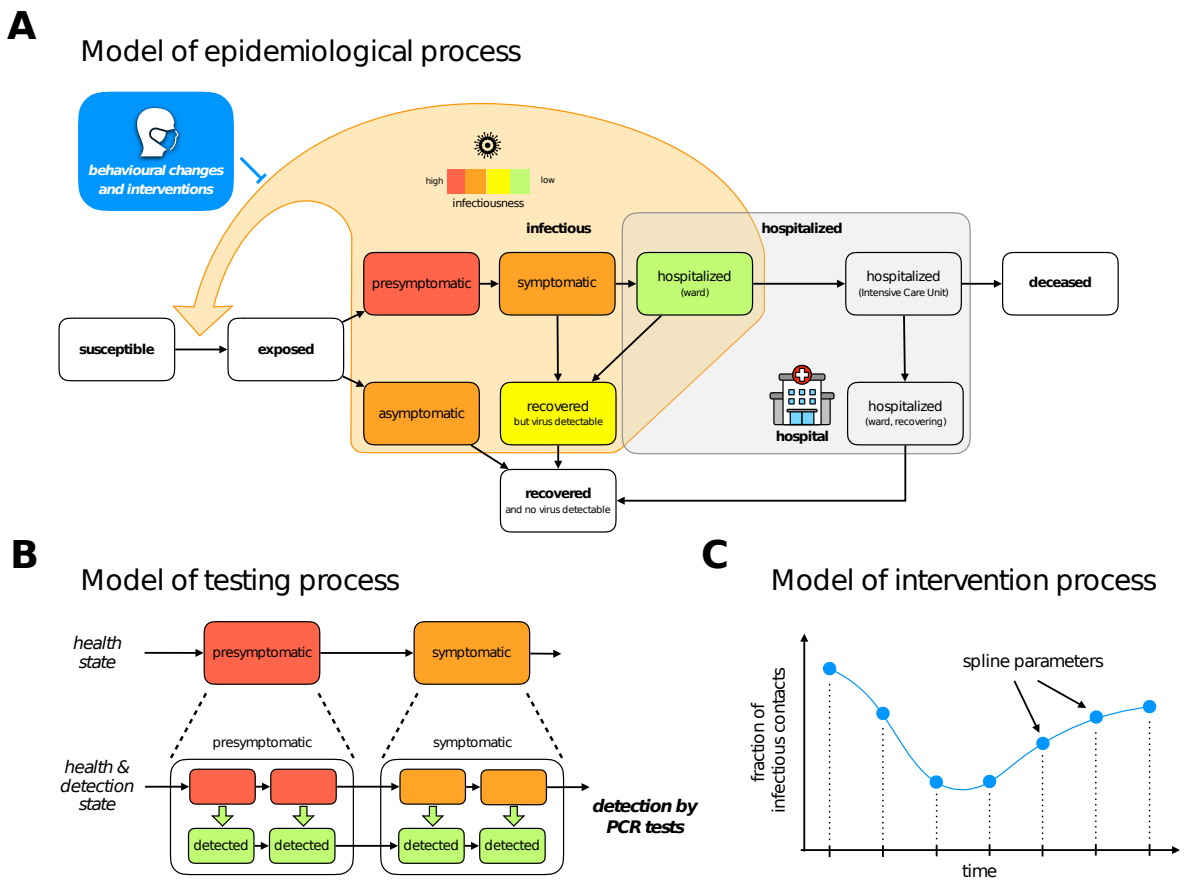



Figure 1: **Structure of the compartment model.** (A) High-level structure indicating possible transitions between various illness phases and hospitalization compartments. Infectious phases are coloured with different colours encoding the degree of infectiousness. (B) Detailed structure of the compartment model. Each compartment is split into several sub-states in order to have Erlang-distributed transition times between compartments, and to explicitly model the testing process by tracking individuals reported to the health care authorities on a parallel but separate branch. (C) Time-dependent parameters (here the viral transmission reduction due to NPIs is used as an example) are modelled by splines which can be encoded inside the parameter vector by their values at the grid points.

behavioural factors, which in turn depend strongly on whether the infected individual has been detected or not (detected individuals are quarantined and thus less likely to spread the disease). The qualitative ordering is indicated by colours in Figure 1A, but the quantitative contributions are considered to be unknown a priori.

## Statistical framework for inferring testing and intervention effects

The compartment model describes possible transitions between states, but the transition rates are unknown and depend on many factors. Particularly relevant is the influence on the infection and detection rates due to the policies set by the government and the healthcare authorities. We model such effects with three parameters and, since policies have evolved over time, we assume these parameters to be time-dependent, modelling them using splines (Figure 1C).

The first two are the detection rates for (i) symptomatic and (ii) asymptomatic/presymptomatic individuals respectively. These rates are influenced by testing capacity, by the effectiveness of contact tracing, and by the criteria to be met in order to be eligible for testing, all of which have undergone considerable changes in the first months of the epidemic. The third is a fractional reduction in the number of infectious contacts compared to the situation at the beginning of the epidemic. Such a reduction is strongly influenced by government restrictions such as business/school closures or lockdowns, but also depends on the compliance of the general population with social distancing and other preventive measures. In all cases time dependency is modelled by Catmull-Rom splines with an inter-node distance fixed at 2 weeks (see *Material and Methods*). Moreover, the evolution of the detection rates has an additional week-periodic component in order to account for the lower case counts during weekends that have been observed in Germany and elsewhere in the world.

To determine the time-dependent infection and detection rates as well as other unknown parameters, we use a Bayesian formulation to integrate different pieces of information, including seroprevalence observed in representative cohort studies, reported case numbers, and prior knowledge on process parameters. As likelihood function we employ the product of the conditional probabilities of observing the individual datasets (see *Material and Methods*). The distribution of the conditional probabilities depends on the characteristics of the different datasets and was determined by testing different competing choices. For the reported number of infected and deceased individuals, the key observation is that their variance increases together with the count values, indicating that a constant-variance additive normal noise model is not appropriate. We opted instead for a negative binomial distribution with constant variance-to-mean ratio, since it is a commonly used choice for count values (Beck and Tolnay 1995) and has been successfully employed in modelling case numbers for the COVID-19 pandemic (Lin et al. 2021; Chan et al. 2021). For the number of hospitalized individuals, we have seen that a constant-variance additive normal noise model works well enough due to their limited dynamic range; while employing a negative binomial distribution in this case too would probably be a superior choice, the gains would not be as strong as in the case of reported infections/deaths and thus not justify the consequent increase in complexity of the likelihood function. As for the seroprevalence data, the number of positive antibody tests (aggregated by equally long time intervals) is directly fit as a result of a binomial random variable, whose probability of success is given by the fraction of seropositive individuals predicted by the model. The priors for the various model parameters were extracted from various published reports (see Supplementary Tables 1 and 2).

The Bayesian posterior distribution over the parameter space was analysed using Markov chain Monte Carlo sampling (see *Material and Methods*), which facilitates a quantitative assessment of the available information, including information about parameter and prediction uncertainties. Due to the high numbers of parameters to be optimized (75) and the relatively high cost of each simulation, sampling was particularly time-consuming.

## Modelling without representative data provides uncertain estimates

As most studies are only based on officially reported case numbers, we first studied the reliability of such an approach. Therefore, we inferred the model parameters from the reported number of infected, hospitalized and deceased individuals for the city of Munich in Germany.

In addition to these commonly used counts, we also employed the publicly available number of reported symptom onsets, which has been rarely used in other modelling efforts but can be easily integrated in our case thanks to our explicit description of the detection process. The number of cases (new infections, deaths and symptom onsets) was extracted from the official report by the Robert Koch Institute (2020a), while the hospital usage was obtained from the web-based information system IVENA. The city of Munich was selected as detailed seroprevalence results are available from the KoCo19 study (Radon, Saathoff, et al. 2020; Pritsch et al. 2021).

Inferred parameter estimates capture correctly the case numbers used for fitting (Figure 2A). Yet, many of the estimated parameters are not well determined, as evinced by the broad credible intervals (Figure 2B). To assess the reliability of the predictions, we employed the posterior samples to predict the seroprevalence and compared it with the KoCo19 data, which was not used for fitting (Figure 2A, top-right panel). The prediction given by the most likely parameters encountered during sampling is compatible with the observed seroprevalances. However, as can be seen from Figure 4, the total number of cases predicted by the model has very large credible intervals. This shows that officially reported case numbers are, even in combination with prior knowledge, insufficient to predict the the actual number of COVID-19 infections during the epidemic with a satisfactory degree of confidence.

## Modelling with representative data reduces uncertainties

In order to quantify the added value provided by prevalence data obtained by extensive serological testing, we extended the previously used dataset with the time-dependent prevalence reported by KoCo19 and performed the same Bayesian parameter estimation procedure. As in the previous analysis, the obtained parameter estimates provide an accurate description of all available data (Figure 3A) and the uncertainty on the values of the single parameters is quite large (Figure 3B, for the full posterior distributions see Supplementary Figure 3). The uncertainty on the hidden states of the model (Figure 4) is instead greatly reduced, in particular for the total number of cases (not surprisingly, since this quantity is closely linked to the seroprevalence level) and for the number of asymptomatic cases, showing the effectiveness of the seroprevalence data in reducing uncertainty. Figure 4 also shows how the actual number of infections is substantially higher than the number of reported cases, highlighting the limitations of the publicly reported case counts.

## Model reveals efficiency of testing strategy

Using the compartment model integrating case reports and representative data, we studied the effectiveness of the testing and NPI strategy. To do that, we computed from the posterior samples the time-dependent detection and number of infectious contacts, along with several secondary properties.

Instead of employing directly the time-dependent detection rates, in order to evaluate the effectiveness of the testing strategy we use a more easily interpretable metric: the probability that an infected individual is reported to the healthcare authorities before the virus is cleared from their system (Figure 5A,B). The effectiveness of testing increased gradually as the epidemic progressed. The model estimates that at the beginning of March only between 20 and 60% of

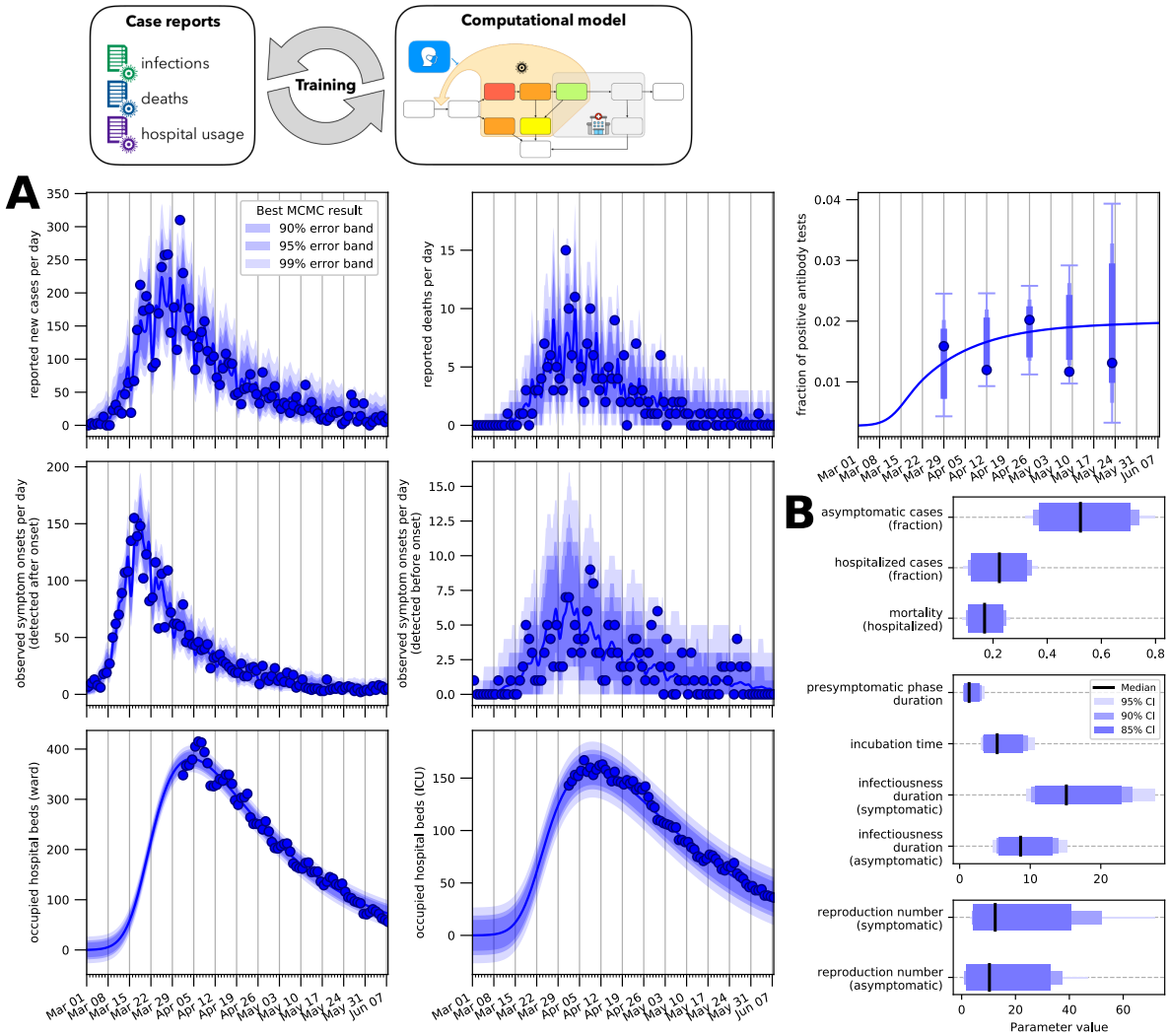



Figure 2: **Modelling and parameter estimation using case reports and hospital usage for Munich, Germany.** (A) Model simulation for the sampled parameter vector with the highest posterior probability compared with the observations used for fitting: case numbers reported by the Robert-Koch Institute and hospital usage for Munich. The error bands show the range of plausible values for the observation, confirming that the noise models used are appropriate. The top-right panel shows the seroprevalence predicted by the model (which was not used for fitting) compared with the observed seroprevalence. Here the error bars are only shown at the observation times since the variance of each observation is linked to the number of total antibody tests carried out in each sub-batch. (B) Credible intervals (bars) and median value (line) for a subset of the model parameters.

symptomatic individuals were reported, while in April and May the fraction was 30% to almost 100%. For asymptomatic cases, the fraction increased from practically 0% to between 10 and 40%. The resulting overall fraction for April and May is estimated to be 25 to 65%.

For the time-dependent reduction in infectious contacts, which models both the effect of government policy and behavioural changes, we observed an opposite effect compared to the detection rate (Figure 5C). Fixed to be 1 on the first day of March, this factor immediately started to drop quickly. Similarly, the effective reproduction number dropped from above three at the beginning

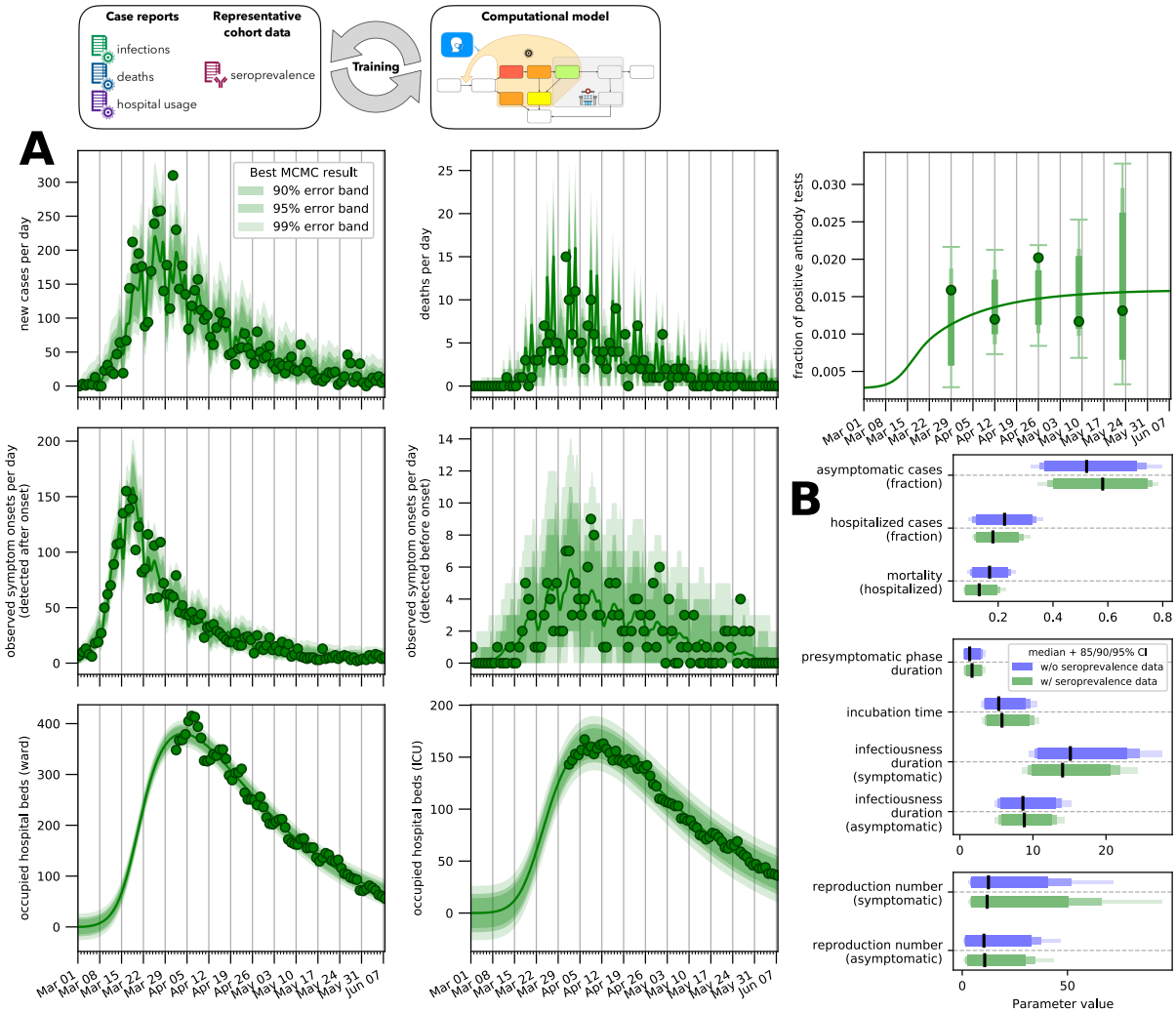



Figure 3: **Modelling and parameter estimation using case reports and hospital usage for Munich, Germany.** (A) Model simulation for the sampled parameter vector with the highest posterior probability compared with the observations used for fitting. The error bands show the range of plausible values for the observation, confirming that the noise models used are appropriate. (B) Credible intervals (bars) and median value (line) for a subset of the model parameters. Results for both the model fitted with seroprevalence data and the one fitted without are reported for ease of comparison.

of March to below one after the middle of March (Figure 5D). This coincides with the raising of public awareness (e.g., a speech by the German chancellor) and various interventions.

The effective reproduction number is influenced by both NPIs and by the testing strategy and it is thus important to deconvolute these two effects. In order to do so, we computed the evolution of the reproduction number for two hypothetical scenarios (Figure 6): one in which NPIs are not applied and one in which testing is not performed (or, equivalently, detected people are not quarantined). This revealed that testing results in a small improvement over what can be achieved with NPIs alone. However, in absence of NPIs such as the lockdown, the testing strategy implemented during the first epidemic wave in Germany would not have been able to lower the reproduction number enough to stop the spreading of the disease.

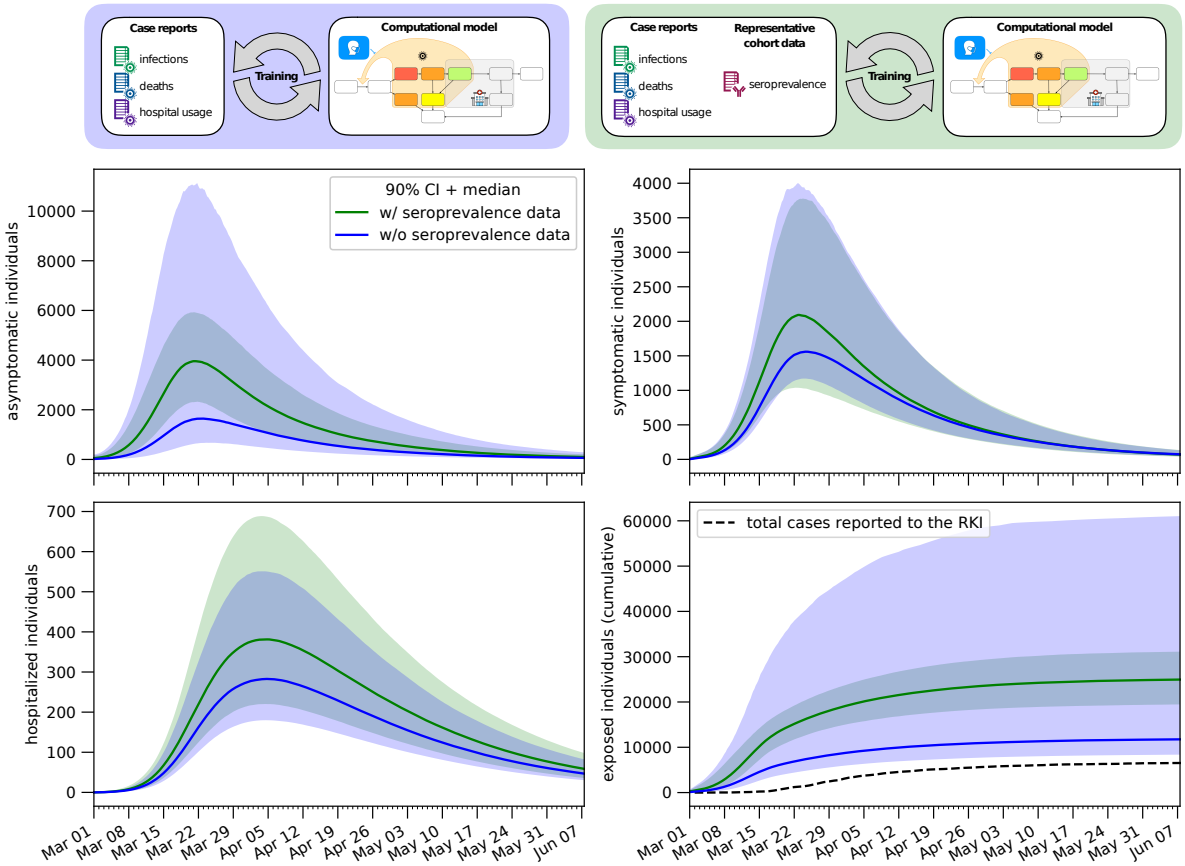



Figure 4: **Estimates for the hidden dynamics of the epidemic.** The number of individuals in different compartments as a function of time is plotted for both the model fitted with seroprevalence data and the one fitted without. In the bottom-left panel the cumulative number of cases detected by the healthcare authorities is also plotted for reference. The bands correspond to 90% posterior credible intervals, while the solid line denotes the median value.

## Discussion

Test availability, testing strategy, governmental interventions and various factors have changed over the course of the COVID-19 epidemic in Germany. This renders the interpretations of the reported case numbers difficult, while creating the need to infer time-dependent characteristics (e.g., to assess the impact of strategies). Here, we approached both points by integrating officially reported case numbers with representative seroprevalence observations using integrative modelling and Bayesian parameter estimation framework. Our analysis revealed that the integration of datasets is critical: The amount of available seroprevalance data was too limited to build models just based on them, while case report numbers on their own resulted in vary large uncertainties on the hidden states of the model, especially on the number of asymptomatic cases (Figure 4).

In addition to the use of various data sources, a strength of our study is the rigorous application of Bayesian uncertainty quantification. The vast majority of models for the COVID-19 epidemic we have found in the literature were not accompanied by uncertainty quantification. Some exceptions exist, such as a study by Lin et al. (2021), in which uncertainty estimates are given for the observed case number and the parameter values. However, they chose not to estimate

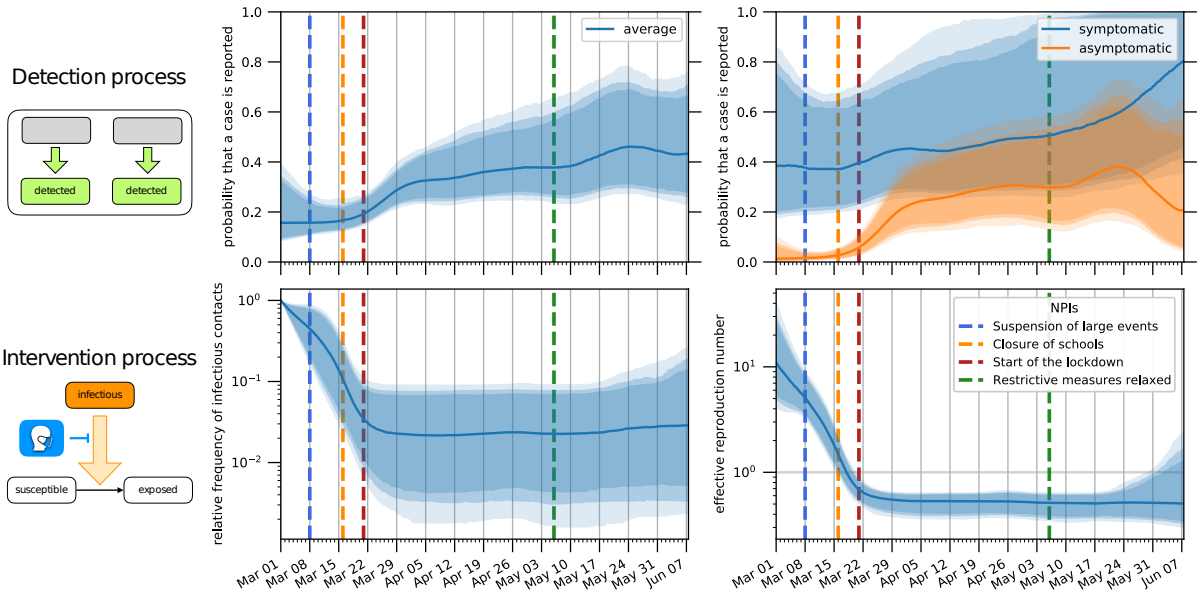



Figure 5: **Estimation of time-dependent parameters.** The bands correspond to 85%/90%/95% posterior credible intervals, while the solid line denotes the median value. Dashed lines indicate at which dates specific NPIs were enforced/lifted. (A, B) Probability that an infected individual is detected and reported to the healthcare authorities before the virus is cleared from their system. (A) contains the average value, while (B) plots separately for symptomatic and asymptomatic/presymptomatic individuals respectively. (C) Reduction in the number of infectious contacts due to NPIs and behavioural changes, relative to the situation on March 1. (D) Effective reproduction number.

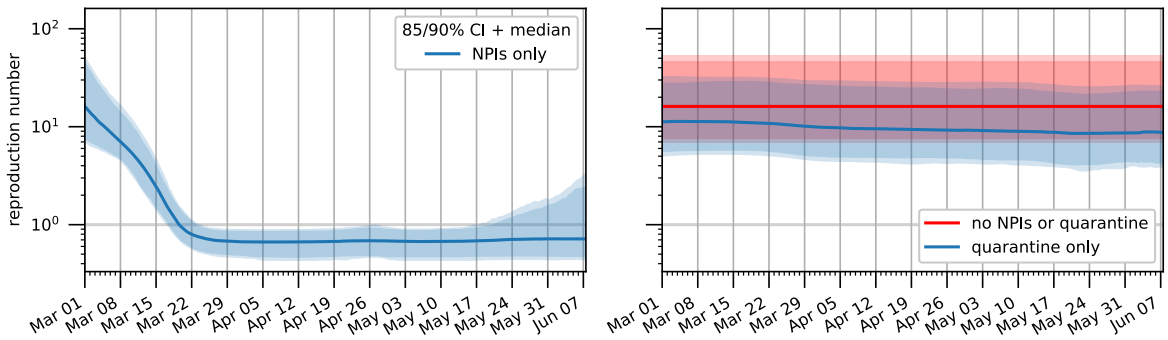



Figure 6: **Relative importance of the detection process and the NPIs on the spread of the epidemic.** Estimates for the reproduction number are plotted for three different scenarios: (A) no testing is performed; (B, blue) no NPIs are applied; (B, red) neither testing nor NPIs are used. The bands correspond to 85%/90% posterior credible intervals, while the solid line denotes the median value.

most illness related parameters, but to fix them to values taken from the literature. This is problematic for two reasons: Firstly, many parameters are region- and situation-specific and can thus lead to wrong estimates of inferred parameters. Secondly, estimates typically have large uncertainty intervals, as also seen in our study and by Raimúndez et al. (2021). Fixing those to single values may lead to an underestimation of the uncertainty of inferred parameters.

Instead, we incorporate pre-existing knowledge from the literature only as prior information and estimate all parameters at the same time.

Our integrative modelling framework is able to estimate the effectiveness of the testing strategies employed during the first wave of the epidemic in Munich (Germany). In particular, it suggests that the fraction of detected asymptomatic SARS-CoV-2 cases quickly saturated. At the beginning of April 24% (90% CI: 13–39) of asymptomatic individuals were detected and the numbers changed afterwards only marginally. Indeed, the model predicts even a small drop at the end of May, but the uncertainty is large (due to a low number of observations).

The proposed model relies on a detailed description of the testing process, with the inclusion of symptom onset data and time-dependent detection rates reflecting the varying test capacity of the health care system and the change of the criteria needed for obtaining a test. While still preliminary, we believe such additional modelling to be very important especially in the initial phase of an outbreak. The proposed formulation could be the basis for future studies and expanded by e.g. including age groups and interactions with neighbouring regions.

Overall, the proposed work highlights the importance of seroprevalence data and thereby complements various existing efforts. We expect that it might contribute to a better understanding of the dynamics of epidemics in a dynamic environment, with changing testing capabilities and NPIs. To facilitate reuse and extension, we openly share the analysis tools with the community.

## Materials and Methods

### Data sources

**Case report data** We use the official case data for Munich (Germany), which is released on a daily basis by the Robert Koch Institute (2020a). We process the data to obtain the total number of new cases and deaths which were communicated to the public health officials for each day. In addition to these two time series which are used in many other studies, we also extract from the raw data the number of symptom onsets for each day, which as far as we know has not been used in any other modelling effort. In particular, we distinguish cases where the patient underwent PCR testing after the symptom onset from cases where the patient was tested before exhibiting any symptoms (presumably thanks to contact tracing efforts). This information on symptom onsets is only known for a subset of the cases, but we believe this data to be important because it gives an indication of how many people are tested before any symptoms are observed and thus of how many reported cases are due to asymptomatic/presymptomatic individuals.

The RKI data has two main peculiarities: Firstly, there is a delay between the time a new case is tested and the time the case is added to the database. However, as we do not consider very recent data, this effect can be ignored in our case. Secondly, there is a non-negligible weekly periodicity in the reported values (both new cases and deaths) which we have to account for in the modelling. The origin of such effect is unclear, but it may be due to reporting delays and/or decreased testing capacity during the weekends. As explained in detail later, we try to explicitly model and estimate such an effect.

**Hospital bed counts** In addition to the RKI data, we employed also the number of hospital beds in Munich occupied by COVID-19 patients (IVENA). Bed counts were aggregated by hospital unit: ward, medium care unit (MCU) and intensive care unit (ICU). Since not all hospitals have a MCU, we further aggregated MCU and ICU bed counts. Due to a non-uniform reporting by the different hospitals (especially in the first days since the reporting system was set up), the data before 25 March 2020 had to be discarded. Some of the smaller hospitals/clinics did not reliably report bed counts even after this date and had to be excluded, leading to a possible, albeit slight, underestimation of the total number of occupied beds.

Another possible problem in the data is that patients may be moved from the surrounding areas to the city, where hospital concentration is higher, resulting in a possible overestimation of the number of occupied beds compared to what can be predicted by a model which does not take into account immigration effects. We thus introduce two under/over-representation factors for the bed counts, one for the occupied beds in ward and the other for the occupied ICU stations. We have to distinguish them since the distribution of patients coming from outside the city may be skewed in favor of severe cases which cannot be treated in smaller hospitals. Evidence for this can be found in the ratio between occupied ICU and ward beds, which differs significantly when comparing the city with the whole of Bavaria (Supplementary Figure 2).

**Seroprevalence data** We use the serological testing results reported by the "Prospective COVID-19 Cohort Munich" study (KoCo19). KoCo19 is organised by the Ludwig Maximilian University (LMU) Hospital and is currently monitoring nearly 3000 households in the Munich city area (Radon, Saathoff, et al. 2020). At regular intervals, blood samples for each household member are gathered and tested for several indicators of the presence of antibodies to SARS-CoV-2 (Olbrich et al. 2021). These tests are then used to impute the lifetime prevalence of COVID-19 in the general population (Pritsch et al. 2021), which due to the potentially large number of asymptomatic cases is impossible to recover from the case counts released by the national health authorities. In this work we employ the results from the first round of testing, spanning the period from April 6th to June 12th, 2020.

## Epidemiological model

**Illness phases and states** At the coarsest level, individuals can be assigned to compartments corresponding to biologically different phases of the infection. From each of these phases an individual can transition to a subset of the other phases and a transition probability can be assigned to each of these possible disease progressions. Such a model can be easily visualized in graph form (see Figure 1A) and converted to a set of ODEs (mostly linear, except for the non-linear infection process). However, as mentioned in the *Results* section, transition times are usually not exponentially distributed and thus each phase is subdivided into distinct model states in order to model Erlang-distributed transition times, in what is usually referred to as the Gamma Chain Trick (Smith 2011). Additionally, the distribution of the transition times may depend not only on the current illness phase, but also on the future phase (e.g., symptomatic individuals who worsen and are hospitalized transition to the next phase faster than symptomatic individuals who are never hospitalized). When this occurs a different branch for each possible progression must be considered, each with its own transition rates and possibly with a different

number of substates for the Gamma Chain Trick.

**Detection process** As mentioned in the *Results* section, the process by which infected individuals are reported to the healthcare authorities (by means of a positive PCR test) is modelled explicitly with additional states that represent detected individuals. A delay is also added between the time a patient dies and the time their death is reported to the healthcare authorities. This can be done by distinguishing deceased but unreported patients and reported deceased individuals, with the rate of the transition between these two states controlling the amount of delay present.

**Reporting of symptom onsets** Individuals who are tested while presymptomatic and whose date of symptoms onset is reported to the healthcare authorities are those who leave the presymptomatic phase while being already detected. Individuals who are tested while symptomatic and whose date of symptoms onset is reported are instead only just a fraction of those who leave the presymptomatic phase while not being detected yet. This is due to the fact that, when the detection rate is rather low, some symptomatic individuals may not be detected before the virus is cleared from their system. Fortunately, the fraction of symptomatic individuals who will eventually be detected can be easily computed from the model structure (all differential equations involved are linear) and can be used to correctly scale the number of onsets. Moreover, since we do not expect all detected individuals to report their symptom onset date, additional scaling factors (whose values are to be estimated) are added in both cases before obtaining the number of symptom onsets per day that are to be compared with the RKI data.

**Infection rate** In a simple SIR model, the equation for the change in the number of susceptible individuals $S$ is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta I}{N} S,$$

where $I$ is the number of infectious individuals, $N$ is the number of individuals who are alive and $\beta$ is the (mean) number of infectious contacts an infectious individual has in a time unit, which we will assume to be equal to one day. The total number of infectious contacts occurring in a day is then $\beta I$ and thus, assuming that such contacts are equally distributed among all alive individuals, each individual will be exposed to the virus an average of $\beta I/N$ times per day. Since only susceptible individuals can get infected, the number of new infections per day will be equal to $\beta IS/N$, as in the above equation. Our model distinguishes several compartments of infectious individuals $I_i$ (e.g., presymptomatic, asymptomatic, symptomatic, quarantined, hospitalized, ...), each with its own level of infectiousness which results in a different value of $\beta_i$. Then, the total number of infectious contacts in a day becomes $\sum_i \beta_i I_i$ and the resulting equation is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\sum_i \beta_i I_i}{N} S.$$

Estimating a different value of $\beta_i$ for each compartment independently would be very difficult. We have thus further decomposed each $\beta_i$ as $\beta_i = \gamma_i \tilde{\beta}_i$, where $\tilde{\beta}_i$ denotes the number of potentially infectious interactions that an individual from $I_i$ has in a day and $\gamma_i$ denotes the fraction of such interactions that are actually infectious. The factor $\gamma_i$ only depends on the biology of

the disease and thus on the illness phase. Some rough relationships between the various $\gamma_i$ are known a priori (e.g., $\gamma_i$ should be lower for asymptomatic individuals than for symptomatic ones) and are encoded in the model with very weak priors. The value $\tilde{\beta}_i$ is instead a measure of the mobility/behaviour of the individuals and can depend on other factors such as the NPIs currently enforced or whether one has been detected by the healthcare authorities (and is thus quarantined). In particular, we consider the values of $\tilde{\beta}_i$ for symptomatic and detected individuals to be constant over time and thus independent of the current NPIs. This is because the mobility of such individuals is always going to be low, since they know they are ill and will limit their contacts accordingly.

For asymptomatic and presymptomatic individuals who do not know they are infected, to improve interpretability we further decompose $\tilde{\beta}_i$ as $\gamma_{\mathrm{NPI}}(t)\tilde{\beta}_0$. Here, $\tilde{\beta}_0$ is the number of potentially infectious interactions in a pre-pandemic setting. The choice of this number is somewhat arbitrary, since any change in it can be compensated by an opposite change in the infectiousness level $\gamma_i$. Thus, we need to fix it (and not estimate it along with the other parameters) and as its value we choose an estimate of daily social interactions for Germany taken from the pan-European POLYMOD survey (Mossong et al. 2008). The factor $\gamma_{\mathrm{NPI}}(t)$ is instead the time-dependent reduction in potentially infectious contacts caused, e.g., by a reduction in mobility or adherence to social distancing norms, which in turn can be influenced by NPIs such as lockdowns. By definition of $\tilde{\beta}_0$, we get that $\gamma_{\mathrm{NPI}}(t)$ must be equal to 1 at the beginning of the epidemic. As an additional constraint, we enforce that the contact number for non-symptomatic individuals can never be smaller than the contact number for sick/quarantined individuals. While this constraint is very reasonable, we have seen that it cannot be learned by the model on its own.

**Time-dependent parameters** Time-dependent parameters are represented in our model by cubic Hermite splines defined on a grid of uniformly spaced points. The interval length has been set to a relatively large value of two weeks: this may cause some small artifacts in regions where changes occur fast, but a trade-off had to be made in order to keep computation time manageable (splines make up for the majority of model parameters). Each spline is encoded in the parameter vector by its values at grid points only and the spline derivatives are computed by finite differences, i.e., we are using a Catmull–Rom spline (Catmull and Rom 1974). A standard interpolating cubic spline would have required the same number of parameters and have been smoother (a cubic Hermite spline is only continuous up to the first derivative), but we believe the potential drawbacks were not worth it. First, a standard interpolating cubic spline requires a more difficult and computationally expensive implementation since a linear system must be solved to compute the polynomial coefficients. Second, the value of a standard spline at any given point depends on all the values at the grid points, while for a Hermite spline it depends only on the values at the two nearest grid points, preventing the unrealistic case of having parameter values at later times influence observations at earlier times (even if such influence would probably be quite small).

Since the time-dependent parameters we use are either transition rates (for the detection process) or multiplicative factors (for the reduction in infectious contacts), they must be positive at all times. This cannot be enforced with a spline because of possible under-shooting and thus we use splines to describe the logarithm of the parameters of interest instead of the parameter values directly. As regularization, we also add zero-mean normal priors for the derivatives and the

curvature of the time-dependent parameter (in its original scale, not in the logarithmic one), encoding our belief that if the data does not strongly suggest otherwise the time-dependent parameters should stay constant in value.

Detection rates have some additional peculiarities. First, the detection rate for asymptomatic individuals is expressed as a fraction of the rate for symptomatic individuals, since we expect asymptomatic individuals to be more difficult to detect; this is done by representing such a fraction as a spline whose values are constrained to the $[0, 1]$ interval. As for the spline encoding the symptomatic detection rate, it actually represents the detection rate relative to its starting value, which is optimized separately. In addition, to obtain the instantaneous detection rate a week-periodic component is added to account for the lower amount of reported cases during the weekends. This additional function is (in logarithmic scale) a periodic Catmull–Rom spline, normalized such that the mean of the values at the grid points (in this case, the seven days of the week) is one. We assume that the delay in death reporting has no long-term trend and is only made up by a periodic component, which is fitted separately from the one used for the detection rates.

## Observation model

Let $\dot{x}(t) = f(x(t), t; \theta)$ be our ODE model of the epidemic, where $x$ is a vector representing the model state and $\theta$ is the parameter vector. In most cases the model state is hidden, in the sense that it cannot be directly (or at least not exactly) measured. Let $\{(t_i, y_i)\}_i$ be the data against which the model must be fitted. In order to link this data to the hidden state of the model, random variables $Y_i \sim p_i(y|x(t_i), \theta)$, known as *observable quantities*, must be introduced. Note that these observable quantities cannot be simple deterministic mappings from the state $x$ to the observed values $y$ since (i) measurement noise cannot be avoided; (ii) even if noise were to be eliminated, a practical model is never a perfect representation of reality. The precise distribution of the observables $Y_i$ depends on the particular data source and is often referred as the *noise model*.

**Case counts (infections/deaths/symptom onsets reported to the RKI)** The RKI releases the total number of cases reported to the health care authorities on any given day. We compute this from the simulated model trajectory using the midpoint rule, using the instantaneous rate of detection at noon of each day. As the noise in the case counts by the RKI seems to increase as the number of cases increase (Supplementary Figure 1), a simple Gaussian noise model with constant variance is not sufficient and we use a negative binomial distribution instead. The negative binomial distribution is a generalization of the Poisson distribution which is over-dispersed, i.e., for which the *variance-to-mean ratio* (VMR, also known as *index of dispersion*) is greater than one (in the limit case in which the VMR tends to one we recover the Poisson distribution). Due to its higher flexibility the negative binomial distribution is often the preferred choice for count data (Beck and Tolnay 1995; Lin et al. 2021; Chan et al. 2021) and we employ it too, parametrizing it by mean (the daily number of cases predicted by the model) and VMR (estimated along the other parameters). While our data (except for the number of new cases) is not significantly over-dispersed, we have observed the negative binomial noise model to perform much better in practice than constant-variance additive noise or log-normally-distributed multiplicative noise.

**Hospitalization data** For the hospitalization data too it can be observed that the variance increases with the mean (Supplementary Figure 1). However, since the bed counts never approach zero significantly, we have found it sufficient to use a simple Gaussian noise model with constant variance instead of employing a more complex noise model.

**Prevalence estimates** The raw data from the serological testing study consists in a test date and test result for each participant. The period corresponding to the first round of testing is split into equal-duration intervals and the antibody test results are aggregated accordingly. Assuming that each of the resulting population subsets is representative of the whole, then the number of positive tests in each interval is a realization of a binomial random variable, with success probability equal to the prevalence (after correcting for the sensitivity and specificity of the test used) and number of trials equal to the total number of tests in that batch.

## Bayesian parameter estimation and priors

Our approach to parameter inference is fully Bayesian: we are interested in sampling probable parameter values according to the posterior distribution

$$p(\theta \mid y) \propto p(y \mid \theta)p(\theta),$$

in which $p(y \mid \theta)$ is the likelihood of the observed data and $p(\theta)$ is a prior distribution on the parameter values. The likelihood can be computed from the probability distributions of the observable quantities as

$$p(y \mid \theta) = \prod_i p_i(y_i|x(t_i), \theta),$$

so that only the prior distribution is left to be defined.

The prior distribution should encode all available information about the parameters not coming directly from the data against which the model is fitted. In our case such information is mainly derived from clinical cases reported in the pre-existing literature. For example, the distribution of the transition times between different illness phases (e.g., incubation time) has been the object of many publications and can be used to obtain a reasonable value for the number of stages in the Gamma Chain Trick expansion. Once the number of stages is set, the same transition time distribution leads to a prior on the transition rates. Often priors can be obtained both for single parameters and for more complex expressions containing several of them. For example, in the case of transition times some overlap may be present: in one paper information about the total hospitalization time may be given, in another the distribution of the time spent in the ICU may be presented and in a third the time from symptom onset to death (which in our model must necessarily transit through the ICU compartment) may be described. We aim to include all possible information in our model in order to exploit as much as possible the data already gathered by other researchers. As consequence, we can only get a non-normalized prior distribution, i.e.,

$$p(\theta) \propto \prod_k p_k(g_k(\theta)),$$

where each $g_k$ maps the parameter vector to a real quantity for which prior information is encoded in the form of a probability distribution $p_k$. Fortunately, computation of the normalization coefficient is not necessary in order to draw from the posterior distribution.

A summary of all priors used in our model is reported in Tables 1 and 2. We have seen that choosing the appropriate source for prior information is difficult. For example, while the biology of the virus is constant across all countries (at least during the first wave of the epidemic, before the rise of any variant) other processes such as infection and hospitalization are extremely country specific. For this reason, we have given more weight to reports on German cases over international studies or meta-analyses. A second problem is that there is no standardization in the medical literature regarding what are the quantities of interest that should be described. For example, mean hospitalization times can be reported separately for patients that need ICU care, patients that are ventilated, depending on the gravity of the symptoms, ... In the worst cases the explanations provided are insufficient to exactly determine what was actually measured, making it difficult if not impossible to map the information to the model. For example, hospitalization can be considered to end with death or discharge or both, but not all sources are clear on which definition they use. Additionally, incomplete information is often reported, such as only giving median transition times instead of more accurate statistics, such as interquartile ranges. A third problem is that often prior information obtained from different sources may appear to be somehow contradictory, a problem which is obviously exacerbated by the previous point. We have tried to only include prior information of whose meaning we were quite confident but still some slight incongruities could not be removed. We opted to employ particularly weakened priors in such cases rather than completely throw the information away. A final problem is that unless the reported values are stratified by age group they may not be applicable to different waves of the epidemic in which the affected demographics differ (e.g., fraction of hospitalized individuals). In our case this is not problematic since we have only modelled the first wave of the epidemic using prior information from studies conducted on patients from the same first wave.

## Computational modelling pipeline

We established a reusable computation pipeline to automate the fitting process. The compartment model is encoded in the Systems Biology Markup Language (SBML, Hucka et al. (2003)), a widely used standard in the systems biology community. The datasets used for fitting, together with parameter priors and bounds, are stored in the PEtab format (Schmiester et al. 2021), a standard format for the formulation of estimation problems. Both formats are supported by various simulation and analysis tools, which aids accessibility and reusability of model and results.

Using the SBML models and the data in PEtab format, we perform numerical simulation and sensitivity calculation with the C++/Python library AMICI (Fröhlich et al. 2021), and parameter estimation using the Python-based tool pyPESTO (Schälte et al. 2021). This combination of tools offers a broad spectrum of functionalities, including advanced gradient-based nonlinear optimization, profile calculation, sampling and ensemble uncertainty analysis. In the current phase, we are using the optimization and the sampling capabilities to infer the unknown process parameters, including the effects of different interventions.

In particular, for the sampling we use the pyPESTO interface to the state-of-the-art gradient-based No-U-Turn sampler (NUTS, Hoffman and Gelman (2014)) implemented in the library PyMC3 (Salvatier, Wiecki, and Fonnesbeck 2016). As parameter estimation with a high number

of parameters is computationally extremely demanding, we employed parallelization on a local high-performance cluster. Yet, as sampling by the NUTS sampler is intrinsically a sequential process, each individual chain can still take weeks before producing a sufficiently high number of samples.

## Funding

This study was funded by the Bavarian State Ministry of Science and the Arts, the University Hospital of Ludwig-Maximilians-University Munich, the Helmholtz Centre Munich, the University of Bonn (via the Transdiciplinary Research Areas), the University of Bielefeld, Munich Center of Health (McHealth) and the German Ministry for Education and Research (MoKoCo19, reference number 01KI20271), German Research Foundation (SEPAN, reference number HA 7376/3-1), Volkswagen Stiftung (reference number: 99 450). This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2047/1 - 390685813 and EXC 2151 - 390873048. The ORCHESTRA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167. The views expressed in this paper are the sole responsibility of the authors and the Commission is not responsible for any use that may be made of the information it contains. The funders had no role in study design, data collection, data analyses, data interpretation, writing, or submission of this manuscript.

## Acknowledgments

We gratefully thank all participants of the KoCo19 study for their trust, time, data, and specimens. This study would also not have been possible without the staff of the Division of Infectious Diseases and Tropical Medicine at the University Hospital of LMU Munich, Helmholtz Centre Munich, and Bundeswehr Institute of Microbiology, as well as all medical students involved. We thank the KoCo19 advisory board members Stefan Endres, Stephanie Jacobs, Bernhard Liebl, Michael Mihatsch, Matthias Tschöp, Manfred Wildner, and Andreas Zapf.

## Institutional Review Board Statement

The study protocol was approved by the Institutional Review Board at the Ludwig-Maximilians-University in Munich, Germany (opinion date 31 March 2020, number 20-275, opinion date amendment: 10 October 2020), prior to study initiation.

## References


Barbarossa, Maria Vittoria et al. (2020). "Modeling the spread of COVID-19 in Germany: Early assessment and possible scenarios". In: *PLOS ONE* 15.9, e0238559. DOI: 10.1371/journal.pone.0238559.

Beck, E. M. and Stewart E. Tolnay (1995). "Analyzing Historical Count Data". In: *Historical Methods* 28.3, pp. 125–131. DOI: 10.1080/01615440.1995.9956360.

Brauner, Jan M. et al. (2021). "Inferring the effectiveness of government interventions against COVID-19". In: *Science* 371.6531. DOI: 10.1126/science.abd9338.

Byrne, Andrew William et al. (2020). "Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases". In: *BMJ Open* 10.8. DOI: 10.1136/bmjopen-2020-039856.

Catmull, Edwin and Raphael Rom (1974). "A class of local interpolating splines". In: *Computer Aided Geometric Design*. Ed. by Robert E. Barnhill and Richard F. Riesenfeld. Academic Press, pp. 317–326. ISBN: 978-0-12-079050-0. DOI: 10.1016/B978-0-12-079050-0.50020-5.

Chan, Stephen et al. (2021). "Count regression models for COVID-19". In: *Physica A* 563, p. 125460. DOI: 10.1016/j.physa.2020.125460.

Courtemanche, Charles et al. (2020). "Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate". In: *Health Affairs* 39.7, pp. 1237–1246. DOI: 10.1377/hlthaff.2020.00608.

Dreher, Michael et al. (2020). "The Characteristics of 50 Hospitalized COVID-19 Patients With and Without ARDS". In: *Deutsches Ärzteblatt International* 117.16, pp. 271–278. DOI: 10.3238/arztebl.2020.0271.

Fröhlich, Fabian et al. (Apr. 2021). "AMICI: high-performance sensitivity analysis for large ordinary differential equation models". In: *Bioinformatics*. btab227. DOI: 10.1093/bioinformatics/btab227.

Giordano, Giulia et al. (June 2020). "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy". en. In: *Nature Medicine* 26.6, pp. 855–860. DOI: 10.1038/s41591-020-0883-7.

Hartl, Tobias, Klaus Wälde, and Enzo Weber (2020). "Measuring the impact of the German public shutdown on the spread of Covid-19". In: *Covid Economics* 1, pp. 25–32.

He, Xi et al. (2020). "Temporal dynamics in viral shedding and transmissibility of COVID-19". In: *Nature Medicine* 26, pp. 672–675. DOI: 10.1038/s41591-020-0869-5.

Hoffman, Matthew D and Andrew Gelman (2014). "The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." In: *Journal of Machine Learning Research* 15.1, pp. 1593–1623.

Hucka, M. et al. (2003). "The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models". In: *Bioinformatics* 19.4, pp. 524–531. DOI: 10.1093/bioinformatics/btg015.

ISARIC (2020). *COVID-19 Report: 03 September 2020*. URL: https://isaric.org/research/covid-19-clinical-research-resources/evidence-reports/.

Isho, Baweleta et al. (2020). "Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients". In: *Science Immunology* 5.52, eabe5511. DOI: 10.1126/sciimmunol.abe5511.

*IVENA eHealth — interdisziplinärer Versorgungsnachweis* (n.d.). Accessed: 2021-09-02. URL: https://www.ivena.de.

Jarvis, Christopher I et al. (2020). "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK". In: *BMC Medicine* 18.124. DOI: 10.1186/s12916-020-01597-8.

Karagiannidis, Christian et al. (2020). "Case characteristics, resource use, and outcomes of 10021 patients with COVID-19 admitted to 920 German hospitals: an observational study". In: 8.9, pp. 853–862. DOI: 10.1016/S2213-2600(20)30316-7.

Latsuzbaia, Ardashel et al. (Aug. 2020). "Evolving social contact patterns during the COVID-19 crisis in Luxembourg". In: *PLOS ONE* 15.8, pp. 1–13. DOI: 10.1371/journal.pone.0237128.

Lauer, Stephen A. et al. (May 5, 2020). "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". In: *Annals of Internal Medicine* 172.9, pp. 577–582. DOI: 10.7326/M20-0504.

Li, Michael Lingzhi et al. (2020). "Forecasting COVID-19 and Analyzing the Effect of Government Interventions". In: *medRxiv*. DOI: 10.1101/2020.06.23.20138693.

Lin, Yen Ting et al. (Jan. 2021). "Daily Forecasting of New Cases for Regional Epidemics of Coronavirus Disease 2019 with Bayesian Uncertainty Quantification". In: *medRxiv*. DOI: 10.1101/2020.07.20.20151506.

Linton, Natalie M. et al. (2020). "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data". In: *Journal of Clinical Medicine* 9.2. DOI: 10.3390/jcm9020538.

Liu, Y et al. (2021). "The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories". In: *BMC Med* 19, p. 40. DOI: 10.1186/s12916-020-01872-8.

Long, Quan-Xin et al. (2020). "Antibody responses to SARS-CoV-2 in patients with COVID-19". In: *Nature Medicine* 26, pp. 845–848. DOI: 10.1038/s41591-020-0897-1.

Lorch, Lars et al. (2021). *Quantifying the Effects of Contact Tracing, Testing, and Containment Measures in the Presence of Infection Hotspots*. arXiv: 2004.07641 [cs.LG].

Lyu, Wei and George L Wehby (2020). "Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US". In: *Health Affairs* 39.8, pp. 1419–1425. DOI: 10.1377/hlthaff.2020.00818.

McAloon, Conor et al. (Aug. 2020). "Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research". In: *BMJ Open* 10 (8), e039652. DOI: 10.1136/BMJOPEN-2020-039652.

Mossong, Joël et al. (2008). "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases". In: *PLOS Medicine* 5.3, e74. DOI: 10.1371/journal.pmed.0050074.

Olbrich, Laura et al. (2021). "A Serology Strategy for Epidemiological Studies Based on the Comparison of the Performance of Seven Different Test Systems - The Representative COVID-19 Cohort Munich". In: *medRxiv*. DOI: 10.1101/2021.01.13.21249735.

Pritsch, Michael et al. (2021). "Prevalence and Risk Factors of Infection in the Representative COVID-19 Cohort Munich". In: *International Journal of Environmental Research and Public Health* 18.7, p. 3572. DOI: 10.3390/ijerph18073572.

Radon, Katja, Abhishek Bakuli, et al. (2021). "From first to second wave: follow-up of the prospective Covid-19 cohort (KoCo19) in Munich (Germany)". In: *medRxiv*. DOI: 10.1101/2021.04.27.21256133.

Radon, Katja, Elmar Saathoff, et al. (June 2020). "Protocol of a population-based prospective COVID-19 cohort study Munich, Germany (KoCo19)". In: *BMC Public Health* 20 (1), p. 1036. DOI: 10.1186/s12889-020-09164-9.

Raimúndez, Elba et al. (2021). "COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modeling". In: *Epidemics* 34, p. 100439. DOI: https://doi.org/10.1016/j.epidem.2021.100439.

Robert Koch Institute (2020a). *COVID-19: Fallzahlen in Deutschland*. URL: https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0.

— (2020b). *Epidemiologischer Steckbrief zu SARS-CoV-2 und COVID-19*. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html.

Sakurai, Aki et al. (2020). "Natural History of Asymptomatic SARS-CoV-2 Infection". In: *New England Journal of Medicine* 383.9, pp. 885–886. DOI: 10.1056/NEJMc2013020.

Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck (Apr. 2016). "Probabilistic programming in Python using PyMC3". In: *PeerJ Computer Science* 2016 (4), e55. DOI: 10.7717/peerj-cs.55. URL: https://github.com/pymc-devs/pymc3.

Schälte, Yannik et al. (2021). *ICB-DCM/pyPESTO: pyPESTO 0.2.5*. DOI: 10.5281/zenodo.2553546.

Schilling, Julia et al. (2020). "Vorläufige Bewertung der Krankheitsschwere von COVID-19 in Deutschland basierend auf übermittelten Fällen gemäß Infektionsschutzgesetz". In: *Epidemiologisches Bulletin* 2020.17, pp. 3–9. DOI: 10.25646/6670.2.

Schmiester, Leonard et al. (Jan. 2021). "PEtab—Interoperable specification of parameter estimation problems in systems biology". In: *PLOS Computational Biology* 17.1, pp. 1–10. DOI: 10.1371/journal.pcbi.1008646.

Siedner, Mark J. et al. (Aug. 2020). "Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study". In: *PLOS Medicine* 17.8, pp. 1–12. DOI: 10.1371/journal.pmed.1003244.

Smith, Hal (2011). "Distributed Delay Equations and the Linear Chain Trick". In: Springer, New York, NY, pp. 119–130. DOI: 10.1007/978-1-4419-7646-8_7.

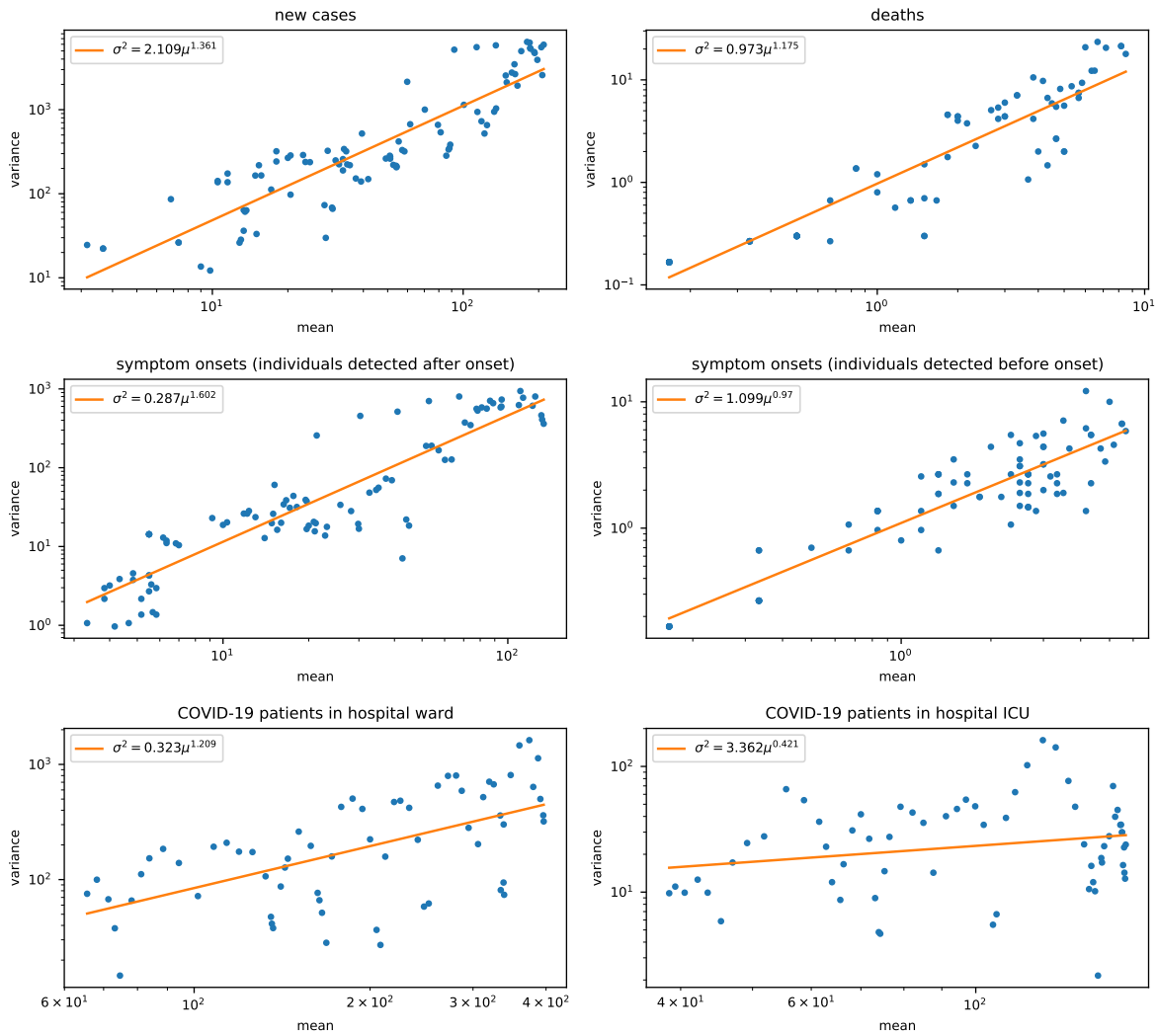
Streeck, Hendrik et al. (2020). "Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany". In: *Nature Communications* 11.1, pp. 1–12.

Sypsa, V et al. (2021). "Effects of Social Distancing Measures during the First Epidemic Wave of Severe Acute Respiratory Syndrome Infection, Greece". In: *Emerging Infectious Diseases* 27.2, pp. 452–462. DOI: 10.3201/eid2702.203412.

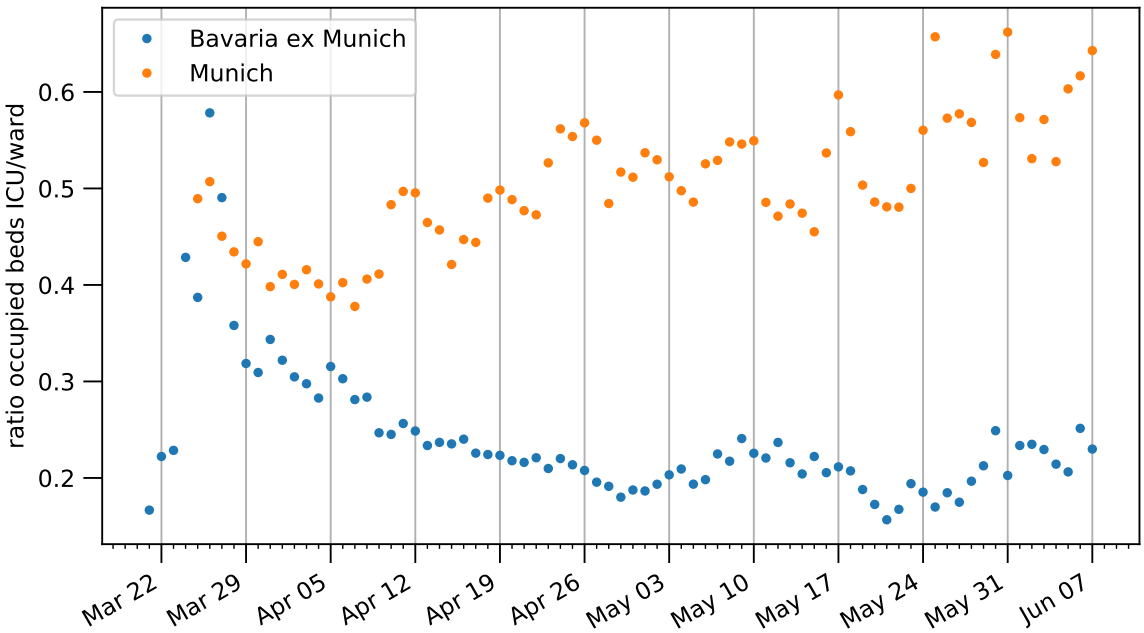
Tolksdorf, Kristin et al. (2020). "Eine höhere Letalität und lange Beatmungsdauer unterscheiden COVID-19 von schwer verlaufenden Atemwegsinfektionen in Grippewellen". In: *Epidemiologisches Bulletin* 41, pp. 3–10. DOI: 10.25646/7111.

Wei, Wycliffe E. et al. (2020). "Presymptomatic Transmission of SARS-CoV-2 – Singapore, January 23-March 16, 2020". In: *Morbidity and Mortality Weekly Report* 69.14, pp. 411–415. DOI: 10.15585/mmwr.mm6914e1.

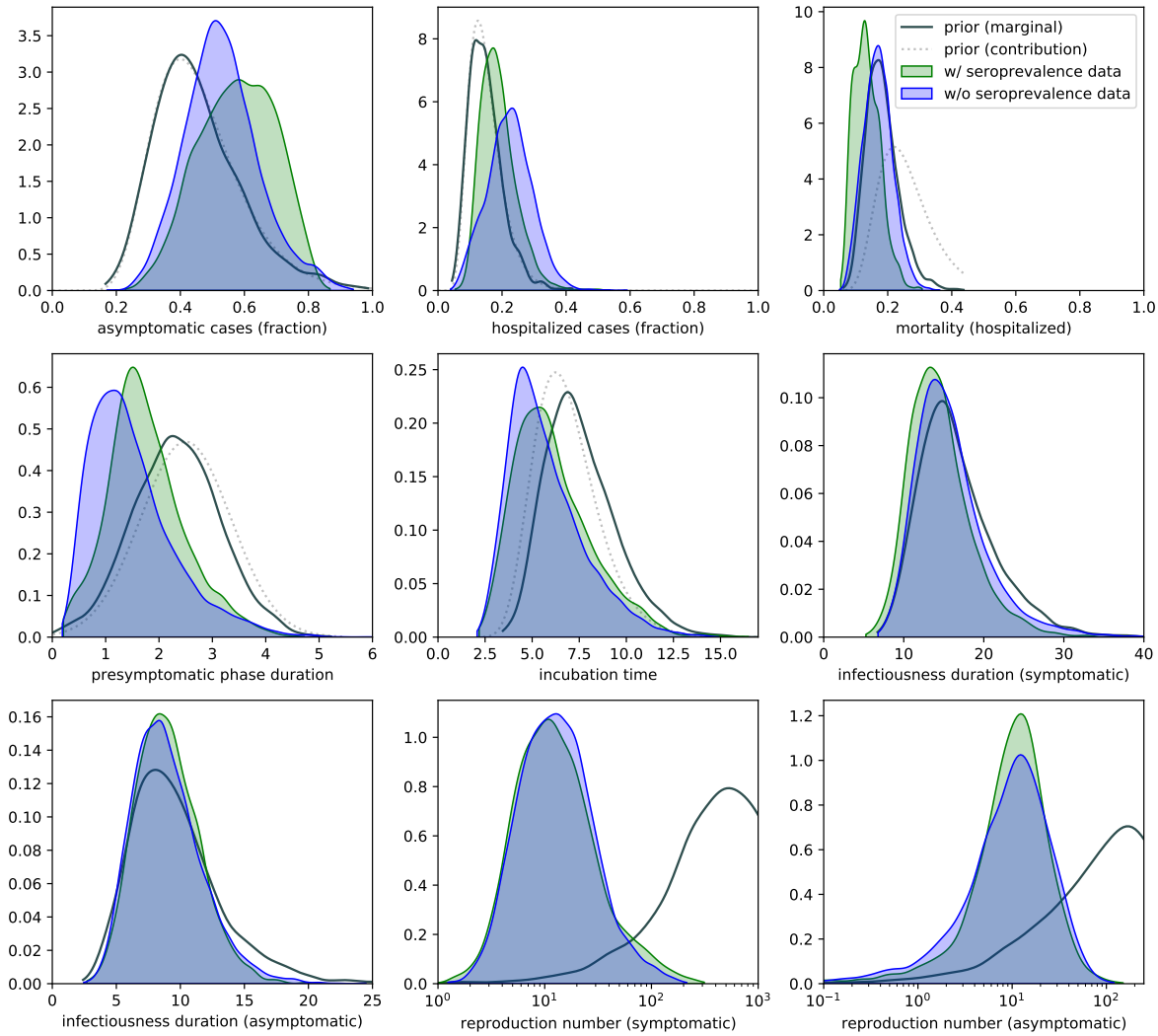
Zhao, S and H Chen (2020). "Modeling the epidemic dynamics and control of COVID-19 outbreak in China". In: *Quant Biol* 8, pp. 11–19. DOI: 10.1007/s40484-020-0199-0.




Supplementary Figure 1: **Relation between mean and variance in case count data**. Means and variances are computed from the timeseries over windows 5 days long.




Supplementary Figure 2: Ratio between beds occupied by COVID-19 patients in ICU and ward, respectively for the city of Munich and the rest of Bavaria.




Supplementary Figure 3: **Posterior distributions for a small subset of significant model parameters.** We have plotted the KDEs obtained for both the case in which seroprevalance data is used for fitting and the one in which it is not. Prior distributions are also plotted: the solid lines are the KDEs obtained from MCMC sampling of the prior, while the dotted lines (where present) are the contributions to the prior density due to the quantity under consideration.

| **Transition time (days)** | **Empirical distribution** | **Notes** | **Source** | **Prior on mean value** |
|---|---|---|---|---|
| Incubation | $\text{Log}\,\mathcal{N}(\mu = 1.63, \sigma = 0.5)$<br>mean 5.6, 95% CI 5.0–6.3<br>$\text{Erlang}(k = 6, \lambda = 0.88)$ | meta-analysis | McAloon et al. (2020)<br>Linton et al. (2020)<br>Lauer et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 1.9, \sigma = 0.25)$ |
| Presymptomatic | mean 2.3, 95% CI 0.8–3.0<br>median 2.5, IQR 2–3 | | He et al. (2020)<br>Wei et al. (2020) | $\mathcal{N}(\mu = 2.5, \sigma = 0.85)$ |
| Post-symptomatic infectiousness | mean 13.4 | meta-analysis; hospitalized patients only;<br>PCR-positive does not imply infectious | Byrne et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 2.56, \sigma = 0.35)$ |
| Asymptomatic infectiousness | Gamma(3) with mean 6<br>median 9, IQR 6–11, range 3–21 | | Byrne et al. (2020)<br>Sakurai et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 2.2, \sigma = 0.35)$ |
| Symptoms | median 7, IQR 3–10 | German study; hospitalized individuals only | Dreher et al. (2020) | $\mathcal{N}(\mu = 7, \sigma = 2.8)$ |
| Symptom onset to hospitalization | mean 7.8, SD 6.2, median 4, IQR 7<br>median 4, IQR 1–8 | international study<br>German study | ISARIC (2020)<br>Dreher et al. (2020) | $\mathcal{N}(\mu = 7.5, \sigma = 2.1)$ |
| Symptom onset to hospitalization (no ICU needed) | median 3, IQR 1–7 | German study;<br>non-ARDS patients (less likely to need ICU) | Dreher et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 1.57, \sigma = 0.425)$ |
| Symptom onset to hospitalization (ICU needed) | median 7, IQR 2–10 | German study;<br>ARDS patients (more likely to need ICU) | Dreher et al. (2020) | $\mathcal{N}(\mu = 7.5, \sigma = 2.0)$ |
| Symptom onset to ICU | median 9, IQR 4–12 | German study | Dreher et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 2.14, \sigma = 0.35)$ |
| Hospitalization to ICU | mean 2.9, SD 6.8, median 1, IQR 3 | international study | ISARIC (2020) | $\mathcal{N}(\mu = 2, \sigma = 1.25)$ |
| Hospitalization | mean 12.6, SD 13, median 9, IQR 12<br>median 8, IQR 5–1<br>median 10, IQR 5–19 | international study<br>German study<br>German study;<br>ARDS patients (possible over-estimation) | ISARIC (2020)<br>Dreher et al. (2020)<br>Tolksdorf et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 2.27, \sigma = 0.265)$ |
| Hospitalization (no ICU needed) | median 7, IQR 5–13 | German study;<br>ARDS patients (possible over-estimation) | Tolksdorf et al. (2020) | $\mathcal{N}(\mu = 7.5, \sigma = 1.575)$ |
| Hospitalization (ICU needed) | median 16, IQR 8–27 | German study;<br>ARDS patients (possible over-estimation) | Tolksdorf et al. (2020) | $\mathcal{N}(\mu = 15.25, \sigma = 4.25)$ |
| Hospitalization (deceased only) | median 10, IQR 4–16 | German study;<br>ARDS patients (more likely to die) | Tolksdorf et al. (2020) | $\mathcal{N}(\mu = 11, \sigma = 4.3)$ |
| ICU stay | mean 13.2, median 8.5, IQR 14.5<br>median 4, IQR 2–10<br>median 5, IQR 2–15 | international study<br>German study<br>German study;<br>ARDS patients (possible over-estimation) | ISARIC (2020)<br>Dreher et al. (2020)<br>Tolksdorf et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = 2.07, \sigma = 0.35)$ |

Supplementary Table 1: Prior information on the epidemiological process (transition times).

| **Fraction** | **Empirical estimate** | **Notes** | **Source** | **Prior** |
|---|---|---|---|---|
| Mortality (hospitalized individuals) | 22% | large-scale German study | Karagiannidis et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = -1.4, \sigma = 0.333)$ |
| | 21% | German study; ARDS patients (more likely to die) | Tolksdorf et al. (2020) | |
| Hospitalization | 10% | German study | Schilling et al. (2020) | $\text{Log}\,\mathcal{N}(\mu = -1.96, \sigma = 0.35)$ |
| | 15% | | Robert Koch Institute (2020b) | |
| ICU (hospitalized individuals) | 17% | international study | ISARIC (2020) | $\text{Log}\,\mathcal{N}(\mu = -2, \sigma = 0.45)$ |
| | 8% | German study | Schilling et al. (2020) | |
| | 17.23% | large-scale German study; non-invasive ventilation | Karagiannidis et al. (2020) | |
| | 13.15% | large-scale German study; invasive ventilation | Karagiannidis et al. (2020) | |
| | 37% | German study; ARDS patients (more likely to need ICU care) | Tolksdorf et al. (2020) | |

Supplementary Table 2: Prior information on the epidemiological process (fractions).